Role of translation in the degradation of antisense long non-coding RNAs in yeast

Sara Andjus

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie
UMR3244 « Dynamique de l'information génétique »

Rôle de la traduction dans la dégradation des longs ARNs non codants antisens chez la levure

Role of translation in the degradation of antisense long non-coding RNAs in yeast

Soutenue par
Sara ANDJUS
Le 20 octobre 2022

Ecole doctorale n° 515
Complexité du vivant

Spécialité
Biologie cellulaire

Composition du jury :

Alice, LEBRETON DR2, IBENS-ENS	*Président*
Anne-Ruxandra, CARVUNIS Assistant Professor, University of Pittsburgh	*Rapporteur*
David, TOLLERVEY Professor, University of Edinburg	*Rapporteur*
Olivier, NAMY DR1, I2BC	*Examinateur*
Antonin, MORILLON DR1, Institut Curie	*Directeur de thèse*
Maxime, WERY IR1, Institut Curie	*Co-Directeur de thèse*

institutCurie | PSL

ACKNOWLEDGEMENTS

First and foremost, I would like to thank all the members of the jury for accepting to review my thesis. Prof. David Tollervey and Dr. Anne-Ruxandra Carvunis for reading and reviewing the manuscript, Dr. Alice Lebreton and Dr. Olivier Namy for examining my thesis defense.

Following, I would like to thank Dr. Claire Torchet, Dr. Gwenaël Badis-Breard and Dr. Antoine Coulon for taking part of my thesis committee, for their expertise, advice, feedback, and substantial encouragement.

I would like to express my gratitude to Dr. Antonin Morillon for giving me the opportunity to perform first the master's internship and then the PhD in his lab. For his never-ending optimism, quick responses, and faith in the experiments we perform. For always supporting the idea that we should conduct a genome-wide analysis. Also, for often zooming out and explaining the importance of the projects despite the PCR not working. I would like especially to thank for the support for my short-term visit to IRB, Barcelona and for being chosen as the organizer of the Unit's Student Club.

My sincere acknowledgement goes to Dr. Maxime Wery for the daily based mentorship over the last four and a half years. In the first place, for teaching what the value of hard work is and for being the best example of it. For explaining and showing all the facets of what science is, from ordering tips, performing an experiment with multiple samples, to managing and funding a project. For sharing his knowledge and skills and devoting considerable time to my scientific growth. A sportsman once said that pressure is privilege. Thank you, Maxime for your critical and rigorous guidance, but also for passionately believing in me. I will always remember the moment I was about to say goodbye to Paris in front of the Louvre when my phone rang and you said shortly ‘Bravo, you are a PhD student’. I would not have reached this level if it wasn’t for your help.

I want to thank Dr. Francesc Posas and Dr. Eulàlia De Nadal, IRB, Barcelona for initiating a collaboration with us and for accepting me in their lab for a short-term visit. I especially thank Dr. Mariona Nadal-Ribelles for sharing her expertise and immerging with me into our single-cell adventure. I have to mention Dr. Lars Steinmetz, thanks to whom this collaboration has initiated.

Morevoer, I would like to thank Dr. Olivier Namy (I2BC, Gif-sur-Yvette) and his lab for the collaboration regarding the ribosome-profiling and the peptidomics part of the project.

This work would not have been performed without the financial support of the ministerial doctoral fellowship, which funded the first three years of my doctoral fellowship, the funding agency FRM, for the fourth-year allocation, and the academic support of the doctoral school Complexité du vivant, Sorbonne University and PSL University. I would like to thank EMBO for the short-term fellowship allowing me to perform my visit at IRB, Barcelona.

I also want to thank other members of Morillon's Team. Especially, Ugo Szachnowski, for performing all the bioinformatic analysis of the data obtained during my PhD. For being the best bioinformatician, involved in basically all the projects in our lab, but also much more. Thank you for your kindness, patience and all the major breakthroughs in science we achieved while eating chocolate. My deep gratitude goes to Marina Pinskaya, for the support she had given me during my PhD. For including me in the organization of the Non-Coding Genome Courses year after year. Together with Xavier Sabate Cadenas, we enjoyed the freedom we had (apart from the poster!) and felt very important to be part of it. Thank you for teaching us all how to be better scientists and especially better people, when in beautiful but also when in hard situations. I admire the superwoman you are with a discreet smile always on board. I thank Dominika Foretek, for all the nice moments we passed together. For her generosity, thinking about others' emotions (and allergies), for all the scientific inputs she proposed, and always having a solution to any problem. Je remercie Nicolas Vogt et Anna Almeida pour leur énorme patience, leur gentillesse, leur écoute et leur volonté d'aider. Également, à Nicolas pour l'aide technique e pour être le meilleur Bob le bricoleur, que tout laboratoire souhaiterait avoir ! Thank you to Nouritza Torrosian, for being our best singing diva and awakening us about the life happening outside the lab. To Marc Gabriel for encouraging me with tobrenole in the right moments and being ready to always wait this ten minutes more to go to the canteen. Grazie Rocco a te, per la tua solita domanda Sare, a te come va?, per l'incoraggiamento dal primo giorno quando eravamo piccoli studenti 'italiani', fino all'ultimo. Daje, che ce l'hai fatta pure tu! I want to thank the former members of our lab, especially Julien Jarroux, I wish we spent more time together in the lab, and Matthieu Lejars for his enormous investing in my french language and the RG tickets experience.

I thank Institut Curie for the unforgettable experience it offered me. I acknowledge all the stuff from the laverie and at the NGS platform. I am grateful to all the members of the UMR3244. I would like to mention Olivia Landre and Marie-France Lavigne for being funny when needed, but also being

serious when that is required. Alexandra and Sandra, for being my role models, Stefano for all the suggestions he gave me as experienced scientists. To Sreelekshmi, for being the best person met in the Institute. I am very grateful to have taken part in the ADIC group and learned that science can be done in many flavors.

Last but not least, I would like to thank my friends Olivera, Jelena, Tamara, Sofia, Guillem, Bertrand, Andrea, Lisa, Valeria, Oguzhan, and Yoann for making this world a better place.

Hvala ti Vojine za svu podrsku, veru i ljubav koju si mi pruzio.

Hvala Vam mama i tata i Isi. Uspeli smo u ovome zajedno.

Four and half years ago I entered Institut Curie in Paris and read the quote of Marie Curie 'Nothing in life is to be feared, it is only to be understood.' Frequently, I was encountered with this notion while performing important experiments. Despite the doubts that science can create, I will do my best to let these words guide my future professional career.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ANS	Anisomycin
as	Antisense
aslncRNA	Antisense Long Non-Coding RNA
bp	Base Pair
CAGE	Cap Analysis Gene Expression
cDNA	Complementary DNA
ChIP	Chromatin Immunoprecipitation
CHX	Cycloheximide
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSM	Complete Supplement Mixture
CUT	Cryptic Unstable Transcript
DECID	Decay Inducing Complex
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dsRNA	Double-Stranded RNA
DUT	Dicer-Sensitive Unstable Transcript
EJC	Exon Junction Complex
ENCODE	The Encyclopedia of DNA Elements
FANTOM	Functional Annotation of The Mouse/Mammalian Genome
FISH	Fluorescence In Situ Hybridization
FISSEQ	Fluorescent In Situ Sequencing
GEMs	Gel Beads-in-Emulsion
GRO-Seq	Global Run-On Sequencing
HDAC	Histone Deacetylases
HLA	Human Leukocyte Antigen
IF	Immunofluorescence
Kb	Kilobase Pair
kDa	Kilodalton
lincRNA	Long Intervening Noncoding RNA
lncRNA	Long Non-Coding RNA
MHC	Major Histocompatibility Complex
miRNA	Micro RNA
mRNA	Messenger RNA

MS	Mass Spectrometry
MUT	Meiotic Unannotated Transcript
NAT	Natural Antisense Transcript
ncDNA	Non-Coding DNA
ncRNA	Non-Coding RNA
NET-Seq	Native Elongating Transcript Sequencing
NGD	No Go Decay
NGS	Next-Generation Sequencing
NMD	Nonsense-Mediated mRNA Decay
NSD	No Stop Decay
nt	Nucleotide
NUT	Nrd1 Unterminated Transcript
ORF	Open Reading Frame
PALR	Promoter-Associated lncRNA
PCG	Protein-Coding Gene
PCR	Polymerase Chain Reaction
piRNA	Piwi-Interacting RNA
PIWI	P-Element-Induced Wimpy
PRC2	Polycomb Repressive Complex 2
PROMPT	Promoter Upstream Transcript
PTC	Premature Termination Codon
RBP	RNA Binding Protein
RISC	RNA-Induced Silencing Complex
RNA	Ribonucleic Acid
RNA-Seq	RNA Sequencing
RNAi	RNA Interference
RNAPII	RNA Polymerase II
RNase	Ribonuclease
RNP	Ribonucleoprotein
RQC	Ribosome-Associated Quality Control Complex
rRNA	Ribosomal RNA
RT-qPCR	Quantitative Reverse Transcription PCR
s	Sense
SAGE	Serial Analysis of Gene Expression
scRNA-Seq	Single Cell RNA-Seq
siRNA	Small Interfering RNA
SLAM-Seq	Thiol(Sh)-Linked Alkylation-Sequencing
smORF	Small ORF
snoRNA	Small Nucleolar RNA
snRNA	Small Nuclear RNA
SURF	Smg-1-Upf1-Erf1-Erf3 Complex

SUT	Stable Unannotated Transcript
TE	Transposable Elements
TIF-Seq	Transcript Isoform Sequencing
TRAMP	Trf4/Air2/Mtr4 Polyadenylation Complex
TRAP	Translating Ribosome Affinity Purification
tRNA	Transfer RNA
TSO	Template-Switching Oligonucleotide
TSS	Transcription Start Site
TTS	Transcription Termination Site
TU	Transcription Unit
UMI	Unique Molecular Identifier
UPF	Upstream Frameshift Protein
UTR	Untranslated Region
WT	Wild-Type
XUT	Xrn1-Sensitive Unstable Transcripts
YPD	Yeast Extract Peptone Dextrose

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Initially thought to be by-products of the pervasive transcription of eukaryotic genomes, long non-coding (lncRNAs) progressively emerged as key players in multiple cellular processes. LncRNAs show tissue-specific expression and respond to diverse stimuli, suggesting that their expression is precisely controlled. Their dysregulated expression has been associated to human diseases, including cancer. Several classes of lncRNAs exist, including antisense (as)lncRNAs that are synthesized from the strand opposite to sense protein-coding genes. Despite their regulatory importance, aslncRNAs have been poorly explored due to their low cellular abundance. In fact, in yeast they are extensively targeted by RNA decay machineries – nuclear exosome and cytoplasmic Xrn1 exoribonuclease.

Recent works in *Saccharomyces cerevisiae* revealed that aslncRNAs are mainly targeted to Xrn1 through translation-dependent Nonsense-Mediated Decay (NMD) pathway. The NMD-sensitivity of aslncRNAs suggests that they are translated, raising the question of their coding potential. On the other hand, aslncRNAs can also form double-stranded RNA with their paired-sense mRNAs, at least in some cells, protecting them from NMD. However, the extent and the regulatory mechanisms governing the fate of the aslncRNA either subjected to translation/decay or pairing/stabilization are unknown.

In this context, to enlighten the metabolism of aslncRNA, the objective of my thesis was to decipher the evolutionary role of decay machineries in controlling aslncRNAs expression, the role of translation in the degradation of aslncRNAs, and the heterogeneity of sense mRNA and aslncRNAs in yeast.

First, we studied aslncRNAs degradation in *Naumovozyma castellii,* a budding yeast endowed with RNAi, unlike *S. cerevisiae*, which lost it during evolution, by examining the interplay between the nuclear exosome, Xrn1 and Dicer. Our data showed that aslncRNAs decay in this species depends on the nuclear exosome and Xrn1 with no major effect of Dicer (Szachnowski, Andjus et al., 2019). They also suggest that the presence of cytoplasmic RNAi machinery in *N. castellii* reinforced nuclear RNA surveillance machinery to temper aslncRNAs expression.

To reveal the role of translation in the degradation of aslncRNAs, we analyzed their fate in conditions of translation inhibition. We show that aslncRNAs accumulate upon translation inhibition, reinforcing the idea that translation controls their decay. Using Ribo-Seq (in collaboration with Dr. Namy's lab at I2BC, Gif sur Yvette) we defined actively translated aslncRNAs. We demonstrate the

molecular bases subjecting aslncRNAs to NMD. Finally, we detected peptides from an NMD-sensitive aslncRNA reporter. The results of this work are on biorXiv (Andjus et al., 2022). The evolutionary role of NMD in modulating lncRNAs, was subject of a published review (Andjus et al., 2021).

Lastly, using single-cell RNA-Seq data (in collaboration with Dr. Posas's lab at IBR, Barcelona), we observed a large heterogeneity of co-expression of sense mRNA/aslncRNAs at the single cell level genome wide, critical for the metabolism of aslncRNAs. Moreover, we showed a direct correlation between aslncRNAs levels and the number of cells containing both sense and as RNA pairs, raising an intriguing hypothesis on the mechanistic impact of duplex formation on the fate of the involving pair.

In conclusion, my project contributed to reconsider that aslncRNAs are devoid of coding potential, highlighting the role of translation in determining their degradation via the NMD. As the NMD factors targeting them are conserved, this work in yeast helps comprehend the aslncRNAs metabolism in higher Eukaryotes. Our work also opens perspectives regarding the possible regulatory roles of the aslncRNA-derived peptides.

INTRODUCTION

Chapter 1. Generalities on Long Non-Coding RNAs

In this chapter, I describe the arrival on the scene of non-coding RNAs and provide a brief overview of the non-coding RNA world. Then, I present the families of non-coding RNAs. Further, I focus on the classification and function of long non-coding RNAs in mammals and yeast and in particular on their decay pathways.

1. The Rise of the Non-Coding RNAs

More than 60 years ago, Francis Crick together with other researchers, defined the Central Dogma of Molecular Biology (Crick, 1970). This Dogma, having its roots in the "one gene-one enzyme" hypothesis proposed by Beadle and Tatum (Beadle and Tatum, 1941), has illustrated the flow of genetic information allowing the expression of genes in a living cell.

Deoxyribonucleic acid (DNA) contains genetic information in the form of genes, tightly packaged in chromosomes. The information from DNA is transferred to a messenger ribonucleic acid (mRNA) via a process called transcription (Figure 1). During transcription, RNA polymerase II (RNAPII) binds to the promoters of genes, opens the DNA, and catalyzes the formation of pre-mRNA. While the pre-mRNA is being synthesized, they undergo the addition of a cap, the 7-methylguanosine triphosphate (m^7-Gppp) at the 5' end, removal of introns by splicing, and polyadenylation at the 3' end, altogether forming a mature mRNA. Those modifications are crucial to protect the RNAs from degradation and to ensure their efficient exportation to the cytoplasm for mRNA translation. Once in this compartment, ribosomes bind the open reading frames (ORF) of mRNAs, serving as a template for translation into proteins. Finally, mRNAs are degraded from the two extremities by different non-specific exonucleases. Importantly, at each step in this process, from transcription to decay, surveillance mechanisms are utilized to safeguard the fidelity and quality of mRNA.

For a long time after the discovery of the Central Dogma, geneticists and molecular biologists have believed that proteins were the only core molecules that controlled cellular identity and phenotype and that RNAs were merely passive carriers of genetic information. The protein-centric view was eventually challenged by the discovery of 'extra' DNA.

In the '80s, a popular opinion was that 'extra' DNA is junk, containing sequences considered to be useless evolutionary fossils with no function (Doolittle and Sapienza, 1980). However, the use of the term junk has been always controversial. Regarding this matter, in the early 2000 Sydney Brenner, the Nobel prize winner in Medicine 2002, proposed to distinguish two terms - junk and

garbage. He explained that the junk is the rubbish one keeps as it might be useful someday, while the garbage is the rubbish one throws away. Thus, claiming that extra DNA by definition cannot be garbage, but only junk because if it was garbage, it would have been discarded by evolution.

Surprisingly (or not), the discovery of the 'junk DNA' signed the start of the Non-Coding RNA Revolution provoking revisions of the concept of the Central Dogma. A large part of genomic sequences previously defined as 'junk DNA', was found instead to be able to produce RNA molecules with some form of functional activity (for review, see (Cech and Steitz, 2014; Morris and Mattick, 2014)). These RNA species attracted a lot of attention and added a new dimension and complexity to the regulation of DNAs, RNAs and proteins.

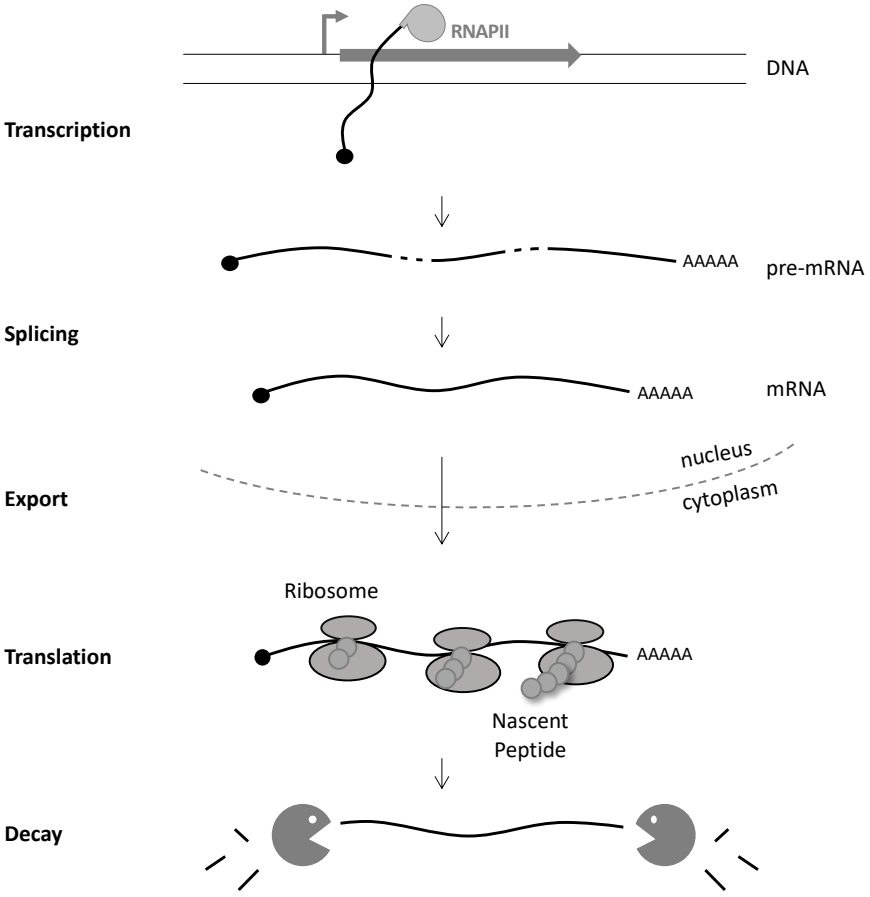


Figure 1 | Schematic Representation of Life Cycle of mRNA in the Cell. The pre-mRNA undergoes capping, splicing and polyadenylation. The mature mRNA in then exported to the cytoplasm where it is subjected to translation and decay. Bent up and block grey arrow represent transcription start site (TSS) and protein-coding gene, respectively. Black ball represents the cap, dashed lines the introns and full lines the exons. Cellular ribonucleases are represented as packmen.

Many studies, from different Eukaryotes, observed that RNAPII can be found at almost any genomic location concluding that their genomes are pervasively transcribed (for review, see (Berretta and Morillon, 2009; Dinger et al., 2009; Jensen et al., 2013)), generating a plethora of cryptic transcripts without coding capability (Djebali et al., 2012) defined as non-coding (nc)RNAs (Figure 2). For instance, even in the simple unicellular Eukaryotic model, *Saccharomyces cerevisiae,* up to 85% of the genome was found to be transcribed (David et al., 2006), with many transcripts arising from intronic and intergenic regions or heterochromatin domains (Steinmetz et al., 2006).

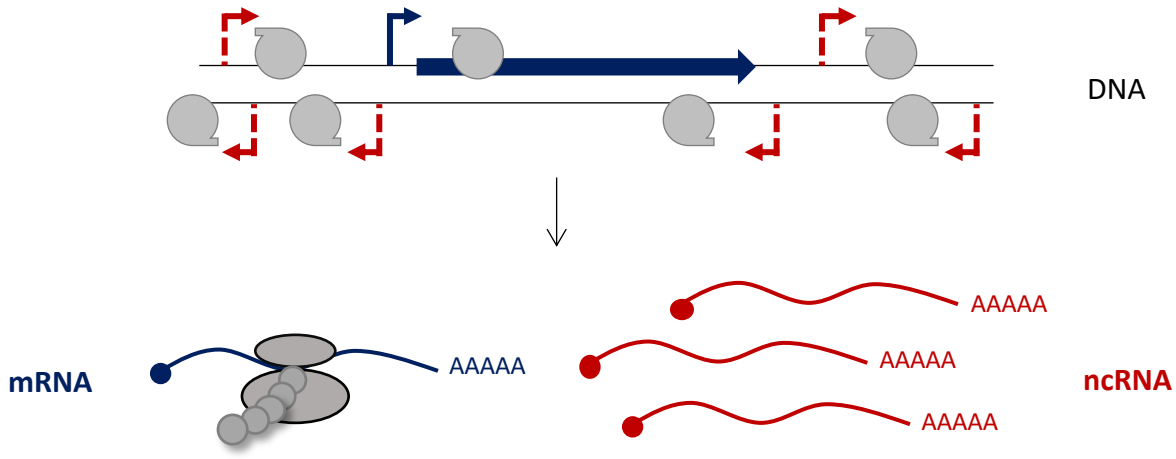


Figure 2 | Simplified Illustration of the Pervasive Transcription Producing NcRNAs. RNAPII can be found at any position in the genome generating mRNA (blue) and ncRNAs (red). mRNAs are translated and produce proteins. Red dashed bend arrows represent TSS of cryptic ncRNA. Other representations same as above.

Two pioneer consortiums focused on the identification and characterization of the transcriptional landscape in mammals. Data obtained by the Functional ANnoTation Of the Mammalian Genome (FANTOM) consortium implied that, in mouse, the number of transcripts was at least 10 times greater than the historically estimated 22,000 genes (Carninci et al., 2005). In parallel, the ENCyclopedia Of DNA Elements (ENCODE) project, launched in 2003, has revealed that about 75% of the human genome is transcribed, while no more than 2% of the whole genome is protein-coding (ENCODE Project Consortium et al., 2007, 2012), concluding again that most of the DNA is non-coding.

Whether these non-coding transcripts have any relevant recognizable purpose or just represent transcriptional noise has been hardly debated (Bakel et al., 2010; for review, see (Clark et al., 2011; Ponting and Haerty, 2022)).

On one hand, many studies described that spurious ncRNAs do not confer any fitness advantage and are thus by-products of transcription (for review, see (Struhl, 2007)). They impose minimal fitness cost, and thus simply tolerating them is more feasible than evolving and maintaining more rigorous control mechanisms that could prevent their production. However, even though some 'junk' may exist and continue to accumulate in our genome in a complete 'selfish' manner (Dawkins, 1989), evolution has reshaped a fraction of this vast quantity of DNA to play interesting new roles (for review, see (Eddy, 2012)).

For instance, it was observed that the proportion of ncDNA increased with organism complexity (less than 25% of prokaryotic genomes, more than 60% of plant and metazoan genomes, and 98,5 % of human genomes (Mattick, 2004)) (Figure 3). This suggested that it's likely that ncRNAs provide an extra layer of regulation of gene expression, allowing for greater complexity and responsiveness.

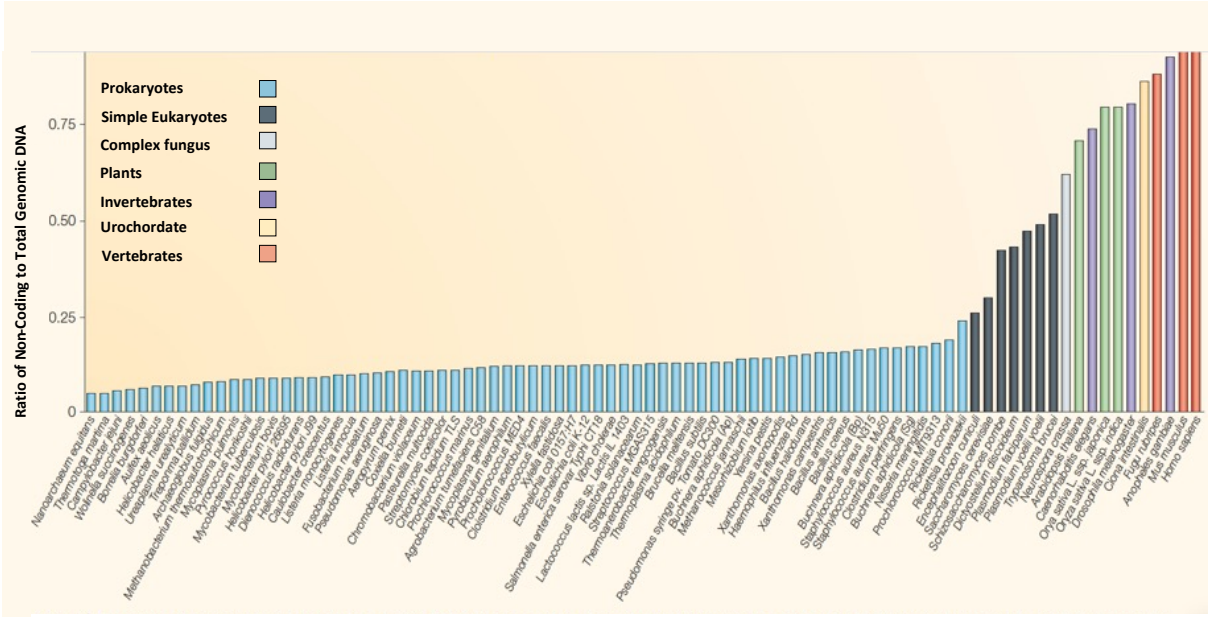


Figure 3 | The Ratio of Non-Coding to Protein-Coding DNA Increases as a Function of Developmental Complexity. The different colours on the left group Prokaryotes (bacteria and archaea) (blue), simple Eukaryotes (black), *Neurospora crassa* (complex fungus) (grey), plants (green), non-chordate invertebrates (nematodes, insects) (purple), *Ciona intestinalis* (urochordate) (yellow) and vertebrates (orange). Adapted from (Mattick, 2004).

Other well-documented studies discovered a growing number of ncRNAs with important cellular functions that evolved from spurious transcripts (for review, see (Palazzo and Koonin, 2020)).

For example, ncRNAs are proved to be involved in regulatory processes, from cell differentiation (Kretz et al., 2013) and development (Ulitsky et al., 2011), chromosome dosage compensation (Penny et al., 1996), genomic imprinting (Latos et al., 2012), to mRNA degradation (Gong and Maquat, 2011) and translation (Yoon et al., 2012), and adaptation to changing environment (Solé et al., 2015). Importantly, consistent with their functional importance, ncRNAs show cell and tissue-specific expression (Cabili et al., 2011; Djebali et al., 2012; Lorenzi et al., 2021; Mercer et al., 2008) and respond to diverse stimuli, suggesting that their expression is precisely controlled.

Moreover, since some ncRNAs have broad impacts on development, their dysregulation has been associated to various diseases including cancer and neurological disorders (for review, see (DiStefano, 2018; Schmitt and Chang, 2016)). Additionally, given that some ncRNAs are highly stable in body fluids (Li et al., 2015), they are emerging as promising candidates in the field of biomarkers and diagnosis (Fattahi et al., 2020; Feng et al., 2018; Renganathan and Felley-Bosco, 2017; Sharma et al., 2020). Therefore, in the near future, it is plausible that targeting ncRNAs will play a crucial role in gene therapy, offering novel options for precision medicine.

2. The Families of Non-Coding RNAs

Non-coding RNAs are broadly spread among the Eukaryotic kingdom and their number and diversity is large. In addition to infrastructural, housekeeping ncRNAs (ribosomal (r)RNAs, transfer (t)RNAs, and small nucle(ol)ar (sn(o))RNAs) with established functions in RNA processing and translation, pervasive transcription generates distinct types of ncRNAs that are commonly classified according to their size into small (<200 nt) and long (≥200 nt) ncRNAs. They are further subdivided in different classes based on differences in their biogenesis, function, or other particular feature (for review, see (Jarroux et al., 2017)).

2.1. The Small Non-Coding RNAs

Small non-coding RNAs can be grouped depending on their function in RNA processing and modification (sn(o)RNAs), protein translation (tRNAs) and regulation of gene expression (short-interfering RNA (siRNA), micro RNA (miRNA) and PIWI-interacting RNAs (piRNAs)) (for review, see (Moazed, 2009)). Small non-coding RNAs involved in the regulation of gene expression consist of 19-30 nt long RNAs responsible for RNA interference (RNAi), the biological process employing small RNAs as guides for sequence-specific gene silencing (Fire et al., 1998; Guo and Kemphues, 1995).

Precursors of siRNAs are long complementary double-stranded (ds)RNAs and hairpin RNAs. They are either produced in the cell itself or can be delivered to cells experimentally, as broadly exploited to manipulate gene expression. Successive endonucleolytic cleavages of dsRNA precursors by the RNaseIII endonuclease Dicer generate 21-25 nt siRNAs, with two-nucleotide 3' overhangs at both 3' ends. The discrete size of the produced dsRNAs and the two-nucleotide 3' overhangs represent hallmarks of RNaseIII-mediated cleavage. The siRNAs, once processed, are loaded into the RNA-induced silencing complex (RISC), containing an effector protein Argonaute. While one of the two strands of the siRNA duplex, the so-called passenger strand is discarded, the remaining strand termed guide RNA strand is retained within the activated RISC complex and directs the cleavage of specific target transcripts, thus inducing transcriptional and post-transcriptional gene silencing (for review, see (Farazi et al., 2008; Tomari and Zamore, 2005)). The targeting is precise because it is determined by base paring between the siRNA and the target mRNA. The cleavage of the target mRNA is catalyzed by Argonaute, leaving the mRNA to be finally degraded by cellular exonucleases. This simplified scheme constitutes the mechanistic basis of the RNAi and presently unites all gene silencing phenomena at transcriptional and post-transcriptional levels.

miRNAs are another type of small RNAs. Their biogenesis is similar but more complex. Most miRNAs derive from RNAs that are cleaved in the nucleus, which then fold into a stem-loop and are processed by an additional RNaseIII enzyme (Drosha) before being exported into the cytoplasm as double-stranded precursor miRNAs (pre-miRNA). Pre-miRNAs are actively transported to the cytoplasm where they are further processed by Dicer, that excises the pre-miRNA terminal stem-loop generating 21-25 nt mature miRNA that are then loaded onto RISC complexes. Usually, only one part of the miRNAs, known as the seed, pairs with the target mRNA. The imprecise matching allows miRNAs to target hundreds of endogenous mRNAs. Thousands of miRNAs participate in modulating diverse biological processes by controlling the expression level of more than 50% of protein-coding genes (for review, see (Bartel, 2009)).

piRNAs are an animal-specific class of small silencing RNAs, distinct from miRNAs and siRNAs (for review, see (Thomson and Lin, 2009)). While miRNAs and siRNAs derive from dsRNA precursors, piRNAs are processed in a Dicer-independent way from long single-stranded precursor transcripts. They are the largest class of small RNAs known to silence transposable elements (TE), regulate gene expression and fight viral infection. piRNAs associate with mammalian PIWI[1]-clade of Argonaute proteins to cleave target RNA, promote heterochromatin assembly and methylate DNA. Many questions regarding the precise molecular mechanisms of piRNA generation and their diverse silencing functions are still unanswered.

[1] The term given to the original *Drosophila* mutant: P-element induced wimpy testis (Piwi).

2.2. The Long Non-Coding RNAs

Even though small ncRNAs remain an active area of investigation, in recent years the attention of the scientific community has shifted towards long non-coding (lnc)RNAs.

In fact, lncRNAs have been identified in all species which have been studied at the genomic level, including animals (Brown et al., 1992; Clemson et al., 1996), plants (Swiezewski et al., 2009), fungi (Houseley et al., 2008), prokaryotes (Bernstein et al., 1993), and even viruses (Reeves et al., 2007). If the exact number of lncRNAs remains to be defined, the ever-expanding high-throughput sequencing technologies allow the establishment of more and more detailed lncRNAs catalogues continuously reshaping their specific features. To give numbers, the latest version (version 41) of the GENCODE, the catalogue of the human and the mouse genome (Frankish et al., 2021), lists 19,095 lncRNA loci. Hence, the number of lncRNA genes are in the same range as the number of human protein-coding genes (~20,000). Another dataset, called NONCODEV5, described an even greater number of lncRNAs (~100,000) in the human genome (Fang et al., 2018). Moreover, other collections involve specific lncRNAs with functional roles in cancer (Vancura et al., 2021).

2.2.1. Techniques For Analyzing Long Non-Coding RNAs

Different methodologies allowed the identification, classification and analysis of lncRNAs. Most of them rely on the detection of transcription from genomic regions that are not annotated as protein-coding and require unbiased RNA detection methods. Among other, these include tiling arrays, serial analysis of gene expression (SAGE) and high-throughput RNA sequencing (RNA-Seq).

In tilling arrays, the complementary (c)DNA is hybridized to microarray slides carrying overlapping oligonucleotides that cover either specific chromosomal regions or a complete genome (Rinn et al., 2003). This approach allows the investigation of global transcription from particular genomic areas and was initially used for the discovery and expression analysis of lncRNA. SAGE was the first method to use sequencing for high-throughput analyses of transcriptomes. It is based on the generation of short stretches of unbiased cDNA sequence by restriction enzymes (Velculescu et al., 1995). Sequencing of transcriptomes by RNA-Seq is one of the most effective methodologies for *de novo* identification and expression analyses of lncRNAs (Mortazavi et al., 2008). This method converts total RNA to a cDNA library that is sequenced by high-throughput sequencing instrument. A single sequencing run can produce millions of reads that are then aligned to a reference genome. After the alignment, the data are transformed into a quantitative measure of gene expression. To deal with the

complexity of eukaryotic pervasive transcription and the fact that many genes produce lncRNAs from the opposite strand of known transcripts, numerous strand-specific sequencing protocols have been developed (Mills et al., 2013).

Other throughput sequencing experiments are used to identify specific regions within lncRNA molecules. These include methods for high-resolution mapping of transcription start sites (TSS), as cap analysis of gene expression (CAGE) (Kodzius et al., 2006), and genome-wide annotation of polyadenylation sites, as poly(A)-position profiling by sequencing (3P-Seq) (Jan et al., 2011; Ulitsky et al., 2012). Moreover, a technique termed Transcript IsoForm sequencing (TIF-Seq), can be used for jointly sequencing both the 5' and 3' ends of lncRNAs (Pelechano et al., 2013).

Today, the latest cutting-edge single-cell transcriptomics, provide a significantly enriched view of transcription dynamics among seemingly identical cells (for review, see (Aldridge and Teichmann, 2020)). The chemistry of these techniques is focusing more and more on optimizing them for the detection of lncRNAs (Isakova et al., 2021). Fluorescent In Situ RNA-Sequencing (FISSEQ) is another technique that enables single-cell transcriptome study, but also the determination of the precise location of each transcript within the cell (Lee et al., 2014).

Most of these high-throughput sequencing techniques generate short reads. The assembly of transcripts from these reads significantly improved by using long reads generated by single-molecule long-read sequencing technologies, such as nanopore sequencing-based methods. Indeed, many studies have used such sequencing (Abdel-Ghany et al., 2016; Treutlein et al., 2014; Wang et al., 2016) and discovered that the resulting full-length sequences can be more accurate for investigating lncRNA characteristics (Hackl et al., 2014; Lagarde et al., 2017; Sharon et al., 2013).

A major limitation of these techniques is the fact that RNA-sequencing techniques measure steady-state RNA levels of fully processed RNA, representing a balance between synthesis and degradation. These levels do not directly reflect transcriptional activity leading to an underestimation of the extent and precise definition of the lncRNA landscape. In addition, given that many lncRNAs are highly unstable they are thus often undetectable by analysis of steady-state RNA levels.

As a solution to those problems, new methods were developed to gain a precise view of the transcription landscape by measuring transcription *per se* rather than steady-state levels of mature lncRNAs (for review, see (Nojima and Proudfoot, 2022)). Thus, measuring the initial transcripts transcribed by RNA polymerase, which has not yet undergone processing. An example of such technique is Global Run-On-sequencing (GRO-Seq) (Booth et al., 2016; Core et al., 2008), thiol(SH)-linked alkylation (SLAM-Seq) (Herzog et al., 2017) or Native Elongating Transcript sequencing (NET-Seq). GRO-Seq and SLAM-Seq map and quantify transcriptionally engaged polymerase density genome-wide using metabolic labeling. NET-Seq is instead based on the immunoprecipitation of

elongating RNAPII followed by deep sequencing of 3' ends of co-precipitated nascent RNAs, providing high resolution while keeping RNA strand information (Churchman and Weissman, 2011).

Although additional technological developments in lncRNA detection will assist in establishing their clearer picture, these techniques have enabled a substantial characterization of the landscape of lncRNAs and how they vary from mRNAs.

2.2.2. Comparison of Long Non-Coding RNA and mRNA Features

Despite having less pronounced characteristics than protein-coding mRNAs, lncRNAs share many features in common with mRNAs challenging researchers to understand their *bona fide* specificities. Once, lncRNAs were clearly distinguished from mRNAs for their disability to code for proteins. However, today, as discussed later in detail, many studies surprisingly contradict this notion, attesting that some non-coding transcripts can also in fact code for micropeptides[2].

Several other features of lncRNAs can be used as criteria to distinguish them from genuine protein-coding genes (for review, see (Ransohoff et al., 2018)). The majority of lncRNAs arise from independent transcriptional units. Their transcripts are typically shorter in size (Derrien et al., 2012; Pauli et al., 2012; Ulitsky et al., 2011), with fewer exons and weaker promoters (Mattioli et al., 2019). They are similar in structure to mRNA since the majority contain the cap structure, poly-adenylated tail, exon-exon splice junctions. LncRNAs are subjected to co-transcriptional splicing that is usually less efficient (Guo et al., 2020; Melé et al., 2017; Tilgner et al., 2012; Zuckerman and Ulitsky, 2019), and their transcription often terminates prematurely (Schlackow et al., 2017). Remarkably, lncRNAs tend to be less abundant (Clark et al., 2012), and their abundance is more often cell, tissue and cancer type-specific if compared to mRNAs (Carlevaro-Fita et al., 2020; Iyer et al., 2015; Lorenzi et al., 2021; Ravasi et al., 2006). Overall, these features result in a level of expression that is typically ten-fold lower than mRNAs (Cabili et al., 2011; Derrien et al., 2012; Guttman et al., 2009, 2010; Pauli et al., 2012; Ravasi et al., 2006; Sigova et al., 2013; Ulitsky et al., 2011). Moreover, the variability of lncRNAs among cells is higher (Cabili et al., 2011; Derrien et al., 2012; Pauli et al., 2012), with many lncRNAs preferentially being expressed in brain and testis (Cabili et al., 2011; Derrien et al., 2012; Ravasi et al., 2006). Globally, lncRNAs exhibit weak evolutionary constraints (Guo et al., 2020; Hezroni et al., 2015; Quinn et al., 2016) with modest primary sequence conservation. For example, fewer than 6% of zebrafish lncRNAs exhibit detectable sequence conservation with human or mouse

[2] Here, for simplicity, I will systematically use the term 'micropeptide' or 'peptide' to refer to the product of the translation of a lncRNA.

lncRNAs (Ulitsky et al., 2011), while less than 12% of human and mouse lncRNAs seem to be preserved in the other species (Cabili et al., 2011; Church et al., 2009).

3. Long Non-Coding RNAs in Mammals

3.1. Classification of Long Non-Coding RNAs in Mammals

Several classes of lncRNAs have been described in mammals (for review, see (Jarroux et al., 2017)), depending on their location with respect to protein-coding genes: Long intergenic/ intervening ncRNAs (lincRNAs), Promoter-Associated LncRNAs (PALRs), Enhancer-associated ncRNA (eRNAs), TElomeric Repeats containing RNAs (TERRA), PROMoter uPstream Transcripts (PROMPT) and Large antisense non-coding RNAs (lancRNAs or Natural Antisense Transcripts, NATs) (Figure 4).

LincRNAs represent the most abundant class of lncRNAs and result from transcription of intergenic regions. LincRNA genes are typically shorter than protein-coding genes and contain only 2-3 exons (for review, see (Clark and Mattick, 2011; Ulitsky and Bartel, 2013)). They are expressed in a tissue-specific manner, to a higher extent than protein-coding genes. The majority of the best-studied lncRNAs belong to lincRNAs. LincRNAs participate in chromosome dosage compensation (Tian et al., 2010; Zhao et al., 2008), development and cellular differentiation, chromatin and chromosome architecture, transcription and many other cellular processes.

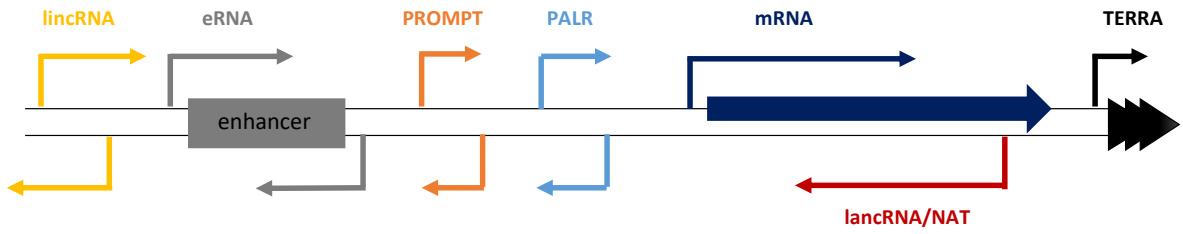


Figure 4 | The LncRNA Landscape in Mammals. lincRNA: Long Intergenic ncRNA (yellow). eRNA: Enhancer RNA (gray). PROMPT: Promoter Upstream Transcripts (orange). PALR: Promoter-associated lncRNA (light blue). LancRNA: Long antisense ncRNA (red). NAT: Natural antisense Transcript (red). TERRA: Telomeric Repeats (black triangles) containing ncRNAs (black). Adapted from (Tisseur et al., 2011).

PALRs are lncRNAs transcribed from promoters of protein-coding genes, and their presence positively correlates with promoter activity. They have been shown to be involved in the regulation of the expression of adjacent genes, as it has been reported for *CCND1* (Wang et al., 2008) or *Six3OS* (Rapicavoli et al., 2011). They are polyadenylated transcripts, longer than 200 nt.

eRNAs or ncRNAs having an enhancer-like function, have been recently discovered. Mostly bidirectional and non-polyadenylated, some of these transcripts have been functionally linked with gene expression. In addition to functioning through pre-existing chromatin conformations, certain eRNAs can promote or directly induce chromatin looping by interacting with scaffold proteins such as the Mediator or the structural maintenance of chromosomes complex cohesin (Melo et al., 2013). These interactions result in regulatory contacts between promoters and enhancers, which can be located close to each other but also be distant (Kim et al., 2018; Li et al., 2013). To cite an example, upon oestrogen receptor (ER) transcription activation, the *NRIP1* enhancer (*eNRIP*) is bi-directionally transcribed into an eRNA, which recruits cohesin to form chromatin loops, thereby promoting contact between the *NRIP1* enhancer and the promoters of *NRIP1* and trefoil factor 1 (*TFF1*), two of the numerous genes activated in response to ER activation (Li et al., 2013).

Telomeres are transcribed from C-rich strands into 300 bp up to 100 kb lncRNAs named TERRAs. These transcripts control telomeres length and chromatin structure and are regulated by the developmental and physiological state of the cell (Azzalin et al., 2007; for review, see (Schoeftner and Blasco, 2010)).

PROMPTs, transcribed from both sense and antisense strands, upstream of the TSS are unstable, polyadenylated transcripts detected in cells depleted for the nuclear RNA exosome (Preker et al., 2008, 2011), a multi-protein complex responsible for decay of various RNAs. In terms of evolution, they are similar to a specific class of exosome-sensitive lncRNAs in yeast, discussed further. The function of PROMPTs is yet to be clarified.

NATs are generated by antisense transcription from DNA strand antisense to genes (for review, see (Pelechano and Steinmetz, 2013)). They were originally discovered in plants (Henz et al., 2007). They can arise from independent promoters, bidirectional promoters of divergent transcription units (Seila et al., 2008; Sigova et al., 2013) or cryptic promoters (Kim et al., 2012). These antisense transcripts can regulate sense transcription directly by transcriptional interference, or indirectly by the production of regulatory antisense (as)lncRNAs. Several aslncRNAs were shown to regulate gene expression (Hu et al., 2011; Martianov et al., 2007; Rinn et al., 2007). For instance, it has been shown that NATs in mouse form RNA:RNA hybrids, that feed into the RNAi machinery and produce endo-siRNA (Carlile et al., 2009; Tam et al., 2008; Watanabe et al., 2008). Examples of aslncRNA gene regulation were shown not only in mammalian cells, but also in other organisms, especially in budding yeast (Berretta et al., 2008; Camblong et al., 2007, 2009; Houseley et al., 2008;

Pinskaya et al., 2009; Uhler et al., 2007; Van Dijk et al., 2011; van Werven et al., 2012), as discussed later.

3.2. Molecular Functions of Long Non-Coding RNAs in Mammals

Despite the large number of mammalian lncRNAs annotated in the recent years, the molecular function of most of them remains to be defined. Nevertheless, several robust examples of lncRNAs defined their versatile mechanisms of action (for review, see (Gourvest et al., 2019; Statello et al., 2021; Yao et al., 2019)). Also, depending on their localization, lncRNA can exert diverse functions (Carlevaro-Fita and Johnson, 2019; Fatica and Bozzoni, 2014). For instance, if located in the nucleus, lncRNAs can be important regulators in gene expression networks by controlling nuclear architecture and transcription, while if located in the cytoplasm can modulate mRNA stability, translation and post-translational modifications (Figure 5). They exert most of these functions by their ability to bind DNA, RNA, and proteins.

Nuclear lncRNAs are involved in epigenetic and transcriptional regulations by recruiting activator or repressor chromatin-modifying complexes and transcription factors onto target genes (for review, see (Huarte and Rinn, 2010; Marchese et al., 2017)) (Figure 5A-B). The main representatives of this functionality are lncRNAs that can recruit and guide the Polycomb Repressive Complex 2 (PRC2), a multiprotein complex that silences target genes by establishing a repressive chromatin state, to regions of interest. For example, *Xist* (X-linked X-Inactive-Specific Transcript) is a *cis*-acting lncRNA transcribed from a specific locus of the X chromosome, the X inactivation center (XIC), and is crucial to achieve X-chromosome inactivation (XCI). During embryonic development, *Xist* molecules spread over one of the two X chromosomes of female mammals and cause the silencing of a large proportion of its genes. A model for *Xist*-mediated XCI states that *Xist* directly recruit components of PRC2, leading to deposition of histone H3 lysine 27 trimethylation (H3K27me3) chromosome-wide to establish repressive chromatin across the inactive X (for review, see (Loda and Heard, 2019)). Another example is illustrated by *HOTAIR* (Hox antisense intergenic RNA), transcribed from the *HoxC* locus, that is known to suppress the expression of the *HoxD* locus also by recruiting PRC2 to establish a repressed chromatin state (Rinn et al., 2007). LncRNAs can regulate transcription and chromatin modifications both, in *cis* or in *trans*, depending if they act close or far from their transcriptional sites, respectively. For example, in mouse extra-embryonal tissues, *Airn* (antisense *Igf2r* RNA non-coding), functions *in trans* as it is directed to the promoters of two distal imprinted target genes, solute carrier family member 2 (*Slc22a2*) and *Slc22a3*, *via* a certain 3D chromosomal conformation. Once reached, *Airn* recruits PRC2, which catalyzes H3K27me3 and gene silencing. *Airn* also functions *in cis*, on its overlapping protein-coding gene insulin-like growth factor 2 receptor

(*Igf2r*). In this case, *Airn* transcription interferes with the recruitment of RNAPII and favors the paternal imprinting of *Igf2r* gene (Latos et al., 2012). Moreover, lncRNAs can form a DNA-RNA duplex/triplex that anchors associated effectors to active chromatin sites such as promoters or enhancers (for review, see (Li et al., 2016)). They are also involved in chromatin remodeling by forming inter/intra chromosomal loops (for review, see (Marchese et al., 2017; Melé and Rinn, 2016)).

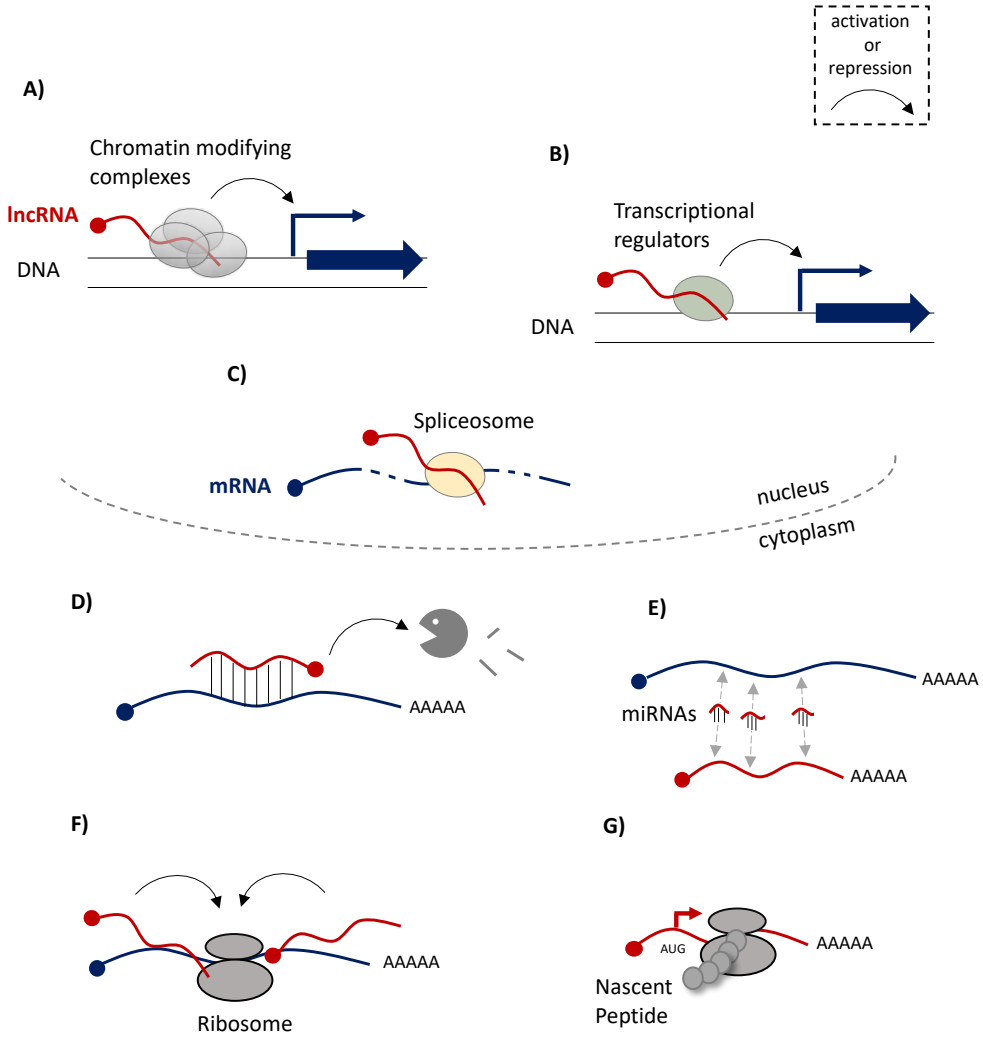


Figure 5 | Schematic Overview of Molecular Functions of LncRNAs. Nuclear lncRNAs are involved in (A) Epigenetic regulations, leading to the recruitment of activator/repressor chromatin modifying complexes on target mRNA promoters, (B) Transcriptional regulations, directing or preventing the recruitment of transcription factors on promoters of mRNA targets or on other active chromatin sites, or in (C) Splicing regulations, recruiting spliceosome partners. Cytoplasmic lncRNAs affect post-transcriptional steps by regulating (D) mRNA stability favoring or preventing their degradation, or by acting as (E) small regulatory RNA sponges. Lastly, lncRNAs regulate (F) mRNA translation and can also be (G) peptide producers. AUG, start codon. Adapted from (Gourvest et al., 2019).

LncRNAs are implicated in post-transcriptional modifications by regulating either splicing or editing in the nucleus (Figure 5C), mRNA stability (Figure 5D-E) or translation (Figure 5F) in the cytoplasm (for review, see (He et al., 2019)).

In most cases lncRNAs regulate gene splicing through interaction with spliceosome partners. The lncRNA metastasis-associated lung adenocarcinoma transcript 1 (*MALAT-1*) is localized at the periphery of nuclear speckles and is required for proper localization of several splicing factors to nuclear speckles, regulating the pre-mRNA splicing (Tripathi et al., 2010).

In the cytoplasm, lncRNAs can directly bind to mRNA and regulate mRNA stability, or competitively bind to mRNA to improve mRNA stability. For instance, lncRNA overexpressed in colon carcinoma-1 (*OCC-1*) can interact with human antigen R (HuR) and recruit ubiquitin ligase to HuR, such that HuR is downregulated by destabilization (Lan et al., 2018). Since HuR serves as a stabilizing factor for a large number of mRNAs, it ultimately causes downregulation of HuR-targeted mRNAs.

Cytoplasmic lncRNAs act as microRNA sponges, also called competitive endogenous RNAs (ceRNAs), containing microRNA binding sites (for review, see (Salmena et al., 2011)). MicroRNA sponges are able to sequestrate microRNAs and keep them away from their mRNA targets leading to the stabilization of their targets. To cite an example, in tumours lncRNA-*PNUTS* is generated by alternative splicing of the *PNUTS* pre-mRNA by heterogeneous nuclear ribonucleoprotein E1 (hnRNPE1) (Grelet et al., 2017). It contains multiple *miR-205* binding sites, which reduce the availability of *miR-205* to bind and suppress the zinc finger E-box-binding homeobox 1 (*ZEB1*) and *ZEB2* mRNAs. *ZEB1* and *ZEB2* are consequently upregulated, promoting the epithelial–mesenchymal transition and breast cancer cell migration and invasion (Grelet et al., 2017).

Involvement of lncRNAs in translation regulation has also been reported. For instance, *lincRNA-p21* interacts with HuR, and such association favors the recruitment of let-7/Ago2 to destabilize *lincRNA-p21* (Yoon et al., 2012). Upon loss of HuR, *lincRNA-p21* accumulates and associates with JunB (*JUNB*) and β-catenin (*CTNNB1*) mRNAs via base pairing to suppress their translation by recruiting the translation repressor Rck (Yoon et al., 2012). LncRNAs can also activate mRNA translation. *Uchl1* (ubiquitin carboxyterminal hydrolase L1) is a gene involved in brain function and neurodegeneration in mice. The lncRNA *as-Uchl1* (antisense to *Uchl1*) enhances the formation of active polysomes on *Uchl1* mRNA and promotes its translation via a SINE B2 segment complementary to a region within the 5’ end of *Uchl1* mRNA (Carrieri et al., 2012).

Lastly, proteomic analyses revealed that despite their acronym, lncRNAs may encode for micropeptides (Figure 5G) (Anderson et al., 2015; Ingolia et al., 2014; Slavoff et al., 2013), some of which functionally relevant, as discussed later.

4. Long Non-Coding RNAs in Yeast

4.1. Classification of Long Non-Coding RNAs in Yeast

In most Eukaryotes small and long ncRNAs co-exist and can co-regulate many cellular processes. However, the budding yeast *S. cerevisiae* represents an exception among Eukaryotes, since it has lost the RNAi system during evolution. Indeed, it is devoid of the RNaseIII endonuclease Dicer that can process dsRNA structures into small ncRNAs (Drinnenberg et al., 2009). Hence, *S. cerevisiae* has become a prominent model to specifically study the effects of lncRNAs, which could partially be hidden due to the action of small ncRNAs.

Several classes of lncRNAs have been described in *S. cerevisiae* (for review, see (Tisseur et al., 2011; Tudek et al., 2015)). They are mainly transcribed by the RNAPII, capped and polyadenylated (Van Dijk et al., 2011; Wyers et al., 2005), as mRNAs. Strikingly, the majority of this transcripts are lowly abundant as the consequence of an extensive degradation by RNA decay machineries. Before characterizing each of these lncRNA classes, I present an overview of the general process of RNA degradation.

As a matter of fact, coding or non-coding, all RNA species in Eukaryotic cells undergo turnover as the last lever of regulation of gene expression (for review, see (Parker, 2012)). Main RNA degradation mechanisms direct RNAs to the cytoplasmic Xrn1 or nuclear Rat1 in yeast/Xrn2 in mammals 5'-3' nucleases, or to the exosome, a conserved cytoplasmic and nuclear complex with 3'-5' exonuclease activity and an endonuclease cleavage site (for review, see (Meyer et al., 2004; Parker and Song, 2004)). The rate at which decay occurs depends on RNA sequence or structural elements and usually requires the RNA to be modified in a way to allow recruitment of the decay enzymes to the transcript. Most *bona fide* transcripts are protected from the decay machineries due to the presence of the cap structure and the poly(A) tail. To initiate the decay, either one of these two structures must be compromised, or the transcripts must be cleaved internally. In general, nonfunctional, aberrant RNAs produced as by-products of transcription, get more rapidly degraded than other RNAs. Thus, the lifespan of an RNA varies depending on the type of the RNA, from three minutes to more than one hundred minutes for yeast mRNAs (Chan et al., 2018; Geisberg et al., 2014; Wang et al., 2002b).

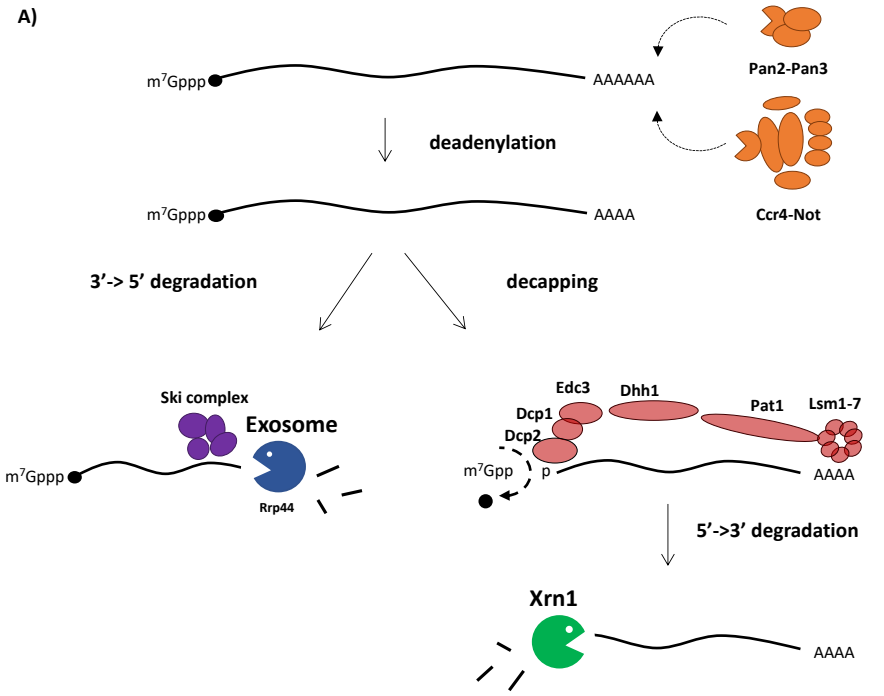
In the cytoplasm, there are two main degradation pathways of Eukaryotic mRNAs (Figure 6A) accomplished by several RNA decay enzymes (Figure 6B). These two main decay pathways exist in all Eukaryotes and their importance varies depending on the species. In general, both pathways start by shortening of the 3' poly(A) tail (Decker and Parker, 1993; Muhlrad and Parker, 1992) during a

process named deadenylation. In yeast, and presumably in other Eukaryotes, this step, is first carried out by the Pan2/Pan3 complex and then by the Ccr4/Pop2/Not complex (Brown and Sachs, 1998; Tucker et al., 2001). The latter complex, which is the main deadenylase complex, consists of two active 3'-5' exonucleases (Ccr4 and Pop2 in yeast/Caf1 in mammals) and includes a large scaffolding protein Not1 and other accessory Not2, Not3, Not4, Not5, Caf40, and Caf130 proteins (for review, see (Denis and Chen, 2003)).

Following deadenylation, mRNAs can be subjected to 3'-5' degradation by the exosome (Anderson and Parker, 1998). The exosome is a multiprotein complex consisting of ten main proteins (Exo10) (Allmang et al., 1999), including the Rrp44 in yeast/Dis3 in mammals protein, which has both an exonuclease and endonuclease domain (Lebreton et al., 2008; Schaeffer et al., 2009) and nine other proteins (Exo9) that form a tunnel allowing RNA to be addressed at the catalytic site of Dis3 (Bonneau et al., 2009). The exosome requires the aid of the Ski complex in order to degrade cytoplasmic mRNAs. The Ski complex is composed of Ski2, Ski3, and Ski8 and interacts with the exosome via Ski7, a protein bridging the two complexes. The Ski complex proteins also form a channel to direct RNA for degradation to the exosome (Halbach et al., 2013). It has been reported that the Ski2 protein is likely an RNA helicase and may allow unwinding of RNA structures to facilitate degradation. The recruitment of the exosome takes place at the 3' unprotected ends as well as on certain mRNAs via sequence specific motifs (Chen et al., 2001).

Alternatively, more frequently, after deadenylation, transcripts are led to decapping, exposing the decapped transcript to 5'-3' degradation by Xrn1 (Decker and Parker, 1993). The decapping step represents indeed a pre-requisite for Xrn1-dependent degradation (Hsu and Stevens, 1993) and in yeast is performed by the decapping enzyme complex formed by Dcp1 and the catalytic subunit Dcp2. This catalytic subunit cleaves the cap structure to release m^7-Gpp and a 5' monophosphate (5' p-RNA) (She et al., 2008). Dcp1 interacts with Dcp2 to stimulate its catalytic activity (Deshmukh et al., 2008; She et al., 2008). Several protein factors, referred to as either decapping enhancers or activators, are known to function to stimulate the rate of decapping *in vivo*. The core set of proteins affecting decapping includes Dhh1, a DEAD-box helicase, Pat1, Edc1, Edc2, Edc3, Scd6, and the Lsm1–7 complex. Pat1 serves as a scaffolding protein for the decapping complex, stimulates Dcp2 and interacts with Lsm1-7 at the 3' end of an mRNA (Sharif and Conti, 2013). Some of these decapping activators promote decapping by inhibiting translation initiation. Evidence in yeast show that each of these factors target a specific subset of mRNA (He et al., 2018), but the molecular mechanism involved remained elusive. A recent study reported that yeast likely contain distinct decapping complexes, and that Dcp2 controls decapping enzyme specificity (He et al., 2022). Upon decapping, mRNAs are degraded in a 5' to 3' direction by the Xrn1 nuclease (Hsu and Stevens, 1993). Genetic investigations indicated that Xrn1 binds to an internal portion of Dcp2 and is recruited

by Dcp2 to the decapping complex (He et al., 2022). Thus, Dcp2-mediated decapping and Xrn1-mediated 5′ to 3′ exoribonucleolytic digestion are physically coupled *in vivo*, indicating that Dcp2 also regulates efficient 5′-3′ exonucleolytic decay. Interestingly, at least some of the degradation activity of Xrn1 was shown to occur co-translationally (Hu et al., 2009; Pelechano et al., 2015), following the translating ribosomes.



B)

Deadenylation		3' -> 5' degradation	5' -> 3' degradation
		Core Exosome	Decapping
Ccr4-Not	Ccr4	Rrp4, Rrp40, Rrp41, Rrp42, Rrp43,	Dcp1
	Pop2	Rrp44, Rrp45, Rrp46	Dcp2
	Not1-5	Csl4	Edc3
	Caf40	Mtr3	Edc4
	Caf130	Cytoplasmic cofactors	Dhh1
Pan2-Pan3	Pan2	Ski2, Ski3,Ski7, Ski8	Pat1
	Pan3		Lsm1-7
			Xrn1

C)

Figure 6 | General Pathways of RNA Degradation in Eukaryotes. (A) Cytoplasmic RNA decay and (B) the main nuclease complexes involved in its different steps. Adapted from (Parker, 2012 & Parker and Sheth, 2007). (C) Nuclear RNA decay factors.

Xrn1 participates in the decay of mRNAs after internal cleavage and in the cytoplasmic mRNA surveillance system that degrades aberrant mRNAs (for review, see (Fourati and Graille, 2014)). Moreover, Xrn1 directs degradation of lncRNAs (Van Dijk et al., 2011), tRNA (Chernyakov et al., 2008), as well as maturation of ribosomal RNAs (rRNAs) (Geerlings et al., 2000).

The nuclear quality control systems (Figure 6C) prevent the function of the aberrant mRNA defective in pre-mRNA splicing, polyadenylation, mRNA export, and a huge number of cryptic lncRNAs arising from pervasive transcription, by inducing their degradation in the nucleus. A paralog of Xrn1 is Rat1, which is predominantly localized to the nucleus, and functions in nuclear RNA processing and/or degradation pathways (Johnson, 1997; Xiang et al., 2009). Rat1 also has a function in the 'torpedo' model of RNAPII transcription termination in degrading nascent mRNA after cleavage and allowing the dissociation of the elongation complex from the RNA (Kim et al., 2004; West et al., 2004). In the nucleus, the exosome associates with the nuclear-specific Rrp6 in yeast/EXOSC10 in mammals subunit carrying an Exo domain with distributive 3' to 5' exoribonuclease activity (Allmang et al., 1999). In order to act both efficiently and specifically on its substrates, the nuclear exosome relies on the interaction with several cofactors. The conserved TRAMP (Trf4/5-Air1/2-Mtr4-Polyadenylation) complex is the main nuclear exosome cofactor of Eukaryotic cells and plays a key role in promoting the degradation of essentially all surveillance targets of the exosome in yeast, including defective pre-tRNA (Kadaba et al., 2004), pre-rRNAs (de la Cruz et al., 1998) and cryptic RNAPII transcripts (Wyers et al., 2005). In budding yeast, the heterotrimeric TRAMP complex consists of either the Trf4 or Trf5 non-canonical poly(A) polymerases, the RNA-binding protein Air1 or Air2, and the DExH-box helicase Mtr4 (LaCava et al., 2005; Vaňáčová et al., 2005; Wyers et al., 2005). TRAMP binding to its substrates is mediated by Air1/2, while Trf4/5 polyadenylates the transcript at the 3' end, which is a preliminary requirement for the efficient degradation by the exosome (LaCava et al., 2005). The addition of the A-tail is supposed to favor substrate engagement by the Mtr4 helicase. Then, the helicase unwinds the RNA, preparing it for the decay by the exosome (Jia et al., 2012). The targeting of RNAs to the TRAMP and exosome complexes appears to be coupled in part with the Nrd1 and Nab3 RNA-binding proteins. The Nrd1 and Nab3 proteins bind to specific elements on nascent RNA (Carroll et al., 2007), and trigger transcription termination by a direct interaction between Nrd1 and the carboxy-terminal domain of RNAPII (Gudipati et al., 2008; Vasiljeva et al., 2008). The Nrd1 complex is also able to interact with the polyadenylation complex TRAMP and the nuclear exosome (Vasiljeva and Buratowski, 2006).

The abovementioned surveillance pathways target and degrade also yeast lncRNAs that result from pervasive transcription. Transcriptome analyses in yeast strain defective for RNA degradation pathways revealed particular classes of yeast lncRNAs in *S. cerevisiae*, many of which are

antisense to protein-coding genes (Figure 7). These lncRNAs include the Cryptic Unstable Transcripts (CUTs) that are sensitive to the nuclear exosome-dependent 3'-5' RNA decay pathway (Neil et al., 2009; Xu et al., 2009), the Xrn1-sensitive Unstable Transcripts (XUTs) that are targeted by the cytoplasmic 5'-3' Xrn1 (Van Dijk et al., 2011) and the Nrd1 Unterminated Transcripts (NUTs) that accumulate upon nuclear depletion of the RNA-binding factor Nrd1 (Schulz et al., 2013). When the meiotic program is activated, yeast diploid cells express an additional type of antisense lncRNAs, the Meiotic Unannotated Transcripts (MUTs) (Lardenois et al., 2011). Finally, Stable Unannotated Transcripts (SUTs) were found to be detectable in wild-type (WT) cells.

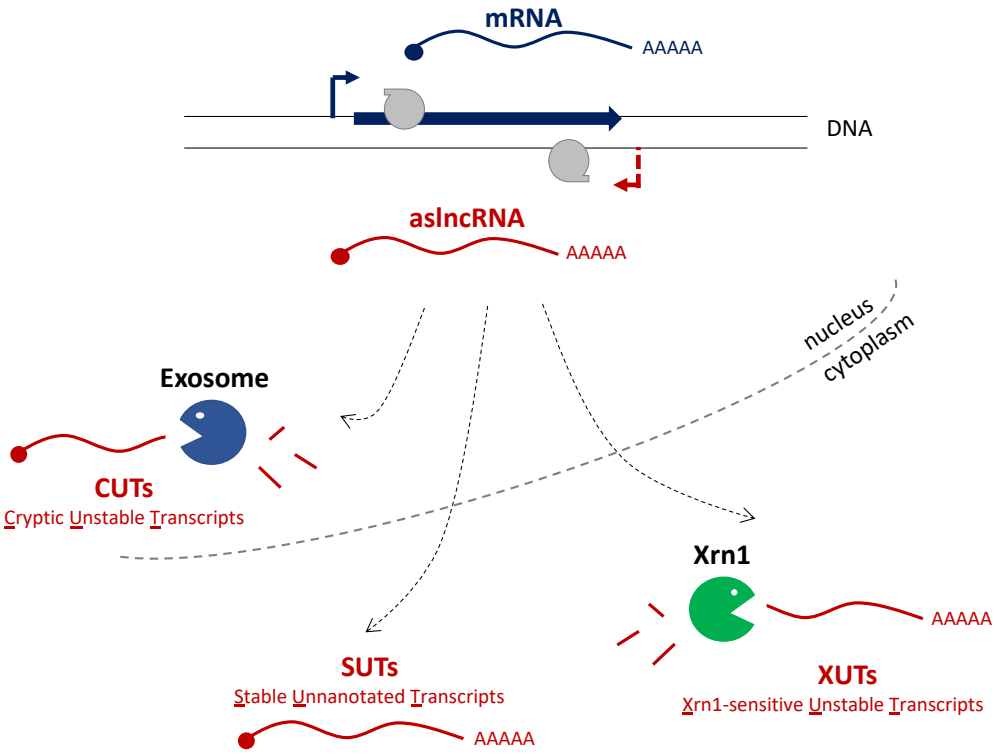


Figure 7 | Schematic Representation of Main Classes of LncRNAs in Yeast. CUTs, SUTs and XUTs are produced by RNAPII, capped and polyadenylated. CUTs are degraded by the nuclear exosome, while XUTs are degraded by the cytoplasmic Xrn1. SUTs are sufficiently stable and readily detectable in WT background. All representations same as above.

CUTs are 200-600 nt long transcripts, produced by the RNAPII, they are capped and polyadenylated by the TRAMP4 complex, consisting of Trf4-Air2 (Tudek et al., 2014). Their instability is due to their mode of transcription termination dependent on the Nrd1-Nab3 pathway (Thiebaut et al., 2006). Nrd1 and Nab3 promote the polyadenylation and subsequent degradation of the transcript by the TRAMP4 and the exosome complexes, respectively. In particular, two studies have reported the first high-resolution genomics maps of CUTs (Neil et al., 2009; Xu et al., 2009). Using a tiling

microarray approach and comparing the transcriptomes of WT yeast growing under diverse conditions with those of mutants lacking Rrp6, Steinmetz and co-workers identified 925 CUTs, corresponding to 13% of identified transcripts (Xu et al., 2009). Instead, Jacquier's group used a 3'-long SAGE approach followed by deep sequencing to compare WT and CUTs-enriched RNA fractions. In such way, they identified and mapped at nucleotide resolution 1,496 CUTs that did not correspond to any annotated feature (Neil et al., 2009). The most abundant fraction of CUTs resulted from cryptic divergent transcription from promoters and nucleosome-free regions. Many CUTs are generated from tandem intergenic regions, mainly in the antisense orientation with respect to the downstream gene. Interestingly, CUTs seem to be the most equivalent to the class of human exosome-sensitive PROMPTs (Preker et al., 2009), therefore suggesting that these unstable transcripts are not unique to yeast.

In the same work mentioned above, Steinmetz's lab identified another class of transcripts, distinguished from the rest of the unstable transcripts. Those transcripts did not correspond to any previously annotated genomic feature and are insensitive to the exosome and detectable in WT cells, hence their definition as stable transcript or SUTs (Xu et al., 2009). SUTs are longer than CUTs, with a median length of 761 nt. They accounted for 12% of the transcripts identified by the tiling microarray and are of unclear function at present. The distinction between CUTs and SUTs is, however, not so clear. Many transcripts defined as CUTs in the second study were classified as SUTs in the first.

Another class of unstable lncRNAs has been discovered as being targeted by the cytoplasmic Xrn1-dependent 5'-3' degradation pathway. The first discovery of such transcript came in 2008, from an intragenic unstable *RTL* ncRNA that is produced in the antisense orientation to the Ty1 mRNA retrotransposon (Berretta et al., 2008). If not destabilized by the cytoplasmic exonuclease Xrn1, the *RTL* ncRNA was found to mediate the transcriptional *trans*-silencing of *TY1* retrotransposon. After this discovery, in 2011, our lab used a single-end strand-specific RNA-Seq in a strain lacking Xrn1 and identified 1,658 Xrn1-sensitive Unstable Non-Coding Transcripts (XUTs) in *S. cerevisiae* (Van Dijk et al., 2011). The majority (66%) of those were antisense to the ORFs and were thus termed antisense (as)XUTs. XUTs were found to be polyadenylated and RNAPII-dependent. They are longer than CUTs as they can reach up to ~750 nt in size. Additionally, this work revealed that XUTs encompassed 75% of SUTs. Few years later, in 2016, our lab discovered using Northern Blot that a XUT can represent a distinct transcript from a stable SUT (Wery et al., 2016). Using CAGE-Seq to precisely map the TSS, the lab revealed that overlapping SUTs and XUTs globally share the same TSS. Instead, the comparison of the annotated 3' coordinates of SUTs and XUTs showed a 3' extension specific to XUTs. In the same work, the analyses of several laboratory strains, by paired-end strand-specific RNA-Seq with high coverage, allowed to define an extended landscape of XUT lncRNAs in *S. cerevisiae*. This refined catalog identified 1,798 XUTs stabilized in *xrn1Δ* mutant cells in at least one *S. cerevisiae*

laboratory strain. Of those XUTs, 1,153 (64%) were antisense to ORFs. In accordance with the decay of mRNAs, to be degraded by Xrn1, XUTs were discovered to undergo decapping by the decapping enzyme Dcp2 (Wery et al., 2016). Currently, our lab is investigating to which extent are XUTs conserved in humans.

Nrd1-Unterminated Transcripts (NUTs), another type of lncRNAs, were discovered in the lab of Cramer, as arising from defective Nrd1-dependent termination of ncRNA transcription (Schulz et al., 2013). These unstable transcripts accumulate upon nuclear depletion of the RNA-binding factor Nrd1.

There is considerable overlap between these classes of unstable lncRNAs. For example, NUTs overlap CUTs, which is not surprising as NUTs were proposed to be extended isoforms of these lncRNAs with defect in termination. Also, SUTs overlap XUTs and some CUTs and XUTs overlap each other. In addition, the majority of the MUTs are sensitive to Rrp6 in mitotic cells, indicating that they might belong to the CUT category. This suggests that there is redundancy among the RNA decay machinery and that the same transcripts might be cooperatively targeted by two distinct RNA surveillance machineries (Marquardt et al., 2011).

4.2. Molecular Functions of Long Non-Coding RNAs in Yeast

LncRNAs in yeast have been involved in a variety of functions, mainly in those regulating gene expression (for review, see (Wery et al., 2011)). The study of gene regulation by aslncRNAs is particularly intriguing, since their genomic position immediately indicates that they may act on their corresponding sense transcripts, through their act of transcription or through the aslncRNA *per se*. In general, the current consensus for CUTs is that the act of cryptic transcription affects gene expression (Jacquier, 2009; Kuehner and Brow, 2008; Martens et al., 2004). One of the first provided mechanisms of non-coding transcriptional interference came from the cryptic *SRG1* transcript. Namely, transcription of *SRG1* prevents the binding of transcription factors to the promoter of the downstream *SER3* stress-responsive gene (Martens et al., 2004). Another example involves the *IMD2* and *URA2* genes, the CUT and the downstream mRNA arise from different TSS but share a common promoter, and possibly compete for the same transcription factors (Kuehner and Brow, 2008). Cryptic antisense transcripts can play roles in the modulation of the expression of the paired sense gene also *via* histone modifications (Berretta et al., 2008; Camblong et al., 2007; Geisler et al., 2012; Houseley et al., 2008; Pinskaya et al., 2009). An example is the *GAL10-GAL1* locus, in which under repressive conditions the lncRNA from the *GAL10* region induces Set2-dependent H3K36 trimethylation (Houseley et al., 2008) and Set1-dependent H3K4 di- and trimethylation (Pinskaya et

al., 2009) across the locus, leading to Rpd3S HDAC recruitment and histone deacetylation finally attenuating *GAL1* induction. Another example of transcriptional interference, functioning after transcription initiation, has been illustrated for the repression of the *IME4* locus by its antisense transcript *RME2* (Gelfand et al., 2011; Hongay et al., 2006). In this case, a 450 bp internal region of the *IME4* gene is required for antisense-mediated repression, which suggests that antisense-mediated transcriptional interference blocks the elongation step of the *IME4* transcript. Moreover, an example of the transcript *per se* regulating gene silencing comes from *PHO84*. Namely, *PHO84* is silenced in *cis* and in *trans* by an aslncRNA that is stabilized in the *rrp6Δ* mutant or in WT aging cells. This stabilization leads to the recruitment of the Hda1 histone deacetylase, histone deacetylation, and *PHO84* transcriptional silencing (Xu et al., 2009). Additional examples, discovered by our lab, showed that stabilization of a subgroup of asXUTs correlates with the transcriptional attenuation of the paired-sense genes suggesting that asXUTs could regulate sense gene expression (Berretta et al., 2008; Van Dijk et al., 2011). The repression mechanism involved the histone methyltransferase (Van Dijk et al., 2011) and histone deacetylase activities (Berretta et al., 2008).

4.3. Long Non-Coding RNAs Conservation in Fission Yeast

After identification of Xrn1-sensitive aslncRNAs in *S. cerevisiae*, our lab also characterized the antisense transcription landscape in fission yeast *Schizosaccharomyces pombe* (Wery et al., 2018a) to understand if the roles of the RNA decay machineries in restricting aslncRNAs levels have been conserved. Note that the budding yeast *S. cerevisiae* and the fission yeast *S. pombe* are as different from each other as either is from animals, their ancestors separated about 420 to 330 million years ago (Sipiczki, 2000). To mention, *S. pombe*, unlike *S. cerevisiae*, shares a functional RNAi machinery with higher Eukaryotes (Volpe et al., 2002) making it an appealing model for studying aslncRNA conservation. Interestingly, using RNA-Seq the lab found out that inactivation of Exo2 (the orthologous gene of Xrn1 (Szankasi and Smith, 1996)) and of Rrp6 lead to the identification of asXUTs (Wery et al., 2018a) and CUTs (Watts et al., 2018), respectively. Consistent with the presence of these RNA species in another work (Atkinson et al., 2018), the results from our lab suggested evolutionary conservation of cryptic aslncRNAs in fission yeast. Similar to the function in *S. cerevisiae*, asXUTs in *S. pombe* have been reported anti-correlating with levels of the paired-sense mRNAs (Wery et al., 2018a), supporting physiological significance to antisense-mediated gene attenuation. These asXUTs were found not to be targets of the RNAi in fission yeast (Wery et al., 2018a) probably as a cause of different locations of Dicer and Exo2, in the nucleus (for review, see (Woolcock and Bühler,

2013)) and cytoplasm, respectively. Apart from these cryptic RNAs, common to *S. cerevisiae*, Dicer-sensitive Unstable Transcripts (DUTs) were identified as a novel class of unstable RNAs in fission yeast, accumulating upon inactivation of Dicer (Atkinson et al., 2018).

Chapter 2. Translation-Dependent mRNA Decay Pathways in Regulating Levels of Long Non-Coding RNAs

Most of XUTs, other than being targeted to Xrn1, have been found to be sensitive to the translation-dependent Nonsense Mediated mRNA Decay (NMD) pathway in budding (Malabat et al., 2015; Wery et al., 2016) and fission yeast (Atkinson et al., 2018), suggesting a conserved role of translation in the metabolism of aslncRNAs. In this chapter, I review the role of NMD and translation in the degradation of cytoplasmic lncRNAs in detail.

In addition to the main RNA decay pathways presented above, specialized cytoplasmic decay pathways act in response to aberrancies in translation to degrade mRNAs.

It is generally taught that such 'aberrant' mRNAs are distinguished from the 'normal' mRNAs due to adaptor proteins that interact with the translation machinery and funnel the aberrant mRNA into a degradation pathway. In these cases, mRNAs can be subjected to either deadenylation independent decapping (Muhlrad and Parker, 1994), rapid 3'-5' degradation (van Hoof et al., 2002), or endonuclease cleavage (Doma and Parker, 2006). Before explaining the biological roles, the specificity in determining the targets and the mechanism of decay for each of these pathways, I present an overview of the general process of mRNA translation.

1. The Mechanism of Translation

After export to the cytoplasm, mRNAs can be translated through a cyclical process by the ribosomes (Figure 8). The ribosome reads the information one codon (three nucleotides) at a time and translate it into a protein through tRNAs that recognize each codon and insert the appropriate amino acid. The process of translation comprises four main phases: initiation, elongation, termination and ribosome recycling.

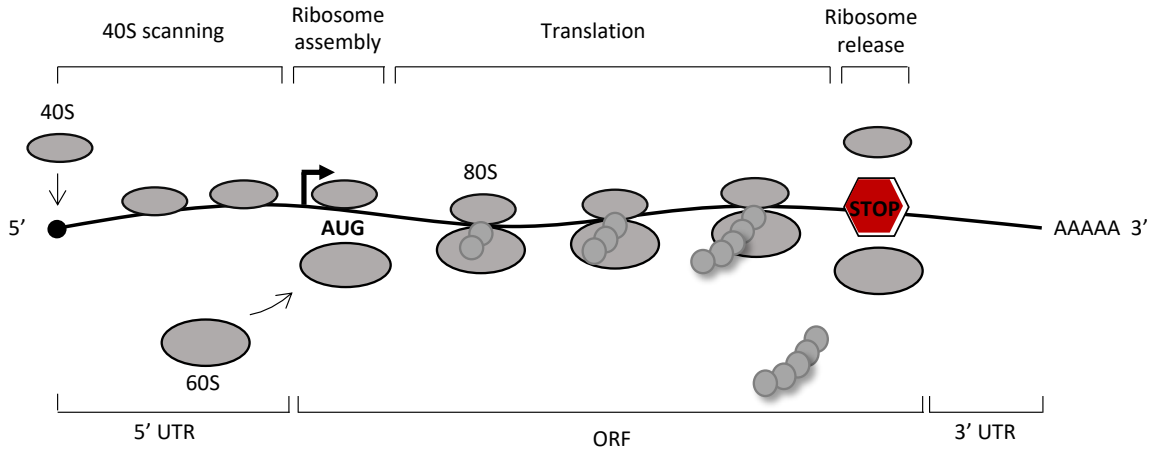


Figure 8 | Schematic Overview of Eukaryotic mRNA Translation. The first step of the initiation phase is the binding of the small ribosomal unit (40S) to the mRNA. The 40S subunit, in association with bound initiator methionyl-tRNA and eukaryotic translation initiation factors (eIFs) (not shown), scans the mRNA to identify the start codon of the ORF. The large ribosomal subunit (60S) then joins the complex, forming a functional ribosome (80S) on which elongation of the polypeptide chain proceeds. When a ribosome encounters a stop codon (STOP), it is actively disassembled and recycled, while the peptide gets released. Bent up black arrow represents AUG, start codon. Adapted from (Guttman et al., 2019).

First, the small subunit of the ribosome (40S in Eukaryotes) assembles on the RNA at the level of its cap. The small subunit is pre-loaded with the initiator transfer RNA (tRNA) that corresponds to the anticodon of methionine AUG (start codon). It is accompanied by numerous pre-initiation factors, including eIF4E, which directly links the m^7-Gppp, stabilizing the formed complex through an interaction with polyadenylate-binding protein (PABPs - Pab1 in yeast/PABPC1 in mammals) (Tarun and Sachs, 1997; Tarun et al., 1997). At this step, the RNA are circularized (Wells et al., 1998), allowing stability of the pre-initiation complex, important to scan the 5' UTR in the 5'-3' direction, in search of the initiator methionine. Although most mRNAs use this scanning mechanism, translation initiation on a few mRNAs is mediated by internal ribosome entry sites (IRES). Such IRES, most known of viral origins, are capable of recruiting the ribosomal complexes to an internal position on the mRNA through a noncanonical mechanism, based on atypical interactions with eIFs and/or 40S subunit, circumventing the scanning process (for review, see (Kieft, 2008)).

Once arrived at the start codon, the large subunit of the ribosome (60S in Eukaryotes) is recruited to form the ribosome (80S in Eukaryotes) which then initiates translation. The control of translation initiation is one of the most fundamental processes in the regulation of gene expression.

In 1978, Kozak proposed that when the AUG codon is in the optimum context of GCCGCC(A/G)CCAUGG (A/G represents A or G and AUG represents the start codon), which is called the 'Kozak consensus sequence', the efficiency of translation initiation is enhanced (for review, see (Kozak, 1978, 2002)). However, despite the sequence is described as a 'consensus' sequence, the extent of its conservation is rather low (Nakagawa et al., 2008). Moreover, translation initiation can occur at non-AUG start codons and/or at AUG in a non-optimal context, resulting in non-canonical translation (for review, see (Andreev et al., 2022)).

After translation initiation, the ribosome enters the elongation phase. At each cycle, a tRNA corresponding to the following codon of the RNA enters the acceptor, aminoacyl site (A site) of the ribosome charged with its amino acid and escorted by the elongation factor eEF1A. The peptide bond between amino acid n-1 and n is carried out within the ribosome at the level of the peptidyl-transferase site (P site), while the deacylated tRNA moves to the ribosomal exit site (E site). The catalysis of this reaction, as well as the action of elongation factor, brings the necessary energy to the ribosome to translocate by three nucleotides, allowing a new n+1 tRNA to position itself at the empty A site.

The ribosome continues translation until reaching the stop codon of RNAs marking the end of the protein-coding sequence. In the classical genetic code, three codons (UAA, UAG, UGA) indicate the end of translation and do not have corresponding tRNA. Instead, the translation termination factors, eRF1 and eRF3 (Eukaryotic Release Factor 1 & 3) are incorporated into the A site of the ribosome, allowing the release of the polypeptide and the initiation of the ribosome recycling.

The termination of the translation is assisted by the PABPs *via* an interaction between eRF1/eRF3 with Pab1. The eRF1/eRF3/GTP complex is recruited to the A site of the ribosome (Mitkevich et al., 2006). eRF1 recognizes the stop codon (Bertram et al., 2000; Blanchet et al., 2015) and together with eRF3 allows the polypeptide, corresponding to the coded protein of the RNA, to detach from the ribosome (Jin et al., 2010). The ribosome remains still attached to the RNA and other mechanisms allow its recycling (for review, see (Hellen, 2018)). Rli1 in yeast/ABCE1 in mammals initiates the separation between the large and the small subunit of the ribosome, allowing the ribosomes to start another cycle of translation on other RNAs.

2. Translation-Dependent mRNA Decay Pathways

The three control mechanisms associated with active translation are the No-Go mRNA Decay (NGD), the Non-Stop mRNA Decay (NSD) and the Nonsense-Mediated mRNA Decay (NMD). They are conserved from yeast to humans. These pathways monitor the activity of translation, that if proceeding smoothly (Figure 9A), will not be activated. However, when elongation is not progressing normally (NSD/NGD) or when translation termination occurs early (NMD) at a premature termination codon (PTC), arising from single nucleotide mutations that convert a canonical triplet nucleotide codon into one of three stop codons, these pathways accelerate the decay of these mRNAs. Once the aberrant mRNAs are detected, these pathways abort translation and have an effect on the decay of the aberrant RNA and on the newly produced proteins which will also be eliminated by quality controls of the associated proteins for NGD and NSD. For NMD, there are no clear results showing specific degradation of newly synthesized truncated proteins.

If an mRNA is devoid of a stop codon (for instance, in the case of truncation, premature 3'-end cleavage and polyadenylation or readthrough of stop codons), it will cause the ribosome to progress to its 3' extremity and stall (Figure 9B). Thus, the ribosome will be unable to terminate translation correctly. Such aberrant mRNAs are rapidly degraded through a process termed NSD (van Hoof et al., 2002; for review, see (Frischmeyer et al., 2002; Klauer and van Hoof, 2012)). NSD requires for the initial cleavage event Dom34 in yeast/Pelota in mammals and Hbs1, paralogs of the translation termination factors eRF1 and eRF3 (Davis and Engebrecht, 1998; Inagaki et al., 2003), and the exosome with the Ski proteins, to rapidly degrade the mRNA in a 3'-5' direction or Xrn1 for 5'-3' degradation.

In contrast, the presence of stable structures such as strong stem loops or damaged nucleotides within an ORF can impede ribosome progression, resulting into ribosome stalling upstream of the stop codon (Figure 9C). In this case, the transcript is targeted to the degradation by the NGD pathway at the translation elongation phase (Doma and Parker, 2006; for review, see (Harigaya and Parker, 2010)). The degradation is triggered by an endonucleolytic cleavage of mRNA near the blocking site (Doma and Parker, 2006). It has been recently reported that Cue2 is the endonuclease that cleaves the targeted RNAs for NGD (D'Orazio et al., 2019). The cut generates a 3' fragment rapidly degraded by Xrn1 and a 5' fragment with the blocked ribosome. At some translation stalls, NGD is promoted by the Dom34 and Hbs1 proteins (Doma and Parker, 2006). The interaction of Dom34 and Hbs1 at the A site of the ribosome allows to recruit Rli1 for the ribosome dissociation. The ribosome-free 5' fragment is then degraded by the cytoplasmic exosome.

In both pathways, the nascent peptide remains attached to the large subunit of the ribosome after ribosome dissociation. This subunit is then recognized by a complex of proteins called Ribosome-based Quality Control (RQC) allowing the extraction of the nascent peptide and the final recycling of the subunit. The nascent released peptide is then subjected to ubiquitin-proteasome-mediated degradation (Defenouillère et al., 2017; for review, see (Defenouillère and Fromont-Racine, 2017)).

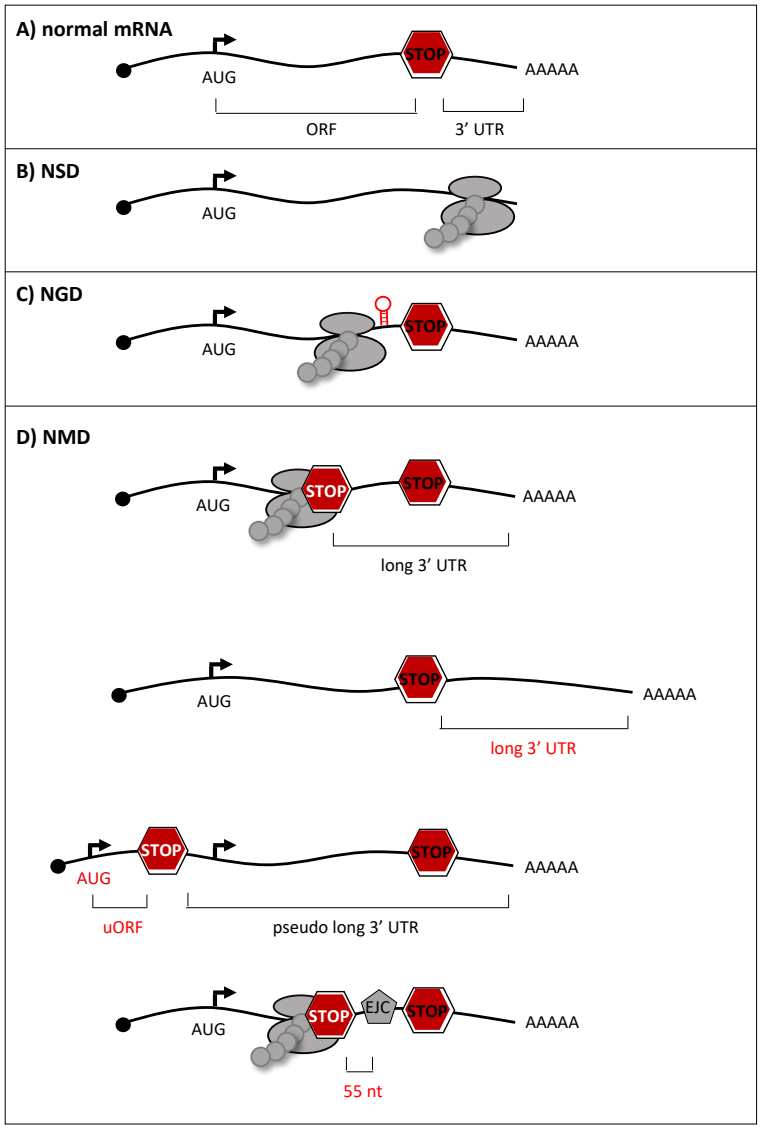


Figure 9 | Schematic Representation of the Inducing Features of Translation-Dependent mRNA Decay Pathways. (A) 'Normal' mRNA. (B) NSD is activated in absence of stop codon causing blockage of ribosome at the end of transcript. (C) NGD is triggered by ribosome block at the level of a blocking sequence (symbol in red). (D) NMD is activated by EJC-independent (premature translation termination at the PTC (STOP written in white), long 3' UTR (or pseudo 3' UTR) or by EJC-dependent pathway (presence of the EJC 55 nt downstream of PTC). uORF, upstream ORF; UTR, untranslated region. Other representations as above.

3. The Nonsense-Mediated mRNA Decay (NMD) Pathway

Note: parts of the following subchapter were adapted from the review (Andjus et al., 2021; see ANNEX II)

The NMD is another quality control pathway targeting transcripts that terminate translation prematurely (for review, see (Behm-Ansmant et al., 2007; Chang et al., 2007)), such as mRNAs harboring a premature termination codon (PTC) within the ORF (Muhlrad and Parker, 1994), as well as PTC-less mRNAs displaying short upstream (u)ORFs (Celik et al., 2017; Oliveira and McCarthy, 1995; Yepiskoposyan et al., 2011) or long 3' untranslated regions (UTRs) (Muhlrad and Parker, 1999; Yepiskoposyan et al., 2011) (Figure 9D). NMD also eliminates mRNAs with a PTC located 55 nt upstream of the last exon–exon junction. The NMD-targeted mRNAs are rapidly degraded (for review, see (Nicholson and Mühlemann, 2010; Rebbapragada and Lykke-Andersen, 2009)), thus preventing the production of truncated, possibly deleterious proteins (Hall and Thein, 1994; Inoue et al., 2004; Pulak and Anderson, 1993).

3.1. Discovery, Conservation and Functions of NMD

As previously mentioned, NMD is a translation-dependent RNA decay pathway (Carter et al., 1995; Thermann et al., 1998; Zhang et al., 1997), which has been evolutionarily conserved (Causier et al., 2017; for review, see (Behm-Ansmant et al., 2007)). It was originally discovered in *S. cerevisiae* by Losson and Lacroute, when they observed that the presence of nonsense mutations reduces the level of a mutant mRNA without affecting its synthesis rate (Losson and Lacroute, 1979). It was uncovered afterwards in humans in the context of $ß^0$-thalassemia, where it was observed that ß-globin mRNAs levels dramatically decrease when carrying nonsense mutations (Baserga and Benz, 1988; Maquat et al., 1981).

UPstream Frameshift proteins (Upfs) 1, 2 and 3 constitute the conserved core components of NMD (Culbertson et al., 1980) and were initially identified in *S. cerevisiae* (Cui et al., 1995; Leeds et al., 1991, 1992). Upf1 is a monomeric, highly regulated superfamily 1 helicase. Its ATPase and helicase activities are essential for NMD (Malabat et al., 2015; Weng et al., 1996). Upf1 has the ability to translocate slowly but with high processivity on nucleic acids and to unwind long dsRNA structures as shown *in vitro* (Fiorini et al., 2015). Upf2 is the second core NMD factor and functions as a bridge between Upf1 and Upf3 (Chamieh et al., 2008; Lykke-Andersen et al., 2000; Melero et al., 2012). Its interaction with Upf1 is a prerequisite for the phosphorylation of Upf1 (Chakrabarti et al., 2011).

However, NMD can be activated independently of Upf2 (Aznarez et al., 2018; Gehring et al., 2005). Upf3 is the least conserved of the three core NMD factors (Culbertson and Leeds, 2003). Vertebrates have two Upf3 paralogs, Upf3A and Upf3B (Lykke-Andersen et al., 2000). In human cells, Upf3B seems to be the main contributor to NMD (Kunz et al., 2006). Like Upf2, Upf3 stimulates the ATPase and helicase activity of Upf1 *in vitro* (Chamieh et al., 2008).

In metazoans, NMD requires four additional factors: Smg1, Smg5, Smg6, and Smg7 (Cali et al., 1999; Pulak and Anderson, 1993; Yamashita et al., 2001; for review, see (Behm-Ansmant et al., 2007)).

Interestingly, there is a correlation between the organism complexity and the dependency on NMD. While Upf1 is essential in *Arabidopsis*, *Drosophila* and vertebrates (Medghalchi et al., 2001; Metzstein and Krasnow, 2006; Wittkopp et al., 2009; Yoine et al., 2006), NMD-deficient mutants in yeast and *C. elegans* are viable (Cui et al., 1995; Hodgkin et al., 1989; Leeds et al., 1991; Pulak and Anderson, 1993).

The initial evidence described that NMD mechanism functions primarily on aberrant mRNAs, although it has become clear that the mRNA quality control represents only one face of the multiple functions of NMD (for review, see (Kurosaki et al., 2019; Lykke-Andersen and Jensen, 2015; Nasif et al., 2018; Peccarelli and Kebaara, 2014; Smith and Baker, 2015)). In yeast, almost half of protein-coding genes can generate NMD-sensitive mRNA isoforms, including truncated mRNAs for which transcription initiation occurs downstream of the canonical translation initiation site (Malabat et al., 2015). NMD also targets intron-containing pre-mRNAs that have escaped splicing and were exported to the cytoplasm (Celik et al., 2017). In addition, NMD regulates 3–10% of physiological, nonmutated mRNAs in yeast, *Drosophila* and humans, including mRNAs with small uORFs (Arribere and Gilbert, 2013; Celik et al., 2017; Gaba et al., 2005; Guan et al., 2006; He et al., 2003; Johansson et al., 2007; Oliveira and McCarthy, 1995; Ruiz-Echevarría and Peltz, 2000; Yepiskoposyan et al., 2011), long 3'-UTRs (Muhlrad and Parker, 1999; Yepiskoposyan et al., 2011), as well as mRNAs displaying low translational efficiency and average codon optimality (Celik et al., 2017). Therefore, considered together, NMD provides a significant contribution to the post-transcriptional regulation of gene expression (Smith and Baker, 2015), particularly for those genes involved in cell-surface dynamics and chromosome structure (Guan et al., 2006; He et al., 2003; Lelivelt and Culbertson, 1999).

Numerous physiological processes rely on the capacity of the cell to adjust NMD activity at global and/or transcript specific levels. NMD factors are essential for embryonic development in vertebrates, as disrupted expression of core NMD factors confers lethality at an early embryonic stage (Medghalchi et al., 2001; Weischenfeldt et al., 2008). NMD is also crucial for the maintenance of hematopoietic stem and progenitor cells (Weischenfeldt et al., 2008), the maturation of T cells (Weischenfeldt et al., 2008), as well as for liver development, function and regeneration in mice

(Thoren et al., 2010). Furthermore, NMD is important for the response to multiple stresses (Karam et al. 2015; Rodríguez-Gabriel et al. 2006; for review, see (Goetz and Wilkinson 2017)), being itself regulated in response to stresses such as hypoxia (Gardner, 2008), and amino acid deprivation (Mendell et al., 2004). In fact, many stress-related mRNAs are targeted by NMD under normal physiological conditions but are stabilized upon stress, due to the inhibition of NMD activity (Usuki et al., 2019). Interestingly, a recent study focused on an alternative mammalian-specific Upf1 isoform, termed $Upf1_{LL}$, which activity is enhanced in response to cellular stress, conditionally altering specific mRNAs (Fritz et al., 2022).

Importantly, NMD appears to be an emerging modulator of neural development (for review, see (Jaffrey and Wilkinson, 2018)) and malignancy (for review, see (Tan et al., 2022)). On one hand, many truncated proteins encoded from NMD targets have been found to be in a stable form in NMD-deficient conditions, and to be involved in cancer development (Liu et al., 2014; Wang et al., 2011). On the other hand, NMD can also downregulate the expression of tumour suppressor genes, suggesting that NMD can benefit tumours, a notion further supported by the finding that mRNAs encoding immunogenic neoantigen peptides are typically targeted for decay by NMD (Pastor et al., 2010). Consequently, therapies modulating the efficiency of NMD have the potential to provide widespread therapeutic benefit against diverse tumour types.

3.2. NMD Factors Act in Concert to Activate mRNA Decay

In many organisms, NMD has been coupled to pre-mRNA splicing (Kerényi et al., 2008; Le Hir et al., 2001; Sun and Maquat, 2000; Sun et al., 2000; Thermann et al., 1998; Zhang et al., 1998). The Exon Junction Complex (EJC) is deposited by the spliceosome at the level of the junction between two exons (Le Hir et al., 2000), and it is normally removed from the coding regions by the translating ribosomes (Dostie and Dreyfuss, 2002). The EJC is formed around four core components: the DEAD-box RNA helicase eIF4A3, MLN51, and the Magoh/Y14 heterodimer (Mabin et al., 2018). The presence of an EJC downstream of a stop codon is recognized as an abnormal situation and enhances the association and activity of Upf1 (Kurosaki and Maquat, 2013). In the EJC-enhanced NMD model, premature translation termination involves the SURF (Smg1–Upf1–eRF1–eRF3) complex, which consists of the Smg1 kinase, Upf1 and the Eukaryotic Termination Factors eRF1 and eRF3, and associates with the ribosome stalled at the PTC (Kashima et al., 2006) (Figure 10A). Upf2 and Upf3 are then recruited to SURF via the proximal EJC, leading to the formation of the DECID (DECayInDucing) complex (Kashima et al., 2006). In particular, it is taught that Upf3 directly interacts with the EJC complex and bridge it to the Upf complex. However, recent studies in mammalian cells demonstrate that Upf3 proteins can also activate NMD independently of their EJC binding (Yi et al.,

2022). The interaction between Upf1 and Upf2 induces a conformational change of Upf1, allowing its phosphorylation by Smg1 and its activation (Kashima et al., 2006). The activated Upf1 recruits the Smg6 endonuclease (Eberle et al., 2009) and the Smg5–7 heterodimer (Ohnishi et al., 2003), which in turn activates RNA deadenylation and decapping. In addition, phosphorylated Upf1 also prevents new translation initiation events by interacting with the translation initiation factor eIF3, inhibiting the formation of a competent translation initiation complex (Isken et al., 2008). Finally, protein phosphatase 2 (PP2A) dephosphorylates Upf1, allowing it to return to its unphosphorylated state for another NMD cycle (Ohnishi et al., 2003).

In addition to the EJC-enhanced NMD, examples of EJC-independent NMD have been described in human cells (Bühler et al., 2006; Metze et al., 2013), as well as in fission yeast (Wen and Brogna, 2010), *C. elegans* (Longman et al., 2007), *Drosophila* (Gatfield et al., 2003) and plants (Kerényi et al., 2008), all of which have orthologs of EJC factors. In contrast, in *S. cerevisiae*, not only is the proportion of intron-containing genes low (4%) (Goffeau et al., 1996), but EJC factors are absent, with the exception of eIF4A3 (Fal1), which acts in pre-rRNA processing in yeast (Alexandrov et al., 2011).

The EJC-independent NMD targets RNAs with extended 3' UTR but lacking EJC downstream of the translation termination codon (Bühler et al., 2006; Eberle et al., 2008; Matsuda et al., 2007; Singh et al., 2008; Wang et al., 2002a; Zhang et al., 1998). Indeed, RNAs, where long EJC-free sequences are inserted downstream of a stop codon, show reduced levels due to accelerated degradation by NMD (Eberle et al., 2008; Singh et al., 2008). This EJC-independent NMD might be a vestige of an ancestral NMD mechanism associated with an abnormally long 3' UTR, referred to as "faux 3' UTR" (Figure 10B), which is still present in *S. cerevisiae* (Amrani et al., 2004). In this model, an altered interaction between the polyadenylate-binding protein Pab1 and the prematurely terminating ribosome results in less efficient termination and enhanced interaction between Upf1 and eRF1/eRF3, triggering NMD. In this context, a recent proteomics-based analysis in yeast characterized the composition of two distinct NMD complexes associated with Upf1 named Upf1-23 (Upf1, Upf2, Upf3) and Upf1-decapping (Dehecq et al., 2018). The latter contained the decapping enzyme Dcp2 and its co-factor Dcp1, the decapping activator Ebs3, and two poorly characterized proteins, Nmd4 and Ebs1. The Upf1-23 complex is recruited and assembled on the RNA substrate, and then a reorganization occurs, with Nmd4, Ebs1, Dcp2 and its co-factors replacing the Upf2/3 heterodimer. Nmd4 and Ebs1 are NMD accessory factors that may be functional homologues of human Smg6 and Smg5/7, respectively (Dehecq et al., 2018; Luke et al., 2007). The identification of these new yeast factors suggests that NMD mechanisms may be more conserved than previously thought.

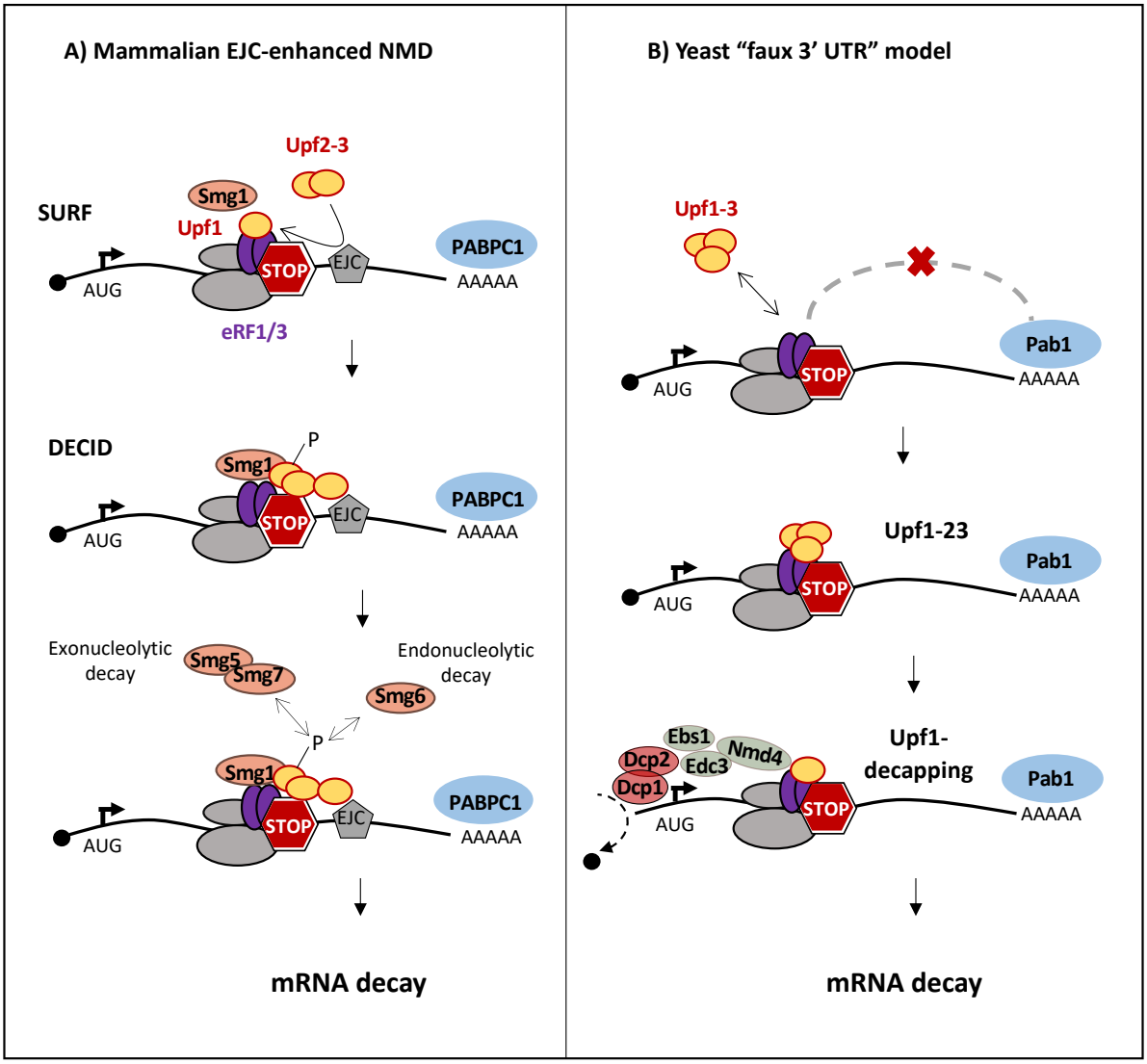


Figure 10 | Models of NMD Activation in Mammals and Yeast. (A) Mammalian EJC-enhanced NMD. EJC, bound downstream of PTC (STOP) interferes with the interaction between PABPC1 and eRF1/eRF3. SURF complex forms. Upf2 and Upf3 get recruited by the EJC and associate with SURF to form the DECID complex. Smg1 phosphorylates Upf1, activating it. Phosphorylated Upf1 promotes RNA decay via Smg6-dependent endonucleolytic cleavage and Smg5–Smg7-dependent triggering of mRNA deadenylation and decapping to further foster mRNA decay. (B) "Faux 3' UTR" model in yeast. A long 3' UTR causes inefficient translation termination and Upf1 interaction with eRF1/eRF3, promoting the formation of the Upf1-23 complex (Upf1, Upf2, Upf3) at the terminating ribosome. Upf2 and Upf3 get replaced by Nmd4, Ebs1, and decapping enhancing factors forming Upf1-decapping complex, leading to RNA decapping and final degradation of the mRNA. Adapted from (Andjus et al., 2021).

The polyadenylate-binding protein 1, Pab1/PABPC1, as previously mentioned, is known to stimulate translation termination efficiency by recruiting the release factors to the ribosome (Ivanov et al., 2016). A long distance between the PTC and Pab1/PABPC1 triggers NMD in all studied species (Bühler et al., 2006; Eberle et al., 2008; Silva et al., 2008; Singh et al., 2008), while tethering it close to the PTC suppresses the NMD sensitivity of the PTC-containing transcripts in yeast (Amrani et al., 2004) and *Drosophila* cells (Behm-Ansmant et al., 2007). Mechanistically, it has been proposed that the long 3' UTR would impede the efficient interaction between Pab1/PABPC1 and eRF1/eRF3, favoring the recruitment of Upf1 by the latter and the formation of a SURF complex at the level of the PTC.

Currently, there are several questions regarding NMD that are not completely understood. They include when and how is Upf1 recruited to its targets and what exactly are the consequences of this recruitment. Also, how the NMD machinery differentiates premature translation termination from normal translation termination on both nonsense mutation-bearing and natural mRNAs remains to be clarified. In the review, in which we focus on the evolutionary role of NMD in the turnover of RNAs, we discuss some of these issues in further depth (Andjus et al., 2021; see ANNEX II).

3.3. Pervasively Translated Long Non-Coding RNA Emerge As Unexpected Class of NMD Substrates

In this subchapter, first I address the notion that lncRNAs have been discovered to associate with the translation machinery and after why consequently some lncRNAs are subjected to the NMD pathway unless engaged in dsRNA formation.

Prokaryotic and Eukaryotic genomes were shown not only to be pervasively transcribed but also to be pervasively translated (Ingolia et al., 2014; Smith et al., 2022). In fact, despite being *a priori* presumed devoid of coding potential (Guttman et al., 2013; Verheggen et al., 2017), some non-coding regions of the genome have been found to be associated to the translation machinery. Given the similar structure for stimulating translation of mRNAs and lncRNAs (*i.e.* cap and polyA tail), and considering the important number of lncRNAs localized in the cytoplasm, it came as no surprise that ribosomes can be a default destination of some lncRNAs. Indeed, many lncRNAs are translated at some level, under at least some condition, or in a particular tissue. This notion started challenging the coding (dis)ability of lncRNAs.

To examine their coding potential, a variety of methods have been proposed (for review, see (Choi et al., 2019; Makarewich and Olson, 2017)). The experimental data test for individual transcripts whether it can yield peptides when translated *in vitro*, whether it associates with polysomes, and/or if its ORFs can produce a protein when fused to a sequence coding for a peptide for which antibodies are available. Other techniques globally investigate translation, such as Ribosome Profiling (Ribo-Seq), a technique utilizing high-throughput sequencing to map RNA regions associated with translating ribosomes (Ingolia et al., 2011), thus, providing a 'snapshot' of all active ribosomes in a cell at a specific time point.

Analysis of Ribo-Seq showed that transcripts produced from non-coding regions of the genome, including intergenic regions and sequences antisense to protein-coding genes, associate with the translation machinery in different models, including *S. cerevisiae* (Brar et al., 2012; Carvunis et al., 2012; Smith et al., 2014; Wery et al., 2016; Wilson and Masel, 2011), fission yeast (Atkinson et al., 2018; Duncan and Mata, 2014), plant (Ruiz-Orera et al., 2014), *Drosophila* (Aspden et al., 2014; Ruiz-Orera et al., 2014), zebrafish (Bazzini et al., 2014; Chew et al., 2013; Ruiz-Orera et al., 2014), mouse (Ingolia et al., 2011; Ruiz-Orera et al., 2014) and human cells (Bazzini et al., 2014; Carlevaro-Fita et al., 2016; Chen et al., 2020; Chothani et al., 2022; Ruiz-Orera et al., 2014; Van Heesch et al., 2014). For instance, Ribo-Seq in mouse embryonic stem cells suggested that as many as half of the lncRNAs expressed in these cells are significantly associated with ribosomes (Ingolia et al., 2011).

Furthermore, a growing body of experimental data indicate that not only lncRNAs but also circular RNAs (Legnini et al., 2017; Yang et al., 2018; Zhang et al., 2018), primary microRNAs transcripts (Dozier et al., 2022; Lauressergues et al., 2015, 2022; Montigny et al., 2021) and transposons (Bonté et al., 2022) can infact be translated.

Non-coding RNAs can contain one or more small (sm)ORFs previously missed since the traditional gene annotation process filtered out ORFs shorter than 100 codons, considering them as noise or false positives. However, as ribosome profiling techniques and proteomics increase in sensitivity and accuracy, it is becoming obvious that at least a fraction of ribosome-bound lncRNA sequences represent genuine small (sm)ORFs, the translation of which can produce functional peptides (Chen et al., 2020; van Heesch et al., 2019; Matsumoto et al., 2017; Prensner et al., 2021; Slavoff et al., 2013; Sun et al., 2021; Wei and Guo, 2020).

In 2020, Weissman's lab published a first catalog of smORFs and functional peptides derived from human lncRNAs, which included the discovery of nearly a thousand of novel lncRNA-associated smORFs. For some of them, CRISPR-mediated knockout of the smORF resulted in a growth phenotype (Chen et al., 2020), indicating that the corresponding peptides are important for cell growth.

Additionally, other studies documented that lncRNA-derived micropeptides are involved in the regulation of RNA decapping (D'Lima et al., 2017), in embryonic development (Kondo et al., 2010;

Zanet et al., 2015), in muscle development (Bi et al., 2017; Lin et al., 2019; Zhang et al., 2018), regeneration (Bi et al., 2017; Matsumoto et al., 2017) or contraction (Magny et al., 2013; Makarewich et al., 2018; Nelson et al., 2016), as well as in tumour development (Huang et al., 2021; Polycarpou-Schwarz et al., 2018; Prensner et al., 2021; Sun et al., 2021) (see Table 1). A recent study provided a list of noncanonical ORF encoding functional proteins essential for cancer cell survival (Prensner et al., 2021). The authors focused on one of them, *GREP1* (glycine-rich extracellular protein-1) and found that it encodes a protein highly expressed in breast cancer proposing it as a potential therapeutic target.

Micropeptide	Species	Target	Function(s)
NoBoDy	Human	mRNA decapping factors	Regulation of mRNA turnover and P-body numbers
CASIMO1	Human	Squalene epoxidase	Carcinogenesis; cell lipid homeostasis
PINT87aa	Human	Polymerase associated factor complex (PAF1c)	Oncogene transcriptional inhibition; tumor suppressive effect
HOXB-AS3	Human	hnRNP A1 splicing factor	Colon cancer growth suppression
RBRP	Human	m^6A reader IGF2BP1	Regulation of m^6A recognition by IGF2BP1 on c-Myc mRNA; tumorigenesis
Minion/Myomixer	Human, mouse	Unknown	Myoblast fusion; muscle formation and development
SPAR	Human, mouse	Lysosomal v-ATPase	Regulation of mTORC1 signaling pathway; muscle regeneration
TUG1-BOAT	Human, mouse	Unknown	Unknown; alters mitochondrial membrane potential when overexpressed
Mtln	Human, mouse	Cardiolipin	Increase of mitochondrial functions
DWORF	Mouse	SERCA	SERCA (sarcoplasmic reticulum Ca^{2+}-ATPase) activation
MLN	Mouse	SERCA	SERCA inhibition
Toddler	Zebrafish	Unknown	Promoting cell migration during embryogenesis
Pri	*Drosophila*	Ubr3 E3 ubiquitin ligase	Proteasome-dependent processing of the developmental Svb transcription factor
Scl	*Drosophila*	Ca-P60A SERCA	Calcium transport regulation

Table 1 | Examples of Functional LncRNA-Derived Micropeptides. Reproduced from (Andjus et al., 2021).

Even if for most of the lncRNA-derived peptides the mechanistic bases are still missing, pioneer studies revealed that peptides can act by binding other proteins and regulating their activity (Anderson et al., 2015; Makarewich et al., 2018; Nelson et al., 2016; Zanet et al., 2015), or as signaling pathway molecules (Pauli et al., 2014). Future works will reveal their additional modes of action.

If found to be beneficial, lncRNA-derived micropeptides may also act as templates for natural selection and birth of new peptides *via* the evolutionary process known as *de novo* gene birth, see discussion (Blevins et al., 2021; Carvunis et al., 2012; Papadopoulos et al., 2021; Schmitz et al., 2018;

Zhao et al., 2014; for review, see (McLysaght and Hurst, 2016; Van Oss and Carvunis, 2019; Parikh et al., 2022)).

Consistent with the observation that lncRNAs actively engage into translation, tilling-array in plants and RNA-Seq studies in yeast using mutants of the NMD pathway revealed lncRNAs as a novel category of cellular RNAs regulated by NMD (Kurihara et al., 2009; Smith et al., 2014). Moreover, NMD inactivation in mouse embryonic stem cells resulted in the stabilization of a subset of lncRNAs (Smith et al., 2014). Further support for NMD-sensitive lncRNAs came from one of the best studied lncRNAs, the Growth Arrest-Specific 5 (*GAS5*) lncRNA. *GAS5* is required for normal growth arrest, it slows down the cell cycle (Mourtada-Maarabouni et al., 2008) and is therefore reduced in numerous cancers (Sun et al., 2014). In actively dividing cells, *GAS5* levels are kept low by the NMD (Mourtada-Maarabouni and Williams, 2013). However, upon exposure to stress such as nutrient deprivation, translation is inhibited, and *GAS5* levels increase (Fleming et al., 1998). Under these conditions, *GAS5* exerts its negative effects on cell proliferation and survival. Other recent transcriptome-wide analyses of RNA binding sites of human, mouse and yeast cells revealed that, in addition to mRNAs, Upf1 can also bind lncRNAs (Colombo et al., 2017; Hurt et al., 2013; Sohrabi-Jahromi et al., 2019; Zünd et al., 2013). Mühlermann's lab used UV crosslinking and immunoprecipitation (CLIP) to identify protein–RNA crosslink sites genome-wide and found a fraction of Upf1 reads mapping to lncRNAs. This suggested that Upf1 can bind any physically accessible transcript in a promiscuous way (Zünd et al., 2013).

In yeast, the notion that cytoplasmic lncRNAs are also translated is manly supported by their sensitivity to NMD. In our lab, we investigate the post-transcriptional regulation of yeast aslncRNAs, specifically cytoplasmic XUTs. Our lab and others discovered that the extensively degraded cytoplasmic XUTs are targeted to Xrn1 *via* NMD (Malabat et al., 2015; Wery et al., 2016). Namely, it was observed at the steady-state level that 73% of XUTs were sensitive to NMD mutants (Wery et al., 2016). Considering that NMD is a process requiring active translation, this result indirectly implied that most XUTs should be translated, and that translation constitutes a prerequisite for their degradation. In fact, analysis of the pioneer Ribo-Seq data (Smith et al., 2014), available at the time, detected for NMD-sensitive XUTs ribosome footprints restricted to their 5' regions, followed by long ribosome-free regions downstream of stop codon (Wery et al., 2016). Given that such long 3' UTR regions represent one of the features targeting mRNAs to NMD, it raised the question of whether it is also the signal for lncRNAs.

On the other hand, genome-wide mapping of dsRNA using small sequencing, in strains with reconstituted RNAi factors from the RNAi-capable *Naumovozyma castellii* yeast, from our lab showed that globally asXUTs form dsRNA structures *in vivo* (Wery et al., 2016). Indeed, small RNA production was present for >80% of asXUTs in Xrn1-lacking cells, deriving from dsRNAs, formed by the asXUTs

paired with their sense mRNAs, at least in some cells. The question of mRNA/aslncRNA co-expression and pairing in *S. cerevisiae* has been a matter of debate over the last years. The same strategy our lab used, was used by others, concluding that a large proportion of mRNAs in yeast are engaged in dsRNA formation *in vivo* (Drinnenberg et al., 2011), including 3'-overlapping mRNAs pairs produced from convergent genes (Sinturel et al., 2015). Other studies using single molecular fluorescent *in situ* hybridization (smFISH) showed that the *PHO84* mRNA and its paired aslncRNAs are expressed in a bimodal manner, never coexisting in the same cell (Castelnuovo et al., 2013). However, for that locus the data from the lab show a massive production of small RNAs from the *PHO84* locus, restricted to the region of overlap between the mRNA and the aslncRNAs, strongly indicating that the two transcripts interact and are therefore co-expressed. More recent studies, also based on single-molecule imaging, confirmed that although transcribed in a bimodal manner, sense and antisense RNAs can coexist within the same cell (Lenstra et al., 2015; Nguyen et al., 2014). However, if RNA:RNA duplexes exist, how heterogeneous their population is remains unknown.

Importantly, formation of dsRNAs, at least in some cells, protected asXUTs from the NMD (Wery et al., 2016), indicating that dsRNAs are a key determinant of the asXUT sensitivity to NMD. In fact, NMD-sensitivity of several intergenic solo XUTs (*i.e.* those without overlap with mRNA) was lost when anti-complementary transcripts were expressed in *trans.* Finally, two RNA helicases, Mtr4 and Dbp2, specifically contributed to the decay of an NMD-sensitive XUT (Wery et al., 2016), postulating that they might help to unwind the dsRNA structures and release the engaged asXUTs as solo, single-stranded transcripts. Then, the single-stranded asXUT could be bound by ribosomes and thus targeted to the NMD.

Together all this data led to a model in which unless blocked by dsRNA structures, that can be dissociated by RNA helicases, ribosomes could rapidly bind smORFs in the 5' region of cytoplasmic lncRNAs (Figure 11). During the pioneer round of translation, the detection of the long 3' UTR region would probably activate the NMD, leading to decapping and final degradation of the XUT by Xrn1. In contrast, the asXUTs fully engaged in RNA duplex would be deprived of translation, protecting the interacting partners from the NMD. However, the extent and the regulatory mechanisms determining the fate of the lncRNAs either subjected to translation/decay or pairing/stabilization are unknown.

One of the reasons that limit our understanding of the metabolism of aslncRNAs is that the major current knowledge on aslncRNAs has been obtained through studies at the level of populations of cells. Even within seemingly homogenous populations of isogenic yeast cells, there is a high degree of heterogeneity that originates from a compact and pervasively transcribed genome. In a population of cells, one could imagine the presence of several subpopulations in terms of expression of each mRNA/aslncRNA pair, generating different configuration for the translation ability and NMD-

sensitivity complicating our understanding of the metabolism of lncRNAs. To tackle the heterogeneity of these subpopulations and get a more precise view of the metabolism of aslncRNAs in yeast, we necessitate going to the unique cell resolution. Implementing single-cell strategies coupled with total RNA measurements is required for defining the non-coding transcriptome heterogeneity within a cell population.

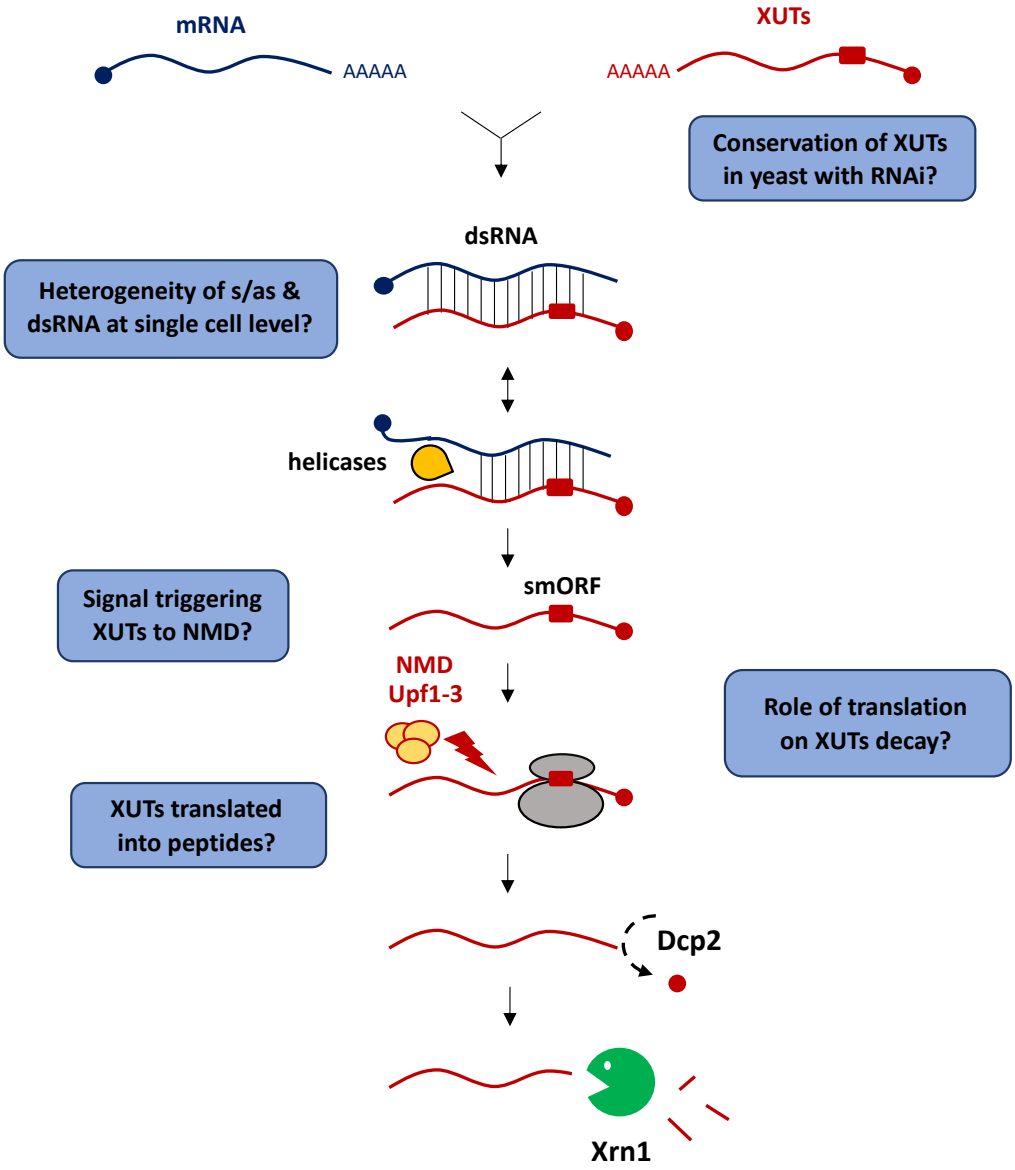


Figure 11 | Working Model. Xrn1-sensitive aslncRNAs (XUTs) once in the cytoplasm could bind with their paired-sense mRNAs to form dsRNAs, possibly dissociated by RNA helicases. Small (sm)ORFs of the released solo XUTs would be rapidly bound by ribosomes for a pioneer round of translation. Subsequently, NMD would be triggered, leading to XUTs decapping by Dcp2, and degradation by Xrn1. In blue boxes are written the main aims of my PhD. Adapted from (Wery et al. 2016).

OBJECTIVES

OBJECTIVES

My thesis took advantage of cryptic aslncRNAs as a paradigm to investigate, at population and also single-cell level, the causes and heterogeneity of RNA (in)stability in a simple Eukaryotic model across its evolution.

Most aslncRNAs are extensively degraded by the nuclear exosome (Neil et al., 2009; Xu et al., 2009) and by the cytoplasmic NMD to the exoribonuclease Xrn1 (Malabat et al., 2015; Van Dijk et al., 2011; Wery et al., 2016) in the budding yeast *S. cerevisiae*, lacking RNAi. On the other hand, aslncRNAs form dsRNA, protecting them from the NMD (Wery et al., 2016). Whether the role of NMD in degrading aslncRNAs is conserved in RNAi-capable relatives, in which dsRNAs are degraded by default, remained unknown. The first task of my project was to set the basis for addressing this question in *N. castellii*, a budding yeast endowed with a cytoplasmic RNAi, by characterizing its aslncRNAs landscape. To that purpose, we performed genome-wide RNA profiling in the mutants of Dcr1, Xrn1, and Rrp6. We described how the different decay machineries and surveillance pathways defined the aslncRNAs landscape and compared it to *S. cerevisiae*. In particular, I showed that Dicer processes aslncRNAs in 22-23 nt small RNAs with a preference for Xrn1-sensitive aslncRNAs. Consequently, I localized Dicer in the cytoplasm. The results of this work, that was a side project at the beginning of my PhD, are presented in form of published (Szachnowski, Andjus et al., 2019) and unpublished work, in the first part of the results section.

In *S. cerevisiae*, the NMD-sensitivity of most Xrn1-sensitive aslncRNAs suggested that aslncRNAs are translated, raising the question of their coding potential. Supporting this idea, a pioneer Ribo-Seq analysis identified putative smORFs bound by ribosomes on a limited number of lncRNAs, providing the proof-of-concept that transcripts annotated as non-coding yet can be translated. However, the extent of yeast aslncRNAs translation, possibly giving rise to peptides, was unknown to date. In this context, the second and main objective of my project was to decipher the function of translation in the degradation of aslncRNAs. For that purpose, I analyzed the fate of aslncRNAs in conditions of translation inhibition (at the level of translation initiation and elongation). We described a comprehensive translational landscape of aslncRNAs, using a high-coverage Ribo-Seq and defined the fraction of aslncRNAs bound by the ribosome in WT and NMD-defective cells (in collaboration with Dr. Olivier Namy's lab at I2BC, Gif-sur-Yvette). Furthermore, I demonstrated the molecular mechanistic bases subjecting an aslncRNA to NMD. Finally, I investigated whether a reporter aslncRNA can produce a novel peptide during the pioneer round of translation, before being eliminated by the NMD. The findings of this work, available currently on biorXiv (Andjus et al., 2022, bioRxiv), as well as its follow up, are presented in the second part in the results section.

According to the model in *S. cerevisiae*, when aslncRNAs form dsRNA with their paired-sense mRNA, at least in some cells, they are protected from the NMD (Wery et al., 2016). Thus, the dsRNA formation represents a crucial step to understand the aslncRNA metabolism. Such duplex formation requires that both mRNA and aslncRNA coexist in the same cell. However, how heterogeneous is this co-expression, at the genome-wide level in single yeast cells remained unknown. To start exploring this idea, lastly my project aimed at characterizing mRNA/aslncRNA expression using yeast single-cell (sc)RNA-Seq. As a preliminary analysis, we analyzed the first strand-specific scRNA-Seq dataset in a hundred of WT cells (Nadal-Ribelles et al., 2019) and Xrn1 mutant cells. Then, I validated and extended this analysis using an independent and complementary scRNA-Seq approach (in collaboration, during my stay in the laboratory of Dr. Francesc Posas and Dr. Eulàlia De Nadal at IRB, Barcelona). I performed the scRNA-Seq analysis in thousands of cells in conditions that stabilize aslncRNAs - genetic context (Xrn1 and Upf1 inactivation) and in a stress condition inhibiting translation (CHX). With both approaches, we detected the co-expression of mRNA/aslncRNA expression at the single cells level. This results further prompted us to reveal, at the population level, to which extent the two transcripts form dsRNAs in conditions stabilizing the aslncRNAs using small RNA-Seq in cells with reconstituted RNAi pathway. The preliminary unpublished results of these experiments are presented in the third part of the results section.

Altogether, in my PhD project, we unveil to which extent aslncRNAs are translated, which could not only contribute to target them *via* NMD but also lead to the production of functional peptides, as well as open perspectives on how their capacity to form dsRNA structures can compete with their translation-dependent degradation. The results here obtained set the basis for the understanding of the aslncRNAs metabolism in higher Eukaryotes, as the factors targeting them are conserved.

Note: For summary of main objectives see blue boxes in Figure 11.

RESULTS

RESULTS

In this section, I present in three chapters the results of my thesis in form of published and unpublished work. The published work involves two publications of which I am co-first author. The publication n°1 entitled "Endogenous RNAi pathway evolutionarily shapes the destiny of the antisense lncRNAs transcriptome" in which we characterize the aslncRNAs population in RNAi-capable yeast N. castellii was published in 2019 in the journal of Life Science Alliance. The publication n°2 "Translation is a key determinant controlling the fate of cytoplasmic long non-coding RNAs" is currently deposited on bioRxiv; doi: https://doi.org/10.1101/2022.05.25.493276. In this work, we study the role of translation in the turnover of cytoplasmic aslncRNAs. After each publication, I present the follow-up of the work with a short discussion. In the final chapter, I present unpublished preliminary data that explore the heterogeneity of sense/antisense RNAs expression at the single-cell level and the capability of the RNA partners to form dsRNAs. Apart from the research articles, presented in the annex section, is the review we published in Non-Coding RNA in 2020 entitled "From Yeast to Mammals, the Nonsense-Mediated mRNA Decay as a Master Regulator of Long Non-Coding RNAs Functional Trajectory".

.

Chapter 3. Conserved Role of Decay Machineries in Shaping Long Non-Coding RNA Landscape in Yeast

1. Introduction

The pervasive transcription of eukaryotic genomes, among others, generates poorly studied antisense (as)lncRNAs. Pioneer studies in *S. cerevisiae* have shown that aslncRNAs are extensively targeted by RNA decay machineries. Therefore, they are undetectable in wild-type (WT) cells but can accumulate upon inactivation of the factors responsible for their decay. In fact, aslncRNAs are extensively degraded by the conserved nuclear exosome and cytoplasmic Xrn1 exoribonuclease. Transcripts sensitive to the exosome (Rrp6) have been termed Cryptic Unstable Transcripts (CUTs) (Neil et al., 2009; Xu et al., 2009), while those sensitive to Xrn1, identified by the lab, Xrn1-sensitvie Unstable transcripts (XUTs) (Van Dijk et al., 2011; Wery et al., 2016). The studies from our lab further revealed that in *S. cerevisiae* the Nonsense-Mediated mRNA Decay (NMD) pathway restricts most cytoplasmic XUTs unless blocked by dsRNA structure (Wery et al., 2016). However, during evolution *S. cerevisiae* has lost the RNA interference (RNAi), allowing the cytoplasmic dsRNAs to accumulate and interfere with the NMD, being an exemption model among Eukaryotes. Whether XUTs are targeted by the NMD in cytoplasmic RNAi-capable yeasts, in which the dsRNAs are degraded by default, remained unknown.

To start investigating this notion, it was necessary to find a model possessing the cytoplasmic RNAi able to interfere with the dsRNAs in the cytoplasm. The fission yeast *S. pombe* was not an optimal model as it harbors a nuclear RNAi pathway shown not to target XUTs (Wery et al., 2018b, 2018a) and the dsRNA structures are not in the cytoplasm. Instead, we studied the effect of NMD and dsRNAs on XUTs decay in the budding yeast *Naumovozyma castellii*, with a cytoplasmic RNAi.

The discovery of RNAi in *N. castellii* was relatively recent (Drinnenberg et al., 2009). Before, RNAi was initially presumed lost in all budding yeasts. However, Argonaute genes were found in budding yeasts (Scannell et al., 2007), including *N. castellii* and *Kluyveromyces polysporus* (both close relatives of *S. cerevisiae*) and *Candida albicans* (the most common yeast pathogen of humans (Berman and Sudbery, 2002)). At the time, the discovery of these genes was enigmatic as other RNAi genes, especially Dicer, have not been found in these species. Then, Bartel's lab discovered the full RNAi pathway in these species (Drinnenberg et al., 2009) that use a noncanonical Dicer to generate 21-23 nt siRNAs, which mostly correspond to repeated sequences, including transposable elements

and subtelomeric repeats. In terms of structure, budding yeast Dicer genes, contrary to *S. pombe* and other fungi, harbors one RNAseIII domain and no helicase nor PAZ domains (Drinnenberg et al., 2009), an important element that confers cleavage accuracy (MacRae and Doudna, 2007). Nevertheless, siRNAs isolated from budding yeasts had the two-nucleotide 3' overhangs, characteristic of the RNaseIII-mediated cleavage, raising a possibility that Dicer achieves this measuring function differently (Drinnenberg et al., 2009). However, whether Dicer affects the aslncRNAs metabolism in this species remained largely unknown.

In this context, it was postulated that the aslncRNA transcriptome expanded since the loss of RNAi in *S. cerevisiae*, in terms of aslncRNA steady-state levels, their lengths, and their degree of overlap with mRNAs if compared to the RNAi-capable budding yeast *N. castellii* (Alcid and Tsukiyama, 2016). However, this study was performed in WT cells in which the majority of aslncRNAs are rapidly degraded, probably leading to missing most of the cryptic lncRNAs. Therefore, an exhaustive annotation of the aslncRNA landscape in *N. castellii* was needed to determine the interactions between aslncRNAs and the cytoplasmic RNAi.

In this work, initiated in the lab prior to my arrival and being a side project of my PhD, we exhaustively annotated the aslncRNA landscape in *N. castellii* by performing genome-wide RNA-Seq in mutants of Dicer, Xrn1 and Rrp6. We showed that aslncRNAs are primarily degraded by the exosome and/or Xrn1, reinforcing the idea that the role of the 3'-5' nuclear and 5'-3' cytoplasmic RNA decay pathways in restricting aslncRNAs levels has been evolutionary conserved. In contrast, the loss of Dicer had almost no effect on the aslncRNAs transcriptome. Comparative analyses between aslncRNAs from *N. castellii* and *S. cerevisiae* revealed an expansion of the exosome-sensitive antisense transcriptome in the RNAi-capable budding yeast, suggesting that the nuclear RNA surveillance machinery has been evolutionarily reinforced for the control of aslncRNAs expression. The discovery of XUTs in this species, paved the way to study the potential effect of NMD in the decay of XUTs.

The results obtained in this work extend the conserved functions of the RNA decay machineries in the regulation of aslncRNA levels in Eukaryotes and open perspectives on the potential evolutionary relevance of RNAi in shaping the aslncRNAs transcriptome.

2. Publication n°1 “Endogenous RNAi Pathway Evolutionarily Shapes the Destiny of the Antisense LncRNAs Transcriptome”

Research Article

TRANSPARENT PROCESS OPEN ACCESS Life Science Alliance

Endogenous RNAi pathway evolutionarily shapes the destiny of the antisense lncRNAs transcriptome

Ugo Szachnowski*, Sara Andjus*, Dominika Foretek, Antonin Morillon, Maxime Wery

Antisense long noncoding (aslnc)RNAs are extensively degraded by the nuclear exosome and the cytoplasmic exoribonuclease Xrn1 in the budding yeast *Saccharomyces cerevisiae*, lacking RNAi. Whether the ribonuclease III Dicer affects aslncRNAs in close RNAi-capable relatives remains unknown. Using genome-wide RNA profiling, here we show that aslncRNAs are primarily targeted by the exosome and Xrn1 in the RNAi-capable budding yeast *Naumovozyma castellii*, Dicer only affecting Xrn1-sensitive aslncRNAs levels in Xrn1-deficient cells. The *dcr1* and *xrn1* mutants display synergic growth defects, indicating that Dicer becomes critical in the absence of Xrn1. Small RNA sequencing showed that Dicer processes aslncRNAs into small RNAs, with a preference for Xrn1-sensitive aslncRNAs. Consistently, Dicer localizes into the cytoplasm. Finally, we observed an expansion of the exosome-sensitive antisense transcriptome in *N. castellii* compared with *S. cerevisiae*, suggesting that the presence of cytoplasmic RNAi has reinforced the nuclear RNA surveillance machinery to temper aslncRNAs expression. Our data provide fundamental insights into aslncRNAs metabolism and open perspectives into the possible evolutionary contribution of RNAi in shaping the aslncRNAs transcriptome.

DOI 10.26508/lsa.201900407 | Received 26 April 2019 | Revised 21 August 2019 | Accepted 22 August 2019 | Published online 28 August 2019

Introduction

Initially considered as by-products of the pervasive transcription of eukaryotic genomes, long noncoding (lnc)RNAs have been progressively recognized as genuine transcripts playing important roles in the regulation of multiple cellular processes (Mercer et al, 2009; Wery et al, 2011; Rinn & Chang, 2012; Jarroux et al, 2017). Supporting the idea that lncRNAs can be functionally important, the dysregulated expression of some of them has been associated to diseases, including cancer and neurological disorders (Schmitt & Chang, 2016; Renganathan & Felley-Bosco, 2017; Saha et al, 2017; Schmitt & Chang, 2017).

Different classes of lncRNAs have been described (Jarroux et al, 2017). Among them, the "antisense" (as)lncRNAs are synthesized from the strand opposite to "sense" protein-coding genes (Pelechano & Steinmetz, 2013) and have attracted a lot of attention given their potential to regulate gene expression (Kopp & Mendell, 2018). In fact, examples of aslncRNA-mediated regulation of gene expression have been reported in different organisms, including the budding yeast *Saccharomyces cerevisiae* (Camblong et al, 2007, 2009; Uhler et al, 2007; Berretta et al, 2008; Houseley et al, 2008; Pinskaya et al, 2009; Van Dijk et al, 2011; van Werven et al, 2012), the fission yeast *Schizosaccharomyces pombe* (Wery et al, 2018a), plants (Swiezewski et al, 2009), and Mammals (Lee & Lu, 1999; Yap et al, 2010).

One of the most striking features of aslncRNAs is their low cellular abundance. Pioneer works in *S. cerevisiae* have revealed that they are extensively degraded by RNA surveillance machineries (Tisseur et al, 2011; Tudek et al, 2015). Consequently, these "cryptic" aslncRNAs cannot be detected in wild-type (WT) cells but accumulate upon inactivation of the factor responsible for their degradation. For example, the cryptic unstable transcripts (CUTs) accumulate in cells lacking Rrp6 (Wyers et al, 2005; Neil et al, 2009; Xu et al, 2009), a nonessential 3′-5′ exoribonuclease of the nuclear exosome (Houseley et al, 2006). On the other hand, the Xrn1-sensitive unstable transcripts (XUTs) are degraded by the cytoplasmic 5′-3′ exoribonuclease Xrn1 (Van Dijk et al, 2011). Despite some of them are produced from intergenic regions, most CUTs and XUTs are antisense to protein-coding genes, at least partially.

This classification into CUTs or XUTs is informative as it provides insights into the RNA decay pathway by which they are degraded. However, it is not exclusive, and there is a non-negligible overlap between the two classes (Van Dijk et al, 2011; Wery et al, 2016). Indeed, the nuclear and the cytoplasmic RNA surveillance pathways can cooperate to target the same transcript, so that a CUT that would escape the nuclear degradation can be targeted by Xrn1 once exported in the cytoplasm. Alternatively, but not exclusively, overlapping lncRNA isoforms produced from the same transcription unit can be degraded by different RNA surveillance pathways (Marquardt et al, 2011).

Both Rrp6 and Xrn1 are conserved across eukaryotes (Houseley et al, 2006; Nagarajan et al, 2013). In this respect, CUTs and XUTs were recently identified in fission yeast (Atkinson et al, 2018; Watts et al, 2018; Wery et al, 2018b), and they are also mainly antisense to

ncRNA, Epigenetic and Genome Fluidity, Institut Curie, Sorbonne Université, CNRS UMR 3244, Paris, France

Correspondence: antonin.morillon@curie.fr; maxime.wery@curie.fr
*Ugo Szachnowski and Sara Andjus contributed equally to this work.

protein-coding genes in this species. This indicates that the roles of the nuclear exosome and Xrn1 in restricting aslncRNAs levels have been conserved across the yeast clade.

However, one singularity that distinguishes *S. cerevisiae* from most other eukaryotes is the loss of the RNAi system during evolution, so it lacks the ribonuclease III Dicer that can process double-stranded (ds)RNA structures into siRNAs (Drinnenberg et al, 2009). However, upon heterologous expression in *S. cerevisiae* of RNAi factors from the close RNAi-capable relative species *Naumovozyma castellii* (Drinnenberg et al, 2009, 2011), we observed a massive production of siRNAs from asXUTs, indicating that they can form dsRNA structures with their paired-sense mRNAs in vivo (Wery et al, 2016). Consistent with this observation, *N. castellii* Dicer was detected in the cytoplasm when expressed in *S. cerevisiae* (Cruz & Houseley, 2014).

S. pombe has a functional RNAi machinery (Volpe et al, 2002), but asXUTs are insulated from it (Wery et al, 2018b). This is probably explained by different subcellular localization, as Dicer is restricted to the nucleus in fission yeast, mainly contributing in heterochromatin formation at centromeric repeats (Woolcock & Buhler, 2013). Yet, Dicer was shown to control a novel class of lncRNAs, referred to as Dicer-sensitive unstable transcripts (DUTs), which are also mainly antisense to protein-coding genes (Atkinson et al, 2018). Thus, in fission yeast, Dicer contributes to the control of aslncRNAs levels.

The discovery of RNAi in budding yeasts, such as *N. castellii*, *Kluyveromyces polysporus*, and *Candida albicans*, is quite recent (Drinnenberg et al, 2009) and whether RNAi plays any role in aslncRNAs metabolism in these species remains largely unknown. In this context, it has recently been proposed that the loss of RNAi in *S. cerevisiae* could have led to an expansion of the aslncRNAs transcriptome (Alcid & Tsukiyama, 2016). This hypothesis was essentially based on the observation that aslncRNAs expression levels, length, and degree of overlap with the paired-sense protein-coding genes are globally reduced in *N. castellii* compared with *S. cerevisiae*. However, these analyses were performed in WT strains, in which most aslncRNAs are likely to be degraded. Furthermore, it was not experimentally demonstrated that aslncRNAs are targeted by the endogenous RNAi machinery in *N. castellii*.

Here, we addressed the question of aslncRNAs degradation in *N. castellii* (Drinnenberg et al, 2009). Using deep transcriptome profiling in mutants of *DCR1*, *XRN1*, and *RRP6*, we showed that aslncRNAs are primarily degraded by the exosome and Xrn1. The loss of Dicer leads to a weak but significant increase in global aslncRNAs levels when combined to the *xrn1* mutation, suggesting that Dicer might become critical in the absence of Xrn1. This idea is supported by genetic evidence showing that the *dcr1* and *xrn1* mutants display synergic growth defects. Using small RNA sequencing, we showed that Dicer can process aslncRNAs into small RNAs, with a preference for asXUTs. Consistently, immunofluorescence experiments revealed that Dicer localizes in the cytoplasm. Finally, comparative analyses between aslncRNAs from *N. castellii* and *S. cerevisiae* revealed an expansion of the exosome-sensitive antisense transcriptome in the RNAi-capable budding yeast, suggesting that the nuclear RNA surveillance machinery has been evolutionarily reinforced for the control of aslncRNAs expression in a context where a Dicer-dependent ribonuclease III activity is present in the cytoplasm, possibly to prevent uncontrolled siRNAs production. Together, our data provide fundamental insights into the aslncRNAs metabolism in a yeast species endowed with cytoplasmic RNAi, further highlighting the conserved roles of the exosome and Xrn1 in the control of aslncRNAs levels in eukaryotes.

Results

AslncRNAs are primarily degraded by Rrp6 and Xrn1 in *N. castellii*

To characterize the population of aslncRNAs in *N. castellii*, we performed genome-wide RNA profiling using RNA-seq data obtained from WT, *dcr1Δ*, *xrn1Δ*, and *rrp6Δ* cells (Fig 1A). For the identification of DUTs and XUTs, we performed RNA-Seq in WT, *dcr1Δ*, and *xrn1Δ* strains, followed by segmentation using the algorithm that we previously developed to annotate CUTs (Watts et al, 2018) and XUTs (Wery et al, 2016, 2018b) in other yeast species. For the identification of *N. castellii* CUTs, we profiled in parallel published RNA-Seq data obtained from *rrp6Δ* cells (Alcid & Tsukiyama, 2016). Among all the ≥200-nt segments not overlapping a coding sequence, tRNA, sn(o)RNA or rRNA on the same strand, using a signal threshold and differential expression analysis between each mutant and its corresponding WT control (Fig 1A; see the Materials and Methods section), we identified 146 stable unannotated transcripts (SUTs, i.e., lncRNAs detected in the WT context but not significantly stabilized in any of the mutant), 10 DUTs, 1,021 XUTs, and 1,280 CUTs (Figs 1B and S1A–C).

At the first glance, the number of DUTs appears to be dramatically low compared with CUTs and XUTs, indicating than the effect of Dcr1 on the lncRNAs transcriptome of *N. castellii* is marginal compared with Rrp6 and Xrn1 (Fig 1C–E). Moreover, these DUTs were also all identified as XUTs, and they are even more sensitive to Xrn1 than to Dcr1 (Fig S1D). Consequently, these 10 lncRNAs, sensitive to both Dcr1 and Xrn1, will only be considered as XUTs hereafter.

As previously observed in *S. cerevisiae* and *S. pombe* (Wery et al 2016, 2018b; Atkinson et al, 2018), many lncRNAs are targeted by both Rrp6 and Xrn1 in *N. castellii* (Fig 1B). Consistently, CUTs and XUTs globally display a moderate sensitivity to Xrn1 and Rrp6, respectively (Fig 1D and E). More precisely, 426 XUTs are stabilized upon inactivation of Rrp6 (Table S1; *rrp6Δ*/WT ratio >2, $P < 0.05$), whereas 610 CUTs accumulate in the absence of Xrn1 (Table S2; *xrn1Δ*/WT ratio >2, $P < 0.05$). Furthermore, 232 CUTs overlap ≥50% of a XUT (Fig S1E). This indicates that Rrp6 and Xrn1 also cooperate to restrict lncRNAs levels in *N. castellii*.

Most of the transcripts we identified are novel (Fig S1F) and are antisense to protein-coding genes, including 93 SUTs (64%), 622 XUTs (61%), and 868 CUTs (68%). These proportions increase when taking into account all the transcripts annotated in *N. castellii* and not only the coding sequences (Fig S1G). Interestingly, we observed that the solo lncRNAs (i.e., those that are not antisense) are globally more expressed than the antisense ones. This is not only the case for the SUT, XUT, and CUT classes in WT cells (Fig 1F) but also for XUTs and CUTs in *xrn1Δ* and *rrp6Δ* cells, respectively (Fig S1H and I).

Overall, we annotated 2,247 lncRNAs in *N. castellii*, 1,583 of which are antisense to protein-coding genes. The vast majority of them are unstable and are primarily degraded by the nuclear exosome

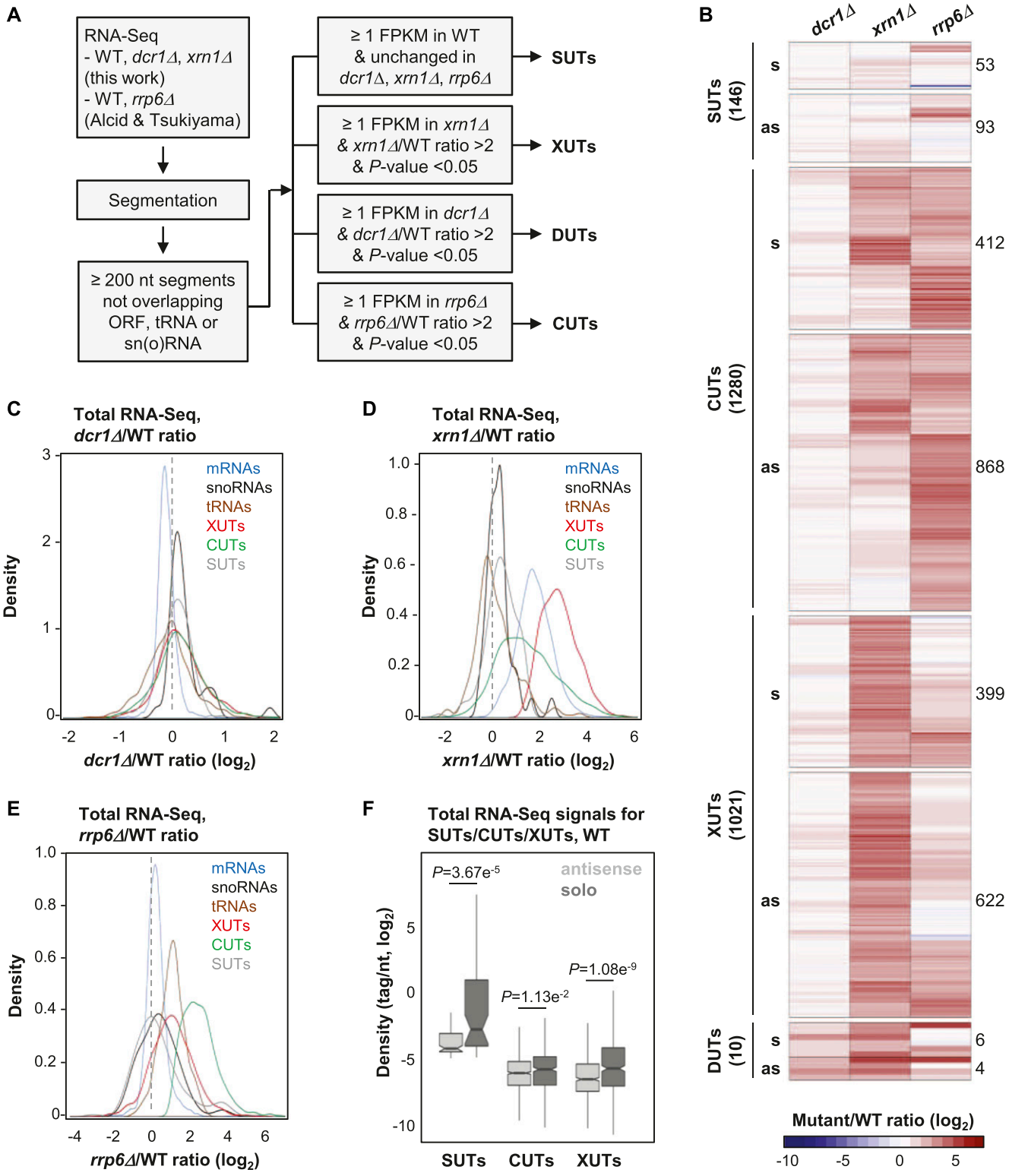


Figure 1. AslncRNAs are primarily degraded by Xrn1 and Rrp6 in *N. castellii*.
(A) Experimental strategy to annotate aslncRNAs in *N. castellii*. RNA-Seq data from biological duplicates of WT, *dcr1Δ*, and *xrn1Δ* cells were segmented using the ZINAR algorithm (Wery et al, 2016). Previously published RNA-Seq data from biological duplicates of *rrp6Δ* cells (Alcid & Tsukiyama, 2016) were segmented in parallel using the same tool. Among the ≥200-nt segments not overlapping an open reading frame (ORF), tRNA, or sn(o)RNAs, we identified 146 SUTs (signal in WT ≥ 1 FPKM; insensitive to Dcr1, Xrn1, or Rrp6), 1021 XUTs (signal in *xrn1Δ* ≥ 1 FPKM; *xrn1Δ*/WT ratio >2, $P < 0.05$), 10 DUTs (signal in *dcr1Δ* ≥ 1 FPKM; *dcr1Δ*/WT ratio >2, $P < 0.05$), and 1,280 CUTs (signal in *rrp6Δ* ≥ 1 FPKM; *rrp6Δ*/WT ratio >2, $P < 0.05$). **(B)** Heat map of the expression fold-change (ratio of tag densities, $\log_2$ scale) for SUTs (146), CUTs (1280), XUTs (1021), and DUTs (10) in

(CUTs) and/or Xrn1 (XUTs), with almost no effect of Dcr1. Fig S2A and B shows snapshots of RNA-Seq signals for illustrative examples of asXUTs and asCUTs.

dcr1 and *xrn1* mutants are synergic

The data above indicate that Dcr1 has no major impact on aslncRNAs levels when Xrn1 and Rrp6 are functional (see Fig 1C). But is it also the case in cells lacking Xrn1 or Rrp6?

Globally, the loss of Dcr1 in *xrn1Δ* cells results into a moderate but significant increase in asXUTs levels compared with the single *xrn1Δ* mutant (Fig 2A; $P = 1.77 \times 10^{-5}$, Wilcoxon rank-sum test; see examples in Fig S2A–D), with no effect on the solo XUTs (Fig 2A; $P = 0.0633$, Wilcoxon rank-sum test).

In contrast, deleting *DCR1* in *rrp6Δ* cells has no significant effect on global CUTs levels, independently of their solo or antisense configuration (Fig 2B; $P = 0.513$ and 0.991, respectively; Wilcoxon rank-sum test).

The marginal effect of Dcr1 inactivation on the coding and noncoding transcriptomes (Figs 1B and C and S2E) is consistent with the normal growth of the *dcr1Δ* mutant, which is undistinguishable from the WT strain (Figs 2C and S2F). Interestingly, the growth of the *dcr1Δ xrn1Δ* double mutant is more affected than the *xrn1Δ* single mutant in rich medium at the optimal temperature 25°C (Figs 2C and S2F). This effect is even stronger at higher (32°C) or lower (18°C) temperatures, or when cells are grown on synthetic medium (Fig 2C).

Thus, Dcr1 significantly impacts asXUTs levels in *xrn1Δ* cells, and the *dcr1* and *xrn1* mutants display synergic growth defects, indicating that Dcr1 becomes critical when Xrn1 is not functional, consistent with the idea that Dcr1 and Xrn1 share similar substrates.

AsXUTs are preferred aslncRNAs targets of Dicer for small RNAs production

We asked whether aslncRNAs are processed into small RNAs by Dicer in *N. castellii*. We sequenced small RNAs from WT, *xrn1Δ*, *dcr1Δ*, and *xrn1Δ dcr1Δ* cells.

In the WT and *xrn1Δ* strains, but not in *dcr1Δ* and *xrn1Δ dcr1Δ*, we observed the accumulation of 22–23-nt small RNAs, with U as the preferred 5′ nucleotide (Fig S3A), which corresponds to the previously described features of siRNAs in *N. castellii* (Drinnenberg et al, 2009). Subsequent bioinformatics analyses filtering 22–23-nt small RNAs revealed that all classes of aslncRNAs are globally targeted by Dcr1 for small RNA production. In fact, small RNA densities are higher for the antisense SUTs, CUTs, and XUTs than their solo counterparts, especially in the *xrn1Δ* context (Figs 3A and S3B–D). Notably, this is also the case in the WT strain, indicating that aslncRNAs can be processed by Dcr1 when Rrp6 and Xrn1 are functional. This suggests that in WT cells, a fraction of aslncRNAs escape the RNA surveillance machineries and interact with the paired-sense mRNAs to form dsRNA that can be processed by Dcr1 into small RNAs. Furthermore, in this condition, the asXUTs appear to be the preferred targets of Dcr1 among the three classes of aslncRNAs (Fig 3A). As illustrative examples, snapshots for the *XUT0527/C05780* and *XUT0213/A12460* pairs show that 22–23-nt small RNAs are produced from the asXUT/mRNA overlapping region in the WT context, with an increase in small RNAs densities in *xrn1Δ* (Figs 3B and S3E). In contrast, for the *CUT0672/C05770* and *CUT0275/A12440* pairs, the levels of 22–23-nt small RNAs in WT cells remain low (Figs 3B and S3E).

Together, these data show that aslncRNAs in *N. castelli* are efficiently targeted by Dcr1 for the production of small RNAs, with a preference for asXUTs.

Dcr1 localizes in the cytoplasm

The observation that aslncRNAs are processed into small RNAs in *N. castellii* indicates that they can form dsRNA structures with the paired-sense mRNAs, which co-localize with Dcr1 into the same subcellular compartment. Because asXUTs (i.e., the aslncRNAs that are degraded in the cytoplasm) constitute the preferred targets of Dcr1 for small RNAs production, we anticipated that Dcr1 localizes in the cytoplasm. Further supporting this hypothesis, Dcr1 was previously detected as cytoplasmic foci when artificially expressed as a fusion with the GFP in *S. cerevisiae* (Cruz & Houseley, 2014).

We constructed a Dcr1-GFP strain in *N. castellii* (Fig S4A). Upon direct visualization in living cells, Dcr1-GFP appeared as individual discrete foci, which are absent not only in the untagged control strain but also in cells expressing the GFP alone (Fig S4B). When detected using GFP nanobody by immunofluorescence in fixed cells, these foci were found in the cytoplasm (Fig 4). Importantly, small RNA sequencing showed that the Dcr1-GFP fusion remains functional for the production of 22–23-nt small RNAs (Fig S4C).

From these observations, we conclude that Dcr1 localizes in the cytoplasmic compartment in *N. castellii*.

Expansion of the exosome-sensitive aslncRNAs transcriptome in *N. castellii*

It has been recently proposed that RNAi could have evolutionarily contributed to restrict the aslncRNAs transcriptome in *N. castellii* (Alcid & Tsukiyama, 2016). This hypothesis was based, for instance, on the observation that 170 aslncRNAs annotated in a WT strain of *N. castellii* are shorter and display a reduced overlap with the paired-sense mRNAs in comparison with the set of aslncRNAs in *S. cerevisiae* (Alcid & Tsukiyama, 2016). As we considerably extended the repertoire of aslncRNAs in *N. castelli*, most of them being unstable because of their extensive degradation by Rrp6 and Xrn1, we decided to repeat this comparative analysis using our catalog of asCUTs and asXUTs. Note that some CUTs in *S. cerevisiae* are smaller

the *dcr1Δ*, *xrn1Δ*, and *rrp6Δ* mutants, relative to the corresponding WT strain. For each class of lncRNA, the number of antisense and solo (i.e., not antisense) transcripts is indicated. **(C)** Density plot of *dcr1Δ*/WT signal ratio for mRNAs (blue), sn(o)RNAs (black), tRNAs (brown), XUTs (red), CUTs (green), and SUTs (grey). **(D)** Density plot of *xrn1Δ*/WT signal ratio for the same classes of transcripts as above. **(E)** Density plot of *rrp6Δ*/WT signal ratio for the same classes of transcripts as above. **(F)** Box plot of densities (tag/nt, $\log_2$ scale) for the antisense (light grey) and solo (dark grey) SUTs, CUTs, and XUTs in WT cells. The *P*-values (adjusted for multiple testing with the Benjamini–Hochberg procedure) obtained upon two-sided Wilcoxon rank-sum test are indicated. Outliers: not shown.

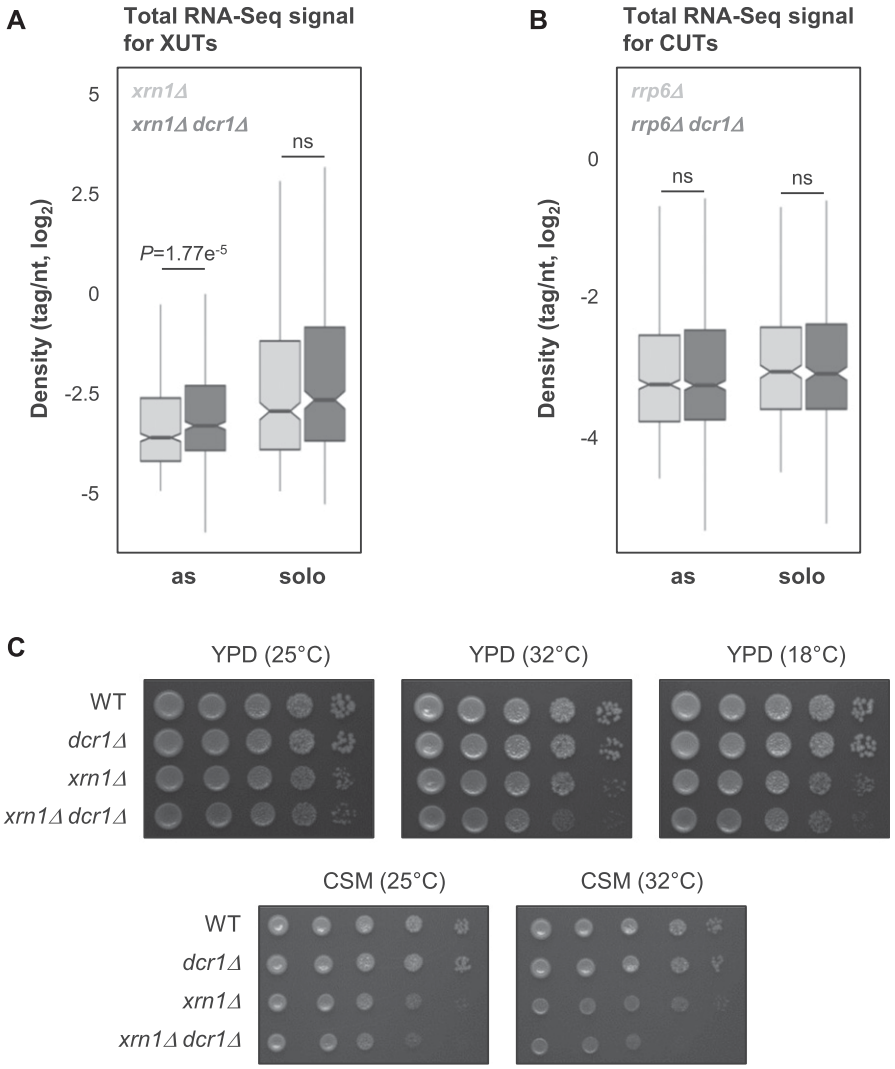


Figure 2. The *dcr1* and *xrn1* mutants display synergic defects.
(A) Box plot of densities (tag/nt, $\log_2$ scale) for the antisense (as) and solo XUTs in the *xrn1Δ* (light grey) and *xrn1Δ dcr1Δ* (dark grey) strains. The significant *P*-value (adjusted for multiple testing with the Benjamini–Hochberg procedure) obtained upon two-sided Wilcoxon rank-sum test is indicated. Outliers: not shown. Ns, not significant. **(B)** Box plot of densities (tag/nt, $\log_2$ scale) for the antisense (as) and solo CUTs in the *rrp6Δ* (light grey) and *rrp6Δ dcr1Δ* (dark grey) strains. Data are presented as above. The raw RNA-Seq data have been previously published (Alcid & Tsukiyama, 2016). **(C)** Effects of *DCR1* and/or *XRN1* deletion on growth. Serial dilutions of YAM2478 (WT), YAM2795 (*dcr1Δ*), YAM2479 (*xrn1Δ*), and YAM2796 (*dcr1Δ xrn1Δ*) cells were dropped on rich medium (YPD) or CSM, then incubated at the indicated temperatures for 2–3 d.

than 200 nt (the threshold commonly used to define lncRNAs). We decided to remove all these <200-nt CUTs from our analysis, to avoid the introduction of a bias in the comparison based on the size of aslncRNAs.

We observed a weak but significant reduction of asCUTs size in *N. castellii* compared with *S. cerevisiae* (Fig 5A; median = 444 and 465 nt, respectively; $P = 9.328 \times 10^{-3}$, Wilcoxon rank-sum test). The size of asXUTs is also reduced in *N. castellii* (Fig 5B; median = 670 versus 709 nt in *S. cerevisiae*), but the difference is not significant ($P = 0.3473$, Wilcoxon rank-sum test). Surprisingly, we noted that the aslncRNAs annotated in this work are globally larger than the 170 previously annotated aslncRNAs (see Fig S5). As a possible explanation of this discrepancy, 54/170 (32%) of the previously annotated aslncRNAs are shorter than the commonly used 200-nt threshold (Fig S5).

Independently of the size of the aslncRNA, the degree of overlap with the paired-sense mRNA is probably more critical to determine its ability to form dsRNA. In this respect, we found no difference between the RNAi-capable and the RNAi-deficient species for the asCUTs (Fig 5C; median length of the overlap = 357 and 370 bp, respectively; $P = 0.5044$, Wilcoxon rank-sum test). In contrast, the overlap between asXUTs and their paired-sense genes is significantly reduced in *N. castellii* (Fig 5D; median = 400, versus 462 bp in *S. cerevisiae*; $P = 1.315 \times 10^{-4}$, Wilcoxon rank-sum test).

Finally, we analyzed the global coverage of the coding transcriptome by aslncRNAs (SUTs and/or CUTs and/or XUTs) in the two yeast species. Overall, aslncRNAs overlap 8.1% of the coding sequences in *N. castellii*, which is reduced in comparison with *S. cerevisiae* (12.9%). However, when we analyzed the asCUTs and asXUTs separately, we observed opposite patterns between the two species. Indeed, the coding transcriptome is mainly overlapped by asCUTs in the RNAi-capable species, whereas in *S. cerevisiae*, it is mainly covered by asXUTs (Fig 5E).

In conclusion, our analysis reveals an expansion of the exosome-sensitive aslncRNAs transcriptome in *N. castellii*, suggesting that the presence of Dicer in the cytoplasm has evolutionarily reinforced the nuclear RNA surveillance machinery to restrict the expression of aslncRNAs in the cytoplasmic compartment. Conversely, the loss of RNAi in *S. cerevisiae* might have allowed an expansion of the Xrn1-sensitive antisense transcriptome, relaxing the pressure to maintain aslncRNAs in the nucleus, away from Dcr1.

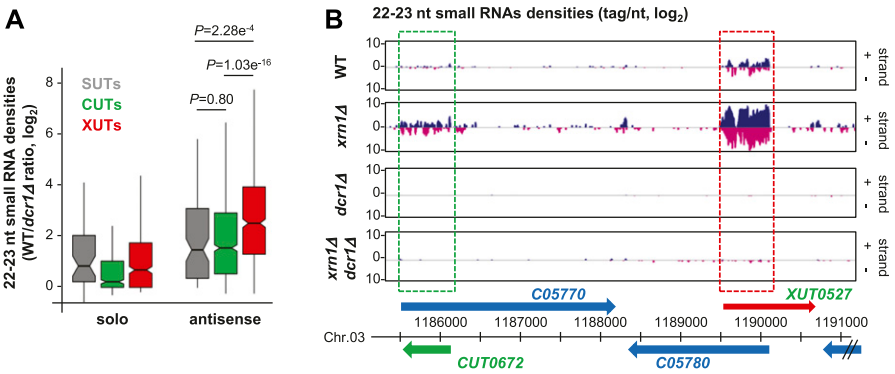


Figure 3. AslncRNAs are processed into 22–23-nt small RNAs by Dcr1 in *N. castellii*.
(A) Box plot of the WT/*dcr1Δ* ratio (log_2 scale) of 22–23-nt uniquely mapped small RNAs densities for the solo and antisense SUTs (grey), CUTs (green), and XUTs (red). The *P*-values (adjusted for multiple testing with the Benjamini–Hochberg procedure) obtained upon two-sided Wilcoxon rank-sum test are indicated. Outliers: not shown. **(B)** Snapshot of small RNAs densities for the *C05770*/*CUT0672* and *XUT0527*/*C05780* pairs. Densities of 22–23-nt small RNAs are shown in a separate panel for each strain. In each panel, signals (tag/nt, log_2) for the + and − strands are shown in blue and pink, respectively. The protein-coding genes, the CUT, and the XUT are represented by blue, green, and red arrows, respectively. The dashed boxes highlight the region of overlap between the aslncRNAs and the paired-sense mRNAs. The snapshot was produced using VING (Descrimes et al, 2015).

Discussion

Previous works in the budding yeast *S. cerevisiae* and the fission yeast *S. pombe* have revealed that aslncRNAs are globally low abundant as they are extensively degraded by RNA surveillance machineries. For instance, the nuclear exosome targets a class of lncRNAs referred to as CUTs (Wyers et al, 2005; Neil et al, 2009; Xu et al, 2009), whereas the cytoplasmic 5′-3′ exoribonuclease Xrn1 degrades the so-called XUTs (Van Dijk et al, 2011), both types of transcripts being mainly antisense to protein-coding genes. However, this classification into CUTs and XUTs is not exclusive, some aslncRNAs being cooperatively targeted by the two RNA decay pathways. In fission yeast, an additional class of aslncRNAs (DUTs) was recently identified. DUTs accumulate in the absence of the ribonuclease III Dicer (Atkinson et al, 2018), highlighting the role of Dicer and RNAi in the control of aslncRNAs expression in fission yeast. This class of transcripts is absent in *S. cerevisiae*, which has lost the RNAi system during evolution. In this respect, *S. cerevisiae* is a notable exception among eukaryotes. In fact, a functional RNAi pathway was discovered in close relatives of *S. cerevisiae*, including *N. castellii* (Drinnenberg et al, 2009), a member of the *sensu lato* group of *Saccharomyces* that diverged from *S. cerevisiae* after the whole genome duplication (Cliften et al, 2006). The role of RNAi on aslncRNAs metabolism remains largely unknown in this species. However, a recent study proposed that the loss of RNAi in *S. cerevisiae* might have allowed the expansion of the aslncRNAs transcriptome (Alcid & Tsukiyama, 2016). This hypothesis was essentially based on the observation that aslncRNAs levels, length and degree of overlap with the paired-sense genes are reduced in the RNAi-capable budding yeast. However, these analyses were performed using a small set of aslncRNAs annotated from a WT strain of *N. castellii*, that is, a context in which most aslncRNAs are likely to be degraded. Furthermore, whether aslncRNAs are directly targeted by the RNAi machinery in *N. castellii* natural context remained unknown.

Using genome-wide RNA profiling in WT, *dcr1Δ*, *xrn1Δ*, and *rrp6Δ* strains of *N. castellii*, here we annotated 2,247 lncRNAs, including 1,583 aslncRNAs. Most of them are unstable and primarily degraded by the nuclear exosome (1,280 CUTs) and/or Xrn1 (1,021 XUTs), reinforcing the idea that the role of the 3′-5′ nuclear and 5′-3′ cytoplasmic RNA decay pathways in restricting aslncRNAs levels has been conserved across the yeast clade. In contrast, the loss of Dcr1 has almost no effect on the aslncRNAs transcriptome. Only 10 DUTs accumulate in *dcr1Δ* cells, and they are also (even more) sensitive to Xrn1 (Fig S1D). This is marginal in comparison with the 1,392 DUTs annotated in fission yeast (Atkinson et al, 2018), raising the question of the function of Dcr1 in *N. castellii*.

DCR1 has been conserved in some budding yeast species (Drinnenberg et al, 2009). However, deleting it in *N. castellii* confers no detectable growth defect, as shown under 50 different conditions (Drinnenberg et al, 2011). As previously proposed, the main role of Dcr1 in budding yeasts might be to silence retrotransposons (Drinnenberg et al, 2009). Consistently, although its genome still contains retrotransposons fragments, which constitute a major source for siRNAs production, no active retrotransposon has been identified in *N. castellii* (Drinnenberg et al, 2009). In addition, the expression of *N. castellii DCR1* and *AGO1* in *S. cerevisiae* leads to the silencing of endogenous retrotransposons (Drinnenberg et al, 2009), as well as to the loss of the dsRNA killer virus (Drinnenberg et al, 2011), with no other major impact on the transcriptome of *S. cerevisiae*.

However, several lines of evidence indicate that Dcr1 becomes critical in the absence of Xrn1. First, the global levels of asXUTs significantly increase in the *xrn1Δ dcr1Δ* mutant, compared with the *xrn1Δ* single mutant (Fig 2A). Second, the number of Dcr1-sensitive protein-coding genes is larger in the *xrn1Δ* context, in comparison with WT and *rrp6Δ* (Fig S2E). Third, the *dcr1Δ* and *xrn1Δ* mutants display synergic growth defects (Fig 2C). This indicates that the presence of Dcr1 becomes important for the cell viability in the absence of Xrn1, that is, when aslncRNAs accumulate in the cytoplasm, presumably forming dsRNA structures with the paired-sense mRNAs. In contrast, *DCR1* deletion was shown to suppress partially the growth defect of the *rrp6Δ* mutant (Alcid & Tsukiyama, 2016), indicating that Dcr1 is deleterious in Rrp6-lacking cells. Whether these opposite effects in the *xrn1Δ* and *rrp6Δ* backgrounds are related to siRNAs production from stabilized asXUTs and asCUTs, respectively, remains unknown. Additional analyses are required to decipher the molecular mechanisms underlying these genetic interactions.

The idea that Dcr1 and Xrn1 functionally interact is reinforced by the observation that Dcr1 localizes in the cytoplasm (Fig 4), which is consistent with previous observations made upon

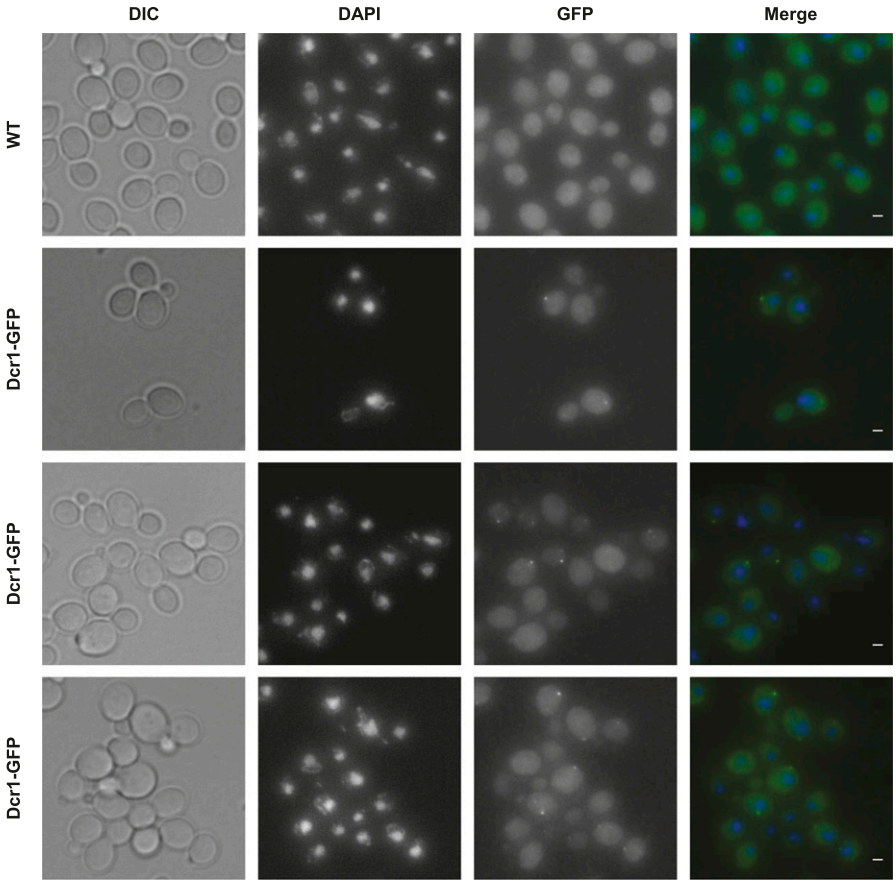


Figure 4. Subcellular localization of Dcr1 in *N. castellii*.
YAM2478 (WT) and YAM2826 (Dcr1-GFP) cells were grown to mid-log phase in CSM, at 25°C. After fixation of cells, the subcellular localization of Dcr1-GFP was performed using immunofluorescence using GFP nanobody (see the Materials and Methods section). DAPI staining was used to visualize DNA. Scale bars: 1 μm.

expression of a Dcr1-GFP fusion in *S. cerevisiae* (Cruz & Houseley, 2014). Moreover, among the different classes of aslncRNAs, the asXUTs constitute the preferred target for small RNAs production (Fig 3A). Notably, these small RNAs are detected in WT cells, indicating that in this context, a fraction of asXUTs can escape Xrn1 to form dsRNA with the paired-sense mRNAs, which can then be processed by Dcr1 into small RNAs. To which extent the generated small RNAs are properly loaded into Argonaute to mediate post-transcriptional gene silencing remains unknown. The resulting effects, if any, are likely to be limited, in keeping with the absence of growth defects of the *dcr1Δ* mutant.

Besides asXUTs, asCUTs are also processed into small RNAs by Dcr1 (Fig 3A and B). As mentioned above, asCUTs (at least a fraction of them) could escape the degradation by Rrp6 and be exported to the cytoplasm. Then, as the asXUTs, they could be processed by Dcr1 upon dsRNA formation, if they are not degraded before by Xrn1. Alternatively, but not exclusively, we cannot exclude the possibility that a small amount of Dcr1 molecules in the cell localize in the nucleus, into levels that are under the detection threshold of our microscope. Perhaps, a more sensitive approach would help definitely answering the question of the subcellular localization of Dcr1 in RNAi-capable budding yeasts, even if all the current data are consistent with a cytoplasmic localization.

Recently, it has been proposed that the loss of RNAi in *S. cerevisiae* might have allowed the expansion of its aslncRNAs transcriptome (Alcid & Tsukiyama, 2016). Conversely, the conservation of a functional RNAi machinery in *N. castellii* would have maintained a negative pressure against aslncRNAs. Among other observations, antisense expression at the *GAL10-GAL1* (*NCAS0E01670-NCAS0E01660*) locus was shown to be very low in WT cells of *N. castellii* (Alcid & Tsukiyama, 2016). Our RNA-Seq data confirmed this observation, further highlighting that despite the genomic organization of the *GAL1-GAL10-GAL7* locus has been conserved between *S. cerevisiae* and *N. castellii*, it is devoid of aslncRNA expression in RNAi-capable species, including in *xrn1Δ* and *rrp6Δ* strains (see the genome-browser associated to this work). Similarly, we confirm the absence of aslncRNA expression for the *PHO84* ortholog of *N. castellii* (*NCAS0B00220*). However, the differences between the RNAi-capable and RNAi-deficient species are more subtle than initially proposed. In fact, we show that more than 1,500 aslncRNAs co-exist with RNAi in *N. castellii*, mainly degraded by the exosome and Xrn1, representing an 8.1% cumulative overlap of the coding sequences by aslncRNAs, which is less than a twofold difference compared with *S. cerevisiae* (12.9%). Strikingly, when we analyzed the degree of overlap with the paired-sense ORFs, we observed that it is significantly reduced in *N. castellii* for the asXUTs but similar between the two species for the asCUTs (Fig 5C and D). Moreover, we observed that globally, the coding regions are mainly overlapped by asCUTs in the RNAi-capable species, whereas in *S. cerevisiae*, they are essentially overlapped by asXUTs. Together, our data suggest that the presence of an active RNAi machinery in the cytoplasm of *N. castellii* has favored the nuclear RNA decay pathway to restrict

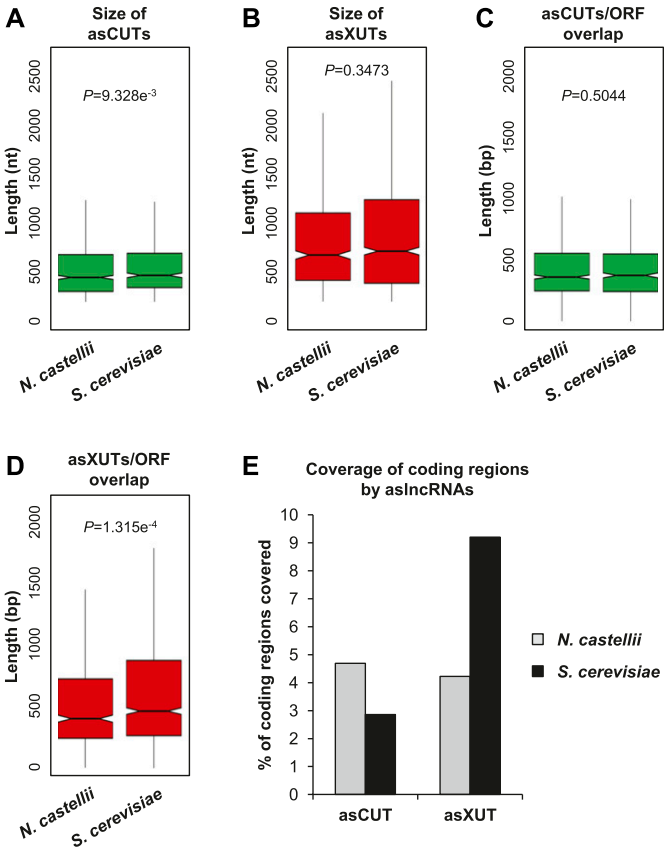


Figure 5. Expansion of the exosome-sensitive aslncRNAs transcriptome in *N. castellii*.
(A) Box plot of asCUTs size (nt) in *N. castellii* (n = 868) and *S. cerevisiae* (n = 535). For *S. cerevisiae*, all the <200-nt CUTs were removed from the analysis. The *P*-value obtained upon two-sided Wilcoxon rank-sum test is indicated. Outliers: not shown. **(B)** Same as above for asXUTs in *N. castellii* (n = 622) and *S. cerevisiae* (n = 1,152). **(C)** Box-plot of the overlap (bp) between asCUTs and the paired-sense ORF in *N. castellii* (n = 889) and *S. cerevisiae* (n = 574). For *S. cerevisiae*, all the <200-nt CUTs were removed from the analysis. The *P*-value obtained upon two-sided Wilcoxon rank-sum test is indicated. Outliers: not shown. **(D)** Same as above for asXUTs/ORF overlap in *N. castellii* (n = 674) and *S. cerevisiae* (n = 1,252). **(E)** Cumulative coverage of the coding regions by asCUTs and asXUTs in *N. castellii* (grey bars) and *S. cerevisiae* (black bars).

aslncRNAs expression, maybe to prevent uncontrolled and deleterious siRNAs production. This last hypothesis is supported by the observation that Dcr1 becomes deleterious in *rrp6Δ* cells (Alcid & Tsukiyama, 2016).

In conclusion, together with our previous studies in *S. cerevisiae* and *S. pombe*, this work in a budding yeast endowed with cytoplasmic RNAi provides fundamental insights into the metabolism and the decay of aslncRNAs in simple eukaryotic models. Our data not only further highlight the conserved roles of the nuclear exosome and Xrn1 in the control of aslncRNAs expression but also open perspectives into the possible evolutionary contribution of RNAi in shaping the aslncRNAs transcriptome. In this respect, the definition of the "cryptic" aslncRNAs landscape in organisms, such as plants and animals, where ribonuclease III activities are found in both the nucleus and the cytoplasm (Lee et al, 2003; Ha & Kim, 2014; Borges & Martienssen, 2015), will be of particular interest.

Materials and Methods

Strains, plasmids, and media

The genotypes of the strains used in this study are listed in Table S3. The YAM2478/DBP005 (WT) and YAM2795/DBP318 (*dcr1Δ*) strains were previously described (Drinnenberg et al, 2009).

The YAM2479 strain (*xrn1Δ::kanMX6*) was constructed by homologous recombination using the *kanMX6* marker flanked by long (>400 bp) *XRN1* targeting sequences. The *XRN1* ortholog in *N. castellii* is *C04170*, according to the Yeast Gene Order Browser (Byrne & Wolfe, 2005). The orthology was confirmed by CLUSTALO alignments (Fig S6A). To construct the *XRN1* deletion cassette, the *kanMX6* marker was first excised from the pFA6a-kanMX6 vector (Longtine et al, 1998) using BamHI and EcoRI digestion and cloned between the BamHI and EcoRI sites into the pCRII-TOPO plasmid (Invitrogen) to give the pCRII-kanMX6 plasmid. The 454-bp region upstream from *XRN1* was amplified by PCR using AMO1964-5 (see Table S4), and then cloned between the KpnI and BamHI sites into pCRII-kanMX6. Finally, the 481 bp downstream to *XRN1* were amplified by PCR using AMO1966-7 (see Table S4), and then cloned between the EcoRI and XbaI sites of the plasmid, giving the pAM376 vector. The deletion cassette was excised using KpnI–XbaI digestion and transformed into the YAM2478 strain. Transformants were selected on yeast extract–peptone–dextrose (YPD) + G418 plates at 25°C and screened by PCR on genomic DNA using oligonucleotides AMO1996-7. One clone was selected to give the YAM2479 strain, which was ultimately validated by Northern blot (Fig S6B).

To construct the YAM2796 strain (*dcr1Δ xrn1Δ*), the *xrn1Δ::kanMX6* cassette was amplified by PCR from YAM2479 genomic DNA using oligonucleotides AMO3227-8 (Table S4) and transformed into YAM2795. Transformants were selected and screened as above.

To construct the YAM2826 strain (Dcr1-GFP), the region corresponding to the last 478 bp of the *DCR1* ORF was amplified by PCR from YAM2478 genomic DNA using oligonucleotides AMO3323 and 3325 (Table S4). In parallel, the region corresponding to 525 bp after the stop codon of the *DCR1* ORF was amplified using oligonucleotides AMO3324 and 3326 (Table S4). After purification on agarose gel, the two PCR products displaying a 42-bp overlap were mixed and used as DNA templates for PCR using oligonucleotides AMO3323 and 3324. The resulting PCR product (1,047 bp long) was cloned between the KpnI and XbaI sites of the pCRII-TOPO plasmid (Invitrogen), to give the pCRII-Dcr1 vector. The GFP(S65T)-kanMX6 cassette was then amplified by PCR from the pFA6a-GFP(S65T)-kanMX6 plasmid using oligonucleotides AMO3327-8 (Table S4). The GFP(S65T)-kanMX6 PCR product was digested by BamHI and EcoRI and cloned between the same sites in the pCRII-Dcr1 vector, to give the pAM566 vector (pCRII-Dcr1-GFP-kanMX6). After verification of absence of mutation by Sanger sequencing, the Dcr1-GFP-kanMX6 construct was excised using NaeI digestion and transformed in the YAM2478 strain. Transformants were selected on YPD + G418 plates at 25°C and screened by PCR on genomic DNA using oligonucleotides AMO3229-30. One clone was selected and validated by Western blot (Fig S4A), giving the YAM2826 strain.

To construct the YAM2842 strain (*dcr1Δ::GFP-kanMX6*), the region corresponding to the *DCR1* promoter was amplified by PCR from

YAM2478 genomic DNA using oligonucleotides AMO3370-1 (Table S4). The resulting PCR product (470 bp) was purified and cloned between the KpnI and BamHI sites of the pAM566 plasmid, replacing the fragment corresponding to the end of the *DCR1* ORF, giving the pAM574 vector (pCRII-dcr1Δ::GFP-kanMX6). The absence of mutation was verified, then the *dcr1Δ::GFP-kanMX6* construct was excised and transformed in YAM2478 cells, as described above. Transformants were selected and screened as above. One clone was validated by Western blot (Fig S4A), giving the YAM2842 strain.

N. castellii strains were grown at 25°C in rich YPD medium to mid-log phase (OD_{600} 0.5). For the microscopy analyses, the cells were grown under the same conditions in complete synthetic medium (CSM).

Total RNA extraction

Total RNA was extracted from exponentially growing (OD_{600} 0.5) cells using standard hot phenol procedure. RNA was resuspended in nuclease-free H_2O (Ambion) and quantified using a NanoDrop 2000c spectrophotometer. Quality and integrity of extracted RNA was checked by Northern blot and/or analysis in a RNA 6000 Pico chip in a 2100 bioanalyzer (Agilent).

Northern blot

10 μg of total RNA were loaded on denaturing 1.2% agarose gel and transferred to Hybond-XL nylon membrane (GE Healthcare). ^{32}P-labelled oligonucleotides (see Table S4) were hybridized overnight at 42°C in ULTRAhyb-Oligo hybridization buffer (Ambion). For detection of the 5′ ITS1 fragment, a double-stranded DNA probe (obtained by PCR amplification using oligonucleotides AMO2002-2003) was ^{32}P-labelled using the Prime-It II Random Primer Labeling Kit (Agilent), and then hybridized overnight at 65°C in PerfectHyb Plus Hybridization Buffer (Sigma-Aldrich).

Total RNA-Seq

Total RNA-Seq analysis was performed from two biological replicates of YAM2478 (WT), YAM2479 (*xrn1Δ*), YAM2795 (*dcr1Δ*), and YAM2796 (*dcr1Δ xrn1Δ*) cells. For each sample, 1 μg of total RNA was mixed with 2 μl of 1:100 diluted ERCC RNA spike-in (Life Technologies), then ribosomal (r)RNAs were depleted using the RiboMinus Eukaryote v2 Kit (Life Technologies). Total RNA-Seq libraries were constructed from 50 ng of rRNA-depleted RNA using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina). Paired-end sequencing (2 × 50 nt) was performed on a HiSeq 2500 system (Illumina).

The *N. castellii* reference genome was retrieved from version 7 of the Yeast Gene Order Browser (Byrne & Wolfe, 2005); snoRNAs were annotated using the *S. cerevisiae* snoRNAs as queries for blastn alignments (E value cutoff e^{-8}). Reads were mapped using version 2.0.9 of TopHat (Kim et al, 2013), with a tolerance of three mismatches and a maximum size for introns of 2 Kb. All bioinformatics analyses used uniquely mapped reads. Tag densities were normalized on the ERCC RNA spike-in signal.

Annotation of lncRNAs

Segmentation was performed using the ZINAR algorithm (Wery et al, 2016). Briefly, the uniquely mapped reads from our WT, *dcr1Δ*, *xrn1Δ*, and *dcr1Δ xrn1Δ* samples were pooled. A signal was computed in a strand-specific manner for each position as the number of times it is covered by a read or the insert sequence between two paired reads. After $\log_2$ transformation, the signal was smoothed using a sliding window (ranging from 5 to 200 nt, with 5-nt increment). All genomic regions showing a smoothed $\log_2$ signal value above a threshold (ranging from 1.44 to 432, with 1.44 increments) were reported as segments. In total, 12,000 segmentations with different sliding window size and threshold parameters were tested in parallel, among which we arbitrarily selected one showing a good compromise between mRNA and novel lncRNAs detection. The parameters for the selected segmentation were: threshold = 27.36; sliding window size = 10 nt. Among the ≥200-nt novel segments that do not overlap ORF, tRNA or sn(o)RNA, we identified 1021 XUTs and 10 DUTs, showing a signal ≥1 FPKM (fragment per kilobase per million mapped reads) and >twofold enrichment in the *xrn1Δ* and *dcr1Δ* mutant, respectively, compared with the WT control, with a *P*-value < 0.05 (adjusted for multiple testing with the Benjamini–Hochberg procedure) upon differential expression analysis using DESeq2 (Love et al, 2014). 262 segments showing a signal ≥1 FPKM in the WT context but no significant enrichment in the *xrn1Δ* or in the *dcr1Δ* mutant were considered as putative SUTs.

For the annotation of CUTs, we used previously published RNA-Seq data from biological duplicates of *rrp6Δ* cells (Alcid & Tsukiyama, 2016). Segmentation was performed following the same procedure as described above, using a threshold of 12.96 and a sliding window of 10 nt. As no ERCC RNA spike-in was included during libraries preparation, tag densities were normalized on the total number of reads uniquely mapped on ORFs. We identified 1,280 CUTs, 116 of which overlapped >50% of transcripts defined as putative SUTs upon segmentation of our RNA-Seq data. Consequently, these 116 transcripts were not considered as SUTs.

Overall, we annotated 10 DUTs, 146 SUTs, 1,021 XUTs, and 1,280 CUTs. An lncRNA was reported as antisense when the overlap with the sense ORF was ≥1 nt.

Small RNA-Seq

Small RNA-Seq analysis was performed from two biological replicates of YAM2478 (WT), YAM2479 (*xrn1Δ*), YAM2795 (*dcr1Δ*), and YAM2796 (*dcr1Δ xrn1Δ*) exponentially growing cells. For the control of Dcr1-GFP functionality (YAM2826 strain), only one library was prepared.

For each sample, 50 μg of total RNA were mixed with 2 μg of total RNA from the YAM2394 (WT) strain of *S. pombe* (Wery et al, 2018b), the 22–23-nt small RNAs derived from the centromeric repeats in the latter species (Djupedal et al, 2009), here constituting RNA spike-in used for the subsequent normalization of the small RNA-Seq signals.

The small RNAs (<80 nt) fraction was purified on 15% TBE–urea polyacrylamide gels. Libraries were constructed from 120 ng of purified small RNAs using the NEBNext Multiplex Small RNA Library Preparation Set for Illumina (New England Biolabs). Single-end

sequencing (50 nt) of libraries was performed on a HiSeq 2500 system (Illumina).

Adapter sequences were removed using the Atropos software (Didion et al, 2017). Reads were then mapped to the *N. castellii* and *S. pombe* reference genomes using the version 2.3.5 of Bowtie (Langmead & Salzberg, 2012), using default parameters, with no mismatch in seed alignment. Subsequent analyses used 22–23-nt reads uniquely mapped on the *N. castellii* genome. Densities were normalized on the levels of the centromeric 22–23-nt small RNAs of *S. pombe*.

Western blot

50 μg of protein extracts were separated on a NuPAGE 4–12% Bis–Tris gel (Invitrogen) and then transferred to a nitrocellulose membrane using an iBlot Dry Blotting System (Invitrogen). The GFP and Pgk1 were detected using mouse anti-GFP (11 814 460 001, Roche) with the SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) and mouse anti-Pgk1 (ab 113687; Abcam) with the SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific), respectively. Images were obtained using a ChemiDoc Imaging System (Bio-Rad).

Microscopy

The cells were grown to mid-log phase (OD_{600} 0.5) in CSM medium, at 25°C. For the live cell analysis, the cells were washed in sterile water and then loaded on a microscope slide. The images were acquired the same day with the same parameters, using a wide-field microscopy system based on an inverted microscope (TE2000; Nikon) equipped with a 100×/1.4 NA immersion objective, a CMOS camera and a collimated white light-emitting diode for the transmission. A Spectra X light engine lamp (Lumencor, Inc) was used to illuminate the samples. The whole system is piloted by the MetaMorph software (Molecular Devices). For z-stacks images, the axial (z) step is at 200 nm, and images shown are a maximum projection of z-stack images. The images were analyzed and processed using the ImageJ software.

Subcellular localization of Dcr1-GFP was performed by immunofluorescence using GFP booster/nanobody (ATTO 488; ChromoTek), according to a previously described procedure (Ries et al, 2012). Briefly, cells were loaded on concanavalin A-coated coverglass and fixed for 15 min in PBS containing 4% paraformaldehyde and 2% of sucrose. After two washes with PBS + 50 mM NH_4Cl, the fixed cells were blocked and permeabilized for 30 min in blocking/hybridization buffer (0.25% Triton X-100, 5% BSA, 0.004% NaN_3 in PBS), under gentle shaking. The cells were then labelled for 90 min with 100 μl of nanobody solution (10 μM ATTO 488 nanobody in blocking/hybridization buffer). Finally, the labelled cells were washed for 5 min in PBS, a drop of VECTASHIELD mounting medium with DAPI (Vector Labs) was added on the cells, and the coverglass was mounted on a microscope slide. Fluorescence images were acquired using the same microscope as described above. The images were analyzed and processed using the ImageJ software, as described above.

Accession numbers and data accessibility

Sequence data generated in this work can be accessed at the NCBI Gene Expression Omnibus using accession number GSE129233. Previously published RNA-Seq data we retrieved from the Sequence Read Archive using accession number SRP056928.

Genome browsers for visualization of processed data are publicly accessible at http://vm-gb.curie.fr/castellii.

Code accessibility

The computational scripts used in this study are accessible at https://github.com/MorillonLab/castellii.

Availability of materials

All unique materials generated in this work are available upon request to the corresponding authors.

Supplementary Information

Supplementary Information is available at https://doi.org/10.26508/lsa.201900407.

Acknowledgements

We thank David Bartel for providing the WT and *dcr1Δ* strains of *N. castellii*. We also would like to thank Patricia Legoix-Né, Virginie Raynal, Benoît Albaud, and Sylvain Baulande from the Next Generation Sequencing (NGS) platform of Institut Curie; Choumouss Kamoun from the bioinformatics platform of Institut Curie; Camille Gautier and Marc Descrimes for preliminary bioinformatics analyses and tools development; Myriam Ruault (UMR3664, Institut Curie) for assistance in microscopy analyses; Hervé Vennin-Rendos for his contribution in the construction of the pAM376 vector; Eve Samani for her enthusiastic participation in validation and characterization of yeast strains; and Ines A Drinnenberg for critical reading of the manuscript. We are grateful to all members of the lab for discussions. High-throughput sequencing was performed by the NGS platform of Institut Curie, supported by the grants ANR-10-EQPX-03 and ANR10-INBS-09-08 from the Agence Nationale de la Recherche (investissements d'avenir) and by the Cancerop̂le Ile-de-France. Antonin Morillon's lab is supported by the "DNA-life" grant (ANR-15-CE12-0007) from the Agence Nationale de la Recherche, and by the "EpincRNA" (starting) and "DARK" (consolidator) grants from the European Research Council.

Author Contributions

U Szachnowski: resources, data curation, software, formal analysis, validation, investigation, and visualization.
S Andjus: formal analysis, investigation, visualization, and writing—review and editing.
D Foretek: formal analysis and investigation.
A Morillon: conceptualization, resources, supervision, funding acquisition, project administration, and writing—original draft, review, and editing.
M Wery: conceptualization, resources, data curation, formal analysis, supervision, validation, investigation, visualization, methodology,

project administration, and writing—original draft, review, and editing.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

References

Alcid EA, Tsukiyama T (2016) Expansion of antisense lncRNA transcriptomes in budding yeast species since the loss of RNAi. *Nat Struct Mol Biol* 23: 450–455. doi:10.1038/nsmb.3192

Atkinson SR, Marguerat S, Bitton DA, Rodriguez-Lopez M, Rallis C, Lemay JF, Cotobal C, Malecki M, Smialowski P, Mata J, et al (2018) Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. *RNA* 24: 1195–1213. doi:10.1261/rna.065524.118

Berretta J, Pinskaya M, Morillon A (2008) A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. *Genes Dev* 22: 615–626. doi:10.1101/gad.458008

Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* 16: 727–741. doi:10.1038/nrm4085

Byrne KP, Wolfe KH (2005) The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* 15: 1456–1461. doi:10.1101/gr.3672305

Camblong J, Beyrouthy N, Guffanti E, Schlaepfer G, Steinmetz LM, Stutz F (2009) Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. *Genes Dev* 23: 1534–1545. doi:10.1101/gad.522509

Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell* 131: 706–717. doi:10.1016/j.cell.2007.09.014

Cliften PF, Fulton RS, Wilson RK, Johnston M (2006) After the duplication: Gene loss and adaptation in Saccharomyces genomes. *Genetics* 172: 863–872. doi:10.1534/genetics.105.048900

Cruz C, Houseley J (2014) Endogenous RNA interference is driven by copy number. *Elife* 3: e01581. doi:10.7554/elife.01581

Descrimes M, Zouari YB, Wery M, Legendre R, Gautheret D, Morillon A (2015) VING: A software for visualization of deep sequencing signals. *BMC Res Notes* 8: 419. doi:10.1186/s13104-015-1404-5

Didion JP, Martin M, Collins FS (2017) Atropos: Specific, sensitive, and speedy trimming of sequencing reads. *PeerJ* 5: e3720. doi:10.7717/peerj.3720

Djupedal I, Kos-Braun IC, Mosher RA, Soderholm N, Simmer F, Hardcastle TJ, Fender A, Heidrich N, Kagansky A, Bayne E, et al (2009) Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. *EMBO J* 28: 3832–3844. doi:10.1038/emboj.2009.351

Drinnenberg IA, Fink GR, Bartel DP (2011) Compatibility with killer explains the rise of RNAi-deficient fungi. *Science* 333: 1592. doi:10.1126/science.1209575

Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP (2009) RNAi in budding yeast. *Science* 326: 544–550. doi:10.1126/science.1176945

Ha M, Kim VN (2014) Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* 15: 509–524. doi:10.1038/nrm3838

Houseley J, LaCava J, Tollervey D (2006) RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* 7: 529–539. doi:10.1038/nrm1964

Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M (2008) A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Mol Cell* 32: 685–695. doi:10.1016/j.molcel.2008.09.027

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

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14: R36. doi:10.1186/gb-2013-14-4-r36

Kopp F, Mendell JT (2018) Functional classification and experimental dissection of long noncoding RNAs. *Cell* 172: 393–407. doi:10.1016/j.cell.2018.01.011

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359. doi:10.1038/nmeth.1923

Lee JT, Lu N (1999) Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* 99: 47–57. doi:10.1016/s0092-8674(00)80061-6

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, et al (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425: 415–419. doi:10.1038/nature01957

Longtine MS, McKenzie A 3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* 14: 953–961. doi:10.1002/(sici)1097-0061(199807)14:10<953::aid-yea293>3.3.co;2-l

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

Marquardt S, Hazelbaker DZ, Buratowski S (2011) Distinct RNA degradation pathways and 3′ extensions of yeast non-coding RNA species. *Transcription* 2: 145–154. doi:10.4161/trns.2.3.16298

Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: Insights into functions. *Nat Rev Genet* 10: 155–159. doi:10.1038/nrg2521

Nagarajan VK, Jones CI, Newbury SF, Green PJ (2013) XRN 5′→3′ exoribonucleases: Structure, mechanisms and functions. *Biochim Biophys Acta* 1829: 590–603. doi:10.1016/j.bbagrm.2013.03.005

Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457: 1038–1042. doi:10.1038/nature07747

Pelechano V, Steinmetz LM (2013) Gene regulation by antisense transcription. *Nat Rev Genet* 14: 880–893. doi:10.1038/nrg3594

Pinskaya M, Gourvennec S, Morillon A (2009) H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J* 28: 1697–1707. doi:10.1038/emboj.2009.108

Renganathan A, Felley-Bosco E (2017) Long noncoding RNAs in cancer and therapeutic potential. *Adv Exp Med Biol* 1008: 199–222. doi:10.1007/978-981-10-5203-3_7

Ries J, Kaplan C, Platonova E, Eghlidi H, Ewers H (2012) A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. *Nat Methods* 9: 582–584. doi:10.1038/nmeth.1991

Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81: 145–166. doi:10.1146/annurev-biochem-051410-092902

Saha P, Verma S, Pathak RU, Mishra RK (2017) Long noncoding RNAs in mammalian development and diseases. *Adv Exp Med Biol* 1008: 155–198. doi:10.1007/978-981-10-5203-3_6

Schmitt AM, Chang HY (2016) Long noncoding RNAs in cancer pathways. *Cancer Cell* 29: 452–463. doi:10.1016/j.ccell.2016.03.010

Schmitt AM, Chang HY (2017) Long noncoding RNAs: At the intersection of cancer and chromatin biology. *Cold Spring Harb Perspect Med* 7: a026492. doi:10.1101/cshperspect.a026492

Solinger JA, Pascolini D, Heyer WD (1999) Active-site mutations in the Xrn1p exoribonuclease of Saccharomyces cerevisiae reveal a specific role in meiosis. *Mol Cell Biol* 19: 5930–5942. doi:10.1128/mcb.19.9.5930

Stevens A, Hsu CL, Isham KR, Larimer FW (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. doi:10.1128/jb.173.21.7024-7028.1991

Swiezewski S, Liu F, Magusin A, Dean C (2009) Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* 462: 799–802. doi:10.1038/nature08618

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

Tudek A, Candelli T, Libri D (2015) Non-coding transcription by RNA polymerase II in yeast: Hasard or necessite? *Biochimie* 117: 28–36. doi:10.1016/j.biochi.2015.04.020

Uhler JP, Hertel C, Svejstrup JQ (2007) A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc Natl Acad Sci U S A* 104: 8011–8016. doi:10.1073/pnas.0702431104

Van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Né P, Loeillet S, et al (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non coding RNA in yeast. *Nature* 475: 114–117. doi:10.1038/nature10118

van Werven FJ, Neuert G, Hendrick N, Lardenois A, Buratowski S, van Oudenaarden A, Primig M, Amon A (2012) Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* 150: 1170–1181. doi:10.1016/j.cell.2012.06.049

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297: 1833–1837. doi:10.1126/science.1074973

Watts BR, Wittmann S, Wery M, Gautier C, Kus K, Birot A, Heo DH, Kilchert C, Morillon A, Vasiljeva L (2018) Histone deacetylation promotes transcriptional silencing at facultative heterochromatin. *Nucleic Acids Res* 46: 5426–5440. doi:10.1093/nar/gky232

Wery M, Descrimes M, Vogt N, Dallongeville AS, Gautheret D, Morillon A (2016) Nonsense-mediated decay restricts LncRNA levels in yeast unless blocked by double-stranded RNA structure. *Mol Cell* 61: 379–392. doi:10.1016/j.molcel.2015.12.020

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

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

Wery M, Kwapisz M, Morillon A (2011) Noncoding RNAs in gene regulation. *Wiley Interdiscip Rev Syst Biol Med* 3: 728–738. doi:10.1002/wsbm.148

Woolcock KJ, Buhler M (2013) Nuclear organisation and RNAi in fission yeast. *Curr Opin Cell Biol* 25: 372–377. doi:10.1016/j.ceb.2013.02.004

Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, 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. doi:10.1016/j.cell.2005.04.030

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457: 1033–1037. doi:10.1038/nature07728

Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* 38: 662–674. doi:10.1016/j.molcel.2010.03.021

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Annotation of novel aslncRNAs in *N. castellii*. (A) Scatter plot of tag density for mRNAs (light grey), sn(o)RNAs (black), SUTs (dark grey), and DUTs (red) in the WT and *dcr1Δ* strains. Results are presented as $\log_2$ of density, expressed in tag/nt. The red line indicates no change (mutant/WT ratio = 1). (B) Scatter plot of tag density for mRNAs (light grey), sn(o)RNAs (black), SUTs (dark grey), and XUTs (red) in the WT and *xrn1Δ* strains. The data are presented as above. (C) Scatter plot of tag density for mRNAs (light grey), sn(o)RNAs (black), SUTs (dark grey), and CUTs (green) in the WT and *rrp6Δ* strains. The raw RNA-Seq data have been previously published (Alcid & Tsukiyama, 2016). The presentation of the results is as above. (D) Scatter plot of tag density for mRNAs (light grey) and DUTs (red) in the *dcr1Δ* and *xrn1Δ* strains. The results are presented as above. (E) Venn diagram showing the overlap (≥50%) between CUTs and XUTs. (F) Overlap (≥1 nt) between the SUTs, XUTs, and CUTs identified in this work and the 170 previously annotated aslncRNAs (Alcid & Tsukiyama, 2016). For each class lncRNAs, the number of transcripts overlapping ≥1 nt of a previously annotated aslncRNA (*) is represented as a light grey bar; the number of novel transcript is represented as a dark grey bar. (G) Proportion of SUTs, CUTs, and XUTs that are antisense (≥1 nt) to an ORF (white bars) or to any annotated transcript (black bars). (H) Box plot of densities (tag/nt, $\log_2$ scale) for the antisense and solo XUTs in *xrn1Δ* cells. The P-value (adjusted for multiple testing with the Benjamini-Hochberg procedure) obtained upon two-sided Wilcoxon rank-sum test is indicated. Outliers: not shown. (I) Same as above for the antisense and solo CUTs in *rrp6Δ* cells. The raw RNA-Seq data have been previously published (Alcid & Tsukiyama, 2016).

Figure S2. Mutants of *dcr1* and *xrn1* display synergic defects. (A) Snapshot of total RNA-Seq signal (tag/nt, $\log_2$ scale) along the C05770/CUT0672 and XUT0527/C05780 loci in WT, *xrn1Δ*, *dcr1Δ*, *xrn1Δ dcr1Δ*, *rr*p6Δ, and *rrp6Δ dcr1Δ* strains. The raw data for *rr*p6Δ and *rrp6Δ dcr1Δ* have been previously published (Alcid & Tsukiyama, 2016). The signals for the + and − strands are visualized as heat maps in the upper and lower panels, respectively, using the VING software (Descrimes et al, 2015). The color turns from yellow to dark blue as the signals increase (see scale under the heat map). The protein-coding genes, the CUT, and the XUT are represented as blue, green, and red arrows, respectively. (B) Same as above for the A12440/CUT0275 and XUT0213/A12460 loci. (C) Histogram of the RNA-Seq signals (tag/nt) shown in Fig S2A for CUT0672 (green bars) and XUT0527 (red bars) in the WT, *xrn1Δ*, *dcr1Δ*, *xrn1Δ dcr1Δ*, *rrp6Δ*, and *rrp6Δ dcr1Δ* strains. Data are presented as mean ± SEM (calculated from the two biological replicates for each strain used in the RNA-Seq analysis). (D) Same as above for CUT0275 (green bars) and XUT0213 (red bars) shown in Fig S2B. (E) Number of Dcr1-sensitive protein-coding genes in the presence or absence of Xrn1 or Rrp6. For each combination of strains, the Dcr1-sensitive protein-coding genes were identified on the basis of a fold-change <0.5 (down-regulated gene) or ≥2 (up-regulated gene), with a significant P-value (<0.05; adjusted for multiple testing with

the Benjamini–Hochberg procedure) upon differential analysis using DESeq2 (Love et al, 2014). The raw data for *rrp6Δ* and *rrp6Δ dcr1Δ* have been previously published (Alcid & Tsukiyama, 2016). (F) Growth curves. Exponentially growing WT (YAM2478), *xrn1Δ* (YAM2479), *dcr1Δ* (YAM2795), and *xrn1Δ dcr1Δ* (YAM2796) cells were diluted to OD_{600} 0.1 in preheated rich medium (YPD), at 25°C. OD_{600} was then measured every hour. OD_{600} at time 0 was set to 1, for each strain. Data are expressed in a log scale.

Figure S3. AslncRNAs are processed into 22–23-nt small RNAs by Dcr1 in *N. castellii*. (A) Size and first base distribution of uniquely mapped small RNA reads in the WT (YAM2478), *xrn1Δ* (YAM2479), *dcr1Δ* (YAM2795), and *xrn1Δ dcr1Δ* (YAM2796) strains. Reads matching tRNAs or rRNAs are excluded. Libraries were constructed using purified small RNAs. (B, C, D) Box plot of 22–23-nt small RNAs densities (tag/nt, $\log_2$ scale) for the antisense (light grey) and solo (dark grey) SUTs (B), XUTs (C), and CUTs (D) in the WT, *xrn1Δ*, *dcr1Δ*, and *xrn1Δ dcr1Δ* strains. Outliers: not shown. (E) Snapshot of small RNAs produced from the A12440/CUT0275 and XUT0213/A12460 pairs. Densities of 22–23-nt small RNAs are shown in a separate panel for each strain. In each panel, signals for the + and − strands are shown in blue and pink, respectively. The protein-coding genes, the CUT, and the XUT are represented by blue, green, and red arrows, respectively. The dashed boxes highlight the region of overlap between the aslncRNAs and the paired-sense mRNAs. The snapshot was produced using VING (Descrimes et al, 2015).

Figure S4. Dcr1-GFP localizes in the cytoplasm. (A) Verification of Dcr1-GFP expression by Western blot. Protein extracts from YAM2478 (WT), YAM2826 (Dcr1-GFP), and YAM2842 (GFP) cells were separated by poly-acrylamide gel electrophoresis and then transferred to a nitrocellulose membrane. The Dcr1-GFP and the GFP bands are indicated by arrows. The size of the protein ladder bands is indicated on the left of the panel. Pgk1 was used as a loading control. (B) Visualization of Dcr1-GFP in living cells. YAM2478 (WT), YAM2826 (Dcr1-GFP), and YAM2842 (GFP) cells were grown to mid-log phase in CSM at 25°C, harvested and then washed in sterile water before direct observation. Scale bars: 1 μm. (C) Size and first base distribution of small RNAs produced in the Dcr1-GFP (YAM2826) strain, computed as described in Fig S3A.

Figure S5. Comparison between SUTs/XUTs and previously annotated aslncRNAs. Box plot representation of the size (nt) distribution for the asSUTs (median = 448 nt), asCUTs (median = 444 nt), and the asXUTs (median = 670 nt) identified in this work and for the 170 aslncRNAs (median = 321 nt) previously annotated in *N. castellii* (Alcid & Tsukiyama, 2016). The red dashed line indicates the 200-nt threshold commonly used as the minimal size of lncRNAs. Outliers: not shown.

Figure S6. Identification and deletion of XRN1 in *N. castellii*. (A) CLUSTALO alignment between the N-terminal region of *S. cerevisiae* and *N. castellii* Xrn1 proteins. Identities are indicated using stars. Residues of the three 5'-3' exonuclease motifs (Solinger et al, 1999) are in bold. The D206 and D208 residues shown to abolish Xrn1 exonuclease activity in vivo upon mutation in alanine (Solinger et al, 1999) are highlighted in red. (B) Validation of XRN1 deletion by Northern blot. Transcripts were detected from total RNA extracted from YAM2478 (WT) and YAM2479 (*xrn1Δ*) cells, using 32P-labelled oligonucleotides (listed in Table S4). The 5' ITS1 fragment is a by-product of pre-rRNA processing which is physiologically targeted by Xrn1 in *S. cerevisiae* (Stevens et al, 1991).

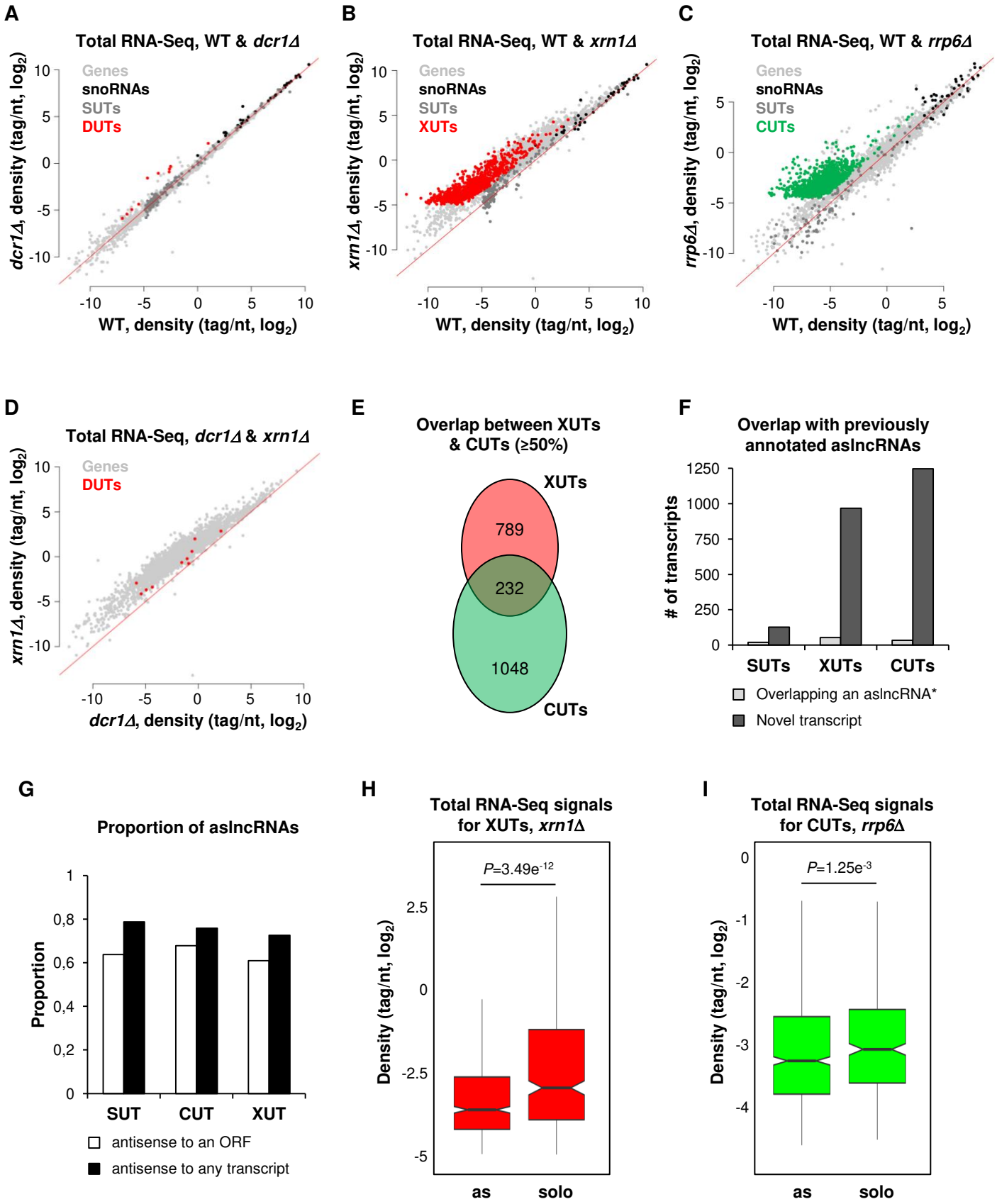

A
Total RNA-Seq, WT & dcr1Δ
Genes
snoRNAs
SUTs
DUTs
dcr1Δ, density (tag/nt, log2)
WT, density (tag/nt, log2)
B
Total RNA-Seq, WT & xrn1Δ
Genes
snoRNAs
SUTs
XUTs
xrn1Δ, density (tag/nt, log2)
WT, density (tag/nt, log2)
C
Total RNA-Seq, WT & rrp6Δ
Genes
snoRNAs
SUTs
CUTs
rrp6Δ, density (tag/nt, log2)
WT, density (tag/nt, log2)
D
Total RNA-Seq, dcr1Δ & xrn1Δ
Genes
DUTs
xrn1Δ, density (tag/nt, log2)
dcr1Δ, density (tag/nt, log2)
E
Overlap between XUTs & CUTs (≥50%)
XUTs
789
232
1048
CUTs
F
Overlap with previously annotated asIncRNAs
# of transcripts
SUTs
XUTs
CUTs
Overlapping an asIncRNA*
Novel transcript
G
Proportion of asIncRNAs
Proportion
SUT
CUT
XUT
antisense to an ORF
antisense to any transcript
H
Total RNA-Seq signals for XUTs, xrn1Δ
P=3.49e-12
Density (tag/nt, log2)
as
solo
I
Total RNA-Seq signals for CUTs, rrp6Δ
P=1.25e-3
Density (tag/nt, log2)
as
solo

Figure S1

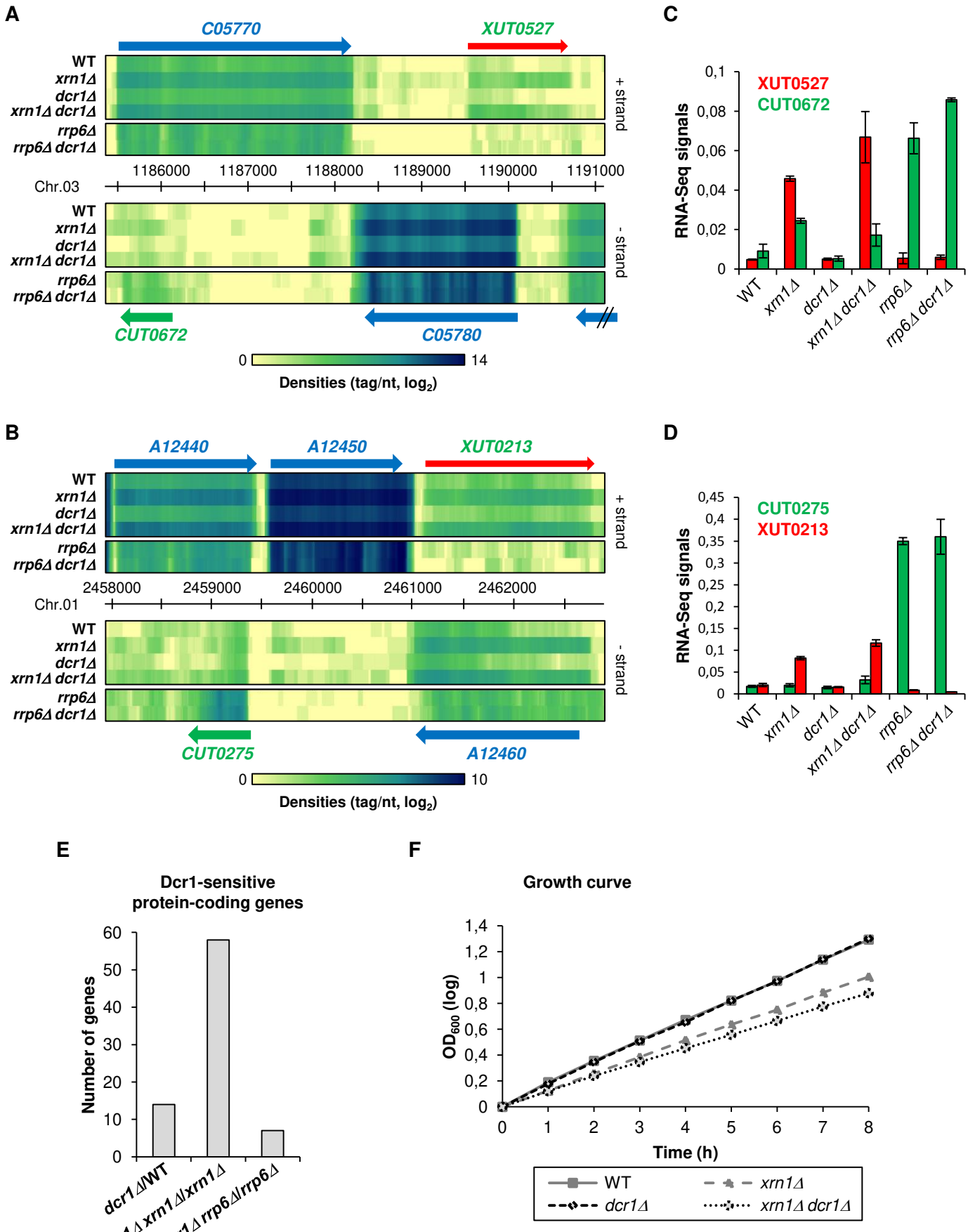

A
C05770
XUT0527
WT
xrn1Δ
dcr1Δ
xrn1Δ dcr1Δ
rrp6Δ
rrp6Δ dcr1Δ
+ strand
1186000
1187000
1188000
1189000
1190000
1191000
Chr.03
- strand
CUT0672
C05780
0
14
Densities (tag/nt, log2)
B
A12440
A12450
XUT0213
2458000
2459000
2460000
2461000
2462000
Chr.01
CUT0275
A12460
0
10
Densities (tag/nt, log2)
C
XUT0527
CUT0672
RNA-Seq signals
0,1
0,08
0,06
0,04
0,02
0
WT
xrn1Δ
dcr1Δ
xrn1Δ dcr1Δ
rrp6Δ
rrp6Δ dcr1Δ
D
CUT0275
XUT0213
RNA-Seq signals
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
E
Dcr1-sensitive protein-coding genes
Number of genes
60
50
40
30
20
10
0
dcr1Δ/WT
dcr1Δ xrn1Δ/xrn1Δ
dcr1Δ rrp6Δ/rrp6Δ
F
Growth curve
OD600 (log)
1,4
1,2
1
0,8
0,6
0,4
0,2
0
0
1
2
3
4
5
6
7
8
Time (h)
WT
xrn1Δ
dcr1Δ
xrn1Δ dcr1Δ

Figure S2

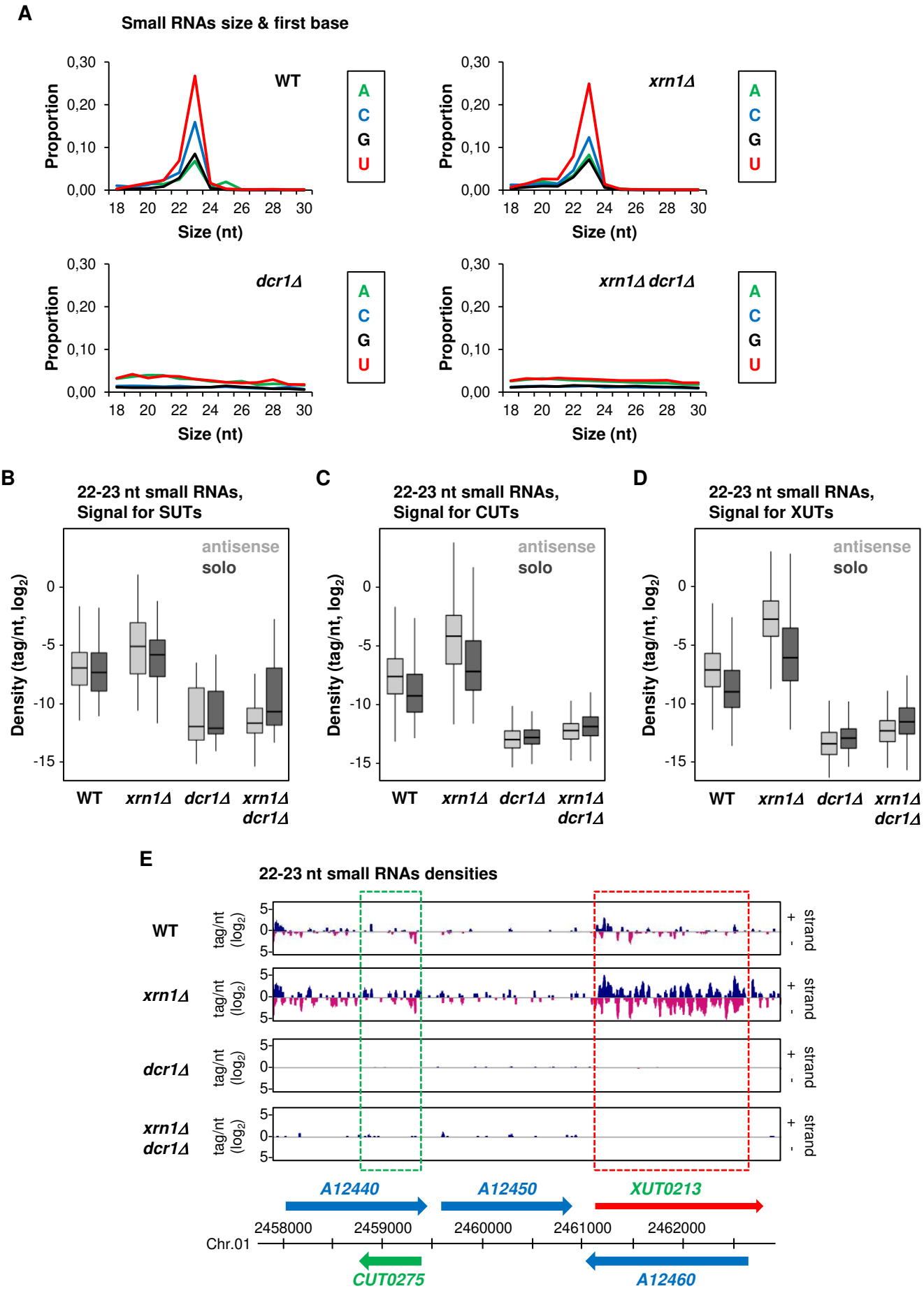

A
Small RNAs size & first base
WT
xrn1Δ
dcr1Δ
xrn1Δ dcr1Δ
Proportion
Size (nt)
A
C
G
U
B
22-23 nt small RNAs,
Signal for SUTs
C
22-23 nt small RNAs,
Signal for CUTs
D
22-23 nt small RNAs,
Signal for XUTs
antisense
solo
Density (tag/nt, log2)
WT
xrn1Δ
dcr1Δ
xrn1Δ
dcr1Δ
E
22-23 nt small RNAs densities
tag/nt (log2)
strand
A12440
A12450
XUT0213
Chr.01
CUT0275
A12460

Figure S3

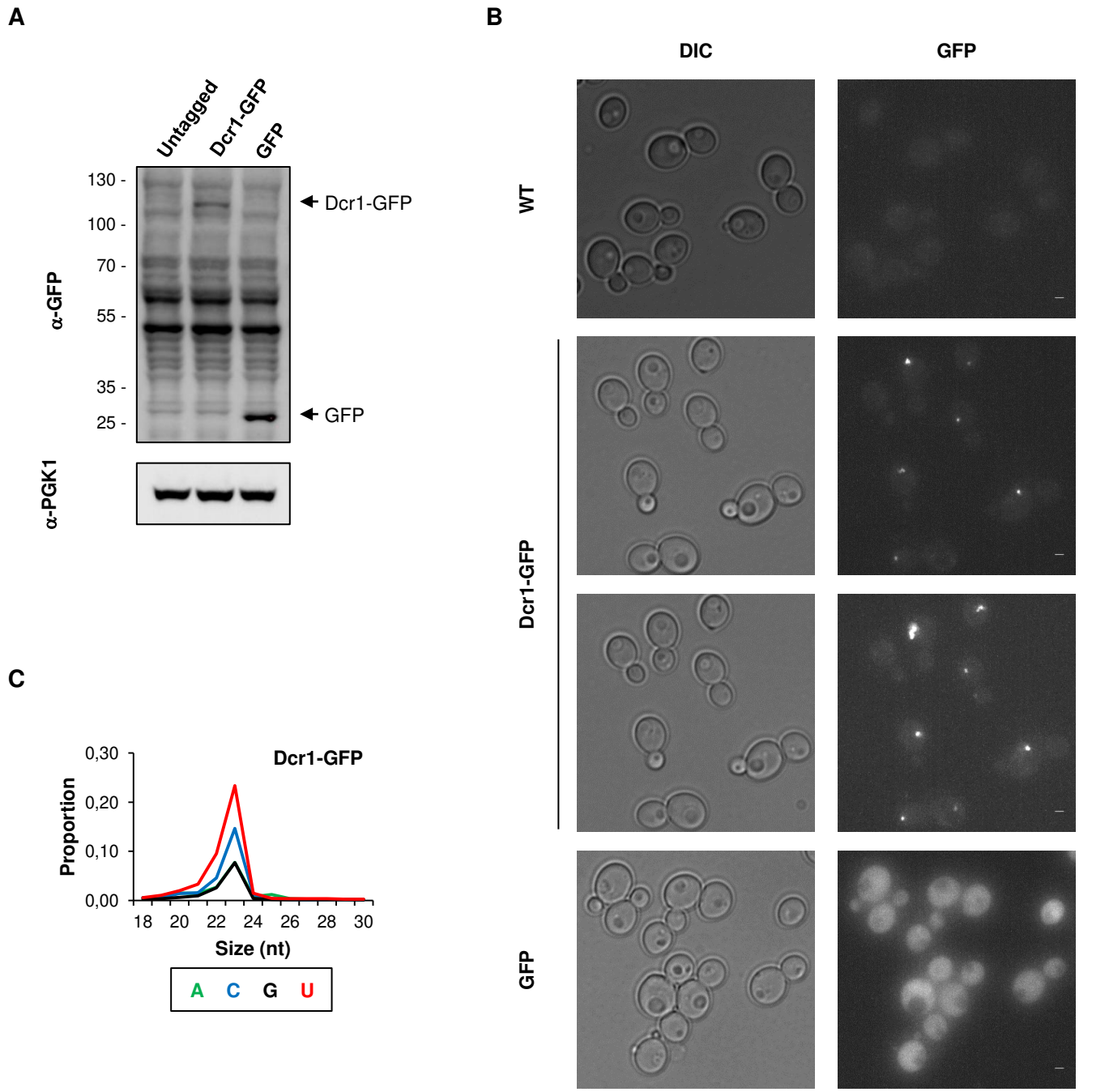

A
Untagged
Dcr1-GFP
GFP
130 -
100 -
70 -
55 -
35 -
25 -
Dcr1-GFP
GFP
α-GFP
α-PGK1
B
DIC
GFP
WT
Dcr1-GFP
GFP
C
Dcr1-GFP
Proportion
0,30
0,20
0,10
0,00
18 20 22 24 26 28 30
Size (nt)
A C G U

Figure S4

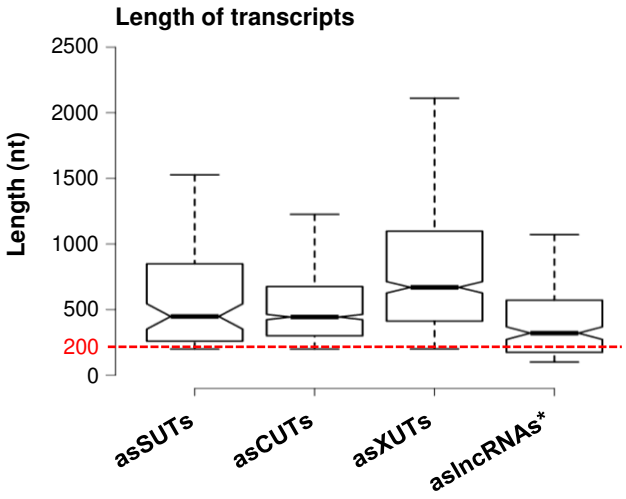

Length of transcripts
Length (nt)
2500
2000
1500
1000
500
200
0
asSUTs
asCUTs
asXUTs
asIncRNAs*

Figure S5

A

```
S.cer.    1 MGIPKFFRYISERWPMILQLIEGTQIPEFDNLYLDMNSILHNCTHGNDDDVTKRLTEEEV 60
N.cas.    1 MGIPKFFRYISERWPMILQLIEGTQIPEFDNLYLDMNSILHTCTHGNDDDVTKRMTEEEV 60
            ************************************** ************ *****

S.cer.   61 FAKICTYIDHLFQTIKPKKIFYMAIDGVAPRAKMNQQRARRFRTAMDAEKALKKAIENGD 120
N.cas.   61 FAKIFTYIDHLFLTIKPKKTFYMAIDGVAPRAKMNQQRSRRFRTAMDAEHALQKAIDHGE 120
            **** ******* ****** ***************** ********** ** ***  *

S.cer.  121 EIPKGEPFDSNSITPGTEFMAKLTKNLQYFIHDKISNDSKWREVQIIFSGHEVPGEGEHK 180
N.cas.  121 EIPKGEPFDSNSITPGTEFMAKLTKNLKYFIHDKISNDAKWREIDIIFSGHEVPGEGEHK 180
            ************************** ********** ****  ***************

S.cer.  181 IMNFIRHLKSQKDFNQNTRHCIYGLDADLIMLGLSTHGPHFALLREEVTFGRRNSEK-KS 239
N.cas.  181 IMDFIRRITAEKDFDENTRHCIYGLDADLIILGLSTHAPHFALLREEVVFGRRNSNKVKT 240
            ** ***     ***  ************** ****** ********** ****** * *

S.cer.  240 LEHQNFYLLHLSLLREYMELEFKEIADEMQFEYNFERILDDFILVMFVIGNDFLPNLPDL 299
N.cas.  241 LENQNFYLLHLSLLREYMELEFSEIADEMQFPFDFERVLDDFILVMFVIGNDFLPNLPDL 300
            ** ******************** ********   *** **********************

S.cer.  300 HLNKGAFPVLLQTFKEALLHTDGYINEHGKINLKRLGVWLNYLSQFELLNFEKDDIDVEW 359
N.cas.  301 HLNKGAFPVILQTFKEALLHLDGYINEHGKINLERLRVWFQYLSQFELLNFEKSDIDVEW 360
            ********* ********** ************ ** **  ************ ******
```

B

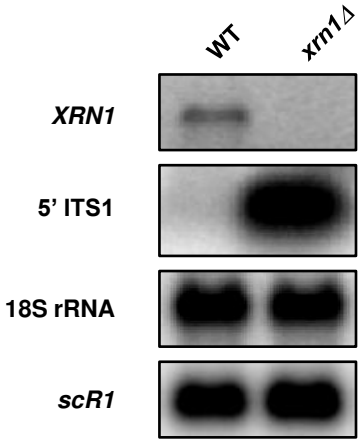


Figure S6

3. Discussion

A characteristic distinguishing *S. cerevisiae* from most other Eukaryotes is the loss of the RNAi system during evolution, which might have allowed the expansion of aslncRNAs (Alcid and Tsukiyama, 2016). This idea motivated us to address the question of asXUTs degradation in *N. castellii*, a budding yeast endowed with RNAi and compare it to *S. cerevisiae*. By examining the interplay between the nuclear exosome, Xrn1 and Dicer, our data showed that aslncRNAs decay in this species depended on both the nuclear exosome and Xrn1 (with no major effect of Dicer). Nevertheless, the mutants of Xrn1 and Dicer displayed synergic growth defects, indicating that Dicer becomes critical in the absence of Xrn1. If compared to *S. cerevisiae*, the presence of cytoplasmic RNAi machinery in *N. castellii* reinforced the nuclear RNA surveillance machinery to prevent aslncRNA-mRNA pairs from becoming RNAi targets. Our results provide insight into adaptation strategies that allow coordination between RNAi and other RNA surveillance pathways.

In the fission yeast *S. pombe*, Dcr1 has been shown to target nearly all sense-antisense RNA pairs when it is overexpressed. Browsing the *N. castellii* genome using our dataset reveals that it is not the case in the budding yeast species. For example, from C05770/C05780 mRNAs no 22-23 nt short RNAs are generated (Fig 3B). The reason for Dcr1 difference in targeting of sense-antisense RNA pairs between fission yeast *S. pombe* and budding yeast *N. castellii* is not clear, but it could be due to differences in the regulation of read-through transcription or in the activity of Dcr1 in these two yeast species.

Dicer has been conserved in some budding yeast species. However, deleting it in *N. castellii* confers no detectable growth defect, as shown under 50 different conditions, suggesting that Dicer may not be essential for the growth and survival of *N. castellii*. However, in our work we found that Dicer become critical in the absence of Xrn1, as the *dcr1Δ* and *xrn1Δ* mutans display synergic growth defects. Potentially, this phenotypic defect could be extended to genome scale analysis by performing RNA-Seq in the tested conditions to see if the asXUTs increase to a higher level in the double mutant, compared with the single mutant. In contrast to this synergic effect, Dcr1 was found to be deleterious in Rrp6-lacking cells as its deletion partially suppressed the growth defect of the *rrp6Δ* mutant. It is unclear whether these contrasting effects of Dcr1 deletion in the *xrn1Δ* and *rrp6Δ* backgrounds are related to siRNA production. Further analysis are needed to understand the molecular mechanisms behind these genetic interactions. For example, one can imagine identifying specific lncRNAs that are stabilized in the *xrn1Δ* and *rrp6Δ* backgrounds and comparing them to the lncRNAs stabilized in the *dcr1Δ* background. Further, one could test if Dcr1, Rrp6 and Xrn1 interact physically and if this interaction is important for their function by using co-immunoprecipitation or yeast two-hybrid assay.

Moreover, works from *S. cerevisiae* showed that the abundant cytoplasmic dsRNA formation interferes with the sensitivity of XUTs to NMD. However, the lack of RNAi in this organism might have allowed an excessive accumulation of dsRNAs. To further test the effect of dsRNA and NMD on asXUTs, *N. castellii* was an excellent model as harboring a cytoplasmic RNAi pathway, in which normally cytoplasmic dsRNAs are eliminated. Therefore, after annotating XUTs in this species, the lab further tested the potential evolutionary conservation of the effect of NMD on aslncRNAs in *N. castellii*.

In line with NMD-sensitive XUTs from *S. cerevisiae*, the preliminary unpublished RNA-Seq data obtained by Maxime Wery and Ugo Szachnowski, performed in cells defective for NMD, showed that 37,5 % of XUTs (383) significantly accumulated in the absence of Upf1 (*upf1Δ*/WT ratio $\geq$2, P -value $\leq$0.05) in *N. castellii*. Interestingly, in this dataset Dicer was found to be an NMD-sensitive mRNA. This preliminary finding is intriguing as the degradation by NMD of Dicer could have an impact on the RNAi pathway. This finding highlights the complexity of the regulation of the XUTs in *N. castellii*, and the interactions between different pathways that are involved in their degradation. Creation of a yeast strain that is defective for both Upf1 and Dcr1 could help to understand the interplay between NMD and RNAi pathways in the regulation of aslncRNAs.

Altogether, the results obtained in this work indicate that the role of the nuclear Exosome, Xrn1 and NMD in aslncRNA decay is conserved in yeast (Figure 12). The conserved role of the decay machineries reinforces the existence of an evolutionary pressure for quality and quantity control of transcripts, including those annotated as 'non-coding'. This conclusion implies that the aslncRNAs have important biological functions and that their stability and decay are tightly regulated to ensure the proper functioning of the cell.

Note: The proposed schematic representation in Figure 12 is further discussed after (see discussion and perspectives).

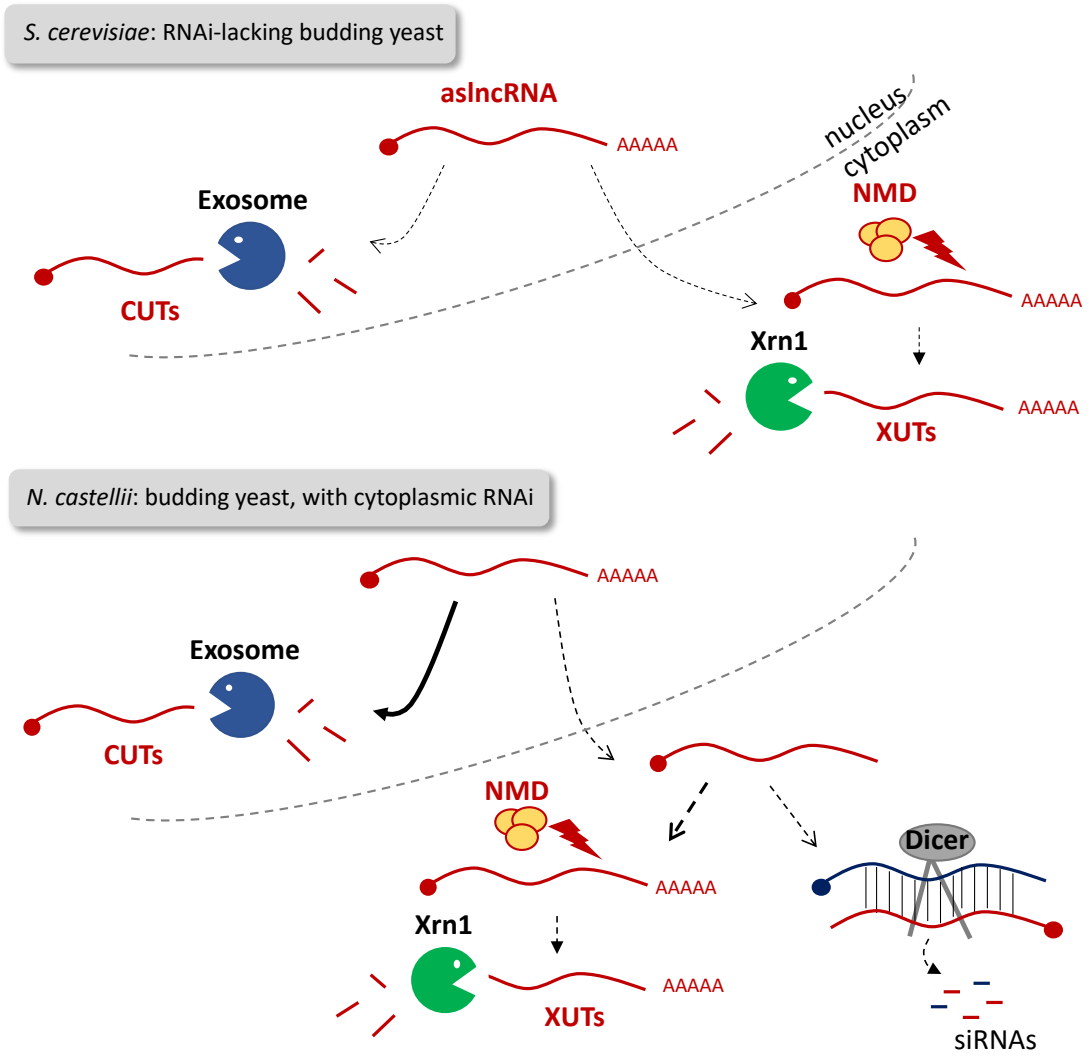


Figure 12 | Schematic Representation of the Antisense Landscape in *S. cerevisiae* and *N. castellii*. Presented are the conserved roles of RNA decay machineries (Exosome, Xrn1 and NMD) in restricting cryptic aslncRNA levels.

Chapter 4. Role of Translation in the Metabolism of Long Non-Coding RNAs

1. Introduction

Previous works from our lab in the budding yeast *S. cerevisiae* have shown that XUTs are synthesized by RNAPII (Van Dijk et al., 2011), capped, and poly-adenylated (Wery et al., 2016), as mRNAs. Importantly, prior to degradation by Xrn1, XUTs are decapped by the decapping enzyme Dcp2 (Wery et al., 2016), a pre-requisite for Xrn1 activity for mRNA. Interestingly, most XUTs (73%) are specifically targeted by NMD (Wery et al., 2016). The NMD is a conserved translation-dependent pathway, it gets activated upon aberrant translation termination, and targets mRNAs with premature stop codon and/or long 3' UTR regions. The sensitivity of XUTs to the mutants of the NMD (*upf1Δ, upf2Δ and upf3Δ*) was a first indication that cytoplasmic XUTs could be templates for translation, which would consequently have an impact on their stability. Supporting this idea, the analysis of the pioneer ribosome profiling (Ribo-Seq) dataset (Smith et al., 2014) mapped ribosomes footprints in the 5' region of XUTs, followed by a downstream long 3' UTR region free of ribosomes, a signal known to activate NMD for mRNAs. This result was one of the first proof-of-concepts that yeast non-coding RNAs can be bound by ribosomes. However, the extent and the biological significance of such pervasive translation was unknown.

In this work, we aimed at deciphering the role of translation on cytoplasmic non-coding RNAs in *S. cerevisiae* using single-gene and genome-wide approaches. We provide and mechanistically describe a model of a translation-dependent decay process for lncRNAs. In collaboration with Dr. Olivier Namy's lab (I2BC, Gif-sur-Yvette), we define the translation landscape of aslncRNAs. Lastly, using reporters we detect a (poly)peptide derived from an NMD-sensitive XUT.

Our results describe mechanistically how the NMD pathway is involved in the clearance of pervasively translated lncRNA transcripts in yeast. Whether peptides derived from ribosome-bound lncRNAs either have any function or represent a source for the evolution of new proteins opens a large conceptual perspective of this study.

2. Publication n°2 “Translation is a Key Determinant Controlling the Fate of Cytoplasmic Long Non-Coding RNAs”

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.

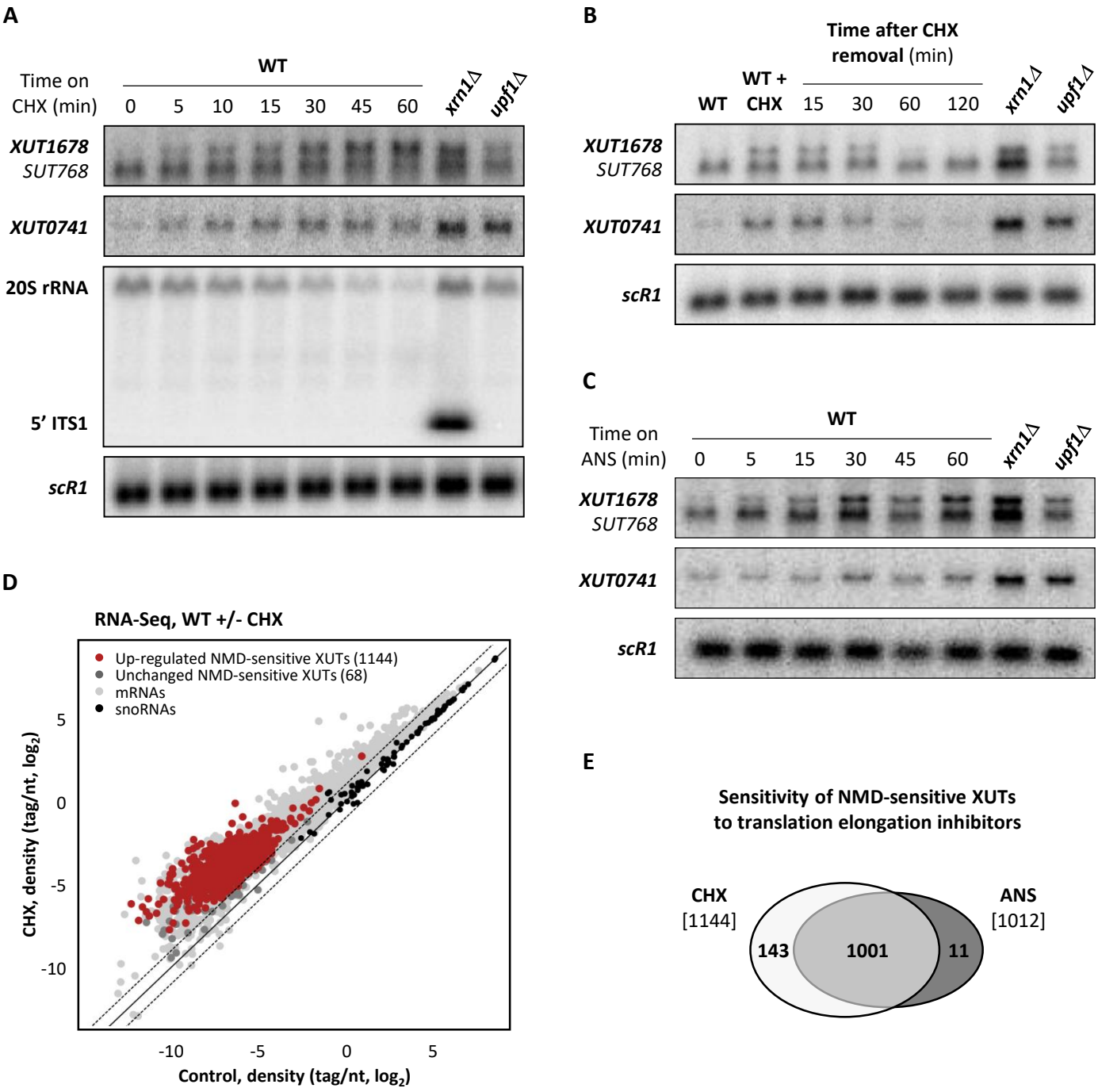

A
Time on CHX (min)
WT
0 5 10 15 30 45 60
xrn1Δ
upf1Δ
XUT1678
SUT768
XUT0741
20S rRNA
5' ITS1
scR1
B
Time after CHX removal (min)
WT
WT + CHX
15 30 60 120
xrn1Δ
upf1Δ
XUT1678
SUT768
XUT0741
scR1
C
Time on ANS (min)
WT
0 5 15 30 45 60
xrn1Δ
upf1Δ
XUT1678
SUT768
XUT0741
scR1
D
RNA-Seq, WT +/- CHX
Up-regulated 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

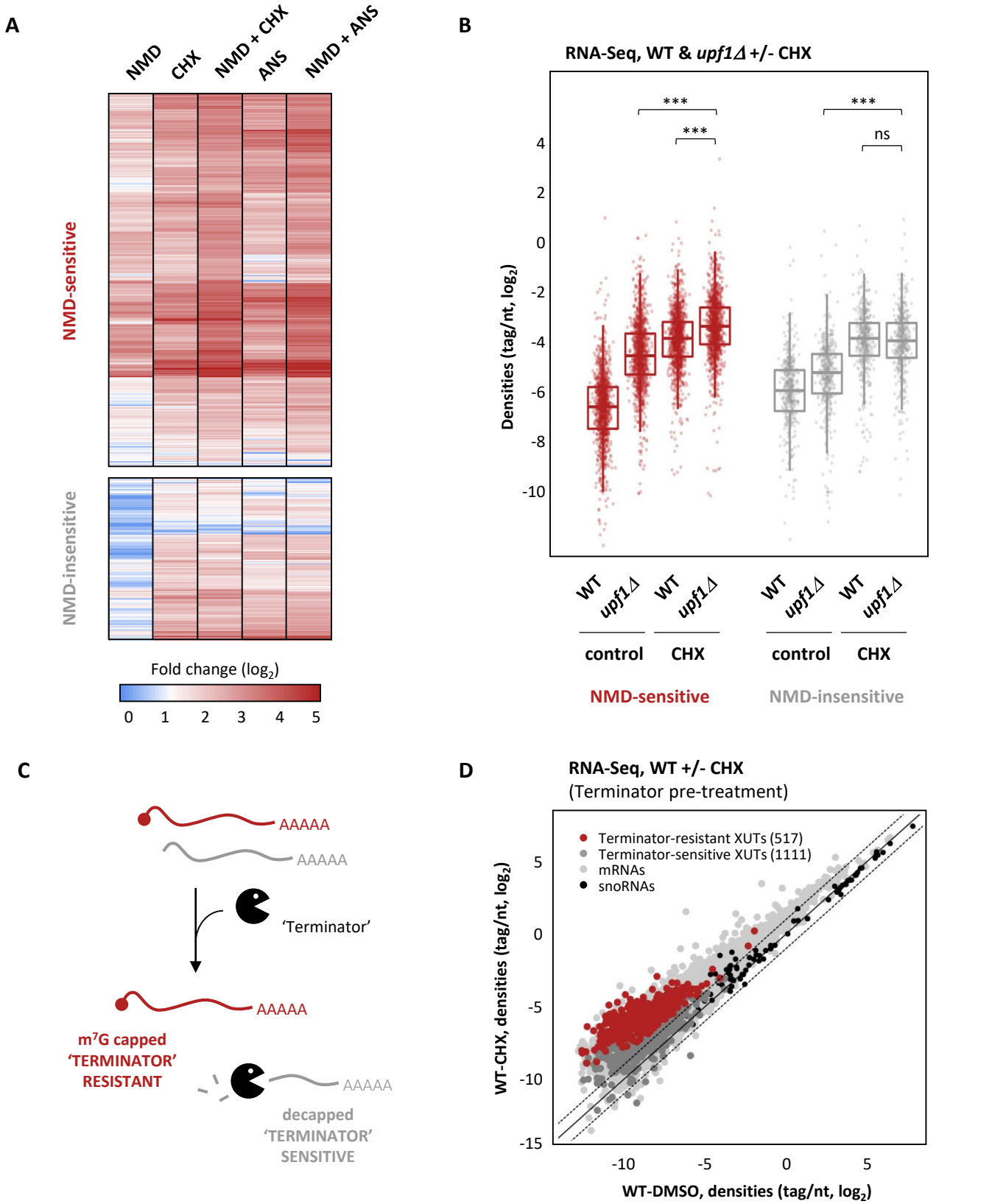

A
NMD
CHX
NMD + CHX
ANS
NMD + ANS
NMD-sensitive
NMD-insensitive
Fold change (log2)
0 1 2 3 4 5
B
RNA-Seq, WT & upf1Δ +/- CHX
***
***
***
ns
Densities (tag/nt, log2)
4 2 0 -2 -4 -6 -8 -10
WT upf1Δ WT upf1Δ WT upf1Δ WT upf1Δ
control CHX control CHX
NMD-sensitive
NMD-insensitive
C
AAAAA
AAAAA
'Terminator'
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

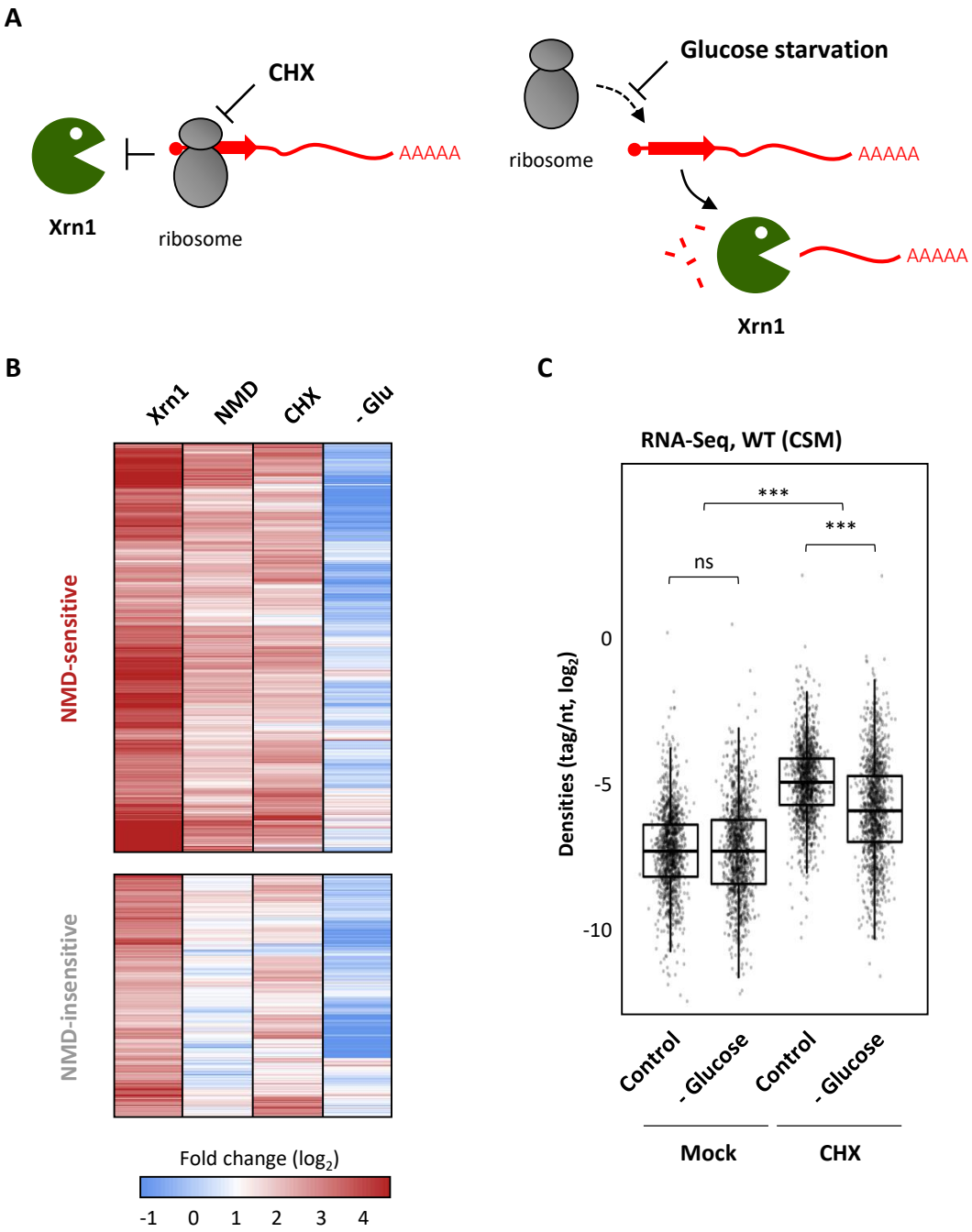

A
CHX
AAAAA
Xrn1
ribosome
Glucose starvation
ribosome
AAAAA
AAAAA
Xrn1
B
Xrn1
NMD
CHX
- Glu
NMD-sensitive
NMD-insensitive
Fold change (log2)
-1 0 1 2 3 4
C
RNA-Seq, WT (CSM)
***
***
ns
Densities (tag/nt, log2)
0
-5
-10
Control
- Glucose
Control
- Glucose
Mock
CHX

Figure 3

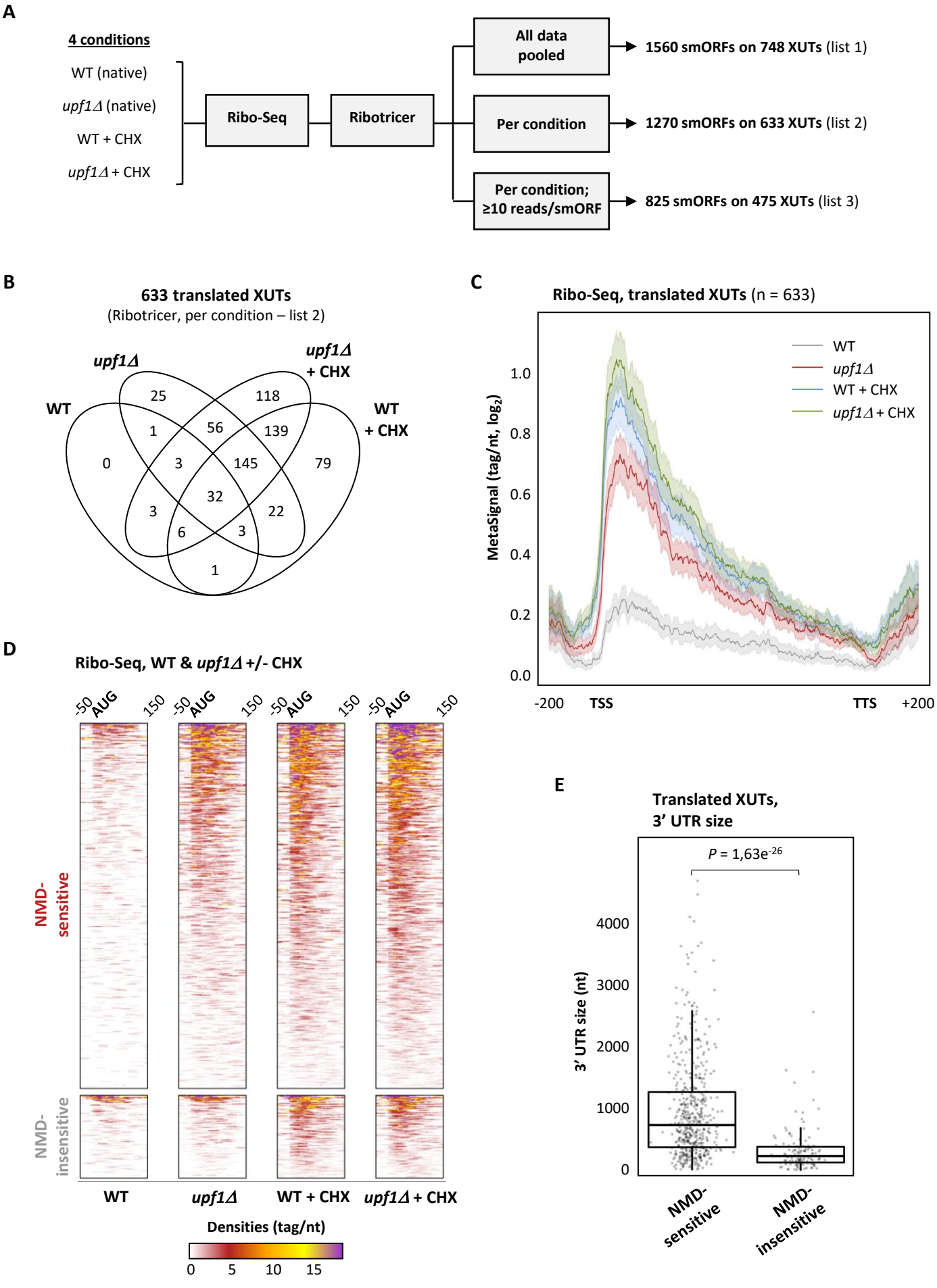

A
4 conditions
WT (native)
upf1Δ (native)
WT + CHX
upf1Δ + CHX
Ribo-Seq
Ribotricer
All data pooled
1560 smORFs on 748 XUTs (list 1)
Per condition
1270 smORFs on 633 XUTs (list 2)
Per condition; ≥10 reads/smORF
825 smORFs on 475 XUTs (list 3)
B
633 translated XUTs
(Ribotricer, per condition – list 2)
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

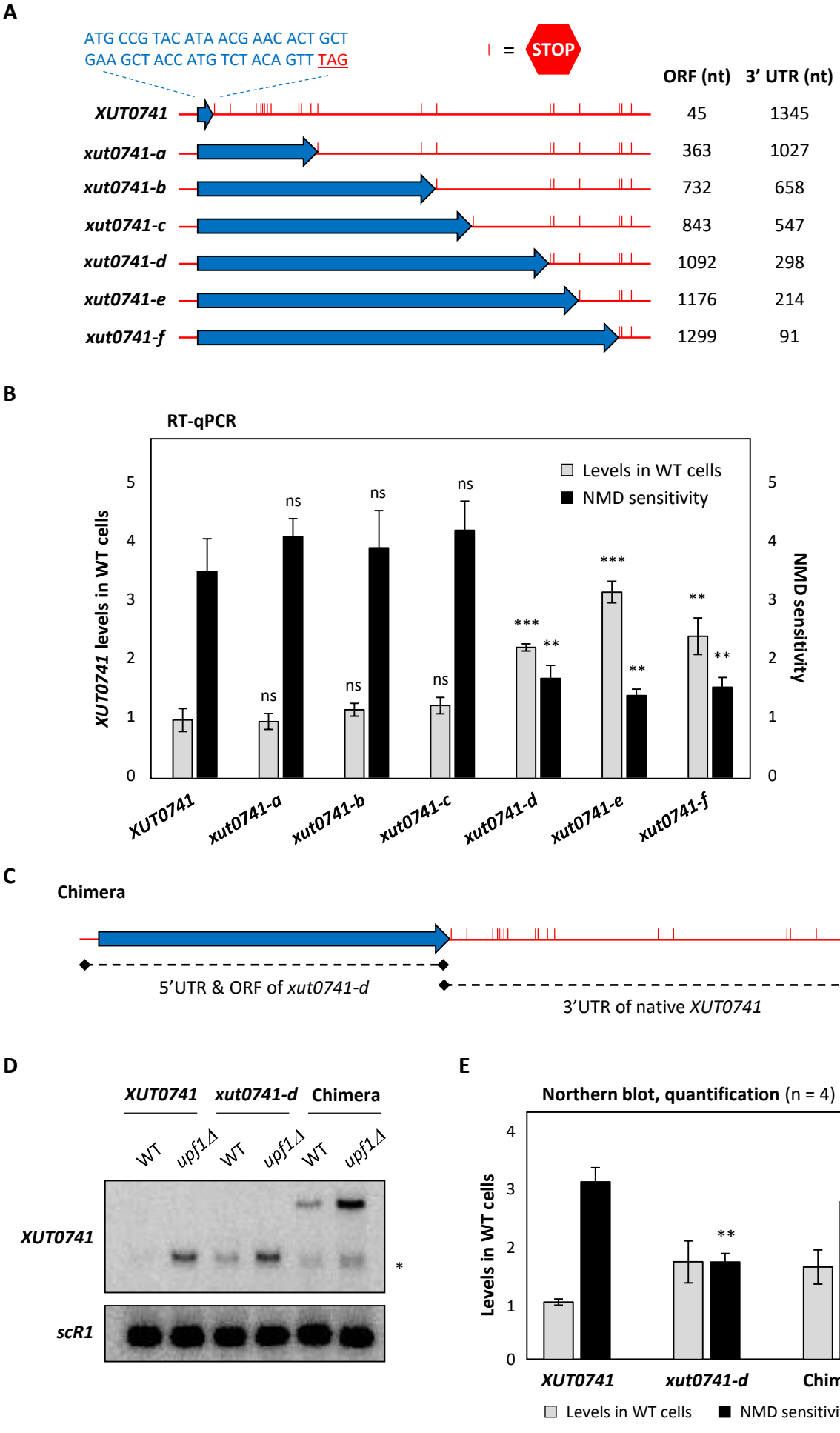

A
ATG CCG TAC ATA ACG AAC ACT GCT
GAA GCT ACC ATG TCT ACA GTT TAG
= STOP
ORF (nt)
3' UTR (nt)
XUT0741 45 1345
xut0741-a 363 1027
xut0741-b 732 658
xut0741-c 843 547
xut0741-d 1092 298
xut0741-e 1176 214
xut0741-f 1299 91
B
RT-qPCR
Levels in WT cells
NMD sensitivity
XUT0741 levels in WT cells
NMD sensitivity
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

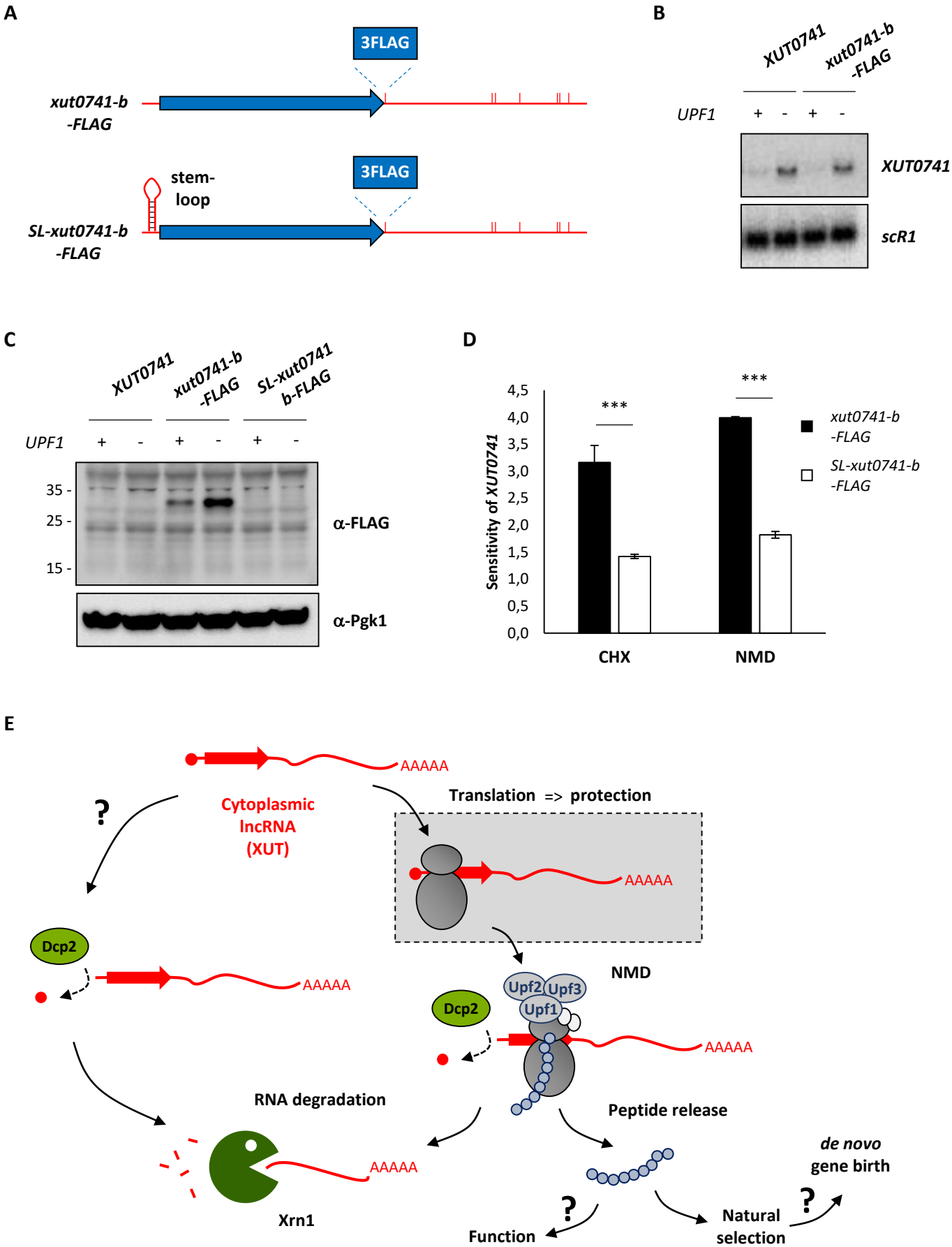
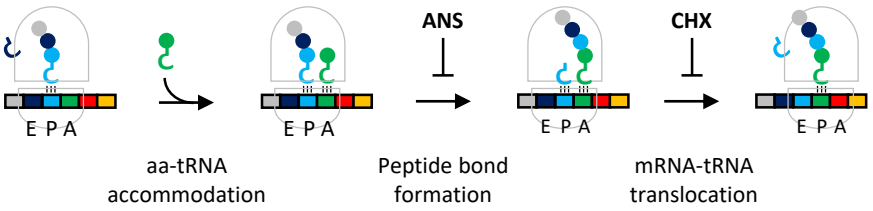
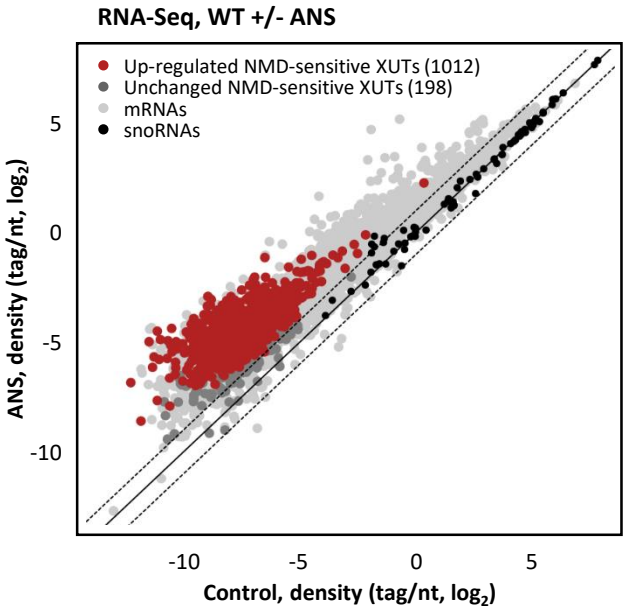

A
3FLAG
xut0741-b
-FLAG
stem-
loop
3FLAG
SL-xut0741-b
-FLAG
B
XUT0741
xut0741-b
-FLAG
UPF1
+ - + -
XUT0741
scR1
C
XUT0741
xut0741-b
-FLAG
SL-xut0741
b-FLAG
UPF1
+ - + - + -
35 -
25 -
15 -
α-FLAG
α-Pgk1
D
Sensitivity of XUT0741
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0
***
***
xut0741-b
-FLAG
SL-xut0741-b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

Figure 6

A



B



C

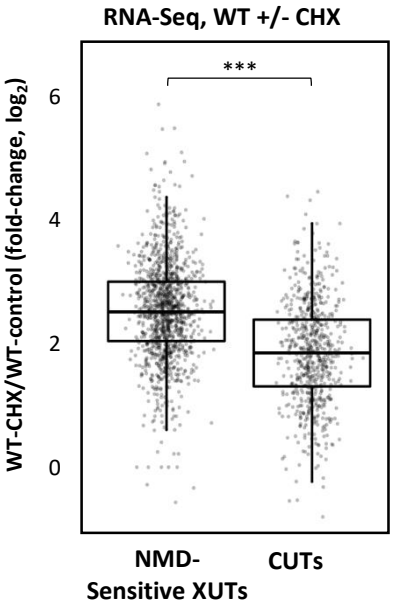


Figure S1

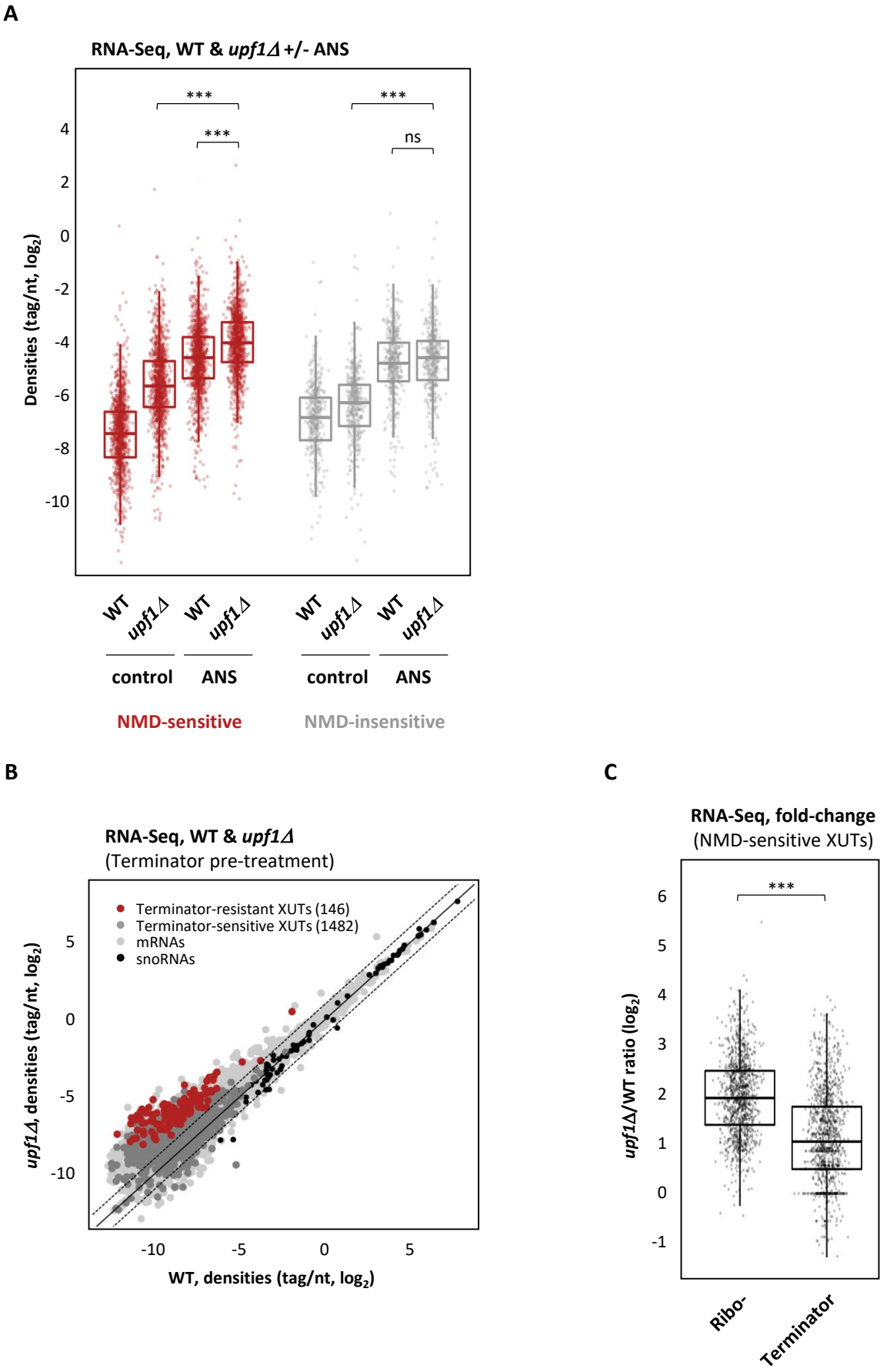
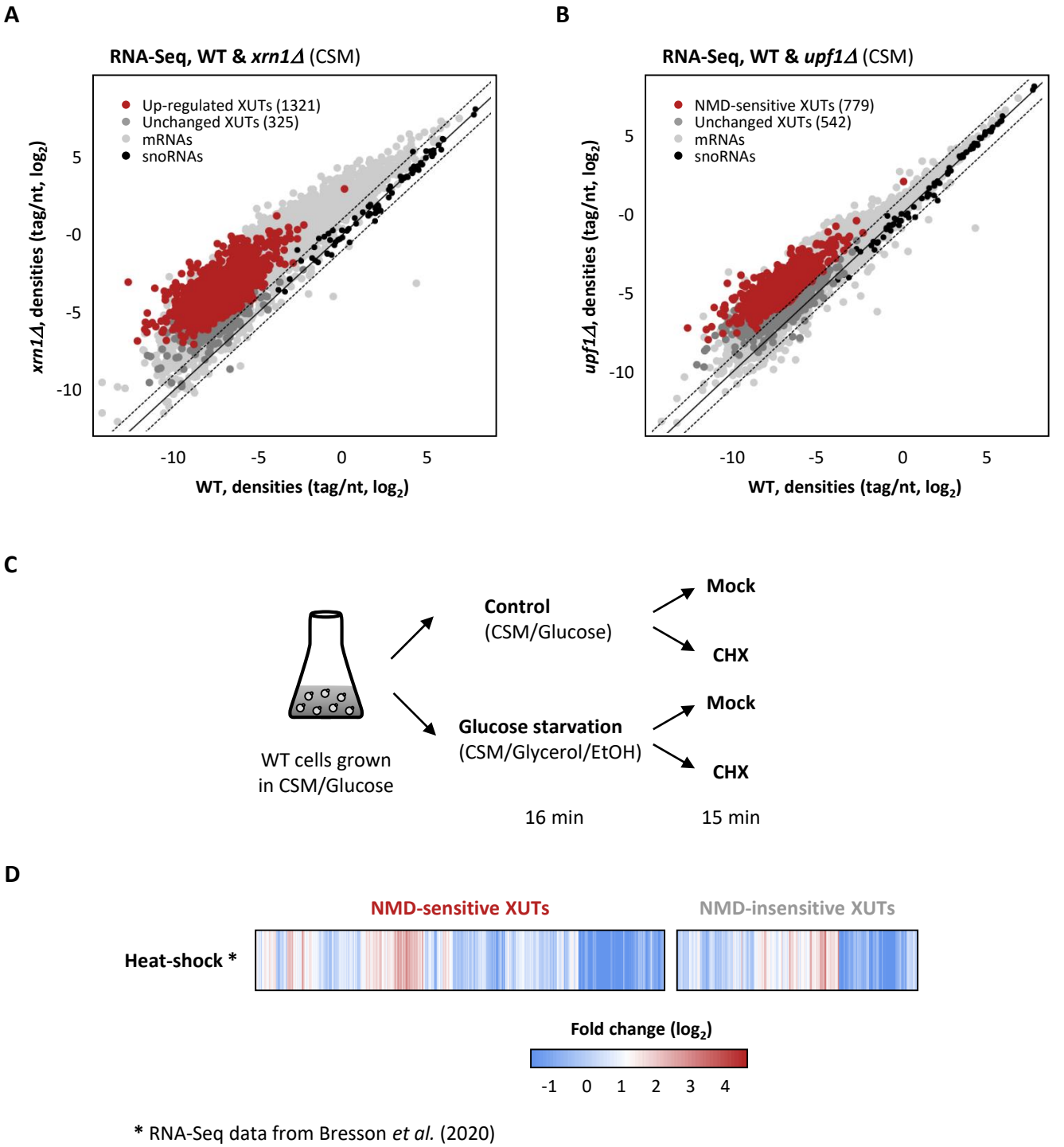

A
RNA-Seq, WT & upf1Δ +/- ANS
Densities (tag/nt, log2)
***
***
***
ns
WT
upf1Δ
WT
upf1Δ
WT
upf1Δ
WT
upf1Δ
control
ANS
control
ANS
NMD-sensitive
NMD-insensitive
B
RNA-Seq, WT & upf1Δ
(Terminator pre-treatment)
Terminator-resistant XUTs (146)
Terminator-sensitive XUTs (1482)
mRNAs
snoRNAs
upf1Δ, densities (tag/nt, log2)
WT, densities (tag/nt, log2)
C
RNA-Seq, fold-change
(NMD-sensitive XUTs)
***
upf1Δ/WT ratio (log2)
Ribo-
Terminator

Figure S2



* RNA-Seq data from Bresson *et al.* (2020)

Figure S3

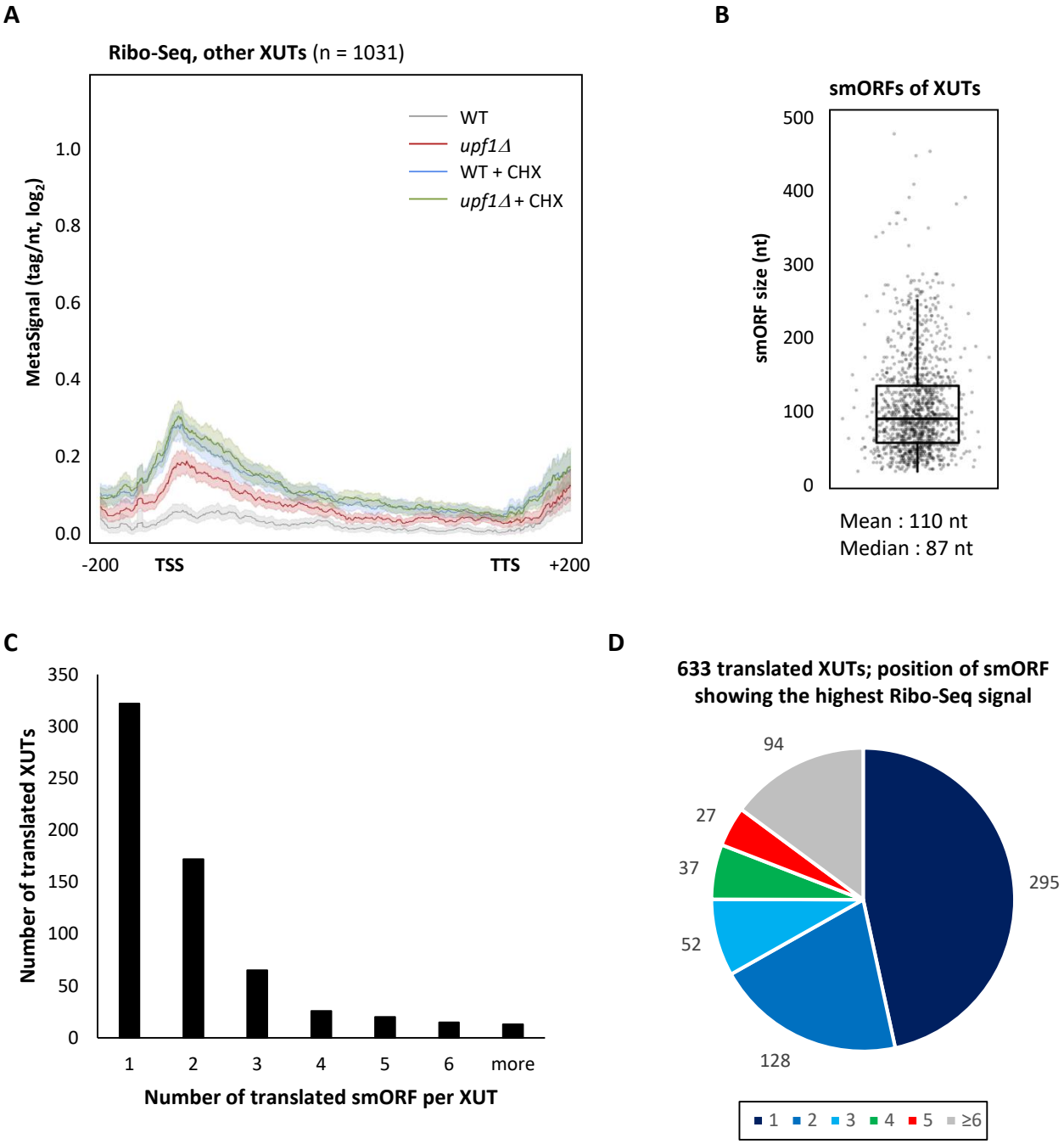
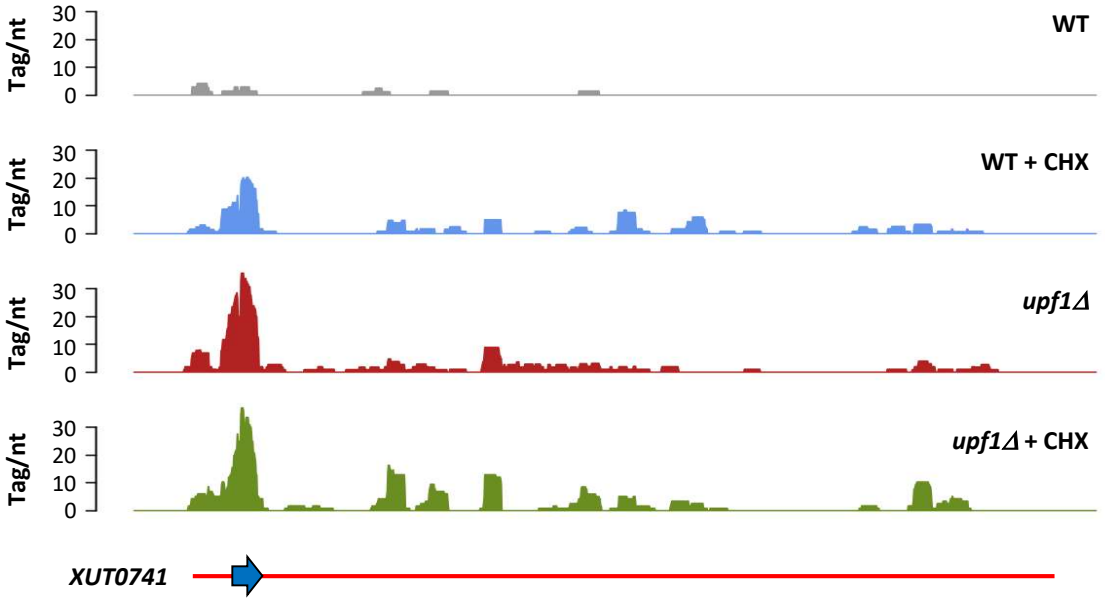

A
Ribo-Seq, other XUTs (n = 1031)
WT
upf1Δ
WT + CHX
upf1Δ + CHX
MetaSignal (tag/nt, log2)
1.0
0.8
0.6
0.4
0.2
0.0
-200
TSS
TTS
+200
B
smORFs of XUTs
smORF size (nt)
500
400
300
200
100
0
Mean : 110 nt
Median : 87 nt
C
Number of translated XUTs
350
300
250
200
150
100
50
0
1
2
3
4
5
6
more
Number of translated smORF per XUT
D
633 translated XUTs; position of smORF showing the highest Ribo-Seq signal
94
27
37
52
128
295
1
2
3
4
5
≥6

Figure S4

A


Ribo-Seq, WT & upf1Δ +/- CHX
Tag/nt
30
20
10
0
WT
WT + CHX
upf1Δ
upf1Δ + CHX
XUT0741

B

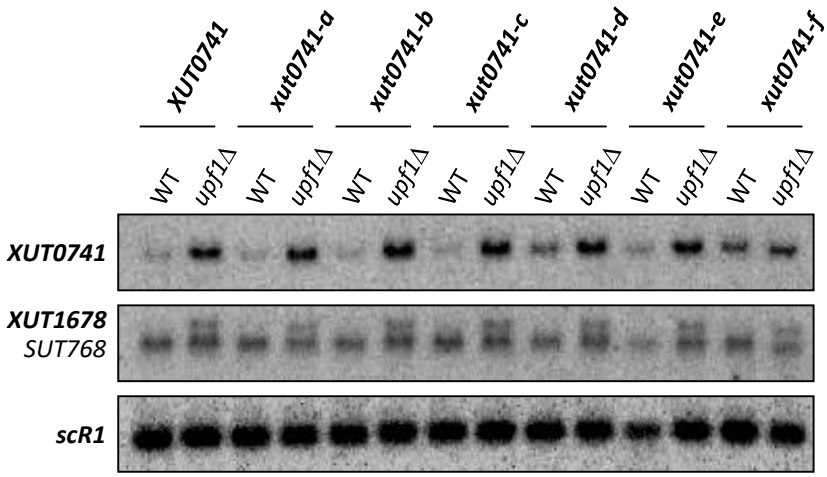

XUT0741
xut0741-a
xut0741-b
xut0741-c
xut0741-d
xut0741-e
xut0741-f
WT
upf1Δ
XUT0741
XUT1678
SUT768
scR1

Figure S5

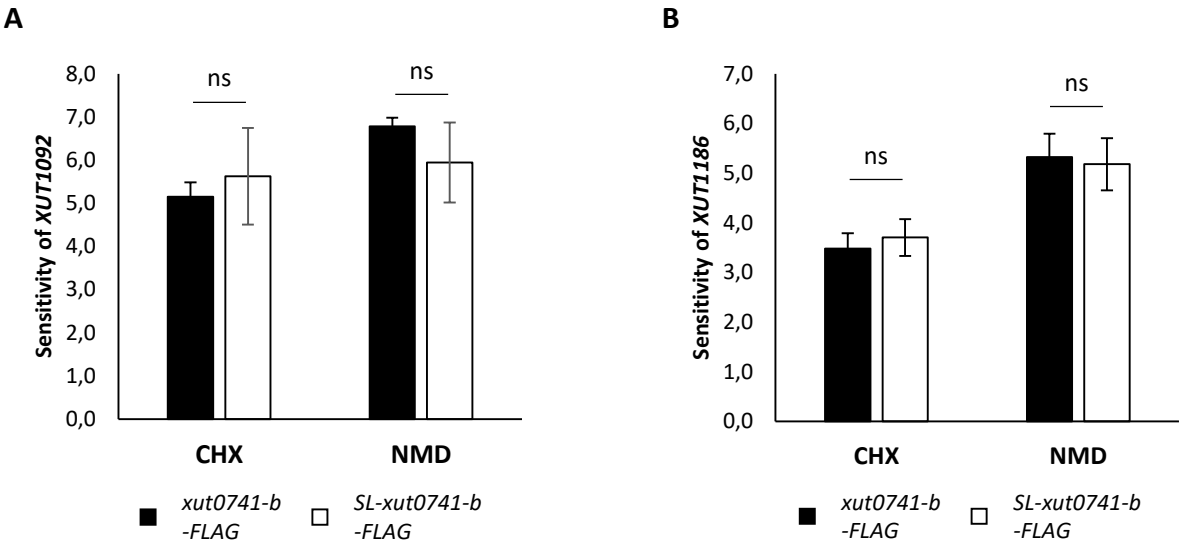
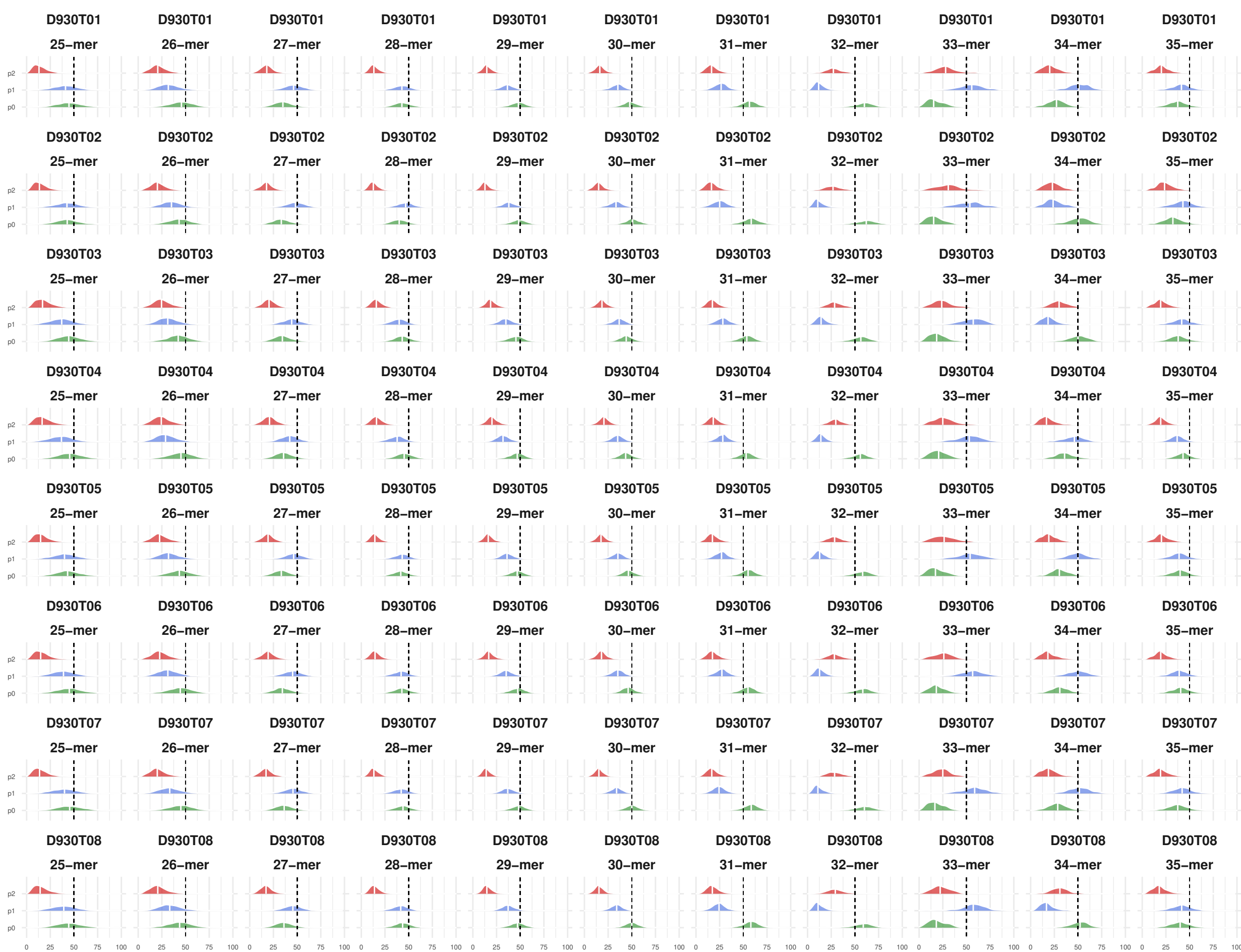

A
B
Sensitivity of XUT1092
Sensitivity of XUT1186
ns
ns
ns
ns
CHX
NMD
CHX
NMD
xut0741-b -FLAG
SL-xut0741-b -FLAG
xut0741-b -FLAG
SL-xut0741-b -FLAG

Figure S6


D930T01 25–mer
D930T01 26–mer
D930T01 27–mer
D930T01 28–mer
D930T01 29–mer
D930T01 30–mer
D930T01 31–mer
D930T01 32–mer
D930T01 33–mer
D930T01 34–mer
D930T01 35–mer
D930T02 25–mer
D930T02 26–mer
D930T02 27–mer
D930T02 28–mer
D930T02 29–mer
D930T02 30–mer
D930T02 31–mer
D930T02 32–mer
D930T02 33–mer
D930T02 34–mer
D930T02 35–mer
D930T03 25–mer
D930T03 26–mer
D930T03 27–mer
D930T03 28–mer
D930T03 29–mer
D930T03 30–mer
D930T03 31–mer
D930T03 32–mer
D930T03 33–mer
D930T03 34–mer
D930T03 35–mer
D930T04 25–mer
D930T04 26–mer
D930T04 27–mer
D930T04 28–mer
D930T04 29–mer
D930T04 30–mer
D930T04 31–mer
D930T04 32–mer
D930T04 33–mer
D930T04 34–mer
D930T04 35–mer
D930T05 25–mer
D930T05 26–mer
D930T05 27–mer
D930T05 28–mer
D930T05 29–mer
D930T05 30–mer
D930T05 31–mer
D930T05 32–mer
D930T05 33–mer
D930T05 34–mer
D930T05 35–mer
D930T06 25–mer
D930T06 26–mer
D930T06 27–mer
D930T06 28–mer
D930T06 29–mer
D930T06 30–mer
D930T06 31–mer
D930T06 32–mer
D930T06 33–mer
D930T06 34–mer
D930T06 35–mer
D930T07 25–mer
D930T07 26–mer
D930T07 27–mer
D930T07 28–mer
D930T07 29–mer
D930T07 30–mer
D930T07 31–mer
D930T07 32–mer
D930T07 33–mer
D930T07 34–mer
D930T07 35–mer
D930T08 25–mer
D930T08 26–mer
D930T08 27–mer
D930T08 28–mer
D930T08 29–mer
D930T08 30–mer
D930T08 31–mer
D930T08 32–mer
D930T08 33–mer
D930T08 34–mer
D930T08 35–mer
p2
p1
p0
0 25 50 75 100

3. Discussion

In this work we investigate the role of translation in the turnover of cytoplasmic lncRNAs. To that purpose, we show that despite being defined as non-coding RNA, XUTs are translated, which impacts their abundance. We detect that XUTs rapidly accumulate upon translation elongation inhibition using two drugs that affect the elongation phase through different modes of action. Instead, XUTs did not accumulate in a condition in which translation initiation (presence of polysomes) was impaired, indicating that the accumulation of XUTs we observe is due to the physical protection of the elongating ribosomes on XUTs. Moreover, our Ribo-Seq experiment revealed ribosomes binding to 38% of the annotated XUTs and identified actively translated small ORFs in their 5'-proximal regions.

One can imagine several potential roles of the ribosome association to smORF of lncRNAs (Figure 13). For instance, smORFs might tether functional factors to the bound ribosome, which could enable downstream lncRNA functions (Figure 13A). In addition, the translation of a smORF can modulate the stability of the lncRNA, by influencing translation dependent RNA decay pathways (Figure 13B). This is consistent with our observation that NMD-sensitive XUTs rapidly accumulate in WT cells following inhibition of translation elongation. Another close example of such case comes from the observation that translation elongation inhibitors result in the stabilization of polysomal lncRNAs in human cells (Carlevaro-Fita et al., 2016), indicating that the role of translation in determining the degradation of cytoplasmic lncRNAs has been evolutionarily conserved. Lastly, the translation of a smORF can also yield peptides (Figure 13C). Such peptides could be deleterious and/or unstable and thus rapidly degraded by the proteasome, as has been proposed for pervasively translated ORFs in *E. coli* (Stringer et al., 2021). Alternatively, peptides could be nonfunctional yet, but still tolerated by the cells, or they could have novel functions that could be even more important than the function of the lncRNAs itself. As mentioned in the introduction of the thesis, the biological relevance of smORF-deriving peptides is increasingly growing (Chen et al., 2020; van Heesch et al., 2019; Matsumoto et al., 2017; Slavoff et al., 2013; Sun et al., 2021; Wei and Guo, 2020). Finally, lncRNAs may serve as reservoir of rapidly evolving smORFs that offer the cell the potential to explore genetic novelty and produce novel peptides.

Note: The impact of the lncRNA translation on the de novo gene birth is discussed after (see discussion and perspectives).

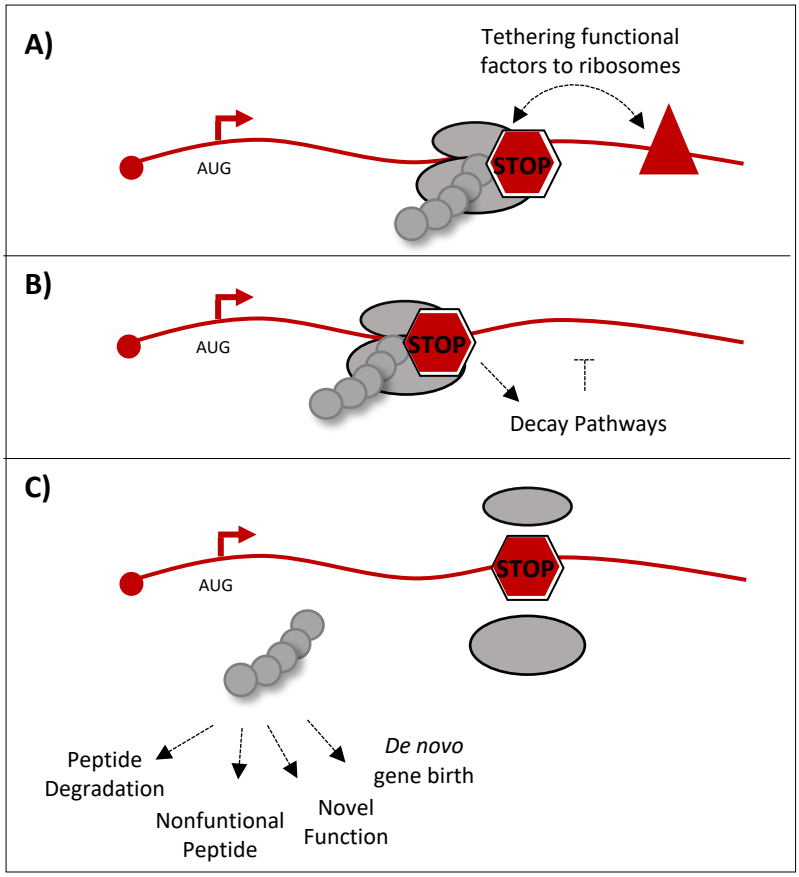


Figure 13 | Potential Role of the Ribosome Association of LncRNAs. (A) LncRNA ORFs might tether functionally important binding factors (red triangle) to ribosomes or (B) modulate the stability of the lncRNA by influencing translation dependent RNA decay pathways. (C) LncRNA-bound by ribosomes can produce peptides that can be rapidly degraded, can be tolerated by the cell even if nonfunctional, can have novel functions or be source of a *de novo* gene birth.

In our study, we detect a peptide derived from the translation of an NMD-sensitive XUT reporter in WT cells, in which the NMD decay pathway is fully functional. For this experiment, due to technical difficulties in detecting by Western blot a 4.5 kDa peptide derived from the native *XUT0741*, we utilized the mutant *XUT0741-b* containing an extended ORF as an NMD-sensitive reporter. This single gene observation paved the way for the characterization of the entire yeast peptidome, searching for all native peptides derived from the translation of XUTs. In this regard, we extended this analysis proteome-wide using Mass Spectrometry (MS) (Figure 14A) in collaboration with Dr. Olivier Namy's lab. Briefly, we prepared the crude extract of NMD-defective cells and subjected the samples to trypsin digestion and Liquid Chromatography (LC)-MS. We subsequently filtered the yeast protein fragments and uniquely matched them to bioinformatically predicted peptide sequence derived from XUT smORFs. This strategy allowed us to detect 100 novel XUTs-derived peptides from

NMD-defective cells. This is exemplified by the peptide produced from smORF of *XUT0541*, aligning to the corresponding ribosome-bound region across *XUT0541* sequence (Figure 14B).

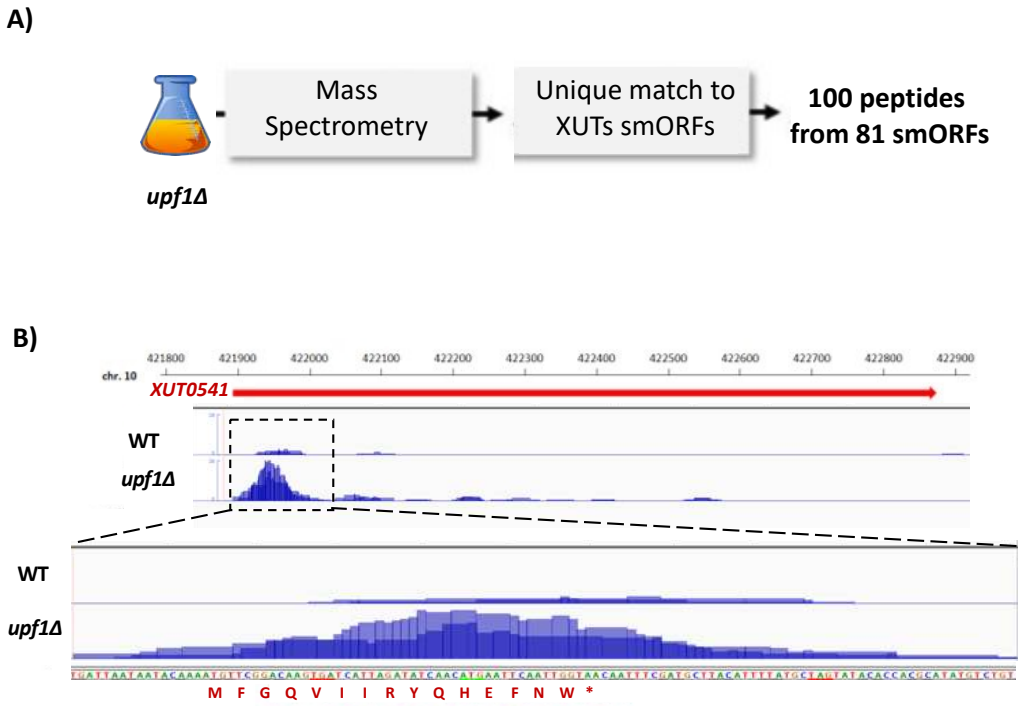


Figure 14 | Detection of Antisense LncRNA-Derived Peptides. (A) Schematic representation of the pilot LC-MS experiments from yeast crude extraction to peptide alignments on the peptide annotation database. (B) Snapshot of the Ribo-Seq signals (Andjus et al., 2022, bioRxiv) along *XUT0541* in WT and *upf1Δ* cells. In red the smORF of the XUT and the corresponding peptide detected by LC-MS.

During this preliminary analysis, we also treated NMD-defective cells with MG-132, a drug inhibiting the proteasome, to potentially prevent the peptide degradation. Interestingly, no peptide enrichment was observed in the treated *vs* the non-treated sample, potentially indicating that the peptides derived from NMD-sensitive XUTs are not immediately directed to the proteasome.

This would be in sharp contrast with other translation-dependent RNA decay pathways, the NSD and the NGD. In these pathways, as mentioned in the introduction, the nascent peptide remains attached to the ribosome, and is then recognized by the RQC complex that targets the peptide to the proteasome for degradation (Defenouillère and Fromont-Racine, 2017; Defenouillère et al., 2017). In the case of the NMD, translation 'normally' terminates at the stop codon (it is the position of the codon that renders the transcript 'abnormal'). We therefore anticipate that the fate of the resulting peptide would not depend on the RQC, and that peptides originating from aslncRNAs smORFs could have a stable expression. This is consistent with a recent study in mammalian cells showing that

NMD-coupled protein quality control is not mediated by canonical RQC factors (Inglis et al., 2022, bioRxiv).

In the close future, we aim to consolidate this promising pilot experiment using independent biological replicates of different WT and NMD-defective cells and improve technical aspects of the analysis by size-selecting exclusively low molecular weight peptides. Our Ribo-Seq analysis revealed that the median size of the smORFs of XUTs is 87 nt, therefore we could except to detect peptides of ∼30 aa. To enrich them, we will use special gel fractionation and/or centrifugal filtering approaches. At the same time, we will perform the yeast growth in the minimal medium (CSM), instead of the previously used rich YPD medium, to exclude the potential contaminating peptides deriving from the components of the rich medium. Also, as a control, we will analyze peptides deriving from the native, non-trypsin digested extracts.

One can speculate whether the XUT-derived peptides are functional. Regarding this idea, recent work in yeast performed a large-scale screening of ncRNAs and provided evidence for four functional SUTs that act *in trans* to regulate target genes involved in respiration (Balarezo-Cisneros et al., 2021). Strikingly, three of the regulatory SUTs overlap an NMD-sensitive XUT, that contains ribosome-binding signal in the 5' region (Ribo-Seq data from Andjus et al., 2022, bioRxiv), suggesting that they are actively translated. To understand whether the regulatory effect of these three NMD-sensitive XUTs could depend on the peptide they potentially produce rather than on the RNA itself, we could imagine expressing *in trans*, under control of an inducible promoter a mutated version of the translated smORF and monitor whether the growth phenotype induced by the regulatory SUTs is altered.

If we succeed to demonstrate that the effect of these NMD-sensitive XUTs is peptide rather than RNA dependent, given the limited number of tested lncRNAs, the significance of this work would be restricted. Nevertheless, important studies showed that the pioneer round of translation of an NMD-sensitive mRNA was found to be necessary for the production of antigenic peptides for the MHC class I pathway (Apcher et al., 2011; Uchihara et al., 2022). Moreover, recently, by integrating Ribo-Seq and Mass Spectrometry data a study identified that cryptic proteins, presented by the HLA, are encoded by the non-coding regions (Ruiz Cuevas et al., 2021). Finally, non-coding regions were shown to be the main targetable source of tumour-specific antigens (Laumont et al., 2018; Zhao et al., 2020).

To conclude, this work opens exciting perspectives of the biological relevance of translation of lncRNAs subjected to NMD, as our understanding of this pathway has now evolved far beyond its initial term, 'Nonsense'-Mediated mRNA decay.

Chapter 5. Heterogeneity of Sense/Antisense RNA Expression and Interaction at the Single Cell Level

1. Introduction

In the previous works the lab showed that aslncRNAs in yeast once in the cytoplasm can be targeted and degraded by the NMD. However, it was also demonstrated that they can create dsRNA with their paired-sense mRNAs, at least in some cells (Wery et al., 2016). Under such dsRNA conformation, aslncRNAs are shielded from the NMD possibly by hiding the access of smORF for the ribosome (Wery et al., 2016). Conceptually, this suggested, that dsRNA production influences the fate of the XUT and is one of the major determinants of their sensitivity to NMD. Mainly, as the dsRNA structure could interfere with their ability to be translated. However, how heterogeneous is the co-expression of mRNA/aslncRNAs and their physical pairing within a population of cells at the genome-wide level, in single-cells remained unknown. Indeed, in a heterogenous population, single cells could be deprived of both RNAs, express one of the pair, or contain both with(out) pairing (Figure 15). Consequently, this generates different configurations for the translation ability and the NMD sensitivity of the aslncRNA thus complicating our understanding of the metabolism of aslncRNAs.

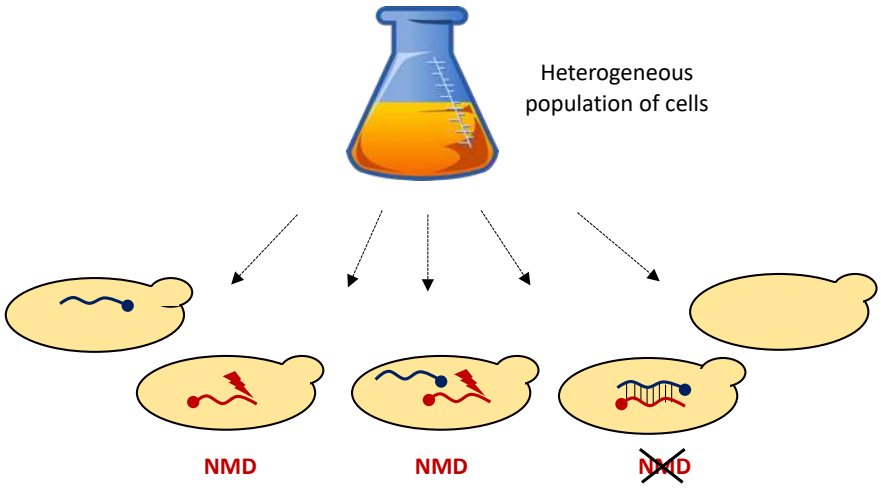


Figure 15 | Heterogeneity of Sense/Antisense RNA Expression and Interaction at the Single-Cell Level. In a heterogeneous population of cells, the NMD sensitivity of an aslncRNA depends on its (dis)ability to be engaged in a dsRNA duplex, itself depending on the co-expression but also on the pairing with its sense partner.

In fact, the heterogeneity of sense/antisense (s/as) expression could explain some of the data that seem to be inconsistent with our working model (Figure 11). In the model, the lab proposed that RNA helicases, Mtr4 and Dbp2 are recruited to the 3' single-stranded extension of some asXUTs, unwind the duplex and release asXUTs as solo, single-stranded transcripts (Wery et al., 2016). Since

several asXUTs fully overlapped by the sense mRNAs were found to be yet NMD-sensitive, our model could not adequately explain its reason. Nonetheless, one can imagine that in an heterogenous population of cells, the fate of such aslncRNA fully engaged in dsRNA would be determined by its ability or failure to engage in a duplex, depending on the co-expression but also the co-localization with its sense partner. In addition, Mtr4 nor Dpb2 loss did not completely recapitulate the effect of NMD inactivation, in terms of XUTs level (Wery et al., 2016), which could be a consequence of the heterogeneous population in which XUTs could be in different configurations. For example, an NMD-sensitive XUT might be single-stranded in some cells and engaged in a dsRNA structure in others. RNA helicases would only act in the second case to provide access to NMD, while Upf1 would target both subpopulations. To characterize these subpopulations, we necessitated going to the unique cell resolution to determine the extent of s/as RNA transcript co-expression and interaction in order to get a more precise view of the metabolism of aslncRNAs in yeast.

To that purpose, we took advantage of the single-cell (sc) approaches. Even if scRNA-Seq is now of common practice in mammalian cell studies, the applications of this technique for yeast was significantly lagging behind since the intrinsic nature of yeast cells disabled the direct use of these protocols (Picelli, 2017). First, small yeast cell size contains at least ten times less amount of RNA per cell than mammalian cells (Miura et al., 2008). Second, the thick cell wall creates a strong barrier for single-cell RNA isolation and therefore standard RNA extraction procedures are incompatible with efficient scRNA-Seq library preparation. Because of these technical challenges, yeast-specific scRNA-Seq methodologies have only recently started to appear, each one of them relying on different cell-isolation and library-preparation methods. Consequently, each sc approach harbors unique strengths and weaknesses that need to be considered.

The first strand-specific scRNA-Seq protocol in yeast (Nadal-Ribelles et al., 2019) came from Dr. Lars Steinmetz's lab and was developed by Dr. Mariona Nadal-Ribelles. This custom made protocol was based on single-cell FACS sorting and isolation in 96-well plates, followed by partial cell wall digestion using zymolyase, a cell wall digestion enzyme (Herrero et al., 1987), and library preparation capturing the transcript start end of poly-adenylated RNAs (Figure 16A). Their work studied 285 single WT yeast cells grown in rich media, and detected on average 3,339 transcripts per cell, ∼10% of those were lowly-abundant ncRNAs. The Gresham's and Verstrepen's labs were the first to develop the droplet based Chromium 10x scRNA-Seq in budding yeast by adapting the protocol to accommodate the zymolyase treatment (Jackson et al., 2020; Jariani et al., 2020). In contrast to the custom method, this state-of-art technology captures the transcript end site of poly-adenylated RNAs and allows to study the transcriptome of thousands of cells (Figure 16B). In addition, it is less laborious and can be performed in parallel for several samples with low cost per

cell if compared to the custom method. However, the number of genes detected per cell tends to be lower in this technology (Skinnider et al., 2019).

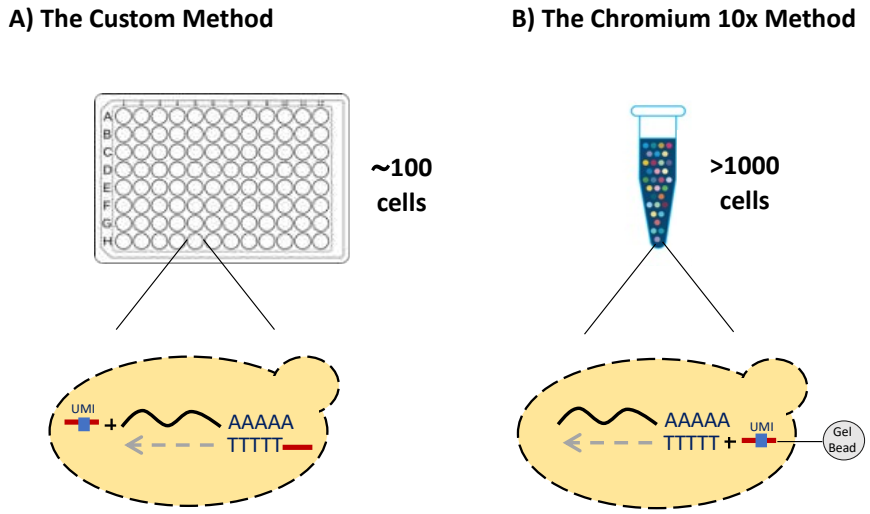


Figure 16 | Simplified Representation of Methods for Yeast ScRNA-Seq Capturing Different Ends of Polyadenylated RNAs. (A) The custom method captures the RNA 5' end of hundred cells separated in single wells of a 96-well plate. Full-length cDNA libraries are produced from biotynilated oligo(dT)- and UMI-containing template-switching oligonucleotide (TSO) while the (B) Chromium 10x method targets the RNA 3' end of thousands of cells in gel beads in emulsion (GEMs), generated by combining barcoded gel beads, a mix containing cells and partitioning oil. Both methods use poly(dT) primer for reverse transcription, although in the 3' assay the poly(dT) sequence is located on the gel bead oligo, while in the 5' assay the poly(dT) is supplied as an RT primer.

Our rationale started with analyzing published and unpublished strand-specific scRNA-Seq data generated by the custom method, provided by Dr. Lars Steinmetz (Stanford University) and Dr. Francesc Posas's and Dr. Eulàlia de Nadal's Team (IRB, Barcelona). We found that in a WT strain, cells expressing both sense and antisense transcripts can be detected, providing the proof-of-concept that cells with both transcripts coexpressed exist. We further validated and extended this analysis, in close collaboration with Dr. Mariona Nadal-Ribelles (Dr. Francesc Posas's & Dr. Eulàlia de Nadal's Team) at IRB, using the independent 10x single cell approach in thousands of cells in conditions that stabilize aslncRNAs, revealing that in these cases the proportion of cells with transcript co-expression is higher. This results further prompted us to start investigating to which extent the sense/as RNA pairs are indeed under dsRNAs structure and how conditions regulating aslncRNAs expression affect the dsRNA formation, finally understanding the impact these events have on the metabolism of aslncRNAs.

2.1. Antisense LncRNAs Coexist with Paired-Sense mRNAs in Wild-Type Single-Cells

To apprehend whether aslncRNAs, being templates of the translation machinery, are also prone to bind paired-sense mRNA, we explored the custom strand-specific scRNA-Seq published dataset (Nadal-Ribelles et al., 2019) in search of the s/as coexistence in individual yeast cells. In this dataset, we were able to define 411 distinct mRNA/aslncRNA pairs detectable in 127 single-cell WT transcriptomes. All the configurations our hypothesis anticipated, in terms of expression of s/as RNA pairs, were present (Figure 17A). On average, in 48,5% of cells none of the transcripts was detected, 41,1% expressed only the sense transcript, while 5,7% expressed only the antisense transcripts. Importantly, 4,7% of cells expressed both s/as pairs (Figure 17B). However, if considering only cells expressing at least one RNA, the proportion of cells expressing both pairs increased to 9,1%.

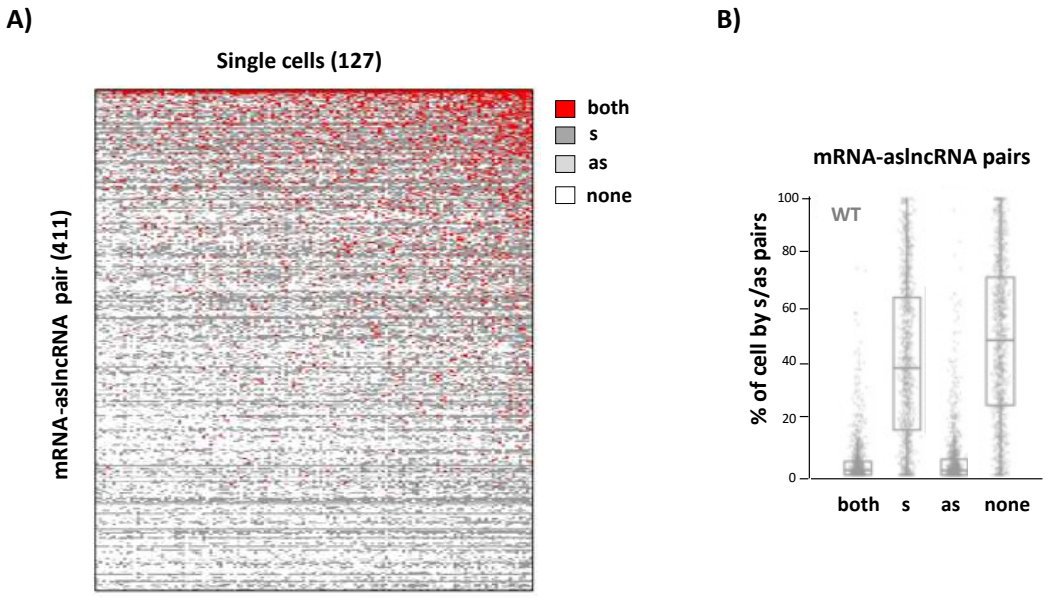


Figure 17 | Sense/Antisense RNAs Coexist in Yeast Single WT Cells. (A) Heatmap representation of s/as expression for 411 mRNA/aslncRNAs pairs in 127 WT cells (data from Nadal-Ribelles et al. 2019). Cells with both, only s, only as, and none transcripts are shown as red, dark grey, light gray, and white boxes, respectively. (B) Box-plot representation of percentage of cells displaying different configuration of s/as RNA in WT cells (n=127). Each dot is a s/as pair.

This result revealed a first proof-of-concept that despite their low expression levels and the technical difficulties inherent to yeast features, aslncRNAs can be detected by the custom scRNA-Seq approach, co-existing in WT cells with the paired-sense partner even in conditions in which the RNA decay pathways are functional.

We anticipated that if the expression of the lowest member of the pair is increased, the number of cells with coexistence might also increase. To test this hypothesis, we first analyzed unpublished scRNA-Seq data, generously shared by Dr. Lars Steinmetz and Dr. Nadal-Ribelles, using the custom method in the mutant of Xrn1, in which aslncRNA levels are stabilized. First, we tested if the custom scRNA-Seq reproduces the bulk RNA-Seq data from the lab in the mutant of Xrn1 (from Wery et al., 2016). We analyzed the expression levels of three classes of RNAs (mRNA, CUT and XUT), represented as a ratio in *xrn1Δ* / WT cells, between bulk RNA-Seq (data from Wery et al., 2016) and pseudo-bulk RNA-Seq[3]. In accordance with the bulk RNA-Seq, XUTs were upregulated in *xrn1Δ*/WT cells in the pseudobulk analysis (Figure 18A-B). However, the expected ratio for mRNAs and CUTs was not exactly the same among the two datasets. This discordance could be explained by the difference in the total RNAs *vs* polyA fraction of RNAs that is captured in RNA-Seq and scRNA-Seq, respectively. Moreover, some CUTs overlap XUTs and therefore some XUTs could be embedded in the fraction of enriched CUTs in the presudobulk dataset.

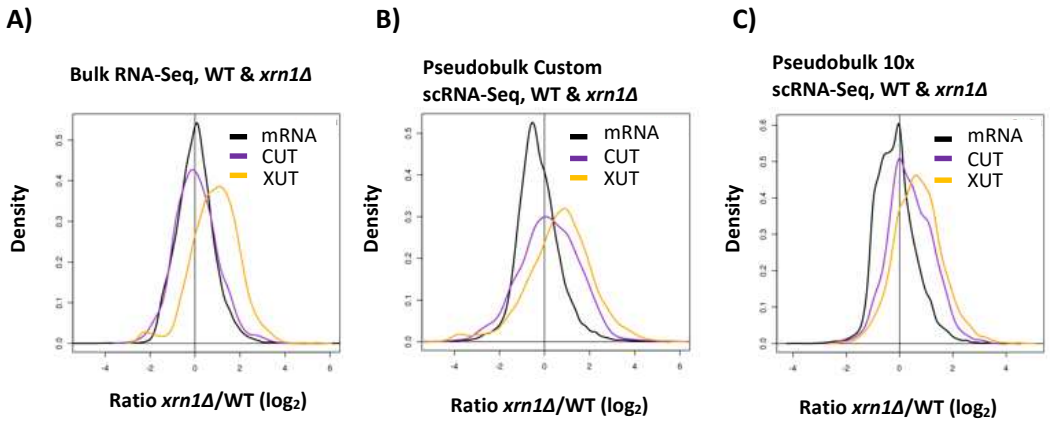


Figure 18 | Single Cell RNA-Seq Correlates With Bulk RNA-Seq Data For XUTs Levels. Density plot of *xrn1Δ* / WT signal ratio in (A) Bulk RNA-Seq (from Wery et al., 2016), (B) Pseudobulk custom scRNA-Seq and (C) Presudobulk 10x scRNA-Seq for mRNA (black), CUT (violet) and XUT (yellow). Data were normalized on mRNAs.

In addition to the custom method, we used an independent and complementary method - the Chromium droplet-based single-cell (10x Genomics) to investigate the s/as coexistence. This method allowed us to significantly increase the number of cells analyzed and thus potentially better

[3] The term pseudobulk RNA-Seq refers to grouping of single cells, where the data from each single cell is combined into a single pseudo sample that resembles a bulk RNA-Seq experiment.

reveal the heterogeneity of the expression of RNAs. We subjected ~2,000 single WT and ~2,000 *xrn1Δ* cells to 10x scRNA-Seq and analyzed the tendencies of the three classes compared to the bulk RNA-Seq. In the case of the 10x pseudobulk analysis, mRNAs and CUTs had similar tendencies to the bulk RNA-Seq (Figure 18A-C). Importantly, again XUTs showed the most important fold change difference in *xrn1Δ*/cells in pseudobulk 10x and bulk RNA-Seq, allowing the use of the dataset for further analysis of XUTs.

2.3. Benchmarking the Custom and the 10x ScRNA-Seq Methods

Subsequently, we benchmarked the sensitivity of the 10x to the custom scRNA-Seq method, by comparing several parameters. First, we noticed that the number of Unique Molecular Identifier (UMI) per cell was higher in the custom method for both WT and *xrn1Δ* cells than in the 10x (see Table 2). Consequently, the custom method allowed the detection of a higher diversity of mRNAs and XUTs detected per cells than the 10x. Indeed, the custom method yielded a high number of mRNAs per WT cell, corresponding to half of the protein-coding genome of yeast, versus the 10x that detected ~40% of mRNAs. In terms of number of XUTs detected per *xrn1Δ* cell, ~30% and ~10% of annotated XUTs were detected by the custom and 10x method, respectively. Therefore, the custom method has a higher sensitivity for detection of XUTs, however the number of cells analyzed in this dataset is rather low.

ScRNA-Seq method	Background	Number of cells analyzed	Number of UMI/cell	Number of mRNA detected/cell	Number of XUTs detected/cell
Custom	WT	127	30061	3468	177
	xrn1Δ	84	54347	4622	512
10x	WT	1435	13979	2489	48
	xrn1Δ	807	30001	3950	196

Table 2 | Comparisons Between the Custom and the 10x ScRNA-Seq Methods. Presented are the number of cells analyzed, and the number of UMI, mRNAs and XUTs detected per cells (on average), in WT and *xrn1Δ* backgrounds for the custom (in blue) and 10x (in red) scRNA-Seq method.

Nevertheless, since the 10x method captures the 3' end of RNAs, in contrast to the custom capturing their 5' end, it potentially provides an added value for the detection on XUTs. Namely, XUTs are defined as the 3' end extended isoforms of stable SUTs (Wery et al., 2016). We checked if the 10x method is able to discriminate reads specifically mapping to XUTs, rather than to overlapping SUTs, which could not be discriminated by any means in the custom method. The 10x scRNA-Seq

signals show for a candidate overlapping *XUT0741/SUT320* that in contrast to the custom method (Figure 19A), the 10x method can map reads exclusively deriving from the XUT (Figure 19B). Indeed, in the custom method the pick present at the 5' region of the *XUT0741/SUT320* cannot be discriminated if deriving from the *XUT0741* or the *SUT320*. In contrast, in the 10x we can nicely see a pick corresponding to the 3' region of the *SUT320* and a pick corresponding to the 3' region of the *XUT0741*. This result, in addition to the higher number of cells analyzed, shows that the 10x method adds a higher precision for quantification of XUTs expression. Altogether, the benchmarking of the two scRNA-Seq methods allowed us to conclude that the custom method has a higher sensitivity in detecting XUTs in a low number of cells, while the 10x method reveals higher precision for the quantification of XUTs in a high number of cells.

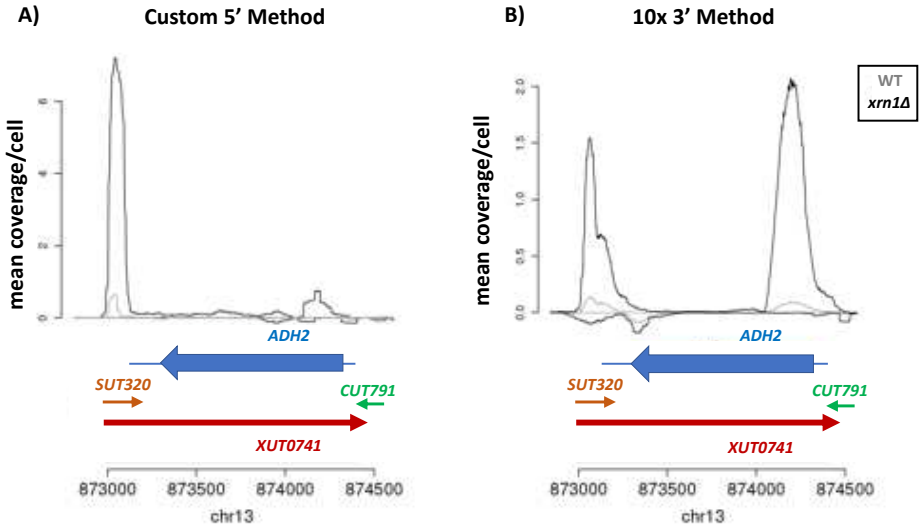


Figure 19 | 10x ScRNA-Seq Reads Differentiate An Overlapping XUT From the SUT. Presented is a snapshot showing distribution of scRNA-Seq reads for (A) Custom and (B) 10x Method across the *XUT0741/SUT320* locus in WT (gray line) and *xrn1Δ* (black line) cells. The positive and the negative values in each panel show the signals along the considered locus and along its antisense strand, respectively. The ORF, XUT, SUT and CUT are represented by blue, red, orange, and green arrows, respectively. The thin blue arrow corresponds to UTR.

2.4. Higher Sense/Antisense RNA Coexistence Upon Antisense LncRNA Stabilization

We proceeded with analyzing whether the number of cells co-expressing s/as RNA is increased in the mutant of Xrn1 in both scRNA-Seq methods. We were able to analyze 1,258 mRNA/asXUT pairs in 84 and 807 *xrn1Δ* single cells using custom and 10x method, respectively. In the custom method, in *xrn1Δ* cells, in 18% median (26% mean) of cells none of the transcripts was detected (grey square), 45% median (44% mean) expressed only the sense transcript (blue square), while 2.4% median (10,35% mean) expressed only the antisense transcripts (green square) (Figure 20A). Importantly,

20% of *xrn1Δ* cells expressed both s/as pairs (red square) (Figure 20A). We observed that on average the percentage of cells expressing asXUT increased from 5,6% to 10,35% in WT and Xrn1-lacking cells in the custom dataset, respectively (Figure 20A; green square). Further, we revealed that in WT and Xrn1-deleted cells the number of cells with two pairs co-expressed on average increased from 4.7% to 20%, respectively (Figure 20A; red square).

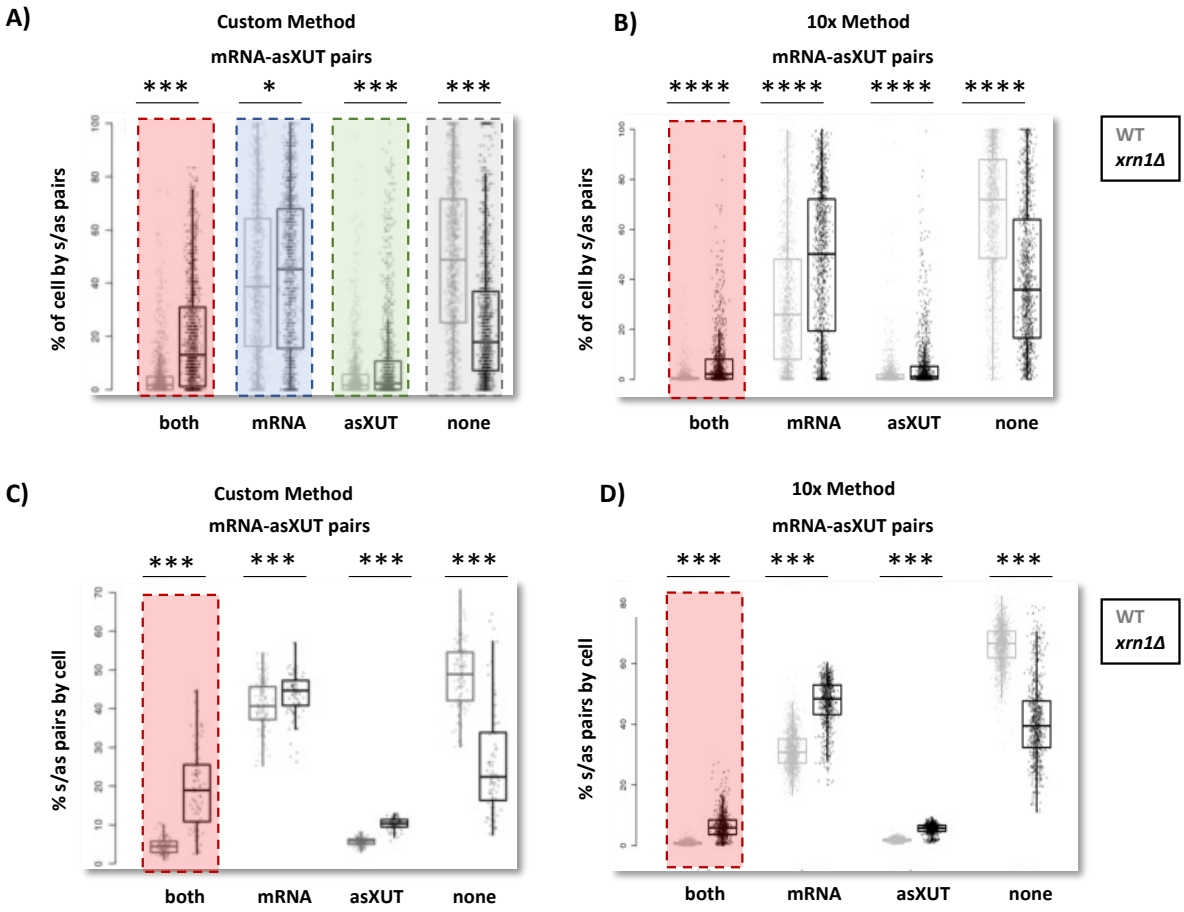


Figure 20 | Higher Number of Cells Coexpressing Both RNAs and More Pairs Expressed in Xrn1-lacking Cells. Box-plot representation of percentage of cells expressing both, mRNA, asXUT and none RNA in WT (grey) and *xrn1Δ* cells (black) in the (A) custom and (B) 10x method; Each dot is a pair. The number of cells are for WT, custom n=127; 10x n=1,435, while for *xrn1Δ*, custom n=84; 10x n=807. Number of analyzed pairs is 1258. Percentage of s/asXUT pairs expressed per cell in WT (gray) and *xrn1Δ* (black) cells in (C) custom and (D) 10x method; Each dot is a cell. * P-value < 0.05; *** P < 0.001; **** P < 0.0001 upon two-sided Wilcoxon rank-sum test.

The increase in the percentage of cells with asXUTs and both transcripts co-expressed in *xrn1Δ vs* WT cells was observed also using the 10x method. In the 10x method, in *xrn1Δ* cells, in 36% median (40% mean) of cells none of the transcripts was detected, 50% median (47% mean) expressed only

the sense transcript, while 1,11% median (5,46% mean) expressed only the antisense transcripts (Figure 20B). The number of cells coexpressing both transcripts increased on average from 0,82% in WT to 6,58% in *xrn1Δ* cells (Figure 20B; red square). The observed increase of cells with both RNAs correlated with a significant decrease of cells without any transcripts (none) in Xrn1-lacking cells in both methods.

In addition, not only did the number of cells with co-expression in *xrn1Δ* cells increase, but also did the number of co-expressed pairs per cell in both methods (Figure 20C-D; red squares), indicating that the observed increase of pairs is not specific to particular pairs.

Furthermore, the analysis of three candidate s/as pairs from the custom method showed that for each, there is an increase of cells with coexistence of the transcripts in Xrn1-deleted cells compared to WT cells (Figure 21; in red).

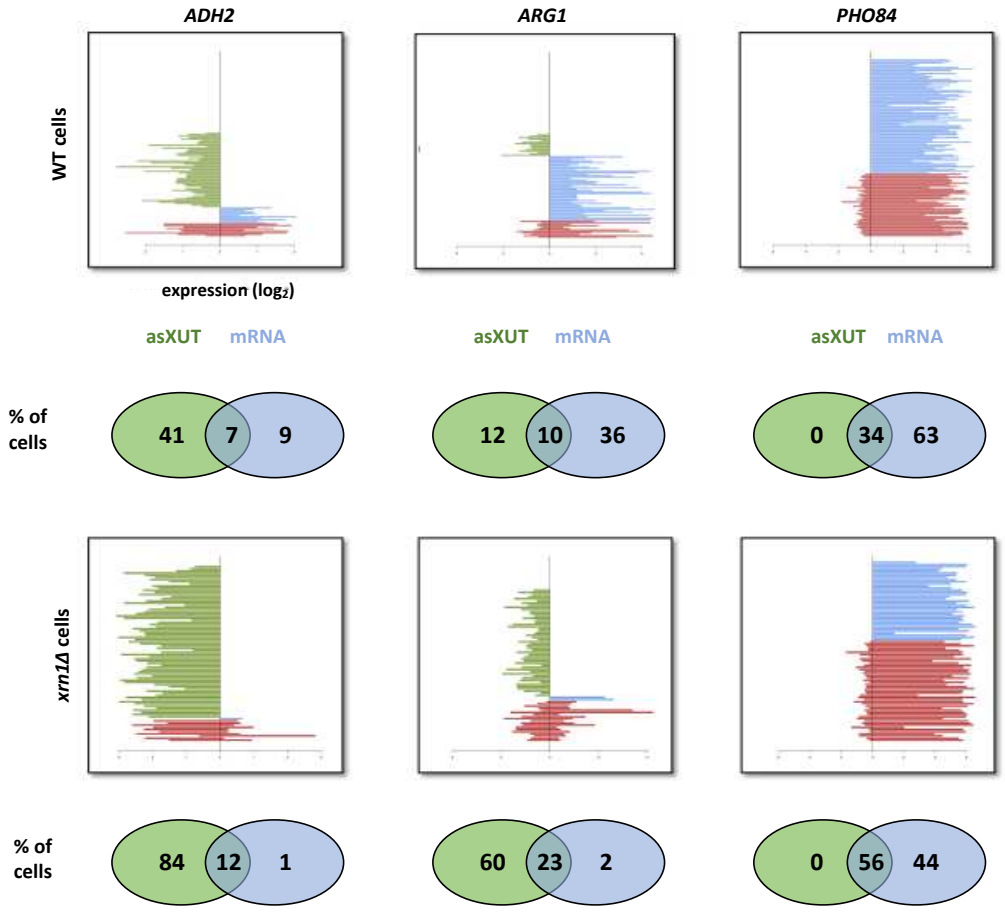


Figure 21| Higher Number of Cells Coexpressing Both RNAs for *ADH2, ARG1* and *PHO84* pairs in Xrn1-lacking Cells. Co-expression of *ADH2, ARG1* and *PHO84* mRNAs and their paired-asXUTs in WT (upper panel; n=127) and *xrn1Δ* (lower panel; n=84) single yeast cells. In each panel, each bar represents the signal detected for the corresponding transcript in single cell. Red, green and blue bars highlight cells where s/asXUT are co-expressed, only asXUT and mRNA are expressed, respectively. Under each panel the percentage of cells with mRNA, asXUT and both is shown. Percentage of empty cells not shown.

For instance, in WT cells, the *ADH2* mRNA is expressed in ~50% of cells expressing its paired *XUT0741* (Figure 21; upper left panel). The *XUT1678*, which is detected in 22% of cells, co-existing with its paired-sense *ARG1* mRNA in ~50% of these cells (Figure 21; upper middle panel), while *XUT06*83 systematically coexists with its pared-sense *PHO84* mRNA (Figure 21; upper right panel). In all three cases, the number of cells with mRNA decreased in Xrn1-lacking cells and the number of cells with both RNA coexpressed increased (Figure 21; lower panels).

The Fisher statistical test showed that the observed increase in the proportion of cells co-expressing *ARG1* and *PHO84* pairs (Figure 22A; in red) and globally for ~60% of all pairs (Figure 22B), was significantly higher in Xrn1-lacking *vs* WT cells.

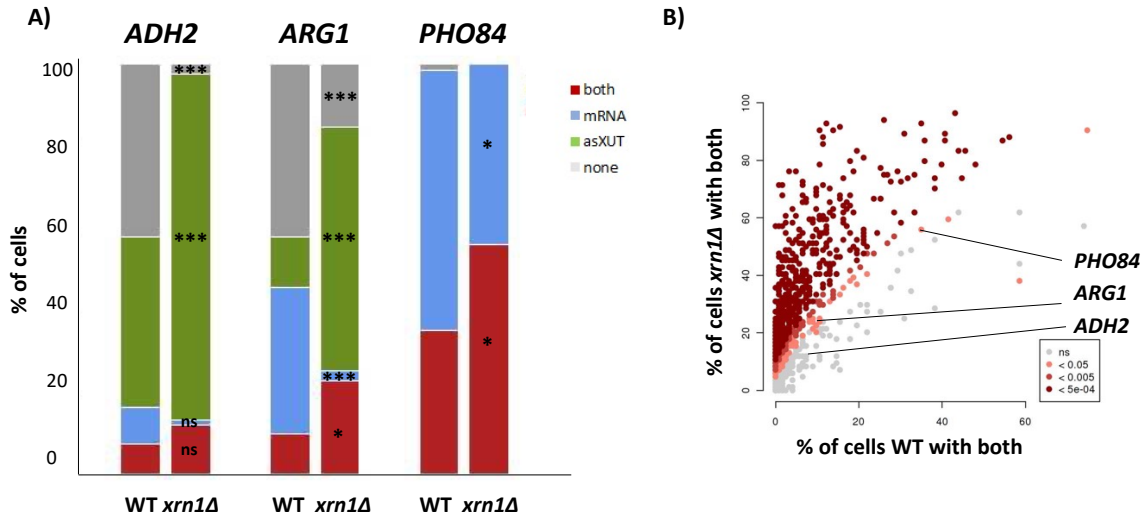


Figure 22 | Statistical Significance of the Increase of Coexpression in Xrn1-lacking Cells. (A) Statistical significance of the increase of the percentage of cells with all configurations in *xrn1Δ* (n=84) *vs* WT (n=127) cells for *ADH2, ARG1* and *PHO84* pairs. * P-value < 0.05; *** P < 0.001; upon Fisher test. (B) Percentage of cells co-expressing both transcripts *in xrn1Δ vs* WT cells for each s/as pair and the statistical significance upon Fisher test. Shades of red color, from intense red to gray, represent high and no statistical significance, respectively.

To further extend the hypothesis that when the lower member of the pair is increased, the number of cells showing coexistence also increases, we subjected another genetic and stress condition, stabilizing aslncRNAs, to 10x scRNA-Seq. In bulk RNA-Seq in NMD-deficient and WT cells treated with cycloheximide (CHX) most XUTs accumulate (Wery et al., 2016, Andjus et al., 2022, bioRxiv, respectively). Therefore, we performed 10x scRNA-Seq in *upf1Δ* cells and WT cells treated for 15 min with CHX. Focusing only on the cells expressing both s/as RNAs, we observed a statistically significant increase of the number of cells expressing both transcripts and of the number of both

pairs expressed per cell in CHX treated and *upf1Δ* cells compared to WT cells (Figure 23). In particular, the percentage of cells expressing both RNAs increased from on average 0,82% in WT to 2,70% in CHX treated and 2,23% in *upf1Δ* cells, respectively (Figure 23A).

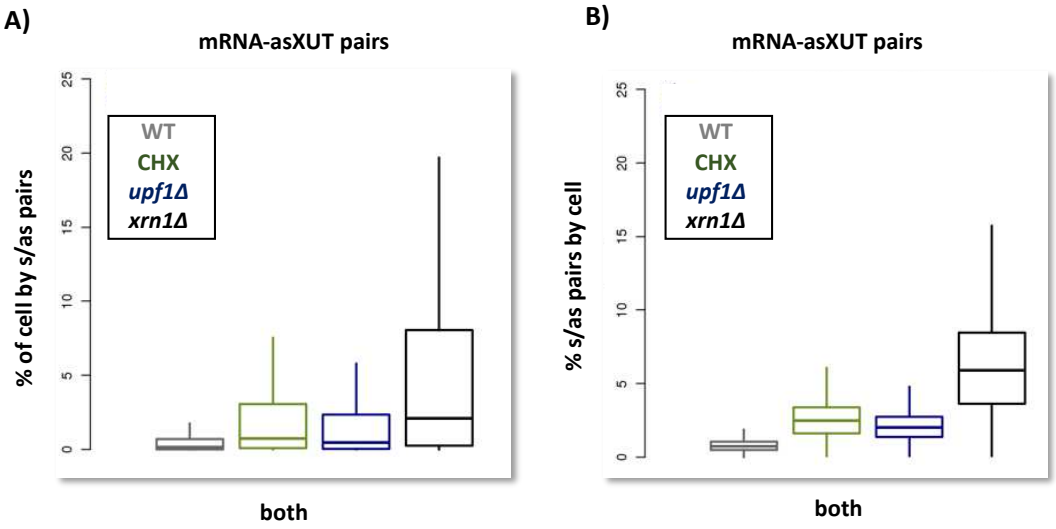


Figure 23 | Increased Number of Cells Coexpressing Both RNAs and More Pairs Expressed Upon AslncRNA Stabilization. Box-plot representation of (A) percentage of cells expressing both s/as RNAs and (B) percentage of s/as pairs coexpressed in WT (grey; n=1,435), CHX-treated WT (green; n=2,420), *upf1Δ* (blue; n=3,027), and *xrn1Δ* (black; 807) cells in 10x scRNA-Seq; Number of analyzed mRNA-asXUT pairs is 1258. Comparisons between all the samples for all the expression profiles are **** P < 0.0001 upon two-sided Wilcoxon rank-sum test.

Furthermore, from this dataset we detected 79 s/as pairs among the top 10% of s/as pairs coexpressed in CHX treated WT, *upf1Δ* and *xrn1Δ* cells (Figure 24; see also Table S5 in ANNEX I for genomic features), indicating that some pairs are constantly present in cells. Future investigations will determine which exact features distinguish these pairs.

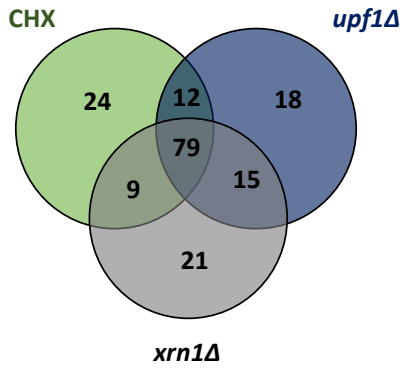


Figure 24 | Venn Diagram Showing the Number of Top 10% Sense/Antisense Pairs Coexpressed in Each of the Indicated Conditions. See also Table S5.

Moreover, one main source of heterogeneity in a clonal population is the cell cycle. Further analysis of the 10x scRNA-Seq will profile XUTs expression across the cell cycle. In particular, we will investigate if the cell cycle phase specific genes overlap asXUTs and whether they have periodic expressions through the cell cycle. In this regard, for instance, two key cell-cycle regulators, *FAR1*, important for the G1/S transition (Vanoni et al., 2005), and *TAF2*, involved in the transition G2/M (Apone et al., 1996), contain an antisense *XUT0521* and *XUT0111*, respectively. Indeed, most cycling antisense transcripts have been located opposite to genes with cell-cycle-related functions (Granovskaia et al., 2010). This data could enable further hypothesis-driven mechanistic studies concerning the roles of asXUTs.

Altogether, using two complementary and independent single-cell approaches, the preliminary analysis we obtain here show that the heterogeneity of sense/antisense RNA expression is high and that they can be co-expressed in the same cell. Moreover, when the expression of the lowest member of the pair is increased, the number of cells showing coexistence also increased. These results set the basis for studying the proportion of cells in which the co-expressed s/as RNAs are involved in dsRNA formation, that protects aslncRNAs from the decay, possibly by blocking their access to ribosomes (Wery et al., 2016).

2.5. Towards Understanding DsRNA Pairing Determinants

The concomitant presence of the aslncRNA with its paired-sense mRNA, detected in single cells, does not necessary result in RNA duplex formation. Indeed, different subcellular localization might prevent the interactions between the two partners. Previously, in the lab dsRNA *in vivo* formation was revealed in population of cells by small RNA sequencing in *S. cerevisiae* cells expressing the Dicer RNaseIII from *N. castellii* (Wery et al., 2016). Indeed, small RNAs from *xrn1Δ* strains expressing RNAi factors, showed that most asXUT engaged in dsRNAs. We applied the same strategy of small RNA-Seq in bulk and extended it to conditions favoring XUTs stabilization (NMD-deficient, in addition to Xrn1-lacking cells) and WT cells treated for 15 min with CHX (inhibitor of translation elongation). Our preliminary results show that in addition to *xrn1Δ*, in *upf1Δ* cells upon RNAi reconstitution ($RNAi^+$), a high accumulation of 19-23 nt small RNAs, deriving from dsRNAs, is present (Figure 25). In contrast, CHX-treated WT cells with reconstituted RNAi, do not produce small RNAs.

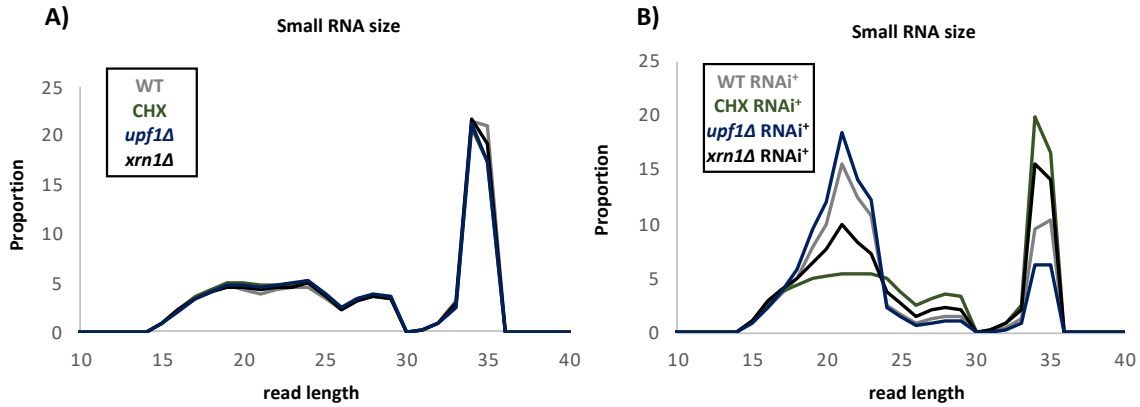


Figure 25 | Small RNAs Are Present in WT, Xrn1- and Upf1-lacking Cells and Lack in CHX-treated Cells Upon RNAi Reconstruction. Presented is size distribution of small RNAs produced (A) without and (B) with RNAi reconstitution in WT, *xrn1Δ* and *upf1Δ* cells and WT cells treated with CHX. Libraries were constructed from purified small RNAs.

Snapshots for the *TAT1/XUT0051* pair confirms the genome-wide results and shows that in WT RNAi$^+$, *xrn1Δ* RNAi$^+$ and *upf1Δ* RNAi$^+$ cells, small RNAs production is mainly restricted to the region of overlap between mRNA and asXUT (Figure 26). Indeed, in WT, *xrn1Δ* and *upf1Δ* cells, with RNAi factors, no or very low signal mapped to the 3' end of *XUT0051*, which is not overlapped by *TAT1* mRNAs.

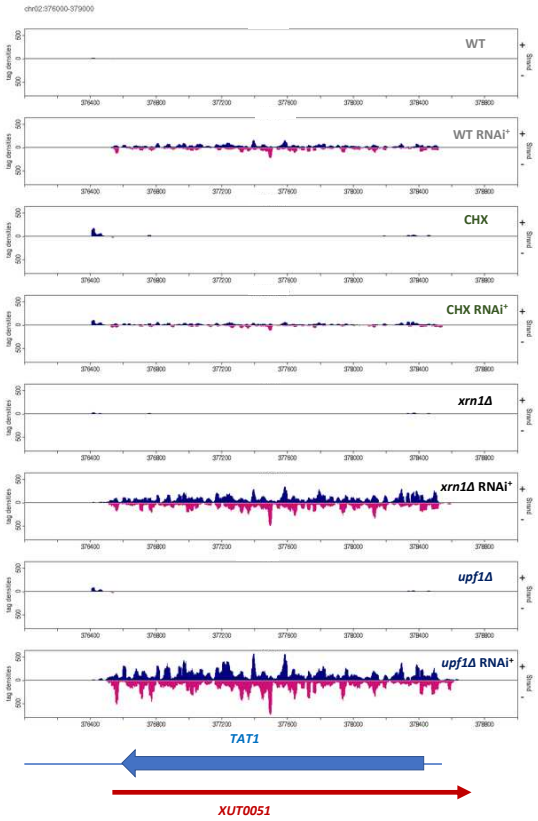


Figure 26 | Snapshot of Small RNAs Along the *TAT1/XUT0051* Locus. The densities of 19-23 nt uniquely mapped reads for the + and – strands are shown (upper and lower), respectively for each condition. The ORF and XUT are represented by blue and red arrows, respectively. The thin blue line corresponds to UTRs. The snapshot was produced using VING (Descimes et al., 2015).

One could imagine that the presence of small RNAs in genetic context upregulating aslncRNAs could be explained by the fact that the ribosomes finished translating the abundant RNAs, that became 'naked' and ready to form dsRNAs. The absence of the dsRNAs upon translation inhibition, could be due to the freezed ribosome bound RNAs that physically impede their binding with a paired sense mRNA. Alternatively, another explanation for the absence of small RNAs in the translation inhibiting condition, could be caused by an inhibitory effect CHX could have on the RNAi pathway. To test this hypothesis, further analysis will confirm whether RNAs that are not translated (*i.e.* overlapping non-translated XUTs, overlapping 3' UTRs, transposons, and repeated sequences) produce small RNAs, excluding the idea that CHX inhibits the RNAi. In alternative to an internal standard, we could also transform from a plasmid a synthetic dsRNA without a start codon and check if it produces small RNAi in cells treated with CHX.

Further investigation will allow us to determine from where the uniquely mapped reads originate in the genome, especially to which extend in respect to the s/as transcription. Also, to which extend are XUTs targeted by the RNAi in the genetic conditions favoring their expression.

As a next step, we want to determine the heterogeneity of the dsRNA formation on individual cells. To achieve that goal, we propose to use the same strains and conditions from the bulk small RNA-Seq and optimize to yeast, the already-existing small scRNA-Seq protocols in mammalian cells (Faridani et al., 2016; VanInsberghe et al., 2021).

Together, these results provide the first genome-wide characterization of the heterogeneity of s/as RNA expression and pairing in single yeast cells, which is crucial to understand the role of dsRNA formation in the life of (aslnc)RNAs, particularly how dsRNA antagonize with their translation-dependent decay pathway.

MATERIAL AND METHODS

Chapter 6. Materials and Methods

For the sake of manuscript clarity, this section includes only the description of the methods used in the third part of my thesis, including mainly the single-cell project. The remaining materials and methods are incorporated in the articles presented in the first two parts of the results section and in the ANNEX I. All NGS data processing has been performed by Ugo Szachnowski (PhD student in our team).

1. Yeast strain construction and growth

YAM1 (WT), YAM6 (*xrn1⊿*) and YAM202 (*upf1⊿*) cell derive from the BY4741 background. Both *XRN1* and *UPF1* were replaced by the *kanMX4* deletion cassette. In order to distinguish the different genotypes, a highly expressed (*TDH3*) gene was in each genotype C-terminally tagged with tags containing different markers. Each tag was previously amplified from a different plasmid using primers S2 and S3 (Janke et al., 2004) and transformed using the standard lithium acetate protocol. YAM1, YAM6 and YAM202 were transformed with the amplified tag from *pYM18* (*9Myc:kanMX4*), *pYM15* (*6HA:HIS3MX6*) and *pYM21* (*9Myc:natNT2*), respectively. Selection for positive transformants, containing the tag cassette downstream *TDH3,* was done on plates containing appropriate antibiotics. Transformants were confirmed using colony PCR with primers detecting *TDH3* gene. All strains were grown overnight at 30°C in rich media (YPD). The next day, cells were diluted to OD_{600} = 0.05 in YPD and grown for two cell divisions. After two cell divisions, WT cells were treated for 15 minutes with DMSO (control) and CHX (100 µg/ml, final concentration).

2. Methanol fixation of cells

Cells were fixed following the manufacturer's protocol from 10x Genomics (CG000136 RevE), except that no SSC Buffer has been used. Briefly, after harvesting 5 ml of the exponentially growing cell culture by centrifugation, they were fixed by adding drop by drop 2 ml of 80% Methanol (MeOH). Then, fixed cells were incubated at -20°C for 10 minutes. Upon equilibration of the methanol fixed cells to 4°C, cells were further rehydrated. First, they were centrifuged, and the supernatant was removed. Then, cells were resuspended in 500 µl of fresh Wash-Resuspension Buffer (PBS-BSA + RNase Inhibitor + actinomycin D). Finally, cells were passed through a Flowmi Cell Stainer to eliminate cell debris and large clumps. Cell concentration was determined using a Countess II

Automated Cell Counter. Cell density (cells/mL) for each condition prior to library preparation was ~1 x 10^6.

3. 10x Single cell library preparation

Single cell library preparation was done using the 10x Genomics Chromium 3' v3.1 Single Cell Gene Expression Dual Index Kit (CG000315 Rev C), following the kit protocol. 38.4 µL of Single Cell Master Mix was prepared to which 36.6 µL H_2O was added in order to recover ~2,000 cells/condition. The microfluidic Chromium Single Cell Chip G was then prepared for use. 11 µL of prepared Zymolyase 100T (100 mg/ml) was added to hydrogel beads. The single-cell master mix, hydrogel beads with zymolyase and the partitioning oil were added according to the manufacturer's protocol, and the cells were encapsulated with hydrogel beads using the 10x Genomics Chromium Controller. Following emulsification, reverse transcription and cleanup was performed according to the manufacturer's protocol. During the RT an additional step was added to allow the zymolyase activity for cell wall digestion at 37°C for 15 minutes. Whole transcriptome amplification was performed using a total of 9 cycles of PCR. Cleanup, fragmentation, adapter ligation, and dual indexing was performed according to the manufacturer's protocol, using 13 cycles of PCR for the indexing reaction. Library fragment sizes were determined using an Agilent Bioanalyzer High Sensitivity Chip. Libraries were quantified by Qubit fluorometer with the Qubit DNA HS Assay Kit. Libraries from each condition were pooled for multiplex sequencing on an Illumina NovaSeq using the sequencing parameters recommended by 10x Genomics (Read 1: 28 bp, Read 2: 91 bp) and standard Illumina read and indexing primers.

4. Single cell data processing and analysis

The FASTq files from Nadal-Ribelles et al., 2019 were downloaded from Gene Expression Omnibus (GEO) under the accession number GSE122392. Reads adaptors were trimmed using Trim Galore (cutadapt v2.10). Reads were aligned on *S. cerevisiae* genome S288C R64-2-1 using bowtie2 (--local parameter) (Langmead and Salzberg, 2012). PCR duplicates were removed using umi_tools dedup (v1.1.1). Count matrix were obtained using feature count (v2.0.0). All subsequent analysis were performed using R software (v3.6.2). After quality controls, cells with < 1.5% of mitochondrial reads and >10 000 UMI were kept. Gene count were normalized using estimateSizeFactorsForMatrix function using protein-coding genes count.

Sequencing results from the 10x experiment were analyzed using the 10x Genomics Cell Ranger v6.1.2. The reference genome used is S288C R64-2-1 including the 2-micron plasmid and the antibiotic resistance marker cassette sequences. The used annotation was R64-1-1 with the addition

of antibiotic resistance markers and SUTs, CUTs and XUTs. mRNAs boundaries (5' and 3') were modified based on TIF-Seq data (Pelechano et al., 2013). All subsequent analysis were performed using Seurat R package v3 with R software v3.6.2. Cell quality control was performed based on antibiotic resistance marker expression, mitochondrial reads (<1%) and number of genes (>500 & <3,500) detected per cell.

5. Small RNA-Seq

Small RNA-Seq analysis was performed from two biological replicates of YAM1725 (RNAi$^+$), YAM1730 (WT), YAM1982 (*xrn1Δ* RNAi$^+$), YAM2271 (*xrn1Δ*), YAM2913 (*upf1Δ*), and YAM2915 (*upf1Δ* RNAi$^+$) of exponentially growing cells. Yeast strains are listed in Table S2 (see ANNEX I). For each sample, 50 µg of total RNA were mixed with 2 µg of total RNA from the YAM2394 (WT) strain of *S. pombe* (Wery et al, 2018b). The *S. pombe* genome contains the 22–23-nt small RNAs derived from the centromeric repeats and was used as a spike-in for normalization of signals. The small RNAs (<80 nt) fraction was purified on 15% TBE–urea polyacrylamide gels. Libraries were constructed from 10 ng of purified small RNAs using the D-Plex Small RNA-Seq kit for Illumina (Diagenode) and the D-Plex Unique Dual Indexes for Illumina – set B (Diagenode). Single-end sequencing (50 nt) of the libraries was performed on a NovaSeq 6000 system (Illumina). Adapter sequences were removed using the Atropos software (Didion et al., 2017). Reads were then mapped to the *S. cerevisiae* and *S. pombe* reference genomes using the version 2.3.5 of Bowtie (Langmead and Salzberg, 2012), using default parameters.

DISCUSSION AND PERSPECTIVES

Chapter 7. Discussion and Perspectives

Despite initially being considered as transcriptional noise, the importance of ncRNAs arose with the observation that an organism's complexity is scarcely correlated with the number of protein-coding mRNAs, yet significantly scaled with the number of ncRNAs (Figure 3) (Mattick, 2004). Since then, some lncRNAs are today well acknowledged as key regulators of numerous cellular activities (for review, see (Jarroux et al., 2017; Yao et al., 2019; Statello et al., 2021)). Importantly, lncRNAs are tissue and cancer-specific (Carlevaro-Fita et al., 2020; Iyer et al., 2015; Lorenzi et al., 2021), and their expression is also tightly regulated to ensure proper cell homeostasis. Increasing studies show that lncRNAs are recognized as new sensitive, non-invasive biomarkers for cancer progression and therapeutical targets for future drug developments (for review, see (Meseure et al., 2015)). Recently, even if being *a priori* defined as non-coding, proteomic analyses revealed a novel function of lncRNAs identifying them as a source of micropeptides (Ruiz-Orera et al., 2014). Interestingly, despite their small size, some of the lncRNA-derived peptides are functional and can trigger cancer phenotypes (Iyer et al., 2015; Carlevaro-Fita et al., 2020; Lorenzi et al., 2021; Almeida et al., 2022). However, the extent, as well as the biological and mechanistic relevance of this pervasive lncRNAs translation remains poorly studied.

In the lab we investigate the post-transcriptional metabolism of RNAs, using yeast unstable cytoplasmic aslncRNAs as a paradigm. These RNAs are efficiently degraded by RNA decay machineries (Tisseur et al., 2011). Therefore, to tackle their low cellular levels, we study them by using mutants in which the decay factors responsible for their degradation are inactive. In particular, we stary a class of cryptic transcripts referred to as XUTs as they are sensitive to the 5'-exoribonuclease cytoplasmic Xrn1 (Van Dijk et al., 2011). Surprisingly, most aslncRNAs prior to degradation by Xrn1 are targeted by the translation-dependent NMD pathway in *S. cerevisiae* (Malabat et al., 2015; Wery et al., 2016), suggesting that they are translated and that translation controls their degradation. However, XUTs can also pair with their sense mRNA into dsRNAs strucutres, protecting them from the NMD (Wery et al., 2016), thus representing another key determinant of the Xrn1 and NMD-sensitivity of XUTs. However, the extent and regulatory mechanisms that determine the fate of lncRNAs subjected to translation/decay or pairing/stabilization are unknown.

In my thesis, we aimed at providing several observations reinforcing these concepts to unveil the impact of translation and dsRNA formation on the fate of cytoplasmic lncRNAs. *S. cerevisiae* is a particular organism as it has lost the RNAi pathway during evolution. Therefore, the lack of RNAi might have allowed the accumulation of the dsRNAs specifically in this organism. To explore this notion, it was important to study the dsRNAs in organisms with RNAi. In the lab, other than *S.*

cerevisiae, we use the fission yeast *S. pombe* and the budding yeast *N. castellii*. In *S. pombe* XUTs were found to accumulate upon Exo2 inactivation (Atkinson et al., 2018; Wery et al., 2018a), showing that XUTs are conserved in fission yeast. In addition, in fission yeast XUTs were found to be sensitive to NMD (Atkinson et al., 2018). However, in *S. pombe* the RNAi pathway is in the nucleus, not interfering with the cytoplasmic XUTs (Wery et al., 2018b). A model to study how the cytoplasmic RNAi machinery could interfere with the cytoplasmic aslncRNAs is *N. castellii*. In this context, we first defined the evolutionary conservation of the aslncRNAs landscape in *N. castellii*. We found out that despite the cytoplasmic RNAi, XUTs are detected and are mainly sensitive to Xrn1 if compared to Dicer (Figure 12). Potentially, one can imagine a scenario in which upon stabilization of asXUTs engaged in dsRNAs, the RNAi pathway should efficiently degrade them and no accumulation would be present. However, in the mutant of Xrn1 and Upf1, XUTs accumulate despite the presence of small RNAs, deriving from dsRNAs. Therefore, we could envision only a small fraction of asXUTs forming dsRNA, feeding the RNAi, while most of asXUTs do not engage in dsRNA structures and thus escape Dicer. This could be caused by the fact that at the single-cell level, XUTs accumulate mainly as single-stranded solo RNAs, not involved in dsRNA, or as co-expressed with the sense mRNAs but not forming dsRNAs. Alternatively, in the Xrn1-lacking cells, most of the stabilized XUTs engage into dsRNAs, that are so abundant that Dicer is not efficient enough to process all of them. Future work, including studies of the heterogeneity of the dsRNAs formation, will decipher the preferential route of aslncRNAs in organisms endowed with endogenous RNAi pathways.

Using S*. cerevisiae*, we investigated the molecular bases targeting aslncRNAs specifically to NMD. We defined to which extent aslncRNAs are bound by ribosomes and are actively translated and observed their accumulation upon translation elongation inhibition. We propose that the physical presence of elongating ribosomes on aslncRNAs protects these RNAs from the decay. We demonstrate the mechanistic elements activating the NMD surveillance pathway of aslncRNAs. Further, we provide a proof-of-principle that peptides deriving from an NMD-sensitive aslncRNA can be detected. Finally, using single-cell RNA-Seq, we tackle the heterogeneity of sense/antisense RNA expression and detect co-expression of both transcripts in a single yeast cell.

In addition to the separate discussions after each chapter in the results section, several other conceptual aspects of our study are yet to be clarified. Among them, how a transcript is recognized as an 'abnormal' substrate for NMD remains unclear. In higher Eukaryotes, the EJC model is proposed to activate the NMD (Le Hir et al., 2001). In yeast the proportion of introns is low and the EJC components are lacking. Studies suggest that a long 3' UTR region in all studied species constitutes a signal for NMD activation (Behm-Ansmant et al., 2007; Bühler et al., 2006; Eberle et al., 2008; Silva et al., 2008; Singh et al., 2008). Our data show that NMD-sensitive XUTs display a longer 3' UTR than the NMD-insensitive ones. However, a work using NMD reporters in yeast revealed that the length of the coding sequence also seems to be critical for NMD activation of mRNAs, excluding

the idea that a long 3' UTR by default leads to NMD (Decourty et al., 2014). Mainly, their data suggest that the ability of ribosomes to signal a defect in translation termination decreases with longer distances between start and stop codons and reveal short ORF length as a feature of NMD substrates in yeast. Another work in yeast showed that a vast majority of NMD-regulated transcripts comprise 'normal' mRNAs with coding regions enriched with non-optimal codons (Celik et al., 2017).

The strategy we used for the chimera construct allowed us to undoubtedly discriminate whether it is the length of the coding or the length of the 3'UTR triggering signal for the NMD of the tested XUT. Not only do we detect that the length of the 3' UTR is critical for its NMD-sensitivity, but also we render a NMD-resistant transcript a NMD-sensitive one 'simply' by adding to it a long native 3' UTR. Additionally, we observe that once the size of the 3' UTR is 298 nt or lower, it becomes less sensitive to NMD. Importantly, the size of this NMD-resistant XUT is in line with the median size 236 nt of the 3' UTR of NMD-insensitive XUTs obtained from our Ribo-Seq data. In addition, this finding is consistent with the observation that mRNAs in yeast with 3' UTR lengths greater than 300 nt are efficiently targeted to NMD (Kebaara and Atkin, 2009).

One could argue that NMD might provide a strong selection for short 3' UTRs by eliminating the expression of RNAs with long 3' UTRs. Nevertheless, some yeast and human mRNAs harbor a long 3' UTR yet escape NMD (Kebaara and Atkin, 2009; Mendell et al., 2004; Singh et al., 2008; Wittmann et al., 2006). This suggests that the increase in 3' UTR has evolved to accommodate regulatory elements that help evade NMD, such as stabilizing elements (Ruiz-Echevarría and Peltz, 2000) and secondary structures that bring the poly(A) tail into proximity of the stop codon (Eberle et al., 2008). However, it is worth mentioning that we failed to identify any specific robust motif in the 3' UTR of NMD-sensitive XUTs, which is consistent with the binding of Upf1 to RNAs without sequence specificity (Hurt et al., 2013; Sohrabi-Jahromi et al., 2019).

We speculate that the presence of the long 3' UTR on XUTs impedes the interaction between translation termination factors eRF1 and eRF3 and Pab1, necessary for efficient translation termination, as explained for mRNAs by the yeast 'faux 3' UTR' model (Amrani et al., 2004). As a consequence of the corrupted interaction, eRF1 and eRF3 would favor the binding with Upf1 at the level of the PTC and the stalled ribosome and activate NMD. Contrary to this classical view, another plausible model involves binding of Upf1 to RNA independently of translation, that can be displaced from the coding region by the translating ribosomes (Zünd et al., 2013). When the translating ribosome gets stalled at the level of a PTC of an NMD target, Upf1 association remains enhanced and proportional to the length of RNA downstream of the PTC. If a 3' UTR of the involved transcript is short, the concentration of bound Upf1 is low. If, however the 3' UTR length is extended, the probability of Upf1 binding is increased, marking the transcripts for NMD decay (Hogg and Goff, 2010; Serdar et al., 2016). Future mechanistic analysis will decipher which mechanism triggers longs 3' UTR of XUTs to NMD.

Despite the low abundance of XUTs, our Ribo-Seq analysis directly assessing the 3-nt periodicity, a hallmark of active translation, allowed us to identify actively translated smORFs for 38% of annotated XUTs. For the remaining portion of XUTs, we imagine that their translation remains fluctuating, probably reflecting the fact that they are rapidly and continuously evolving, thus challenging their detection using typical Ribo-Seq approach.

Apart from Ribo-Seq, other state-of-the-art techniques such as ribosome fractionation, and translating ribosome affinity purification (TRAP) discovered the presence of numerous non-canonical ORFs on cytoplasmic smORFs lncRNAs. Spectacularly, a recent study provided the first human atlas of RNA translation across 11 primary cells and tissues and revealed 7,767 small regions actively translated outside of the protein-coding genome, mostly exclusive to primates and cell or tissue specific (Chothani et al., 2022). Furthermore, even a method has been recently developed for mapping of microproteins to subcellular localizations (Na et al., 2022).

These and other studies unveiled several distinctive features between classic and non-canonical ORFs. Mainly, the non-canonical ones, apart from being more challenging to detect, are shorter in size, have lower transcription and translation levels, can initiate on near-cognate codons (different from AUG) (for review, see (Andreev et al., 2022)), are transcribed in RNAs that are less stable *in vivo* (Chen et al., 2020; Erhard et al., 2018; Fields et al., 2015; Ivanov et al., 2011; Lu et al., 2019; Samandi et al., 2017; Starck et al., 2012) and seem not to be conserved, rather are evolutionary transient with no signatures of selection (Wacholder et al., 2021, bioRxiv). Nevertheless, consistent with the notion that lack of conservation does not imply lack of function (Pang et al., 2006), a subset of evolutionary transient ORFs in yeast were reported in the literature to confer phenotypes (Wacholder et al., 2021, bioRxiv).

Our preliminary analysis using a classical Mass Spectrometry technique detected 100 peptides deriving from smORFs of XUTs (Figure 14). Alternative to the potential function of XUT-derived peptides described in the discussion of the result section (Figure 13), we propose the hypothesis that the NMD-sensitive lncRNA-derived peptides could be exposed to the natural selection and contribute to *de novo* gene birth. In this process, new genes can arise from ancestral non-coding transcripts and serve as raw material for the birth of novel genuine protein-coding genes, represent building blocks for protein construction in progress.

In principle, for a non-coding DNA to initiate functioning as a protein-coding gene, an ORF must originate, the DNA must be transcribed and the mRNA translated, and the deriving protein should ultimately become integrated into a cellular process. Such d*e novo* gene does not possess a homologous gene in any sister lineage. Once its expression at mRNA and protein level is confirmed, one can use genetic approaches to detect a specific phenotype or change in fitness upon disruption of a particular sequence. Several non-mutually exclusive mechanisms of *de novo* gene birth have been described: an ‘ORF first’ and ‘transcription first’ model. In the former an intergenic ORF gains

transcription, while in the latter an intergenic sequence gains transcription before evolving an ORF, which has been shown for lncRNAs. Since yeast XUT possess the characteristics mentioned above, indeed we could imagine a 'transcription first' model proposing how the XUT could have evolved in a *de novo* gene (Figure 27). Considering that *de novo* gene birth has been associated with adaptation to environmental stress (for review, see (Arendsee et al., 2014)), and that NMD can be downregulated under stress conditions (Gardner, 2008; Mendell et al., 2004), we speculate that a stress repressing NMD would allow the NMD-sensitive lncRNA to expose its peptide potential to selection. This could be fulfilled by a mutation of the stop codon, allowing the smORF to gain in length to a genuine ORFs. NMD would not target lncRNAs with a short 3' UTR thus allowing them to become stable and resistant to NMD. However, we can't exclude that the pressure of selection was at the level of the transcription unit of the RNAs, reducing the length of the transcript produced, thus containing a shorter 3' UTR and being protected from NMD. This could have been the case for NMD-insensitive aslncRNAs that are shorter than the NMD-sensitive ones. If we consider the smORF of a lncRNAs becoming a genuine ORFs, and producing a protein that if beneficial, would become positively selected, while the gene from which it derives would become a *bona fide* gene. During adaptive evolution, the importance of the gene function could increase due to positive selection, and the gene may become essential.

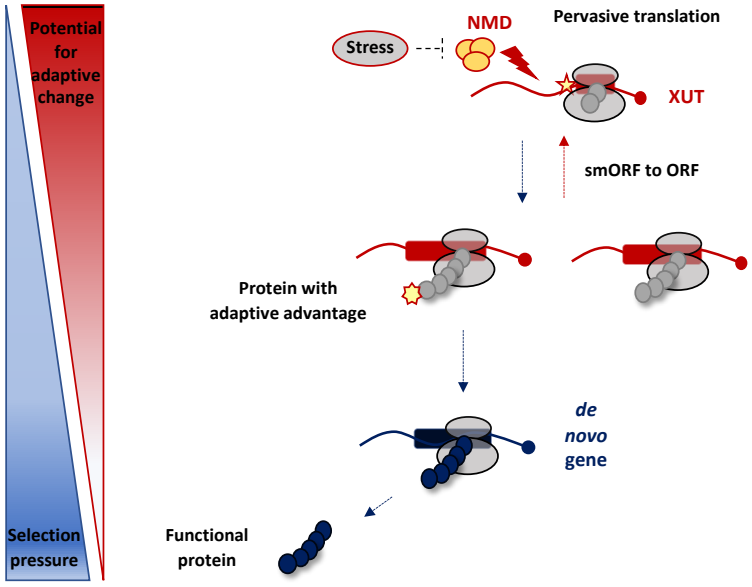


Figure 27 | Speculative Model of XUTs Evolving in a *De Novo* Gene. Ancestral pervasively transcribed and translated non-coding XUT with a smORF is rapidly degraded by NMD. Upon stress, NMD is inhibited, smORF of the XUT can evolve to a complete translatable ORF after a mutation of a stop codon (yellow star), thus becoming immune to NMD and stable. If the full-translated protein of the new gene is beneficial and increases the fitness of an organism (yellow heptagram), the gene could be co-opted for this adaptive function. Boxes and thin lines represent (sm)ORF and the UTRs, respectively. Left part of the graph from (Vakirlis et al., 2020).

The emergence of new genes in yeast has been linked to specific characteristics and possible physiological implications. Studies have reported condition-specific expression of *de novo* ORFs in response to various stresses, leading to the question of whether these translated elements represent a rapidly evolving part of the cell's response to stress. A study showed that overexpression of young *S. cerevisiae de novo* ORFs with predicted transmembrane domains can increase colony growth under nitrogen or carbon limitation (Vakirlis et al., 2020). Transmembrane domains are overrepresented among annotated *de novo* ORFs in *S. cerevisiae*, but the mechanisms behind this adaptation to starvation stress are yet to be fully understood. The study also proposed that the evolutionary mechanisms giving rise to *de novo* ORFs with transmembrane domains is likely a direct result of codon biases in the genetic code where transmembrane residues tend to be encoded by thymine-rich codons. The 'transmembrane-first' model proposes that translation of intergenic sequences that are rich in thymine have a high propensity to generate transmembrane peptides, which in turn are more likely to be adaptive and preserved by natural selection. We find that a another potential mechanism of regulation of *de novo* ORFs could involve the NMD pathway. Testing the fitness of the cells overexpressing those NMD-sensitive XUTs in conditions known to inhibit NMD would confirm our hypothesis. If the XUTs are involved in adaptation to stress, then inhibiting NMD should lead to an improvement in the fitness of the cells under stress. One could also test if those XUTs-derived peptides contain thymine-rich codons by analyzing the sequence of the coding region and calculating the codon usage.

Interestingly, the process of *de novo* evolution can also be reverted, and coding genes could become non-coding. There are several well-documented examples of swings between coding and non-coding gene capacities throughout evolution. For instance, *Xist* gene is a lncRNA partially evolved from a previously coding gene. Indeed, sequence comparison between species revealed that evolution of *Xist* gene was accompanied by pseudogenization of a coding gene located within a region of synteny between chicken and human and by integration of mobile elements that allowed the silencing function of *Xist* (Duret et al., 2006; Elisaphenko et al., 2008). The other way around, transposable elements that used to reside into introns got inserted into coding regions as novel exons in human (Nekrutenko and Li, 2001).

In addition, few examples of *de novo* genes perform an essential biological purpose. In *Drosophila,* knockdown of candidate *de novo* genes have suggested effects on viability and male fertility (Chen et al., 2010; Reinhardt et al., 2013). A recent study revealed a mouse-specific *de novo* gene involved in reproduction (Xie et al., 2019). Male fertility was also found to be affected for *Pldi* in mice, which codes for a lncRNA. In this case the knockout was shown to affect sperm motility and testis weight (Heinen et al., 2009). A human-specific *de novo* gene FLJ33706 was discovered to be highly expressed in Alzheimer's disease brain tissue (Li et al., 2010a). In yeast, the *de novo* evolved gene *BSC4* was reported to be involved in DNA repair (Cai et al., 2008) and *MDF1* (Li et al., 2010b,

2014) was associated to suppressing mating and to promoting fermentation. Originally, *BSC4* was identified as a translated ORF displaying Sup35-dependent translational readthrough (Namy et al., 2003). After, it was revealed that *BSC4* recently developed in *S. cerevisiae* lineage *via* point mutations in a previously non-coding gene (Cai et al., 2008). The *MDF1* ORF emerged *de novo* in *S. cerevisiae* in the previously non-coding sequence antisense to conserved protein-coding gene *ADF1* (Li et al., 2010b, 2014).

Considering that some XUTs have been related to regulatory functions (Berretta et al., 2008; Van Dijk et al., 2011; Wery et al., 2018b, 2018a) one could imagine a scenario in which they can have both coding and non-coding functions, thus classify as bifunctional RNAs. *Xenopus*, *E. coli*, *Drosophila*, mammals and other organisms each provide examples of bifunctional RNAs, in which the identical mature RNA embodies both coding and regulatory functions (Chooniedass-Kothari et al., 2004; Kloc et al., 2005; Rongo et al., 1995; Wadler and Vanderpool, 2007), the choice of which can be dynamically altered depending on cell states/types or on environmental cues, subcellular localization, or at the single-cell level determined by their (dis)ability to bind a partner RNA. To cite a particular example that could be explain by the latter case is SgrS (SuGar transport-Related small RNA), a bifunctional RNA discovered in *E. coli*, that can, upon excessive accumulation of sugar-phosphate, inhibit the translation of mRNAs encoding glucose transporters through base-pairing by hidding their ribosome binding site, but can also encode a functional polypeptide SgrT that prevents glucose transport using a mechanistically distinct pathway (Wadler and Vanderpool, 2007).

In the final part of my thesis, we investigate whether aslncRNAs can, other than translated, be also prone to bind their sense-paired mRNAs, forming dsRNAs. If so, they would impede the NMD to degrade them and become stable.

In mammals, the production of endogenous dsRNA carries, however, a substantial danger for the cell. Their presence conflicts with the cell's defense system against viruses, which is tailored to recognize and eliminate dsRNA. Namely, during viral infection there is a cytosolic accumulation of dsRNAs, reminiscent of the endogenous dsRNA, that triggers a strong inflammatory response in vertebrate systems (for review, see (Wang and Carmichael, 2004)). Accordingly, specific sensor proteins detect the dsRNA and induce the activation of interferons (Müller et al., 1994), which in return, induce an antiviral state in cells, together finally suppressing the translation of viral mRNAs and degrading the dsRNA. Despite the danger of eliciting this cascade, the presence of RNAi and the widespread antisense transcription of eukaryotic genomes have indicated that, endogenous dsRNA formation occurs in yeast (Drinnenberg et al., 2011; Sinturel et al., 2015), *Drosophila* (Ghildiyal et al., 2008), plants (Borsani et al., 2005) and mammals (Carlile et al., 2008; Watanabe et al., 2008; Werner et al., 2021). In particular, mouse testis (Werner et al., 2021), rich in antisense RNAs, and liver (Gao et al., 2020) are organs with widespread formation of dsRNA, as recently reported, using an antibody pull-down of dsRNAs (Werner et al., 2021).

Some of RNA:RNA duplexes, formed between sense and antisense RNAs, can trigger, other than immune response, various other mechanisms and regulatory cascades, such as RNA regulation, RNA masking, the establishment of chromatin marks and RNAi. The formation of transient dsRNA can alter the stability and translation of both interacting partners. Such example comes from the non-coding antisense *BACE1* transcript (*BACE1-AS*), particularly expressed in Alzheimer's disease, that increases the stability of the *BACE1* mRNA caused by dsRNA formation altering its secondary or tertiary structure (Faghihi et al., 2008). RNA masking is yet another proposed function of dsRNAs where the antisense transcript occludes a regulatory motif in the sense RNA by direct base pairing, altering the stability of the mRNA (for review, see (Faghihi and Wahlestedt, 2009)). Such regulatory motifs can be ribosome entry binding sites, splicing elements or miRNA-binding sites. A well-documented case of translation inhibition is the antisense for *PU.1* mRNA. The translation of the *PU.1* mRNA is inhibited by the non-coding antisense transcript, proposed to bind the sense transcript and stall its translation (Ebralidze et al., 2008). On the other hand, the link between dsRNA and their involvement in RNAi, processing by Dicer into endo-siRNAs, is well-established in *C. elegans* (Duchaine et al. 2006; Vasale et al. 2010) and *Drosophila* (Czech et al. 2008; Ghildiyal et al. 2008; Lucchetta et al. 2009), and *A. thaliana* (Borsani et al., 2005), less so in vertebrates (Watanabe et al. 2006, 2008; Carlile et al. 2009). In fission yeast, Dicer is restricted to the nucleus and associates to specific chromosomal regions, mainly the centromeric repeats (for review, see (Bühler and Moazed, 2007; Woolcock and Bühler, 2013)). The small RNAs that are Dicer-dependent are involved in directing and maintaining the formation of heterochromatin in fission yeast (Djupedal et al., 2009).

Altogether, the abovementioned examples from the literature highlight the important biological role of dsRNA structures that can, in addition to the coding-potential, be a key determinant of the fate of RNAs. Nevertheless, our single cell approach suggsts that the balance for the detectable aslncRNAs goes on the side of the translation-dependent degradation, rather than the dsRNA formation.

The decay of lncRNAs in human cells is more complex in comparison to yeast. Recent studies revealed that multiple lncRNAs have nuclear residence and undergo exosome-dependent 3'-5' degradation (Davidson et al., 2019; Szczepińska et al., 2015). Other studies have also proven that there are many lncRNAs that are exported and have important functions in the cytoplasm (Mukherjee et al., 2017). However, until recently, there was no clear identification of cytoplasmic surveillance actors degrading lncRNAs in that subcellular compartment. The high sequence similarity of Xrn1 between yeast and human suggested that lncRNAs could also be regulated by human hXrn1. Another actor of RNA degradation in humans is Dis3 with both nuclear and cytoplasmic residence (for review, see (Chlebowski et al., 2013)). In our lab, using an auxin inducible system to deplete Xrn1 in human cells, the preliminary results showed that low number of lncRNAs are stabilized in the cytoplasm (Foretek et al., in preparation). However, analysis of the cytoplasmic RNA content after

Dis3 depletion revealed substantially higher number of upregulated lncRNAs, suggesting that Dis3, in contrast to yeast Xrn1, has a major role in the turnover of human cytoplasmic lncRNAs. Our lab also performed Ribo-Seq in Dis3-depleted cells and found that an important fraction of Dis3-sensitive lncRNAs is bound to ribosomes, suggesting that these cytoplasmic lncRNAs could be potential source of peptides (Foretek et al., in preparation). Importantly, mutations of the Dis3 decay factor, are found in 10% of patients with multiple myeloma with poor prognosis (for review, see (Fasken et al., 2020)). The lab is currently investigating whether in those patients the upregulated Dis3-sensitive cytoplasmic lncRNAs are found and are a potential source of peptides (Foretek et al., in preparation), specific neoantigens, that could drive future therapeutical drug design. The mechanism underlying the control of lncRNA translation we provided in yeast will have unexpected implications in this regard.

To conclude, in my thesis, we aimed at describing the mechanisms of post-transcriptional regulation of lncRNA, enlightening the role of translation and dsRNA formation in modulating RNA stability *in vivo*. In particular, we uncover factors that control aslncRNAs metabolism, from RNA processing machineries to physiological stresses, from population to single cells. We determine that the translation activity is a major regulator of the aslncRNAs turnover and start investigating how their capacity to form dsRNA structures can antagonize with their translation-dependent degradation. Our results pave the way for the understanding of the fundamental bases controlling translation of both coding and non-coding transcriptomes in yeast and higher species.

Altogether, the results obtained during my PhD contribute to understanding the conceptual question of the still challenging *raison d'être* of the lncRNAs in the cell. We might soon undercover that in contrast to coding transcripts, whose duty is ‘only’ to produce proteins, the ‘non-coding’ transcripts include both, those that are translated relying on the peptide function and also those operating as RNA regulators.

REFERENCES

Abdel-Ghany, S.E., Hamilton, M., Jacobi, J.L., Ngam, P., Devitt, N., Schilkey, F., Ben-Hur, A., and Reddy, A.S.N. (2016). A survey of the sorghum transcriptome using single-molecule long reads. Nat. Commun. *7*, 11706. https://doi.org/10.1038/ncomms11706.

Alcid, E.A., and Tsukiyama, T. (2016). Expansion of antisense lncRNA transcriptomes since the loss of RNAi. Nat. Struct. Mol. Biol. *23*, 450–455. https://doi.org/10.1038/nsmb.3192.

Aldridge, S., and Teichmann, S.A. (2020). Single cell transcriptomics comes of age. Nat. Commun. *11*, 4307. https://doi.org/10.1038/s41467-020-18158-5.

Alexandrov, A., Colognori, D., and Steitz, J.A. (2011). Human eIF4AIII interacts with an eIF4G-like partner, NOM1, revealing an evolutionarily conserved function outside the exon junction complex. Genes Dev. *25*, 1078–1090. https://doi.org/10.1101/gad.2045411.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999). The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. Genes Dev. *13*, 2148–2158. https://doi.org/10.1101/gad.13.16.2148.

Almeida, A., Gabriel, M., Firlej, V., Martin-Jaular, L., Lejars, M., Cipolla, R., Petit, F., Vogt, N., San-Roman, M., Dingli, F., et al. (2022). Urinary extracellular vesicles contain mature transcriptome enriched in circular and long noncoding RNAs with functional significance in prostate cancer. J. Extracell. Vesicles *11*, e12210. https://doi.org/10.1002/jev2.12210.

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*. https://doi.org/10.1038/nature03060.

Anderson, J.S., and Parker, R.P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J. *17*, 1497–1506. https://doi.org/10.1093/emboj/17.5.1497.

Anderson, D.M., Anderson, K.M., Chang, C.-L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R., et al. (2015). A Micropeptide Encoded by a Putative Long Non-coding RNA Regulates Muscle Performance. Cell *160*, 595–606. https://doi.org/10.1016/j.cell.2015.01.009.

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. Non-Coding RNA *7*, 44. https://doi.org/10.3390/ncrna7030044.

Andjus, S., Szachnowski, U., Vogt, N., Hatin, I., Papadopoulos, C., Lopes, A., Namy, O., Wery, M., and Morillon, A. (2022). Translation is a key determinant controlling the fate of cytoplasmic long non-coding RNAs. 2022.05.25.493276. https://doi.org/10.1101/2022.05.25.493276.

Andreev, D.E., Loughran, G., Fedorova, A.D., Mikhaylova, M.S., Shatsky, I.N., and Baranov, P.V. (2022). Non-AUG translation initiation in mammals. Genome Biol. *23*, 111. https://doi.org/10.1186/s13059-022-02674-2.

Apcher, S., Daskalogianni, C., Lejeune, F., Manoury, B., Imhoos, G., Heslop, L., and Fåhraeus, R. (2011). Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. Proc. Natl. Acad. Sci. U. S. A. *108*, 11572–11577. https://doi.org/10.1073/pnas.1104104108.

Apone, L.M., Virbasius, C.M., Reese, J.C., and Green, M.R. (1996). Yeast TAF(II)90 is required for cell-cycle progression through G2/M but not for general transcription activation. Genes Dev. *10*, 2368–2380. https://doi.org/10.1101/gad.10.18.2368.

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

Arribere, J.A., and Gilbert, W.V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Res. *23*, 977–987. https://doi.org/10.1101/gr.150342.112.

Aspden, J.L., Eyre-Walker, Y.C., Phillips, R.J., Amin, U., Mumtaz, M.A.S., Brocard, M., and Couso, J.P. (2014). Extensive translation of small open reading frames revealed by poly-ribo-seq. ELife *3*. https://doi.org/10.7554/eLife.03528.

Atkinson, S.R., Marguerat, S., Bitton, D.A., Rodríguez-López, M., Rallis, C., Lemay, J.-F., Cotobal, C., Malecki, M., Smialowski, P., Mata, J., et al. (2018). Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. RNA N. Y. N *24*, 1195–1213. https://doi.org/10.1261/rna.065524.118.

Aznarez, I., Nomakuchi, T.T., Tetenbaum-Novatt, J., Rahman, M.A., Fregoso, O., Rees, H., and Krainer, A.R. (2018). Mechanism of Nonsense-Mediated mRNA Decay Stimulation by Splicing Factor SRSF1. Cell Rep. *23*, 2186–2198. https://doi.org/10.1016/j.celrep.2018.04.039.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science *318*, 798–801. https://doi.org/10.1126/science.1147182.

Bakel, H. van, Nislow, C., Blencowe, B.J., and Hughes, T.R. (2010). Most "Dark Matter" Transcripts Are Associated With Known Genes. PLOS Biol. *8*, e1000371. https://doi.org/10.1371/journal.pbio.1000371.

Balarezo-Cisneros, L.N., Parker, S., Fraczek, M.G., Timouma, S., Wang, P., O'Keefe, R.T., Millar, C.B., and Delneri, D. (2021). Functional and transcriptional profiling of non-coding RNAs in yeast reveal context-dependent phenotypes and in trans effects on the protein regulatory network. PLOS Genet. *17*, e1008761. https://doi.org/10.1371/journal.pgen.1008761.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215–233. https://doi.org/10.1016/j.cell.2009.01.002.

Baserga, S.J., and Benz, E.J. (1988). Nonsense mutations in the human beta-globin gene affect mRNA metabolism. Proc. Natl. Acad. Sci. *85*, 2056–2060. https://doi.org/10.1073/pnas.85.7.2056.

Bazzini, A.A., Johnstone, T.G., Christiano, R., MacKowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., et al. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J. *33*. https://doi.org/10.1002/embj.201488411.

Beadle, G.W., and Tatum, E.L. (1941). Genetic Control of Biochemical Reactions in Neurospora. Proc. Natl. Acad. Sci. U. S. A. *27*, 499–506. https://doi.org/10.1073/pnas.27.11.499.

Behm-Ansmant, I., Kashima, I., Rehwinkel, J., Saulière, J., Wittkopp, N., and Izaurralde, E. (2007). mRNA quality control: An ancient machinery recognizes and degrades mRNAs with nonsense codons.

Berman, J., and Sudbery, P.E. (2002). Candida Albicans: a molecular revolution built on lessons from budding yeast. Nat. Rev. Genet. *3*, 918–930. https://doi.org/10.1038/nrg948.

Bernstein, H.D., Zopf, D., Freymann, D.M., and Walter, P. (1993). Functional substitution of the signal recognition particle 54-kDa subunit by its Escherichia coli homolog. Proc. Natl. Acad. Sci. U. S. A. *90*, 5229–5233. https://doi.org/10.1073/pnas.90.11.5229.

Berretta, J., and Morillon, A. (2009). Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO Rep. *10*, 973–982. https://doi.org/10.1038/embor.2009.181.

Berretta, J., Pinskaya, M., and Morillon, A. (2008). A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev. *22*, 615–626. https://doi.org/10.1101/gad.458008.

Bertram, G., Bell, H.A., Ritchie, D.W., Fullerton, G., and Stansfield, I. (2000). Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. RNA N. Y. N *6*, 1236–1247. https://doi.org/10.1017/s1355838200000777.

Bi, P., Ramirez-Martinez, A., Li, H., Cannavino, J., McAnally, J.R., Shelton, J.M., Sánchez-Ortiz, E., Bassel-Duby, R., and Olson, E.N. (2017). Control of muscle formation by the fusogenic micropeptide myomixer. Science *356*, 323–327. https://doi.org/10.1126/science.aam9361.

Blanchet, S., Rowe, M., Von der Haar, T., Fabret, C., Demais, S., Howard, M.J., and Namy, O. (2015). New insights into stop codon recognition by eRF1. Nucleic Acids Res. *43*, 3298–3308. https://doi.org/10.1093/nar/gkv154.

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

Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell *139*, 547–559. https://doi.org/10.1016/j.cell.2009.08.042.

Bonté, P.-E., Arribas, Y.A., Merlotti, A., Carrascal, M., Zhang, J.V., Zueva, E., Binder, Z.A., Alanio, C., Goudot, C., and Amigorena, S. (2022). Single-cell RNA-seq-based proteogenomics identifies glioblastoma-specific transposable elements encoding HLA-I-presented peptides. Cell Rep. *39*, 110916. https://doi.org/10.1016/j.celrep.2022.110916.

Booth, G.T., Wang, I.X., Cheung, V.G., and Lis, J.T. (2016). Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. Genome Res. *26*, 799–811. https://doi.org/10.1101/gr.204578.116.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.-K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell *123*, 1279–1291. https://doi.org/10.1016/j.cell.2005.11.035.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling.

Brown, C.E., and Sachs, A.B. (1998). Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. Mol. Cell. Biol. *18*, 6548–6559. https://doi.org/10.1128/MCB.18.11.6548.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527–542. https://doi.org/10.1016/0092-8674(92)90520-m.

Bühler, M., and Moazed, D. (2007). Transcription and RNAi in heterochromatic gene silencing. Nat. Struct. Mol. Biol. *14*, 1041–1048. https://doi.org/10.1038/nsmb1315.

Bühler, M., Steiner, S., Mohn, F., Paillusson, A., and Mühlemann, O. (2006). EJC-independent degradation of nonsense immunoglobulin-μ mRNA depends on 3′ UTR length. Nat. Struct. Mol. Biol. *13*. https://doi.org/10.1038/nsmb1081.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927. https://doi.org/10.1101/gad.17446611.

Cai, J., Zhao, R., Jiang, H., and Wang, W. (2008). De novo origination of a new protein-coding gene in Saccharomyces cerevisiae. Genetics *179*, 487–496. https://doi.org/10.1534/genetics.107.084491.

Cali, B.M., Kuchma, S.L., Latham, J., and Anderson, P. (1999). smg-7 is required for mRNA surveillance in Caenorhabditis elegans. Genetics *151*, 605–616. https://doi.org/10.1093/genetics/151.2.605.

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell *131*, 706–717. https://doi.org/10.1016/j.cell.2007.09.014.

Camblong, J., Beyrouthy, N., Guffanti, E., Schlaepfer, G., Steinmetz, L.M., and Stutz, F. (2009). Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. Genes Dev. *23*, 1534–1545. https://doi.org/10.1101/gad.522509.

Carlevaro-Fita, J., and Johnson, R. (2019). Global Positioning System: Understanding Long Noncoding RNAs through Subcellular Localization. Mol. Cell *73*, 869–883. https://doi.org/10.1016/j.molcel.2019.02.008.

Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L.A., and Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. RNA N. Y. N *22*, 867–882. https://doi.org/10.1261/rna.053561.115.

Carlevaro-Fita, J., Lanzós, A., Feuerbach, L., Hong, C., Mas-Ponte, D., Pedersen, J.S., and Johnson, R. (2020). Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. Commun. Biol. *3*, 1–16. https://doi.org/10.1038/s42003-019-0741-7.

Carlile, M., Nalbant, P., Preston-Fayers, K., McHaffie, G.S., and Werner, A. (2008). Processing of naturally occurring sense/antisense transcripts of the vertebrate Slc34a gene into short RNAs. Physiol. Genomics *34*, 95–100. https://doi.org/10.1152/physiolgenomics.00004.2008.

Carlile, M., Swan, D., Jackson, K., Preston-Fayers, K., Ballester, B., Flicek, P., and Werner, A. (2009). Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene Slc34a1 in murine tissues. Nucleic Acids Res. *37*, 2274–2282. .

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559–1563. https://doi.org/10.1126/science.1112014.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature *491*, 454–457. https://doi.org/10.1038/nature11508.

Carroll, K.L., Ghirlando, R., Ames, J.M., and Corden, J.L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. RNA *13*, 361–373. https://doi.org/10.1261/rna.338407.

Carter, M.S., Doskow, J., Morris, P., Li, S., Nhim, R.P., Sandstedt, S., and Wilkinson, M.F. (1995). A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. J. Biol. Chem. *270*, 28995–29003. https://doi.org/10.1074/jbc.270.48.28995.

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*. https://doi.org/10.1038/nature11184.

Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F., and Zenklusen, D. (2013). Bimodal expression of PHO84 is modulated by early termination of antisense transcription. Nat. Struct. Mol. Biol. *20*, 851–858. https://doi.org/10.1038/nsmb.2598.

Causier, B., Li, Z., De Smet, R., Lloyd, J.P.B., Van De Peer, Y., and Davies, B. (2017). Conservation of Nonsense-Mediated mRNA Decay Complex Components Throughout Eukaryotic Evolution. Sci. Rep. *7*. https://doi.org/10.1038/s41598-017-16942-w.

Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. Cell *157*, 77–94. https://doi.org/10.1016/j.cell.2014.03.008.

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 N. Y. N *23*, 735–748. https://doi.org/10.1261/rna.060541.116.

Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011). Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. Mol. Cell *41*, 693–703. https://doi.org/10.1016/j.molcel.2011.02.010.

Chamieh, H., Ballut, L., Bonneau, F., and Le Hir, H. (2008). NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat. Struct. Mol. Biol. *15*, 85–93. https://doi.org/10.1038/nsmb1330.

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*, e32536. https://doi.org/10.7554/eLife.32536.

Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The Nonsense-mediated decay RNA surveillance pathway.

Chen, C.Y., Gherzi, R., Ong, S.E., Chan, E.L., Raijmakers, R., Pruijn, G.J., Stoecklin, G., Moroni, C., Mann, M., and Karin, M. (2001). AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell *107*, 451–464. https://doi.org/10.1016/s0092-8674(01)00578-5.

Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., et al. (2020). Pervasive functional translation of noncanonical human open reading frames. Science *367*, 1140–1146. https://doi.org/10.1126/science.aay0262.

Chen, S., Zhang, Y.E., and Long, M. (2010). New Genes in Drosophila Quickly Become Essential. Science *330*, 1682–1685. https://doi.org/10.1126/science.1196380.

Chernyakov, I., Whipple, J.M., Kotelawala, L., Grayhack, E.J., and Phizicky, E.M. (2008). Degradation of several hypomodified mature tRNA species in *Saccharomyces cerevisiae* is mediated by Met22 and the 5'–3' exonucleases Rat1 and Xrn1. Genes Dev. *22*, 1369–1380. https://doi.org/10.1101/gad.1654308.

Chew, G.L., Pauli, A., Rinn, J.L., Regev, A., Schier, A.F., and Valen, E. (2013). Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. Dev. Camb. *140*. https://doi.org/10.1242/dev.098343.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: the exosome. Biochim. Biophys. Acta *1829*, 552–560. https://doi.org/10.1016/j.bbagrm.2013.01.006.

Choi, S.-W., Kim, H.-W., and Nam, J.-W. (2019). The small peptide world in long noncoding RNAs. Brief. Bioinform. *20*, 1853–1864. https://doi.org/10.1093/bib/bby055.

Chooniedass-Kothari, S., Emberley, E., Hamedani, M. k., Troup, S., Wang, X., Czosnek, A., Hube, F., Mutawe, M., Watson, P. h., and Leygue, E. (2004). The steroid receptor RNA activator is the first functional RNA encoding a protein. FEBS Lett. *566*, 43–47. https://doi.org/10.1016/j.febslet.2004.03.104.

Chothani, S.P., Adami, E., Widjaja, A.A., Langley, S.R., Viswanathan, S., Pua, C.J., Zhihao, N.T., Harmston, N., D'Agostino, G., Whiffin, N., et al. (2022). A high-resolution map of human RNA translation. Mol. Cell *82*, 2885-2899.e8. https://doi.org/10.1016/j.molcel.2022.06.023.

Church, D.M., Goodstadt, L., Hillier, L.W., Zody, M.C., Goldstein, S., She, X., Bult, C.J., Agarwala, R., Cherry, J.L., DiCuccio, M., et al. (2009). Lineage-Specific Biology Revealed by a Finished Genome Assembly of the Mouse. PLoS Biol. *7*, e1000112. https://doi.org/10.1371/journal.pbio.1000112.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373. https://doi.org/10.1038/nature09652.

Clark, M.B., and Mattick, J.S. (2011). Long noncoding RNAs in cell biology. Semin. Cell Dev. Biol. *22*, 366–376. https://doi.org/10.1016/j.semcdb.2011.01.001.

Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., Morillon, A., et al. (2011). The Reality of Pervasive Transcription. PLOS Biol. *9*, e1000625. https://doi.org/10.1371/journal.pbio.1000625.

Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., Fox, A.H., Fortini, E., Moscato, P., Dinger, M.E., and Mattick, J.S. (2012). Genome-wide analysis of long noncoding RNA stability. Genome Res. *22*, 885–898. https://doi.org/10.1101/gr.131037.111.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275. https://doi.org/10.1083/jcb.132.3.259.

Colombo, M., Karousis, E.D., Bourquin, J., Bruggmann, R., and Mühlemann, O. (2017). Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy

between SMG6- and SMG7-mediated degradation pathways. RNA N. Y. N *23*, 189–201. https://doi.org/10.1261/rna.059055.116.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848. https://doi.org/10.1126/science.1162228.

Crick, F. (1970). Central Dogma of Molecular Biology. Nature *227*, 561–563. https://doi.org/10.1038/227561a0.

de la Cruz, J., Kressler, D., Tollervey, D., and Linder, P. (1998). Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae. EMBO J. *17*, 1128–1140. https://doi.org/10.1093/emboj/17.4.1128.

Cui, Y., Hagan, K.W., Zhang, S., and Peltz, S.W. (1995). Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. Genes Dev. *9*, 423–436. https://doi.org/10.1101/gad.9.4.423.

Culbertson, M.R., and Leeds, P.F. (2003). Looking at mRNA decay pathways through the window of molecular evolution. Curr. Opin. Genet. Dev. *13*, 207–214. https://doi.org/10.1016/s0959-437x(03)00014-5.

Culbertson, M.R., Underbrink, K.M., and Fink, G.R. (1980). Frameshift suppression Saccharomyces cerevisiae. II. Genetic properties of group II suppressors. Genetics *95*, 833–853. https://doi.org/10.1093/genetics/95.4.833.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci. U. S. A. *103*, 5320–5325. https://doi.org/10.1073/pnas.0601091103.

Davidson, L., Francis, L., Cordiner, R.A., Eaton, J.D., Estell, C., Macias, S., Cáceres, J.F., and West, S. (2019). Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. Cell Rep. *26*, 2779-2791.e5. https://doi.org/10.1016/j.celrep.2019.02.012.

Davis, L., and Engebrecht, J. (1998). Yeast dom34 mutants are defective in multiple developmental pathways and exhibit decreased levels of polyribosomes. Genetics *149*, 45–56. https://doi.org/10.1093/genetics/149.1.45.

Dawkins, R., 1941- (1989). The selfish gene (New edition. Oxford ; New York : Oxford University Press, 1989.).

Decker, C.J., and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. Genes Dev. *7*, 1632–1643. https://doi.org/10.1101/gad.7.8.1632.

Decourty, L., Doyen, A., Malabat, C., Frachon, E., Rispal, D., Séraphin, B., Feuerbach, F., Jacquier, A., and Saveanu, C. (2014). Long Open Reading Frame Transcripts Escape Nonsense-Mediated mRNA Decay in Yeast. Cell Rep. *6*. https://doi.org/10.1016/j.celrep.2014.01.025.

Defenouillère, Q., and Fromont-Racine, M. (2017). The ribosome-bound quality control complex: from aberrant peptide clearance to proteostasis maintenance. Curr. Genet. *63*, 997–1005. https://doi.org/10.1007/s00294-017-0708-5.

Defenouillère, Q., Namane, A., Mouaikel, J., Jacquier, A., and Fromont-Racine, M. (2017). The ribosome-bound quality control complex remains associated to aberrant peptides during their proteasomal targeting and interacts with Tom1 to limit protein aggregation. Mol. Biol. Cell *28*, 1165–1176. https://doi.org/10.1091/mbc.E16-10-0746.

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*. https://doi.org/10.15252/embj.201899278.

Denis, C.L., and Chen, J. (2003). The CCR4-NOT complex plays diverse roles in mRNA metabolism. Prog. Nucleic Acid Res. Mol. Biol. *73*, 221–250. https://doi.org/10.1016/s0079-6603(03)01007-9.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. *22*. https://doi.org/10.1101/gr.132159.111.

Deshmukh, M.V., Jones, B.N., Quang-Dang, D.-U., Flinders, J., Floor, S.N., Kim, C., Jemielity, J., Kalek, M., Darzynkiewicz, E., and Gross, J.D. (2008). mRNA decapping is promoted by an RNA-binding channel in Dcp2. Mol. Cell *29*, 324–336. https://doi.org/10.1016/j.molcel.2007.11.027.

Didion, J.P., Martin, M., and Collins, F.S. (2017). Atropos: specific, sensitive, and speedy trimming of sequencing reads. PeerJ *5*, e3720. https://doi.org/10.7717/peerj.3720.

Dinger, M.E., Amaral, P.P., Mercer, T.R., and Mattick, J.S. (2009). Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. Brief. Funct. Genomic. Proteomic. *8*, 407–423. https://doi.org/10.1093/bfgp/elp038.

DiStefano, J.K. (2018). The Emerging Role of Long Noncoding RNAs in Human Disease. Methods Mol. Biol. Clifton NJ *1706*, 91–110. https://doi.org/10.1007/978-1-4939-7471-9_6.

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. https://doi.org/10.1038/nature11233.

Djupedal, I., Kos-Braun, I.C., Mosher, R.A., Soderholm, N., Simmer, F., Hardcastle, T.J., Fender, A., Heidrich, N., Kagansky, A., Bayne, E., et al. (2009). Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. EMBO J. *28*, 3832–3844. https://doi.org/10.1038/emboj.2009.351.

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. https://doi.org/10.1038/nchembio.2249.

Doma, M.K., and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature *440*, 561–564. https://doi.org/10.1038/nature04530.

Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature *284*, 601–603. https://doi.org/10.1038/284601a0.

D'Orazio, K.N., Wu, C.C.-C., Sinha, N., Loll-Krippleber, R., Brown, G.W., and Green, R. (2019). The endonuclease Cue2 cleaves mRNAs at stalled ribosomes during No Go Decay. ELife *8*, e49117. https://doi.org/10.7554/eLife.49117.

Dostie, J., and Dreyfuss, G. (2002). Translation Is Required to Remove Y14 from mRNAs in the Cytoplasm. Curr. Biol. *12*, 1060–1067. https://doi.org/10.1016/S0960-9822(02)00902-8.

Dozier, C., Montigny, A., Viladrich, M., Culerrier, R., Combier, J.-P., Besson, A., and Plaza, S. (2022). Small ORFs as New Regulators of Pri-miRNAs and miRNAs Expression in Human and Drosophila. Int. J. Mol. Sci. *23*, 5764. https://doi.org/10.3390/ijms23105764.

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*. https://doi.org/10.1126/science.1176945.

Drinnenberg, I.A., Fink, G.R., and Bartel, D.P. (2011). Compatibility with killer explains the rise of RNAi-deficient fungi. Science *333*, 1592. https://doi.org/10.1126/science.1209575.

Duncan, C.D.S., and Mata, J. (2014). The translational landscape of fission-yeast meiosis and sporulation. Nat. Struct. Mol. Biol. *21*. https://doi.org/10.1038/nsmb.2843.

Duret, L., Chureau, C., Samain, S., Weissenbach, J., and Avner, P. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science *312*, 1653–1655. https://doi.org/10.1126/science.1126316.

Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z., and Mühlemann, O. (2008). Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. PLoS Biol. *6*. https://doi.org/10.1371/journal.pbio.0060092.

Eberle, A.B., Lykke-Andersen, S., Mühlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat. Struct. Mol. Biol. *16*. https://doi.org/10.1038/nsmb.1530.

Ebralidze, A.K., Guibal, F.C., Steidl, U., Zhang, P., Lee, S., Bartholdy, B., Jorda, M.A., Petkova, V., Rosenbauer, F., Huang, G., et al. (2008). PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. Genes Dev. *22*, 2085–2092. https://doi.org/10.1101/gad.1654808.

Eddy, S.R. (2012). The C-value paradox, junk DNA and ENCODE. Curr. Biol. CB *22*, R898-899. https://doi.org/10.1016/j.cub.2012.10.002.

Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N., and Zakian, S.M. (2008). A Dual Origin of the Xist Gene from a Protein-Coding Gene and a Set of Transposable Elements. PLOS ONE *3*, e2521. https://doi.org/10.1371/journal.pone.0002521.

ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816. https://doi.org/10.1038/nature05874.

Faghihi, M.A., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol. *10*, 637–643. https://doi.org/10.1038/nrm2738.

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St. Laurent III, G., Kenny, P.J., and Wahlestedt, C. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. Nat. Med. *14*, 723–730. https://doi.org/10.1038/nm1784.

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, X., Teng, X., Sun, X., et al. (2018). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. Nucleic Acids Res. *46*, D308–D314. https://doi.org/10.1093/nar/gkx1107.

Farazi, T.A., Juranek, S.A., and Tuschl, T. (2008). The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Dev. Camb. Engl. *135*, 1201–1214. https://doi.org/10.1242/dev.005629.

Faridani, O., Abdullayev, I., Hagemann-Jensen, M., Schell, J., Lanner, F., and Sandberg, R. (2016). Single-cell sequencing of the small-RNA transcriptome. Nat. Biotechnol. *34*. https://doi.org/10.1038/nbt.3701.

Fasken, M.B., Morton, D.J., Kuiper, E.G., Jones, S.K., Leung, S.W., and Corbett, A.H. (2020). The RNA Exosome and Human Disease. Methods Mol. Biol. Clifton NJ *2062*, 3–33. https://doi.org/10.1007/978-1-4939-9822-7_1.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. *15*, 7–21. https://doi.org/10.1038/nrg3606.

Fattahi, S., Kosari-Monfared, M., Golpour, M., Emami, Z., Ghasemiyan, M., Nouri, M., and Akhavan-Niaki, H. (2020). LncRNAs as potential diagnostic and prognostic biomarkers in gastric cancer: A novel approach to personalized medicine. J. Cell. Physiol. *235*, 3189–3206. https://doi.org/10.1002/jcp.29260.

Feng, L., Liao, Y.-T., He, J.-C., Xie, C.-L., Chen, S.-Y., Fan, H.-H., Su, Z.-P., and Wang, Z. (2018). Plasma long non-coding RNA BACE1 as a novel biomarker for diagnosis of Alzheimer disease. BMC Neurol. *18*, 4. https://doi.org/10.1186/s12883-017-1008-x.

Fiorini, F., Bagchi, D., Le Hir, H., and Croquette, V. (2015). Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. Nat. Commun. *6*, 7581. https://doi.org/10.1038/ncomms8581.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806–811. https://doi.org/10.1038/35888.

Fleming, J.V., Hay, S.M., Harries, D.N., and Rees, W.D. (1998). Effects of nutrient deprivation and differentiation on the expression of growth-arrest genes (gas and gadd) in F9 embryonal carcinoma cells. Biochem. J. *330*, 573–579. .

Fourati, Z., and Graille, M. (2014). Cytoplasmic mRNA Surveillance Pathways. In Fungal RNA Biology, A. Sesma, and T. von der Haar, eds. (Cham: Springer International Publishing), pp. 195–216.

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Armstrong, J., Barnes, I., et al. (2021). GENCODE 2021. Nucleic Acids Res. *49*, D916–D923. https://doi.org/10.1093/nar/gkaa1087.

Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R., and Dietz, H.C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science *295*, 2258–2261. https://doi.org/10.1126/science.1067338.

Fritz, S.E., Ranganathan, S., Wang, C.D., and Hogg, J.R. (2022). An alternative UPF1 isoform drives conditional remodeling of nonsense-mediated mRNA decay. EMBO J. *41*, e108898. https://doi.org/10.15252/embj.2021108898.

Gaba, A., Jacobson, A., and Sachs, M.S. (2005). Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. Mol. Cell *20*. https://doi.org/10.1016/j.molcel.2005.09.019.

Gao, Y., Vasic, R., Song, Y., Teng, R., Liu, C., Gbyli, R., Biancon, G., Nelakanti, R., Lobben, K., Kudo, E., et al. (2020). m6A Modification Prevents Formation of Endogenous Double-Stranded RNAs and Deleterious Innate Immune Responses during Hematopoietic Development. Immunity *52*, 1007-1021.e8. https://doi.org/10.1016/j.immuni.2020.05.003.

Gardner, L.B. (2008). Hypoxic Inhibition of Nonsense-Mediated RNA Decay Regulates Gene Expression and the Integrated Stress Response. Mol. Cell. Biol. *28*. https://doi.org/10.1128/mcb.02284-07.

Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. EMBO J. *22*, 3960–3970. https://doi.org/10.1093/emboj/cdg371.

Geerlings, T.H., Vos, J.C., and Raué, H.A. (2000). The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by 5'-->3' exonucleases. RNA N. Y. N *6*, 1698–1703. https://doi.org/10.1017/s1355838200001540.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Mol. Cell *20*, 65–75. https://doi.org/10.1016/j.molcel.2005.08.012.

Geisberg, J.V., Moqtaderi, Z., Fan, X., Ozsolak, F., and Struhl, K. (2014). Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. Cell *156*, 812–824. https://doi.org/10.1016/j.cell.2013.12.026.

Geisler, S., Lojek, L., Khalil, A.M., Baker, K.E., and Coller, J. (2012). Decapping of long non-coding RNAs regulates inducible genes. Mol. Cell *45*, 279–291. https://doi.org/10.1016/j.molcel.2011.11.025.

Gelfand, B., Mead, J., Bruning, A., Apostolopoulos, N., Tadigotla, V., Nagaraj, V., Sengupta, A.M., and Vershon, A.K. (2011). Regulated Antisense Transcription Controls Expression of Cell-Type-Specific Genes in Yeast ▿. Mol. Cell. Biol. *31*, 1701–1709. https://doi.org/10.1128/MCB.01071-10.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L.W., Zapp, M.L., Weng, Z., et al. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077–1081. https://doi.org/10.1126/science.1157396.

Goetz, A.E., and Wilkinson, M. (2017). Stress and the nonsense-mediated RNA decay pathway.

Goffeau, A., Barrell, G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. (1996). Life with 6000 genes. Science *274*. https://doi.org/10.1126/science.274.5287.546.

Gong, C., and Maquat, L.E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature *470*, 284–288. https://doi.org/10.1038/nature09701.

Gourvest, M., Brousset, P., and Bousquet, M. (2019). Long Noncoding RNAs in Acute Myeloid Leukemia: Functional Characterization and Clinical Relevance. Cancers *11*, 1638. https://doi.org/10.3390/cancers11111638.

Granovskaia, M.V., Jensen, L.J., Ritchie, M.E., Toedling, J., Ning, Y., Bork, P., Huber, W., and Steinmetz, L.M. (2010). High-resolution transcription atlas of the mitotic cell cycle in budding yeast. Genome Biol. *11*, R24. https://doi.org/10.1186/gb-2010-11-3-r24.

Grelet, S., Link, L.A., Howley, B., Obellianne, C., Palanisamy, V., Gangaraju, V.K., Diehl, J.A., and Howe, P.H. (2017). A regulated PNUTS mRNA to lncRNA splice switch mediates EMT and tumour progression. Nat. Cell Biol. *19*, 1105–1115. https://doi.org/10.1038/ncb3595.

Guan, Q., Zheng, W., Tang, S., Liu, X., Zinkel, R.A., Tsui, K.-W., Yandell, B.S., and Culbertson, M.R. (2006). Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. PLoS Genet. *2*, e203. https://doi.org/10.1371/journal.pgen.0020203.

Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat. Struct. Mol. Biol. *15*, 786–794. https://doi.org/10.1038/nsmb.1460.

Guo, S., and Kemphues, K.J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611–620. https://doi.org/10.1016/0092-8674(95)90082-9.

Guo, C.-J., Ma, X.-K., Xing, Y.-H., Zheng, C.-C., Xu, Y.-F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G.G., et al. (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. Cell *181*, 621-636.e22. https://doi.org/10.1016/j.cell.2020.03.006.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227. https://doi.org/10.1038/nature07672.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. *28*, 503–510. https://doi.org/10.1038/nbt.1633.

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell *154*. https://doi.org/10.1016/j.cell.2013.06.009.

Hackl, T., Hedrich, R., Schultz, J., and Förster, F. (2014). proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinforma. Oxf. Engl. *30*, 3004–3011. https://doi.org/10.1093/bioinformatics/btu392.

Halbach, F., Reichelt, P., Rode, M., and Conti, E. (2013). The Yeast Ski Complex: Crystal Structure and RNA Channeling to the Exosome Complex. Cell *154*, 814–826. https://doi.org/10.1016/j.cell.2013.07.017.

Hall, G.W., and Thein, S. (1994). Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia. Blood *83*, 2031–2037. .

Harigaya, Y., and Parker, R. (2010). No-go decay: a quality control mechanism for RNA in translation. Wiley Interdiscip. Rev. RNA *1*, 132–141. https://doi.org/10.1002/wrna.17.

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. https://doi.org/10.1016/s1097-2765(03)00446-5.

He, F., Celik, A., Wu, C., and Jacobson, A. (2018). General decapping activators target different subsets of inefficiently translated mRNAs. ELife *7*, e34409. https://doi.org/10.7554/eLife.34409.

He, F., Wu, C., and Jacobson, A. (2022). Dcp2 C-terminal cis-binding elements control selective targeting of the decapping enzyme by forming distinct decapping complexes. ELife *11*, e74410. https://doi.org/10.7554/eLife.74410.

He, R.-Z., Luo, D.-X., and Mo, Y.-Y. (2019). Emerging roles of lncRNAs in the post-transcriptional regulation in cancer. Genes Dis. *6*, 6–15. https://doi.org/10.1016/j.gendis.2019.01.003.

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.e29. https://doi.org/10.1016/j.cell.2019.05.010.

Hellen, C.U.T. (2018). Translation Termination and Ribosome Recycling in Eukaryotes. Cold Spring Harb. Perspect. Biol. *10*, a032656. https://doi.org/10.1101/cshperspect.a032656.

Henz, S.R., Cumbie, J.S., Kasschau, K.D., Lohmann, J.U., Carrington, J.C., Weigel, D., and Schmid, M. (2007). Distinct expression patterns of natural antisense transcripts in Arabidopsis. Plant Physiol. *144*, 1247–1255. https://doi.org/10.1104/pp.107.100396.

Herrero, E., Sanz, P., and Sentandreu, R. (1987). Cell Wall Proteins Liberated by Zymolyase from Several Ascomycetous and Imperfect Yeasts. Microbiol.-Sgm *133*, 2895–2903. https://doi.org/10.1099/00221287-133-10-2895.

Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nat. Methods *14*, 1198–1204. https://doi.org/10.1038/nmeth.4435.

Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P., and Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep. *11*, 1110–1122. https://doi.org/10.1016/j.celrep.2015.04.023.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics *123*, 301–313. https://doi.org/10.1093/genetics/123.2.301.

Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell *143*, 379–389. https://doi.org/10.1016/j.cell.2010.10.005.

Hongay, C.F., Grisafi, P.L., Galitski, T., and Fink, G.R. (2006). Antisense transcription controls cell fate in Saccharomyces cerevisiae. Cell *127*, 735–745. https://doi.org/10.1016/j.cell.2006.09.038.

van Hoof, A., Frischmeyer, P.A., Dietz, H.C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science *295*, 2262–2264. https://doi.org/10.1126/science.1067272.

Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., and Vogelauer, M. (2008). A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol. Cell *32*, 685–695. https://doi.org/10.1016/j.molcel.2008.09.027.

Hsu, C.L., and Stevens, A. (1993). Yeast cells lacking 5'-->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. Mol. Cell. Biol. *13*, 4826–4835. https://doi.org/10.1128/mcb.13.8.4826-4835.1993.

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. https://doi.org/10.1038/nature08265.

Hu, W., Yuan, B., Flygare, J., and Lodish, H.F. (2011). Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. Genes Dev. *25*, 2573–2578. https://doi.org/10.1101/gad.178780.111.

Huang, N., Li, F., Zhang, M., Zhou, H., Chen, Z., Ma, X., Yang, L., Wu, X., Zhong, J., Xiao, F., et al. (2021). An Upstream Open Reading Frame in Phosphatase and Tensin Homolog Encodes a Circuit Breaker of Lactate Metabolism. Cell Metab. *33*, 454. https://doi.org/10.1016/j.cmet.2021.01.008.

Huarte, M., and Rinn, J.L. (2010). Large non-coding RNAs: missing links in cancer? Hum. Mol. Genet. *19*, R152–R161. https://doi.org/10.1093/hmg/ddq353.

Hurt, J.A., Robertson, A.D., and Burge, C.B. (2013). Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. Genome Res. *23*. https://doi.org/10.1101/gr.157354.113.

Inagaki, Y., Blouin, C., Susko, E., and Roger, A.J. (2003). Assessing functional divergence in EF-1α and its paralogs in eukaryotes and archaebacteria. Nucleic Acids Res. *31*, 4227–4237. .

Inglis, A.J., Guna, A., Merchán, Á.G., Pal, A., Esantsi, T.K., Keys, H.R., Frenkel, E.M., Oania, R., Weissman, J.S., and Voorhees, R.M. (2022). Coupled protein quality control during nonsense mediated mRNA decay. 2021.12.22.473893. https://doi.org/10.1101/2021.12.22.473893.

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*. https://doi.org/10.1016/j.cell.2011.10.002.

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., and Weissman, J.S. (2014). Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. Cell Rep. *8*, 1365–1379. https://doi.org/10.1016/j.celrep.2014.07.045.

Inoue, K., Khajavi, M., Ohyama, T., Hirabayashi, S., Wilson, J., Reggin, J.D., Mancias, P., Butler, I.J., Wilkinson, M.F., Wegner, M., et al. (2004). Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. Nat. Genet. *36*, 361–369. https://doi.org/10.1038/ng1322.

Isakova, A., Neff, N., and Quake, S.R. (2021). Single-cell quantification of a broad RNA spectrum reveals unique noncoding patterns associated with cell types and states. Proc. Natl. Acad. Sci. *118*, e2113568118. https://doi.org/10.1073/pnas.2113568118.

Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W.B., and Maquat, L.E. (2008). Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. Cell *133*, 314–327. https://doi.org/10.1016/j.cell.2008.02.030.

Ivanov, A., Mikhailova, T., Eliseev, B., Yeramala, L., Sokolova, E., Susorov, D., Shuvalov, A., Schaffitzel, C., and Alkalaeva, E. (2016). PABP enhances release factor recruitment and stop codon recognition during translation termination. Nucleic Acids Res. *44*, 7766–7776. https://doi.org/10.1093/nar/gkw635.

Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. *47*, 199–208. https://doi.org/10.1038/ng.3192.

Jackson, C.A., Castro, D.M., Saldi, G.-A., Bonneau, R., and Gresham, D. (2020). Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. ELife *9*, e51254. https://doi.org/10.7554/eLife.51254.

Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat. Rev. Genet. *10*, 833–844. https://doi.org/10.1038/nrg2683.

Jaffrey, S.R., and Wilkinson, M.F. (2018). Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease. Nat. Rev. Neurosci. *19*, 715–728. https://doi.org/10.1038/s41583-018-0079-z.

Jan, C.H., Friedman, R.C., Ruby, J.G., and Bartel, D.P. (2011). Formation, Regulation and Evolution of Caenorhabditis elegans 3'UTRs. Nature *469*, 97–101. https://doi.org/10.1038/nature09616.

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., et al. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast Chichester Engl. *21*, 947–962. https://doi.org/10.1002/yea.1142.

Jariani, A., Vermeersch, L., Cerulus, B., Perez-Samper, G., Voordeckers, K., Van Brussel, T., Thienpont, B., Lambrechts, D., and Verstrepen, K.J. (2020). A new protocol for single-cell RNA-seq reveals stochastic gene expression during lag phase in budding yeast. ELife *9*, e55320. https://doi.org/10.7554/eLife.55320.

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

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. Mol. Cell *52*, 473–484. https://doi.org/10.1016/j.molcel.2013.10.032.

Jia, H., Wang, X., Anderson, J.T., and Jankowsky, E. (2012). RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. Proc. Natl. Acad. Sci. U. S. A. *109*, 7292–7297. https://doi.org/10.1073/pnas.1201085109.

Jin, H., Kelley, A.C., Loakes, D., and Ramakrishnan, V. (2010). Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release. Proc. Natl. Acad. Sci. *107*, 8593–8598. https://doi.org/10.1073/pnas.1003995107.

Johansson, M.J.O., He, F., Spatrick, P., Li, C., and Jacobson, A. (2007). Association of yeast Upf1p with direct substrates of the NMD pathway. Proc. Natl. Acad. Sci. U. S. A. *104*, 20872–20877. https://doi.org/10.1073/pnas.0709257105.

Johnson, A.W. (1997). Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. Mol. Cell. Biol. *17*, 6122–6130. .

Kadaba, S., Krueger, A., Trice, T., Krecic, A.M., Hinnebusch, A.G., and Anderson, J. (2004). Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae. Genes Dev. *18*, 1227–1240. https://doi.org/10.1101/gad.1183804.

Karam, R., Lou, C., Kroeger, H., Huang, L., Lin, J.H., and Wilkinson, M.F. (2015). The unfolded protein response is shaped by the NMD pathway. EMBO Rep. *16*. https://doi.org/10.15252/embr.201439696.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev. *20*. https://doi.org/10.1101/gad.1389006.

Kebaara, B.W., and Atkin, A.L. (2009). Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. Nucleic Acids Res. *37*, 2771–2778. https://doi.org/10.1093/nar/gkp146.

Kerényi, Z., Mérai, Z., Hiripi, L., Benkovics, A., Gyula, P., Lacomme, C., Barta, E., Nagy, F., and Silhavy, D. (2008). Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. EMBO J. *27*, 1585–1595. https://doi.org/10.1038/emboj.2008.88.

Kieft, J.S. (2008). Viral IRES RNA structures and ribosome interactions. Trends Biochem. Sci. *33*, 274–283. https://doi.org/10.1016/j.tibs.2008.04.007.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517–522. https://doi.org/10.1038/nature03041.

Kim, T., Xu, Z., Clauder-Münster, S., Steinmetz, L.M., and Buratowski, S. (2012). Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. Cell *150*, 1158–1169. https://doi.org/10.1016/j.cell.2012.08.016.

Kim, Y.J., Xie, P., Cao, L., Zhang, M.Q., and Kim, T.H. (2018). Global transcriptional activity dynamics reveal functional enhancer RNAs. Genome Res. *28*, 1799–1811. https://doi.org/10.1101/gr.233486.117.

Klauer, A.A., and van Hoof, A. (2012). Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. Wiley Interdiscip. Rev. RNA *3*, 649–660. https://doi.org/10.1002/wrna.1124.

Kloc, M., Wilk, K., Vargas, D., Shirato, Y., Bilinski, S., and Etkin, L.D. (2005). Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of Xenopus oocytes. Dev. Camb. Engl. *132*, 3445–3457. https://doi.org/10.1242/dev.01919.

Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., et al. (2006). CAGE: cap analysis of gene expression. Nat. Methods *3*, 211–222. https://doi.org/10.1038/nmeth0306-211.

Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., Kobayashi, S., Payre, F., and Kageyama, Y. (2010). Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. Science *329*, 336–339. https://doi.org/10.1126/science.1188158.

Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell *15*, 1109–1123. https://doi.org/10.1016/0092-8674(78)90039-9.

Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. Gene *299*, 1–34. https://doi.org/10.1016/s0378-1119(02)01056-9.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature *493*, 231–235. https://doi.org/10.1038/nature11661.

Kuehner, J.N., and Brow, D.A. (2008). Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. Mol. Cell *31*, 201–211. https://doi.org/10.1016/j.molcel.2008.05.018.

Kunz, J.B., Neu-Yilik, G., Hentze, M.W., Kulozik, A.E., and Gehring, N.H. (2006). Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. RNA *12*, 1015–1022. https://doi.org/10.1261/rna.12506.

Kurihara, Y., Matsui, A., Hanada, K., Kawashima, M., Ishida, J., Morosawa, T., Tanaka, M., Kaminuma, E., Mochizuki, Y., Matsushima, A., et al. (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. *106*. https://doi.org/10.1073/pnas.0808902106.

Kurosaki, T., and Maquat, L.E. (2013). Rules that govern UPF1 binding to mRNA 3′ UTRs. Proc. Natl. Acad. Sci. *110*, 3357–3362. https://doi.org/10.1073/pnas.1219908110.

Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell *121*, 713–724. https://doi.org/10.1016/j.cell.2005.04.029.

Lagarde, J., Uszczynska-Ratajczak, B., Carbonell, S., Pérez-Lluch, S., Abad, A., Davis, C., Gingeras, T.R., Frankish, A., Harrow, J., Guigo, R., et al. (2017). High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. Nat. Genet. *49*, 1731–1740. https://doi.org/10.1038/ng.3988.

Lan, Y., Xiao, X., He, Z., Luo, Y., Wu, C., Li, L., and Song, X. (2018). Long noncoding RNA OCC-1 suppresses cell growth through destabilizing HuR protein in colorectal cancer. Nucleic Acids Res. *46*, 5809–5821. https://doi.org/10.1093/nar/gky214.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

Lardenois, A., Liu, Y., Walther, T., Chalmel, F., Evrard, B., Granovskaia, M., Chu, A., Davis, R.W., Steinmetz, L.M., and Primig, M. (2011). Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proc. Natl. Acad. Sci. U. S. A. *108*, 1058–1063. https://doi.org/10.1073/pnas.1016459108.

Latos, P.A., Pauler, F.M., Koerner, M.V., Şenergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science *338*, 1469–1472. https://doi.org/10.1126/science.1228110.

Lauressergues, D., Couzigou, J.-M., Clemente, H.S., Martinez, Y., Dunand, C., Bécard, G., and Combier, J.-P. (2015). Primary transcripts of microRNAs encode regulatory peptides. Nature *520*, 90–93. https://doi.org/10.1038/nature14346.

Lauressergues, D., Ormancey, M., Guillotin, B., Clemente, H.S., Camborde, L., Duboé, C., Tourneur, S., Charpentier, P., Barozet, A., Jauneau, A., et al. (2022). Characterization of plant microRNA-encoded peptides (miPEPs) reveals molecular mechanisms from the translation to activity and specificity. Cell Rep. *38*. https://doi.org/10.1016/j.celrep.2022.110339.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J. *19*, 6860–6869. https://doi.org/10.1093/emboj/19.24.6860.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J. *20*. https://doi.org/10.1093/emboj/20.17.4987.

Lebreton, A., Tomecki, R., Dziembowski, A., and Séraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature *456*, 993–996. https://doi.org/10.1038/nature07480.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. Science *343*, 1360–1363. https://doi.org/10.1126/science.1250212.

Leeds, P., Peltz, S.W., Jacobson, A., and Culbertson, M.R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. Genes Dev. *5*, 2303–2314. https://doi.org/10.1101/gad.5.12a.2303.

Leeds, P., Wood, J.M., Lee, B.S., and Culbertson, M.R. (1992). Gene products that promote mRNA turnover in Saccharomyces cerevisiae. Mol. Cell. Biol. *12*, 2165–2177. https://doi.org/10.1128/mcb.12.5.2165-2177.1992.

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., et al. (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. Mol. Cell *66*, 22-37.e9. https://doi.org/10.1016/j.molcel.2017.02.017.

Lelivelt, M.J., and Culbertson, M.R. (1999). Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. Mol. Cell. Biol. *19*, 6710–6719. https://doi.org/10.1128/MCB.19.10.6710.

Lenstra, T.L., Coulon, A., Chow, C.C., and Larson, D.R. (2015). Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. Mol. Cell *60*, 597–610. https://doi.org/10.1016/j.molcel.2015.09.028.

Li, C.-Y., Zhang, Y., Wang, Z., Zhang, Y., Cao, C., Zhang, P.-W., Lu, S.-J., Li, X.-M., Yu, Q., Zheng, X., et al. (2010a). A human-specific de novo protein-coding gene associated with human brain functions. PLoS Comput. Biol. *6*, e1000734. https://doi.org/10.1371/journal.pcbi.1000734.

Li, D., Dong, Y., Jiang, Y., Jiang, H., Cai, J., and Wang, W. (2010b). A de novo originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. Cell Res. *20*, 408–420. https://doi.org/10.1038/cr.2010.31.

Li, D., Yan, Z., Lu, L., Jiang, H., and Wang, W. (2014). Pleiotropy of the de novo-originated gene MDF1. Sci. Rep. *4*, 7280. https://doi.org/10.1038/srep07280.

Li, Q., Shao, Y., Zhang, X., Zheng, T., Miao, M., Qin, L., Wang, B., Ye, G., Xiao, B., and Guo, J. (2015). Plasma long noncoding RNA protected by exosomes as a potential stable biomarker for gastric cancer. Tumor Biol. *36*, 2007–2012. https://doi.org/10.1007/s13277-014-2807-y.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516–520. https://doi.org/10.1038/nature12210.

Li, Y., Syed, J., and Sugiyama, H. (2016). RNA-DNA Triplex Formation by Long Noncoding RNAs. Cell Chem. Biol. *23*, 1325–1333. https://doi.org/10.1016/j.chembiol.2016.09.011.

Lin, Y.-F., Xiao, M.-H., Chen, H.-X., Meng, Y., Zhao, N., Yang, L., Tang, H., Wang, J.-L., Liu, X., Zhu, Y., et al. (2019). A novel mitochondrial micropeptide MPM enhances mitochondrial respiratory activity and promotes myogenic differentiation. Cell Death Dis. *10*, 528. https://doi.org/10.1038/s41419-019-1767-y.

Liu, C., Karam, R., Zhou, Y., Su, F., Ji, Y., Li, G., Xu, G., Lu, L., Wang, C., Song, M., et al. (2014). The UPF1 RNA surveillance gene is commonly mutated in pancreatic adenosquamous carcinoma. Nat. Med. *20*, 596–598. https://doi.org/10.1038/nm.3548.

Loda, A., and Heard, E. (2019). Xist RNA in action: Past, present, and future. PLoS Genet. *15*, e1008333. https://doi.org/10.1371/journal.pgen.1008333.

Longman, D., Plasterk, R.H.A., Johnstone, I.L., and Cáceres, J.F. (2007). Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. Genes Dev. *21*, 1075–1085. https://doi.org/10.1101/gad.417707.

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. https://doi.org/10.1038/s41587-021-00936-1.

Losson, R., and Lacroute, F. (1979). Interference of nonsense mutations with eukaryotic messenger RNA stability. Proc. Natl. Acad. Sci. U. S. A. *76*, 5134–5137. .

Luke, B., Azzalin, C.M., Hug, N., Deplazes, A., Peter, M., and Lingner, J. (2007). Saccharomyces cerevisiae Ebs1p is a putative ortholog of human Smg7 and promotes nonsense-mediated mRNA decay. Nucleic Acids Res. *35*, 7688–7697. https://doi.org/10.1093/nar/gkm912.

Lykke-Andersen, S., and Jensen, T.H. (2015). Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes.

Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2000). Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell *103*, 1121–1131. https://doi.org/10.1016/s0092-8674(00)00214-2.

Mabin, J.W., Woodward, L.A., Patton, R.D., Yi, Z., Jia, M., Wysocki, V.H., Bundschuh, R., and Singh, G. (2018). The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity. Cell Rep. *25*, 2431-2446.e7. https://doi.org/10.1016/j.celrep.2018.11.046.

MacRae, I.J., and Doudna, J.A. (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Curr. Opin. Struct. Biol. *17*, 138–145. https://doi.org/10.1016/j.sbi.2006.12.002.

Magny, E.G., Pueyo, J.I., Pearl, F.M.G., Cespedes, M.A., Niven, J.E., Bishop, S.A., and Couso, J.P. (2013). Conserved Regulation of Cardiac Calcium Uptake by Peptides Encoded in Small Open Reading Frames. Science *341*, 1116–1120. https://doi.org/10.1126/science.1238802.

Makarewich, C.A., and Olson, E.N. (2017). Mining for Micropeptides. Trends Cell Biol. *27*, 685–696. https://doi.org/10.1016/j.tcb.2017.04.006.

Makarewich, C.A., Munir, A.Z., Schiattarella, G.G., Bezprozvannaya, S., Raguimova, O.N., Cho, E.E., Vidal, A.H., Robia, S.L., Bassel-Duby, R., and Olson, E.N. (2018). The DWORF micropeptide enhances contractility and prevents heart failure in a mouse model of dilated cardiomyopathy. ELife *7*. https://doi.org/10.7554/eLife.38319.

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. https://doi.org/10.7554/eLife.06722.

Maquat, L.E., Kinniburgh, A.J., Rachmilewitz, E.A., and Ross, J. (1981). Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia. Cell *27*, 543–553. https://doi.org/10.1016/0092-8674(81)90396-2.

Marchese, F.P., Raimondi, I., and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. Genome Biol. *18*, 206. https://doi.org/10.1186/s13059-017-1348-2.

Marquardt, S., Hazelbaker, D.Z., and Buratowski, S. (2011). Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. Transcription *2*, 145–154. https://doi.org/10.4161/trns.2.3.16298.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature *429*, 571–574. https://doi.org/10.1038/nature02538.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature *445*, 666–670. https://doi.org/10.1038/nature05519.

Matsuda, D., Hosoda, N., Kim, Y.K., and Maquat, L.E. (2007). Failsafe nonsense-mediated mRNA decay does not detectably target eIF4E-bound mRNA. Nat. Struct. Mol. Biol. *14*. https://doi.org/10.1038/nsmb1297.

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. https://doi.org/10.1038/nature21034.

Mattick, J.S. (2004). RNA regulation: a new genetics? Nat. Rev. Genet. *5*, 316–323. https://doi.org/10.1038/nrg1321.

Mattioli, K., Volders, P.-J., Gerhardinger, C., Lee, J.C., Maass, P.G., Melé, M., and Rinn, J.L. (2019). High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. Genome Res. *29*, 344–355. https://doi.org/10.1101/gr.242222.118.

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. https://doi.org/10.1038/nrg.2016.78.

Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M., and Dietz, H.C. (2001). Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. Hum. Mol. Genet. *10*, 99–105. https://doi.org/10.1093/hmg/10.2.99.

Melé, M., and Rinn, J.L. (2016). "Cat's Cradling" the 3D Genome by the Act of LncRNA Transcription. Mol. Cell *62*, 657–664. https://doi.org/10.1016/j.molcel.2016.05.011.

Melé, M., Mattioli, K., Mallard, W., Shechner, D.M., Gerhardinger, C., and Rinn, J.L. (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. Genome Res. *27*, 27–37. https://doi.org/10.1101/gr.214205.116.

Melero, R., Buchwald, G., Castaño, R., Raabe, M., Gil, D., Lázaro, M., Urlaub, H., Conti, E., and Llorca, O. (2012). The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3' end. Nat. Struct. Mol. Biol. *19*, 498–505, S1-2. https://doi.org/10.1038/nsmb.2287.

Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A.F., Elkon, R., Melo, S.A., Léveillé, N., Kalluri, R., et al. (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol. Cell *49*, 524–535. https://doi.org/10.1016/j.molcel.2012.11.021.

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. https://doi.org/10.1038/ng1429.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proc. Natl. Acad. Sci. U. S. A. *105*, 716–721. https://doi.org/10.1073/pnas.0706729105.

Meseure, D., Drak Alsibai, K., Nicolas, A., Bieche, I., and Morillon, A. (2015). Long Noncoding RNAs as New Architects in Cancer Epigenetics, Prognostic Biomarkers, and Potential Therapeutic Targets. BioMed Res. Int. *2015*, 320214. https://doi.org/10.1155/2015/320214.

Metze, S., Herzog, V.A., Ruepp, M.D., and Mühlemann, O. (2013). Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. RNA *19*. https://doi.org/10.1261/rna.038893.113.

Metzstein, M.M., and Krasnow, M.A. (2006). Functions of the nonsense-mediated mRNA decay pathway in Drosophila development. PLoS Genet. *2*, e180. https://doi.org/10.1371/journal.pgen.0020180.

Meyer, S., Temme, C., and Wahle, E. (2004). Messenger RNA turnover in eukaryotes: pathways and enzymes. Crit. Rev. Biochem. Mol. Biol. *39*, 197–216. https://doi.org/10.1080/10409230490513991.

Mills, J.D., Kawahara, Y., and Janitz, M. (2013). Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling. Curr. Genomics *14*, 173–181. https://doi.org/10.2174/1389202911314030003.

Mitkevich, V.A., Kononenko, A.V., Petrushanko, I.Y., Yanvarev, D.V., Makarov, A.A., and Kisselev, L.L. (2006). Termination of translation in eukaryotes is mediated by the quaternary eRF1*eRF3*GTP*Mg2+ complex. The biological roles of eRF3 and prokaryotic RF3 are profoundly distinct. Nucleic Acids Res. *34*, 3947–3954. https://doi.org/10.1093/nar/gkl549.

Miura, F., Kawaguchi, N., Yoshida, M., Uematsu, C., Kito, K., Sakaki, Y., and Ito, T. (2008). Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs. BMC Genomics *9*, 574. https://doi.org/10.1186/1471-2164-9-574.

Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. Nature *457*, 413–420. https://doi.org/10.1038/nature07756.

Montigny, A., Tavormina, P., Duboe, C., San Clémente, H., Aguilar, M., Valenti, P., Lauressergues, D., Combier, J.-P., and Plaza, S. (2021). Drosophila primary microRNA-8 encodes a microRNA-

encoded peptide acting in parallel of miR-8. Genome Biol. *22*, 118. https://doi.org/10.1186/s13059-021-02345-8.

Morris, K.V., and Mattick, J.S. (2014). The rise of regulatory RNA. Nat. Rev. Genet. *15*, 423–437. https://doi.org/10.1038/nrg3722.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628. https://doi.org/10.1038/nmeth.1226.

Mourtada-Maarabouni, M., and Williams, G.T. (2013). Growth arrest on inhibition of nonsense-mediated decay is mediated by noncoding RNA GAS5. BioMed Res. Int. *2013*, 358015. https://doi.org/10.1155/2013/358015.

Mourtada-Maarabouni, M., Hedge, V.L., Kirkham, L., Farzaneh, F., and Williams, G.T. (2008). Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). J. Cell Sci. *121*, 939–946. https://doi.org/10.1242/jcs.024646.

Muhlrad, D., and Parker, R. (1992). Mutations affecting stability and deadenylation of the yeast MFA2 transcript. Genes Dev. *6*, 2100–2111. https://doi.org/10.1101/gad.6.11.2100.

Muhlrad, D., and Parker, R. (1994). Premature translational termination triggers mRNA decapping. Nature *370*. https://doi.org/10.1038/370578a0.

Muhlrad, D., and Parker, R. (1999). Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. RNA *5*. https://doi.org/10.1017/S1355838299990829.

Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M., and Ohler, U. (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. Nat. Struct. Mol. Biol. *24*, 86–96. https://doi.org/10.1038/nsmb.3325.

Müller, U., Steinhoff, U., Reis, L.F., Hemmi, S., Pavlovic, J., Zinkernagel, R.M., and Aguet, M. (1994). Functional role of type I and type II interferons in antiviral defense. Science *264*, 1918–1921. https://doi.org/10.1126/science.8009221.

Na, Z., Dai, X., Zheng, S.-J., Bryant, C.J., Loh, K.H., Su, H., Luo, Y., Buhagiar, A.F., Cao, X., Baserga, S.J., et al. (2022). Mapping subcellular localizations of unannotated microproteins and alternative proteins with MicroID. Mol. Cell *82*, 2900-2911.e7. https://doi.org/10.1016/j.molcel.2022.06.035.

Nadal-Ribelles, M., Islam, S., Wei, W., Latorre, P., Nguyen, M., de Nadal, E., Posas, F., and Steinmetz, L.M. (2019). Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast populations. Nat. Microbiol. *4*, 683–692. https://doi.org/10.1038/s41564-018-0346-9.

Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H., and Miura, K. (2008). Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. Nucleic Acids Res. *36*, 861–871. https://doi.org/10.1093/nar/gkm1102.

Namy, O., Duchateau-Nguyen, G., Hatin, I., Hermann-Le Denmat, S., Termier, M., and Rousset, J. (2003). Identification of stop codon readthrough genes in Saccharomyces cerevisiae. Nucleic Acids Res. *31*, 2289–2296. https://doi.org/10.1093/nar/gkg330.

Nasif, S., Contu, L., and Mühlemann, O. (2018). Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. Semin. Cell Dev. Biol. *75*, 78–87. https://doi.org/10.1016/j.semcdb.2017.08.053.

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. https://doi.org/10.1038/nature07747.

Nekrutenko, A., and Li, W.H. (2001). Transposable elements are found in a large number of human protein-coding genes. Trends Genet. TIG *17*, 619–621. https://doi.org/10.1016/s0168-9525(01)02445-3.

Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R., Chen, X., Kavalali, E.T., et al. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science *351*, 271–275. https://doi.org/10.1126/science.aad4076.

Nguyen, T., Fischl, H., Howe, F.S., Woloszczuk, R., Serra Barros, A., Xu, Z., Brown, D., Murray, S.C., Haenni, S., Halstead, J.M., et al. (2014). Transcription mediated insulation and interference direct gene cluster expression switches. ELife *3*, e03635. https://doi.org/10.7554/eLife.03635.

Nicholson, P., and Mühlemann, O. (2010). Cutting the nonsense: the degradation of PTC-containing mRNAs. Biochem. Soc. Trans. *38*, 1615–1620. https://doi.org/10.1042/BST0381615.

Nojima, T., and Proudfoot, N.J. (2022). Mechanisms of lncRNA biogenesis as revealed by nascent transcriptomics. Nat. Rev. Mol. Cell Biol. 1–18. https://doi.org/10.1038/s41580-021-00447-6.

Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K.R., Grimson, A., Hachiya, T., Hentze, M.W., Anderson, P., and Ohno, S. (2003). Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. Mol. Cell *12*. https://doi.org/10.1016/S1097-2765(03)00443-X.

Oliveira, C.C., and McCarthy, J.E. (1995). The relationship between eukaryotic translation and mRNA stability. A short upstream open reading frame strongly inhibits translational initiation and greatly accelerates mRNA degradation in the yeast Saccharomyces cerevisiae. J. Biol. Chem. *270*, 8936–8943. https://doi.org/10.1074/jbc.270.15.8936.

Palazzo, A.F., and Koonin, E.V. (2020). Functional Long Non-coding RNAs Evolve from Junk Transcripts. Cell *183*, 1151–1161. https://doi.org/10.1016/j.cell.2020.09.047.

Pang, K.C., Frith, M.C., and Mattick, J.S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet. *22*, 1–5. https://doi.org/10.1016/j.tig.2005.10.003.

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. https://doi.org/10.1101/gr.275638.121.

Parikh, S.B., Houghton, C., Branden Van Oss, S., Wacholder, A., and Carvunis, A.-R. (2022). Origins, evolution, and physiological implications of de novo genes in yeast. Yeast Chichester Engl. https://doi.org/10.1002/yea.3810.

Parker, R. (2012). RNA degradation in Saccharomyces cerevisae. Genetics *191*. https://doi.org/10.1534/genetics.111.137265.

Parker, R., and Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. Nat. Struct. Mol. Biol. *11*, 121–127. https://doi.org/10.1038/nsmb724.

Pastor, F., Kolonias, D., Giangrande, P.H., and Gilboa, E. (2010). Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. Nature *465*, 227–230. https://doi.org/10.1038/nature08999.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., et al. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. *22*, 577–591. https://doi.org/10.1101/gr.133009.111.

Pauli, A., Norris, M.L., Valen, E., Chew, G.-L., Gagnon, J.A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D., et al. (2014). Toddler: an embryonic signal that promotes cell movement via Apelin receptors. Science *343*, 1248636. https://doi.org/10.1126/science.1248636.

Peccarelli, M., and Kebaara, B.W. (2014). Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. Eukaryot. Cell *13*. https://doi.org/10.1128/EC.00090-14.

Pelechano, V., and Steinmetz, L.M. (2013). Gene regulation by antisense transcription. Nat. Rev. Genet. *14*, 880–893. https://doi.org/10.1038/nrg3594.

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

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

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137. https://doi.org/10.1038/379131a0.

Picelli, S. (2017). Single-cell RNA-sequencing: The future of genome biology is now. RNA Biol. *14*, 637–650. https://doi.org/10.1080/15476286.2016.1201618.

Pinskaya, M., Gourvennec, S., and Morillon, A. (2009). H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. EMBO J. *28*, 1697–1707. https://doi.org/10.1038/emboj.2009.108.

Polycarpou-Schwarz, M., Groß, M., Mestdagh, P., Schott, J., Grund, S.E., Hildenbrand, C., Rom, J., Aulmann, S., Sinn, H.-P., Vandesompele, J., et al. (2018). The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation. Oncogene *37*, 4750–4768. https://doi.org/10.1038/s41388-018-0281-5.

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

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851–1854. https://doi.org/10.1126/science.1164096.

Preker, P., Nielsen, J., Schierup, M.H., and Jensen, T.H. (2009). RNA polymerase plays both sides: vivid and bidirectional transcription around and upstream of active promoters. Cell Cycle Georget. Tex *8*, 1106–1107. .

Preker, P., Almvig, K., Christensen, M.S., Valen, E., Mapendano, C.K., Sandelin, A., and Jensen, T.H. (2011). PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. Nucleic Acids Res. *39*, 7179–7193. https://doi.org/10.1093/nar/gkr370.

Prensner, J.R., Enache, O.M., Luria, V., Krug, K., Clauser, K.R., Dempster, J.M., Karger, A., Wang, L., Stumbraite, K., Wang, V.M., et al. (2021). Noncanonical open reading frames encode functional proteins essential for cancer cell survival. Nat. Biotechnol. *39*, 697–704. https://doi.org/10.1038/s41587-020-00806-2.

Pulak, R., and Anderson, P. (1993). mRNA Surveillance by the Caenorhabditis elegans smg genes. Genes Dev. *7*. https://doi.org/10.1101/gad.7.10.1885.

Quinn, J.J., Zhang, Q.C., Georgiev, P., Ilik, I.A., Akhtar, A., and Chang, H.Y. (2016). Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. Genes Dev. *30*, 191–207. https://doi.org/10.1101/gad.272187.115.

Ransohoff, J.D., Wei, Y., and Khavari, P.A. (2018). The functions and unique features of long intergenic non-coding RNA. Nat. Rev. Mol. Cell Biol. *19*, 143–157. https://doi.org/10.1038/nrm.2017.104.

Rapicavoli, N.A., Poth, E.M., Zhu, H., and Blackshaw, S. (2011). The long noncoding RNA Six3OS acts in trans to regulate retinal development by modulating Six3 activity. Neural Develop. *6*, 32. https://doi.org/10.1186/1749-8104-6-32.

Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M., et al. (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res. *16*, 11–19. https://doi.org/10.1101/gr.4200206.

Rebbapragada, I., and Lykke-Andersen, J. (2009). Execution of nonsense-mediated mRNA decay: what defines a substrate? Curr. Opin. Cell Biol. *21*, 394–402. https://doi.org/10.1016/j.ceb.2009.02.007.

Reeves, M.B., Davies, A.A., McSharry, B.P., Wilkinson, G.W., and Sinclair, J.H. (2007). Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. Science *316*, 1345–1348. https://doi.org/10.1126/science.1142984.

Reinhardt, J.A., Wanjiru, B.M., Brant, A.T., Saelao, P., Begun, D.J., and Jones, C.D. (2013). De Novo ORFs in Drosophila Are Important to Organismal Fitness and Evolved Rapidly from Previously Non-coding Sequences. PLOS Genet. *9*, e1003860. https://doi.org/10.1371/journal.pgen.1003860.

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

Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., Hartman, S., Harrison, P.M., Nelson, F.K., Miller, P., Gerstein, M., et al. (2003). The transcriptional activity of human Chromosome 22. Genes Dev. *17*, 529–540. https://doi.org/10.1101/gad.1055203.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323. https://doi.org/10.1016/j.cell.2007.05.022.

Rodríguez-Gabriel, M.A., Watt, S., Bähler, J., and Russell, P. (2006). Upf1, an RNA Helicase Required for Nonsense-Mediated mRNA Decay, Modulates the Transcriptional Response to Oxidative Stress in Fission Yeast. Mol. Cell. Biol. *26*. https://doi.org/10.1128/mcb.00286-06.

Rongo, C., Gavis, E.R., and Lehmann, R. (1995). Localization of oskar RNA regulates oskar translation and requires Oskar protein. Dev. Camb. Engl. *121*, 2737–2746. https://doi.org/10.1242/dev.121.9.2737.

Ruiz Cuevas, M.V., Hardy, M.-P., Hollý, J., Bonneil, É., Durette, C., Courcelles, M., Lanoix, J., Côté, C., Staudt, L.M., Lemieux, S., et al. (2021). Most non-canonical proteins uniquely populate the proteome or immunopeptidome. Cell Rep. *34*, 108815. https://doi.org/10.1016/j.celrep.2021.108815.

Ruiz-Echevarría, M.J., and Peltz, S.W. (2000). The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. Cell *101*, 741–751. https://doi.org/10.1016/s0092-8674(00)80886-7.

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*. https://doi.org/10.7554/eLife.03523.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell *146*, 353–358. https://doi.org/10.1016/j.cell.2011.07.014.

Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M., and Wolfe, K.H. (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc. Natl. Acad. Sci. U. S. A. *104*, 8397–8402. https://doi.org/10.1073/pnas.0608218104.

Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G., Sanchez-Rotunno, M., Arraiano, C.M., and van Hoof, A. (2009). The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. Nat. Struct. Mol. Biol. *16*, 56–62. https://doi.org/10.1038/nsmb.1528.

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. Mol. Cell *65*, 25–38. https://doi.org/10.1016/j.molcel.2016.11.029.

Schmitt, A.M., and Chang, H.Y. (2016). Long Noncoding RNAs in Cancer Pathways. Cancer Cell *29*, 452–463. https://doi.org/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. https://doi.org/10.1038/s41559-018-0639-7.

Schoeftner, S., and Blasco, M.A. (2010). Chromatin regulation and non-coding RNAs at mammalian telomeres. Semin. Cell Dev. Biol. *21*, 186–193. https://doi.org/10.1016/j.semcdb.2009.09.015.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell *155*, 1075–1087. https://doi.org/10.1016/j.cell.2013.10.024.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851. https://doi.org/10.1126/science.1162253.

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. https://doi.org/10.1038/ncomms14021.

Sharif, H., and Conti, E. (2013). Architecture of the Lsm1-7-Pat1 Complex: A Conserved Assembly in Eukaryotic mRNA Turnover. Cell Rep. *5*, 283–291. https://doi.org/10.1016/j.celrep.2013.10.004.

Sharma, U., Barwal, T.S., Malhotra, A., Pant, N., Vivek, null, Dey, D., Gautam, A., Tuli, H.S., Vasquez, K.M., and Jain, A. (2020). Long non-coding RNA TINCR as potential biomarker and therapeutic target for cancer. Life Sci. *257*, 118035. https://doi.org/10.1016/j.lfs.2020.118035.

Sharon, D., Tilgner, H., Grubert, F., and Snyder, M. (2013). A single-molecule long-read survey of the human transcriptome. Nat. Biotechnol. *31*, 1009–1014. https://doi.org/10.1038/nbt.2705.

She, M., Decker, C.J., Svergun, D.I., Round, A., Chen, N., Muhlrad, D., Parker, R., and Song, H. (2008). Structural basis of Dcp2 recognition and activation by Dcp1. Mol. Cell *29*, 337–349. https://doi.org/10.1016/j.molcel.2008.01.002.

Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C., et al. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. *110*, 2876–2881. https://doi.org/10.1073/pnas.1221904110.

Silva, A.L., Ribeiro, P., Inácio, Â., Liebhaber, S.A., and Romão, L. (2008). Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. RNA *14*, 563–576. https://doi.org/10.1261/rna.815108.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol. *6*. https://doi.org/10.1371/journal.pbio.0060111.

Sinturel, F., Navickas, A., Wery, M., Descrimes, M., Morillon, A., Torchet, C., and Benard, L. (2015). Cytoplasmic Control of Sense-Antisense mRNA Pairs. Cell Rep. *12*, 1853–1864. https://doi.org/10.1016/j.celrep.2015.08.016.

Sipiczki, M. (2000). Where does fission yeast sit on the tree of life? Genome Biol. *1*, reviews1011.1-reviews1011.4. .

Skinnider, M.A., Squair, J.W., and Foster, L.J. (2019). Evaluating measures of association for single-cell transcriptomics. Nat. Methods *16*, 381–386. https://doi.org/10.1038/s41592-019-0372-4.

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. https://doi.org/10.1038/nchembio.1120.

Smith, J.E., and Baker, K.E. (2015). Nonsense-mediated RNA decay - a switch and dial for regulating gene expression.

Smith, C., Canestrari, J.G., Wang, A.J., Champion, M.M., Derbyshire, K.M., Gray, T.A., and Wade, J.T. (2022). Pervasive translation in Mycobacterium tuberculosis. ELife *11*, e73980. https://doi.org/10.7554/eLife.73980.

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*. https://doi.org/10.1016/j.celrep.2014.05.023.

Sohrabi-Jahromi, S., Hofmann, K.B., Boltendahl, A., Roth, C., Gressel, S., Baejen, C., Soeding, J., and Cramer, P. (2019). Transcriptome maps of general eukaryotic rna degradation factors. ELife *8*. https://doi.org/10.7554/eLife.47040.

Solé, C., Nadal-Ribelles, M., de Nadal, E., and Posas, F. (2015). A novel role for lncRNAs in cell cycle control during stress adaptation. Curr. Genet. *61*, 299–308. https://doi.org/10.1007/s00294-014-0453-y.

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. https://doi.org/10.1038/s41580-020-00315-9.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol. Cell *24*, 735–746. https://doi.org/10.1016/j.molcel.2006.10.023.

Stringer, A., Smith, C., Mangano, K., and Wade, J.T. (2021). Identification of novel translated small ORFs in Escherichia coli using complementary ribosome profiling approaches. J. Bacteriol. JB0035221. https://doi.org/10.1128/JB.00352-21.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol. *14*, 103–105. https://doi.org/10.1038/nsmb0207-103.

Sun, X., and Maquat, L.E. (2000). mRNA surveillance in mammalian cells: the relationship between introns and translation termination. RNA N. Y. N *6*, 1–8. https://doi.org/10.1017/s1355838200991660.

Sun, L., Wang, W., Han, C., Huang, W., Sun, Y., Fang, K., Zeng, Z., Yang, Q., Pan, Q., Chen, T., et al. (2021). The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. Mol. Cell *81*, 4493-4508.e9. https://doi.org/10.1016/j.molcel.2021.08.033.

Sun, M., Jin, F., Xia, R., Kong, R., Li, J., Xu, T., Liu, Y., Zhang, E., Liu, X., and De, W. (2014). Decreased expression of long noncoding RNA GAS5 indicates a poor prognosis and promotes cell proliferation in gastric cancer. BMC Cancer *14*, 319. https://doi.org/10.1186/1471-2407-14-319.

Sun, X., Moriarty, P.M., and Maquat, L.E. (2000). Nonsense-mediated decay of glutathione peroxidase 1 mRNA in the cytoplasm depends on intron position. EMBO J. *19*, 4734–4744. https://doi.org/10.1093/emboj/19.17.4734.

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature *462*, 799–802. https://doi.org/10.1038/nature08618.

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*. https://doi.org/10.26508/lsa.201900407.

Szankasi, P., and Smith, G.R. (1996). Requirement of S. pombe exonuclease II, a homologue of S. cerevisiae Sep1, for normal mitotic growth and viability. Curr. Genet. *30*, 284–293. https://doi.org/10.1007/s002940050134.

Szczepińska, T., Kalisiak, K., Tomecki, R., Labno, A., Borowski, L.S., Kulinski, T.M., Adamska, D., Kosinska, J., and Dziembowski, A. (2015). DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. Genome Res. *25*, 1622–1633. https://doi.org/10.1101/gr.189597.115.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534–538. https://doi.org/10.1038/nature06904.

Tan, K., Stupack, D.G., and Wilkinson, M.F. (2022). Nonsense-mediated RNA decay: an emerging modulator of malignancy. Nat. Rev. Cancer https://doi.org/10.1038/s41568-022-00481-2.

Tarun, S.Z., and Sachs, A.B. (1997). Binding of eukaryotic translation initiation factor 4E (eIF4E) to eIF4G represses translation of uncapped mRNA. Mol. Cell. Biol. *17*, 6876–6886. .

Tarun, S.Z., Wells, S.E., Deardorff, J.A., and Sachs, A.B. (1997). Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. Proc. Natl. Acad. Sci. U. S. A. *94*, 9046–9051. https://doi.org/10.1073/pnas.94.17.9046.

Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. EMBO J. *17*, 3484–3494. https://doi.org/10.1093/emboj/17.12.3484.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, 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. https://doi.org/10.1016/j.molcel.2006.07.029.

Thomson, T., and Lin, H. (2009). The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. Annu. Rev. Cell Dev. Biol. *25*, 355–376. https://doi.org/10.1146/annurev.cellbio.24.110707.175327.

Thoren, L.A., Nørgaard, G.A., Weischenfeldt, J., Waage, J., Jakobsen, J.S., Damgaard, I., Bergström, F.C., Blom, A.M., Borup, R., Bisgaard, H.C., et al. (2010). UPF2 Is a Critical Regulator of Liver Development, Function and Regeneration. PLOS ONE *5*, e11650. https://doi.org/10.1371/journal.pone.0011650.

Tian, D., Sun, S., and Lee, J.T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell *143*, 390–403. https://doi.org/10.1016/j.cell.2010.09.049.

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res. *22*, 1616–1625. https://doi.org/10.1101/gr.134445.111.

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

Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. Genes Dev. *19*, 517–529. https://doi.org/10.1101/gad.1284105.

Treutlein, B., Gokce, O., Quake, S.R., and Südhof, T.C. (2014). Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. Proc. Natl. Acad. Sci. U. S. A. *111*, E1291-1299. https://doi.org/10.1073/pnas.1403244111.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. Mol. Cell *39*, 925–938. https://doi.org/10.1016/j.molcel.2010.08.011.

Tucker, M., Valencia-Sanchez, M.A., Staples, R.R., Chen, J., Denis, C.L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. Cell *104*, 377–386. https://doi.org/10.1016/s0092-8674(01)00225-2.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., et al. (2014). Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. Mol. Cell *55*, 467–481. https://doi.org/10.1016/j.molcel.2014.05.031.

Tudek, A., Candelli, T., and Libri, D. (2015). Non-coding transcription by RNA polymerase II in yeast: Hasard or nécessité? Biochimie *117*, 28–36. https://doi.org/10.1016/j.biochi.2015.04.020.

Uchihara, Y., Permata, T.B.M., Sato, H., Kawabata-Iwakawa, R., Katada, S., Gu, W., Kakoti, S., Yamauchi, M., Kato, R., Gondhowiardjo, S., et al. (2022). DNA damage promotes HLA class I presentation by stimulating a pioneer round of translation-associated antigen production. Mol. Cell *82*, 2557-2570.e7. https://doi.org/10.1016/j.molcel.2022.04.030.

Uhler, J.P., Hertel, C., and Svejstrup, J.Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. Proc. Natl. Acad. Sci. U. S. A. *104*, 8011–8016. https://doi.org/10.1073/pnas.0702431104.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46. https://doi.org/10.1016/j.cell.2013.06.020.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550. https://doi.org/10.1016/j.cell.2011.11.055.

Ulitsky, I., Shkumatava, A., Jan, C.H., Subtelny, A.O., Koppstein, D., Bell, G.W., Sive, H., and Bartel, D.P. (2012). Extensive alternative polyadenylation during zebrafish development. Genome Res. *22*, 2054–2066. https://doi.org/10.1101/gr.139733.112.

Usuki, F., Yamashita, A., and Fujimura, M. (2019). Environmental stresses suppress nonsense-mediated mRNA decay (NMD) and affect cells by stabilizing NMD-targeted gene expression. Sci. Rep. *9*. https://doi.org/10.1038/s41598-018-38015-2.

Vakirlis, N., Acar, O., Hsu, B. et al. (2020). De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences. Nat Commun 11, 781. https://doi.org/10.1038/s41467-020-14500-z

Van Dijk, E.L., Chen, C.L., Daubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Ne, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*. https://doi.org/10.1038/nature10118.

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*. https://doi.org/10.1186/gb-2014-15-1-r6.

Van Oss, S.B., and Carvunis, A.-R. (2019). De novo gene birth. PLOS Genet. *15*, e1008160. https://doi.org/10.1371/journal.pgen.1008160.

Vaňáčová, Š., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A New Yeast Poly(A) Polymerase Complex Involved in RNA Quality Control. PLOS Biol. *3*, e189. https://doi.org/10.1371/journal.pbio.0030189.

Vancura, A., Lanzós, A., Bosch-Guiteras, N., Esteban, M.T., Gutierrez, A.H., Haefliger, S., and Johnson, R. (2021). Cancer LncRNA Census 2 (CLC2): an enhanced resource reveals clinical features of cancer lncRNAs. NAR Cancer *3*, zcab013. https://doi.org/10.1093/narcan/zcab013.

VanInsberghe, M., van den Berg, J., Andersson-Rolf, A., Clevers, H., and van Oudenaarden, A. (2021). Single-cell Ribo-seq reveals cell cycle-dependent translational pausing. Nature *597*, 561–565. https://doi.org/10.1038/s41586-021-03887-4.

Vanoni, M., Rossi, R.L., Querin, L., Zinzalla, V., and Alberghina, L. (2005). Glucose modulation of cell size in yeast. Biochem. Soc. Trans. *33*, 294–296. https://doi.org/10.1042/BST0330294.

Vasiljeva, L., and Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. Mol. Cell *21*, 239–248. https://doi.org/10.1016/j.molcel.2005.11.028.

Vasiljeva, L., Kim, M., Terzi, N., Soares, L.M., and Buratowski, S. (2008). Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. Mol. Cell *29*, 313–323. https://doi.org/10.1016/j.molcel.2008.01.011.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science *270*, 484–487. https://doi.org/10.1126/science.270.5235.484.

Verheggen, K., Volders, P.-J., Mestdagh, P., Menschaert, G., Van Damme, P., Gevaert, K., Martens, L., and Vandesompele, J. (2017). Noncoding after All: Biases in Proteomics Data Do Not Explain Observed Absence of lncRNA Translation Products. J. Proteome Res. *16*, 2508–2515. https://doi.org/10.1021/acs.jproteome.7b00085.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837. https://doi.org/10.1126/science.1074973.

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

Wadler, C.S., and Vanderpool, C.K. (2007). A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc. Natl. Acad. Sci. U. S. A. *104*, 20454–20459. https://doi.org/10.1073/pnas.0708102104.

Wang, Q., and Carmichael, G.G. (2004). Effects of Length and Location on the Cellular Response to Double-Stranded RNA. Microbiol. Mol. Biol. Rev. *68*, 432–452. https://doi.org/10.1128/MMBR.68.3.432-452.2004.

Wang, B., Tseng, E., Regulski, M., Clark, T.A., Hon, T., Jiao, Y., Lu, Z., Olson, A., Stein, J.C., and Ware, D. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nat. Commun. *7*, 11708. https://doi.org/10.1038/ncomms11708.

Wang, D., Zavadil, J., Martin, L., Parisi, F., Friedman, E., Levy, D., Harding, H., Ron, D., and Gardner, L.B. (2011). Inhibition of nonsense-mediated RNA decay by the tumor microenvironment promotes tumorigenesis. Mol. Cell. Biol. *31*, 3670–3680. https://doi.org/10.1128/MCB.05704-11.

Wang, J., Gudikote, J.P., Olivas, O.R., and Wilkinson, M.F. (2002a). Boundary-independent polar nonsense-mediated decay. EMBO Rep. *3*, 274–279. https://doi.org/10.1093/embo-reports/kvf036.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature *454*, 126–130. https://doi.org/10.1038/nature06992.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. (2002b). Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. U. S. A. *99*. https://doi.org/10.1073/pnas.092538799.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539–543. https://doi.org/10.1038/nature06908.

Watts, B.R., Wittmann, S., Wery, M., Gautier, C., Kus, K., Birot, A., Heo, D.-H., Kilchert, C., Morillon, A., and Vasiljeva, L. (2018). Histone deacetylation promotes transcriptional silencing at facultative heterochromatin. Nucleic Acids Res. *46*, 5426–5440. https://doi.org/10.1093/nar/gky232.

Wei, L.-H., and Guo, J.U. (2020). Coding functions of “noncoding” RNAs. Science *367*, 1074–1075. https://doi.org/10.1126/science.aba6117.

Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Mönch, K., Thoren, L.A., Nielsen, F.C., Jacobsen, S.E.W., Nerlov, C., and Porse, B.T. (2008). NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. Genes Dev. *22*, 1381–1396. https://doi.org/10.1101/gad.468808.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Mol. Cell *2*, 135–140. https://doi.org/10.1016/s1097-2765(00)80122-7.

Wen, J., and Brogna, S. (2010). Splicing-dependent NMD does not require the EJC in Schizosaccharomyces pombe. EMBO J. *29*, 1537–1551. https://doi.org/10.1038/emboj.2010.48.

Weng, Y., Czaplinski, K., and Peltz, S.W. (1996). Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. Mol. Cell. Biol. *16*, 5477–5490. https://doi.org/10.1128/mcb.16.10.5477.

Werner, A., Clark, J.E., Samaranayake, C., Casement, J., Zinad, H.S., Sadeq, S., Al-Hashimi, S., Smith, M., Kotaja, N., and Mattick, J.S. (2021). Widespread formation of double-stranded RNAs in testis. Genome Res. *31*, 1174–1186. https://doi.org/10.1101/gr.265603.120.

van Werven, F.J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M., and Amon, A. (2012). Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. Cell *150*, 1170–1181. https://doi.org/10.1016/j.cell.2012.06.049.

Wery, M., Kwapisz, M., and Morillon, A. (2011). Noncoding RNAs in gene regulation. Wiley Interdiscip. Rev. Syst. Biol. Med. *3*, 728–738. https://doi.org/10.1002/wsbm.148.

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. Mol. Cell *61*. https://doi.org/10.1016/j.molcel.2015.12.020.

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

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

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522–525. https://doi.org/10.1038/nature03035.

Wilson, B.A., and Masel, J. (2011). Putatively noncoding transcripts show extensive association with ribosomes. Genome Biol. Evol. *3*. https://doi.org/10.1093/gbe/evr099.

Wittkopp, N., Huntzinger, E., Weiler, C., Saulière, J., Schmidt, S., Sonawane, M., and Izaurralde, E. (2009). Nonsense-Mediated mRNA Decay Effectors Are Essential for Zebrafish Embryonic Development and Survival. Mol. Cell. Biol. *29*. https://doi.org/10.1128/mcb.00177-09.

Wittmann, J., Hol, E.M., and Jäck, H.-M. (2006). hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. Mol. Cell. Biol. *26*, 1272–1287. https://doi.org/10.1128/MCB.26.4.1272-1287.2006.

Woolcock, K.J., and Bühler, M. (2013). Nuclear organisation and RNAi in fission yeast. Curr. Opin. Cell Biol. *25*, 372–377. https://doi.org/10.1016/j.ceb.2013.02.004.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, 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. https://doi.org/10.1016/j.cell.2005.04.030.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. (2009). Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. Nature *458*, 784–788. https://doi.org/10.1038/nature07731.

Xie, C., Bekpen, C., Künzel, S., Keshavarz, M., Krebs-Wheaton, R., Skrabar, N., Ullrich, K.K., and Tautz, D. (2019). A de novo evolved gene in the house mouse regulates female pregnancy cycles. ELife *8*, e44392. https://doi.org/10.7554/eLife.44392.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, 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. https://doi.org/10.1038/nature07728.

Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and Ohno, S. (2001). Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. Genes Dev. *15*, 2215–2228. https://doi.org/10.1101/gad.913001.

Yang, Y., Gao, X., Zhang, M., Yan, S., Sun, C., Xiao, F., Huang, N., Yang, X., Zhao, K., Zhou, H., et al. (2018). Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. J. Natl. Cancer Inst. *110*. https://doi.org/10.1093/jnci/djx166.

Yao, R.W., Wang, Y., and Chen, L.L. (2019). Cellular functions of long noncoding RNAs.

Yepiskoposyan, H., Aeschimann, F., Nilsson, D., Okoniewski, M., and Mühlemann, O. (2011). Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. RNA *17*, 2108–2118. https://doi.org/10.1261/rna.030247.111.

Yi, Z., Arvola, R.M., Myers, S., Dilsavor, C.N., Abu Alhasan, R., Carter, B.N., Patton, R.D., Bundschuh, R., and Singh, G. (2022). Mammalian UPF3A and UPF3B can activate nonsense-mediated mRNA decay independently of their exon junction complex binding. EMBO J. *41*, e109202. https://doi.org/10.15252/embj.2021109202.

Yoine, M., Nishii, T., and Nakamura, K. (2006). Arabidopsis UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. Plant Cell Physiol. *47*, 572–580. https://doi.org/10.1093/pcp/pcj035.

Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 Suppresses Target mRNA Translation. Mol. Cell *47*, 648–655. https://doi.org/10.1016/j.molcel.2012.06.027.

Zanet, J., Benrabah, E., Li, T., Pélissier-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. https://doi.org/10.1126/science.aac5677.

Zhang, J., Sun, X., Qian, Y., and Maquat, L.E. (1998). Intron function in the nonsense-mediated decay of β-globin mRNA: Indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. RNA *4*. https://doi.org/10.1017/S1355838298971849.

Zhang, M., Zhao, K., Xu, X., Yang, Y., Yan, S., Wei, P., Liu, H., Xu, J., Xiao, F., Zhou, H., et al. (2018). A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. Nat. Commun. *9*, 4475. https://doi.org/10.1038/s41467-018-06862-2.

Zhang, S., Welch, E.M., Hogan, K., Brown, A.H., Peltz, S.W., and Jacobson, A. (1997). Polysome-associated mRNAs are substrates for the nonsense-mediated mRNA decay pathway in Saccharomyces cerevisiae. RNA *3*, 234–244. .

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750–756. https://doi.org/10.1126/science.1163045.

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. https://doi.org/10.1126/science.1248286.

Zuckerman, B., and Ulitsky, I. (2019). Predictive models of subcellular localization of long RNAs. RNA N. Y. N *25*, 557–572. https://doi.org/10.1261/rna.068288.118.

Zünd, D., Gruber, A.R., Zavolan, M., and Mühlemann, O. (2013). Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3′ UTRs. Nat. Struct. Mol. Biol. *20*. https://doi.org/10.1038/nsmb.2635.

(2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

1. Supplementary Tables Publication n°1

Table S1. Yeast strains used in Publication n°1.

Strain ID	Genotype	Source/Reference
YAM2478	*MATα ura3-1 hoΔ*	DBP005 (Drinnenberg et al, 2009)
YAM2479	*MATα ura3-1 hoΔ xrn1Δ::kanMX6*	This work
YAM2795	*MATα ura3-1 hoΔ dcr1Δ*	DBP318 (Drinnenberg et al, 2009)
YAM2796	*MATα ura3-1 hoΔ dcr1Δ xrn1Δ::kanMX6*	This work
YAM2826	*MATα ura3-1 hoΔ dcr1-GFP(S65T)-kanMX6*	This work
YAM2842	*MATα ura3-1 hoΔ dcr1Δ::GFP(S65T)-kanMX6*	This work

Table S4. Oligonucleotides used in Publication n°1.

ID	Sequence 5'-3'	Target	Use
AMO1964	GGGGTACCAAAAATTGAAAAATTCTGGGC	*XRN1*	PCR (Cloning)
AMO1965	CGGGATCCGATTAAAATGAATGTAAATTTATGTTACA	*XRN1*	PCR (Cloning)
AMO1966	CGGAATTCGAGTATCCGTTGAATGACATTTAAA	*XRN1*	PCR (Cloning)
AMO1967	GCTCTAGATATTGATTTGAGAGAAGAAGCG	*XRN1*	PCR (Cloning)
AMO1996	CAGTACTACTCAATTGCTCTCGAGC	*XRN1*	PCR
AMO1997	GTTGAAGAAAGAGCAGGAACTCTCC	*XRN1*	PCR
AMO1998	TCTTGGTTACATCGTCGTCGTTACC	*XRN1*	Northern-blot
AMO2000	GAAACGTGCAATCCATGTCTGACCG	*scR1*	Northern-blot
AMO2001	CCAGAAGGAAAGGCCCGGTTGGA	18S rRNA	Northern-blot
AMO2002	AAATTTAATAATTGGGTCGAATCGTAAGGG	5' ITS1	PCR (probe)
AMO2003	TTTGTATTTCATAACGAAATTGGTTTTGAC	5' ITS1	PCR (probe)
AMO3227	TGTAAAATCCTACATTTAAATAGTGC	*XRN1*	PCR
AMO3228	GTCGTAAACTTACAGTTGATGAGG	*XRN1*	PCR
AMO3229	ATGAGTGTTCGAGGTTTAATTAGCG	*DCR1*	PCR
AMO3230	TACGATATCCTAAGTACAGATGCC	*DCR1*	PCR
AMO3323	CGGGGTACCGCCGGCTGTTTCAAATGCACTTGG	*DCR1*	PCR (Cloning)
AMO3324	GCTCTAGAGCCGGCATTCTATAAAGAAAATACTGATGAG	*DCR1*	PCR (Cloning)
AMO3325	AAATATCATTTGAATTCAAAGCTTTGGATCCCCAGATTGTTGC AATGCCTCAAGTATTCC	*DCR1*	PCR (Cloning)
AMO3326	CAACAATCTGGGGATCCAAAGCTTTGAATTCAAATGATATTT ATGCACCTTTTATTTATC	*DCR1*	PCR (Cloning)
AMO3327	TACGCTGCAGGTCGACGGATCC	*GFP(S65T)-kanMX6*	PCR (Cloning)
AMO3328	TGGATCTGATATCATCGATGAATTCGAGC	*GFP(S65T)-kanMX6*	PCR (Cloning)
AMO3370	CGGGGTACCGCCGGCTTTAAAACCATGGAATAGACATAG	*DCR1*	PCR (Cloning)
AMO3371	CGCGGATCCCCATTTGTATAATTGCGTGTAGGTCAC	*DCR1*	PCR (Cloning)

2. Supplementary Tables Publication n°2

Table S3. Yeast strains used in Publication n°2.

Strain ID	Background	Genotype	Source/Reference
YAM1	BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Euroscarf
YAM6	BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xrn1Δ::kanMX4*	Euroscarf
YAM202	BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 upf1Δ::kanMX4*	Euroscarf
YAM2831	YAM1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 adh2Δ::URA3*	This work
YAM2852	YAM2831	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-a*	This work
YAM2853	YAM2831	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-b*	This work
YAM2854	YAM2831	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-d*	This work
YAM2855	YAM2831	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-f*	This work
YAM2862	YAM2852	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-a upf1Δ::kanMX4*	This work
YAM2863	YAM2853	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-b upf1Δ::kanMX4*	This work
YAM2864	YAM2854	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-d upf1Δ::kanMX4*	This work
YAM2865	YAM2855	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-f upf1Δ::kanMX4*	This work
YAM2893	YAM2831	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-e*	This work
YAM2896	YAM2893	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-e upf1Δ::kanMX4*	This work
YAM2898	YAM2831	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-c*	This work
YAM2901	YAM2898	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-c upf1Δ::kanMX4*	This work
YAM2903	YAM2831	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-chimera*	This work
YAM2904	YAM2903	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-chimera upf1Δ::kanMX4*	This work
YAM2908	YAM2831	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-b-3FLAG*	This work
YAM2911	YAM2908	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xut0741-b-3FLAG upf1Δ::kanMX4*	This work
YAM2934	YAM2831	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 SL-xut0741-b-3FLAG*	This work
YAM2937	YAM2934	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 SL-xut0741-b-3FLAG upf1Δ::kanMX4*	This work

Table S4. Oligonucleotides used in Publication n°2.

ID	Sequence 5'-3'	Target	Use
AMO190	GGCGAACTCCGTAATTCGCC	*UPF1*	PCR
AMO193	GGCTGTAATGGCTTTCTGG	*scR1*	qPCR
AMO415	GTGCGGAATAGAGAACTATCC	*scR1*	RT + qPCR
AMO496	TCTTGCCCAGTAAAAGCTCTCATG	ITS1 & 20S pre-rRNA	Northern blot
AMO1482	ATCCCGGCCGCCTCCATCAC	*scR1*	Northern blot
AMO1595	GGGAAAAGTTTGTTTGGCTTATTCTGGTGGTTTAG	*XUT1678/SUT768*	Northern blot
AMO1762	GACAGTGTTCGAAGTTTCACGA	*XUT0741*	Northern blot
AMO2710	TGGGAGGGACACCTTTATACGC	*UPF1*	PCR
AMO2711	CTAGGATATCAAGTCCATGCCC	*UPF1*	PCR
AMO2712	CTTTATTACGCATTGCAGTGCG	*UPF1*	PCR
AMO2752	TTCAACGTGAAATTGGTGGA	*XUT1092*	qPCR
AMO2753	AGTGACATCTGGCGCTGATA	*XUT1092*	RT + qPCR
AMO2776	TCCAGTGATGTGGACGAGAA	*XUT1186*	qPCR
AMO2777	AAGCCGTTATGAAGACTCCAA	*XUT1186*	RT + qPCR
AMO3350	GAAGTCGTTCGTACTAGCAACATGG	*XUT0741/ADH2*	PCR
AMO3351	CAGGCGGGAAACCATCCACTTCAC	*XUT0741/ADH2*	
AMO3354	GGCTGGAAGATCGGTGACTA	*XUT0741*	RT
AMO3359	CAACTTGAGAGCAGGCCACT	*XUT0741*	Cloning
AMO3379	GGGGTACCGCCGGCCCGCTATATTTTGG	*XUT0741*	Cloning
AMO3382	GCTCTAGAGCCGGCATCTCCAACTTATAAGTTGG	*XUT0741*	Cloning
AMO3471	GGGGTACCAGATCTGAATTCAGGAATGGGTACAAC TCACAGG	*XUT0741*	Cloning
AMO3497	TACCCATTCCTGAATTCTTAGTTGGTGGTCACGAAG GTGCCGGT	*XUT0741*	Cloning
AMO3530	CCTTGTAATCGATGTCATGATCTTTATAATCACCGTC ATGGTCTTTGTAGTCGGTTCTTTGGCTGTTCAATATG	*XUT0741 + 3FLAG*	Cloning
AMO3531	TATAAAGATCATGACATCGATTACAAGGATGACGAT GACAAGCTAGGATCCTAGACCACCAGCAGCACCAG AAA	*XUT0741 + 3FLAG*	Cloning
AMO3549	ACTGATCCCGCGGTTCGCCGCGGTTTGTTGGAGGA TGCCGTA	*XUT0741*	Cloning
AMO3550	AAACCGCGGCGAACCGCGGGATCAGTCGGTTATAG TTTGTCC	*XUT0741*	Cloning
AMO3558	CGAACACTGCTGAAGCTACC	*XUT0741*	qPCR
AMO3559	CTACTTTTTGCTCCCACCGC	*XUT0741*	qPCR
5S_44	/5BiotinTEG/ACT+ACTCGGTCAGGCTCT+TACCAGC TTAACT+ACAGTT	5S rRNA	Ribo-Seq (ribo-depletion)
5.8S_125	/5BiotinTEG/AA+ATGACGCTCA+AACAGGCATGCC CCCTGGA+AT	5.8S rRNA	Ribo-Seq (ribo-depletion)
18S_1712	/5BiotinTEG/AA+ATGACCAAGT+TTGTCCAA+ATTC TCCGCTC	18S rRNA	Ribo-Seq (ribo-depletion)
18S_rDNA1	/5BiotinTEG/TGATGCCCCCGACCGTCCCTAT+TAAT CATTACGACCA+AGTTTGTCCAA+ATTCTCCGCTCTG AGA	18S rRNA	Ribo-Seq (ribo-depletion)

25S_557	/5BiotinTEG/GACTT+ACGTCGCAGTCCTC+AGTCCC AGCTGGCAGT+ATTCCCACAGGCTA	25S rRNA	Ribo-Seq (ribo-depletion)
25S_698	/5BiotinTEG/CGAGGCCCCA+ACCTACGTTCACT+TT CATT+ACGCGT	25S rRNA	Ribo-Seq (ribo-depletion)
25S_rDNA2	/5BiotinTEG/GCTAGCCTGCT+ATGGTTCAGCGACG CCACAACTGATCA+AATGCCCTTCCCTTTCAACAA+T TTCACG	25S rRNA	Ribo-Seq (ribo-depletion)
25S_rDNA3	/5BiotinTEG/TTCCAGCTCCGCTTCATTGAATAAGTA +AAGAACTAT+TTTGCCGACTTCCCTTATCTACATT+ATTCTA	25S rRNA	Ribo-Seq (ribo-depletion)
25S_rDNA5	/5BiotinTEG/ATTCTATT+ATTCCATGCTAAT+ATATT CGAGCAAGCGGTTATCAGTACGACCTGGCATGAA+ AAC	25S rRNA	Ribo-Seq (ribo-depletion)

Note: + corresponds to a Locked nucleic acid base.

3. Supplementary Sequences of *XUT0741* alleles used in Publication n°2.

> XUT0741

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTGGAGGATGCCGTACATAA
CGAACACTGCTGAAGCTACCATGTCTACAGTTTAGAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCTAG
CAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACTA
ATATAGGCATACTTGATAATGAAAACTATAAATCGTAAAGACATAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTA
CCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGTAAACTGGATAAGCCAACTACCTTTATTGGAGACTTGACTAG
ACCTCTGGCAAAGAAATCTAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAA
CGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTG
GTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCT
AACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCAT
CAATACCTAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCTAGACCACCAGCAGCACCAGAAATGGCC
GCCCAGTGGCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGC
CAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAAC
CAGACAAGTCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTG
ATACCGGCGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAAC
TAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGA
TTAACAATTCGTTGGGCTTTGGCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGGCT
TTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTGTAGCTTGATAT
TCTATTTACCAAGAAGAAACAAGAAGTGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGAAAAAAGTCGCT
ACTGGCACTCTATTTATATGTGATAGGCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTCCTCTGCCTTTT
TTTTTTTTTTTTTTTTCATTCTCTCAATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGGC

Color code

Black = *XUT0741* sequence
Grey = flancking sequences used for the integration by homologous recombination (NaeI sites underlined)
Yellow = *XUT0741* smORF Red = in-frame stop codons

> pAM594 insert (xut0741-a)

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTGGAGGATGCCGTACATAA
CGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAG
CAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACcA
ATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTA
CCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAG
ACCTCTGGCAAAGAAATCTAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAA
CGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTG
GTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCT
AACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCAT
CAATACCTAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCTAGACCACCAGCAGCACCAGAAATGGCC
GCCCAGTGGCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGC
CAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAAC
CAGACAAGTCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTG
ATACCGGCGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAAC
TAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGA
TTAACAATTCGTTGGGCTTTGGCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGGCT
TTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTGTAGCTTGATAT
TCTATTTACCAAGAAGAAACAAGAAGTGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGAAAAAAGTCGCT
ACTGGCACTCTATTTATATGTGATAGGCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTCCTCTGCCTTTT
TTTTTTTTTTTTTTTTTCATTCTCTCAATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGGC

> pAM596 insert (xut0741-b)

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTGGAGGATGCCGTACATAA
CGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAG
CAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACcA
ATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTA
CCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAG
ACCTCTGGCAAAGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAA
CGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTG
GTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCT
AACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCAT
CAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCTAGACCACCAGCAGCACCAGAAATGGCC
GCCCAGTGGCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGC
CAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAAC
CAGACAAGTCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTG
ATACCGGCGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAAC
TAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGA
TTAACAATTCGTTGGGCTTTGGCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGGCT
TTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTGTAGCTTGATAT
TCTATTTACCAAGAAGAAACAAGAAGTGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGAAAAAAGTCGCT
ACTGGCACTCTATTTATATGTGATAGGCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTCCTCTGCCTTTT
TTTTTTTTTTTTTTTTTCATTCTCTCAATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGGC

> pAM724 insert (xut0741-c)

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTGGAGGATGCCGTACATAA
CGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAG
CAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACcA
ATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTA
CCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAG
ACCTCTGGCAAAGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAA
CGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTG
GTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCT
AACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCAT
CAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCgAGACCACCAGCAGCACCAGAAATGGCC
GCCCAGTGGCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGC
tAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAAC
CAGACAAGTCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTG
ATACCGGCGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAAC
TAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGA
TTAACAATTCGTTGGGCTTTGGCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGGCT
TTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTGTAGCTTGATAT
TCTATTTACCAAGAAGAAACAAGAAGTGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGAAAAAAGTCGCT
ACTGGCACTCTATTTATATGTGATAGGCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTCCTCTGCCTTTT
TTTTTTTTTTTTTTTTCATTCTCTCAATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGGC

> pAM598 insert (xut0741-d)

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTGGAGGATGCCGTACATAA
CGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAG
CAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACcA
ATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTA
CCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAG
ACCTCTGGCAAAGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAA
CGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTG
GTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCT
AACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCAT
CAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCgAGACCACCAGCAGCACCAGAAATGGCC
GCCCAGTGGCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGC
CAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAAC
CAGACAAGTCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTG
ATACCGGCGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAAC
TAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGA
TTAACAATTCGTTGGGCTTTGGCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGGCT
TTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTGTAGCTTGATAT
TCTATTTACCAAGAAGAAACAAGAAGTGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGAAAAAAGTCGCT
ACTGGCACTCTATTTATATGTGATAGGCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTCCTCTGCCTTTT
TTTTTTTTTTTTTTTTCATTCTCTCAATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGGC

> pAM723 insert (xut0741-e)

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTGGAGGATGCCGTACATAA
CGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAG
CAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACcA
ATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTA
CCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAG
ACCTCTGGCAAAGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAA
CGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTG
GTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCT
AACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCAT
CAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCgAGACCACCAGCAGCACCAGAAATGGCC
GCCCAGTGGCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGC
CAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAAC
CAGACAAGTCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTG
ATACCGGCGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAAC
cAATGGcAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGA
TTAACAATTCGTTGGGCTTTGGCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGGCT
TTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTGTAGCTTGATAT
TCTATTTACCAAGAAGAAACAAGAAGTGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGAAAAAAGTCGCT
ACTGGCACTCTATTTATATGTGATAGGCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTCCTCTGCCTTTT
TTTTTTTTTTTTTTTTTCATTCTCTCAATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGGC

> pAM600 insert (xut0741-f)

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTGGAGGATGCCGTACATAA
CGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAG
CAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACcA
ATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTA
CCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAG
ACCTCTGGCAAAGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAA
CGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTG
GTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCT
AACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCAT
CAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCgAGACCACCAGCAGCACCAGAAATGGCC
GCCCAGTGGCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGC
CAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAAC
CAGACAAGTCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTG
ATACCGGCGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAAC
cAATGGcAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGA
TcAACAATTCGTTGGGCTTTGGCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGGCT
TTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTGTAGCTTGATAT
TCTATTTACCAAGAAGAAACAAGAAGTGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGAAAAAAGTCGCT
ACTGGCACTCTATTTATATGTGATAGGCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTCCTCTGCCTTTT
TTTTTTTTTTTTTTTTTCATTCTCTCAATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGGC

> pAM726 insert (chimera) – EcoRI site

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTGGAGGATGCCGTACATAA
CGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAG
CAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACcA
ATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTA
CCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAG
ACCTCTGGCAAAGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAA
CGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTG
GTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCT
AACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCAT
CAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCgAGACCACCAGCAGCACCAGAAATGGCC
GCCCAGTGGCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGC
CAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAAC
CAGACAAGTCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTG
ATACCGGCGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAAC
TAAGAATTCAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCTAGCAAACGCGGTGGGAGCAAAAAGTAGA
ATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACTAATATAGGCATACTTGATAATGAAAAC
TATAAATCGTAAAGACATAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTACCAGCAATTTGGCCCTTCTCCATCTT
TTCGTAAATTTCTGGTAAACTGGATAAGCCAACTACCTTTATTGGAGACTTGACTAGACCTCTGGCAAAGAAATCTAAGGCTT
CTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAACGTGGTTGAAGACATCAGAGGAGCAC
TTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTGGTAGAAGCTTCGATAGCGGCTTCGGA
AACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCTAACAATGTCCTTCTCTTTGGTGAAGT
CGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCATCAATACCTAAGACTCTGTAACCCATC
GCCTTAGCATATTGAACAGCCAAAGAACCTAGACCACCAGCAGCACCAGAAATGGCCGCCCAGTGGCCTGCTCTCAAGTTGGC
AGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGCCAAGTCAGTACCTTGAGGAATGTGAG
CGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAACCAGACAAGTCAGCGTGAGGACAGTTG
GATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTGATACCGGCGTAGTCACCGATCTTCCA
GCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAACTAATGGTAACTTAGTTGGCAATGGCC
AGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGATTAACAATTCGTTGGGCTTTGGCTTT
GGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGGCTTTTTGAGTTTCTGGAATAGACATTGT
GTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTGTAGCTTGATATTCTATTTACCAAGAAGAAACAAGAAG
TGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGAAAAAAGTCGCTACTGGCACTCTATTTATATGTGATAG
GCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTCCTCTGCCTTTTTTTTTTTTTTTTTTCATTCTCTCA
ATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGGC

> pAM728 insert (xut0741-b-3FLAG)

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTGGAGGATGCCGTACATAA
CGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGGGTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAG
CAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATACcA
ATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTTATTTAGAAGTGTCAACAACGTATCTA
CCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAG
ACCTCTGGCAAAGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGATAGACTTGACAA
CGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACCGTTCGCCCTACAGTATCTG
GTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCT
AACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCTTCCTTTCCTGGACCACCAT
CAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCGACTACAAAGACCATGACGGTGATTATA
AAGATCATGACATCGATTACAAGGATGACGATGACAAGCTAGGATCCTAGACCACCAGCAGCACCAGAAATGGCCGCCCAGTG
GCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGACTTCAGCCAAGTCAG
TACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTCGTGGGTGTAACCAGACAAG
TCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTGATACCGGC
GTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCGTGACCACCAACTAATGGTA
ACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGATTAACAAT
TCGTTGGGCTTTGGCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGGCTTTTTGAGT
TTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTGTAGCTTGATATTCTATTTA
CCAAGAAGAAACAAGAAGTGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGAAAAAAGTCGCTACTGGCAC
TCTATTTATATGTGATAGGCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTCCTCTGCCTTTTTTTTTTTT
TTTTTTTTCATTCTCTCAATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGGC

> pAM741 insert (SL-xut0741-b-3FLAG)

GCCGGCCCGCTATATTTTGGTTTTAGATCCTGTCAATACTGAGTTCATCTTTCATTTTCTCAATATAACATTACCGTTATCTC
CCTTATACTTCTCAAATTCCCATCTACGGAACCCTGATCAAGCCCTGAGAAACTATATGAGGGTGTGTACATTGCAGTGCATC
ATTTGTGAGGGTTCAATAATTGAAATTATAGGGTGGACGTCAAGACGAAAAGTGAAAAATTACATCCGTATAGAATTATATAA
CTTGATGAGATGAGATGAGTAAATGACAGAAGAATTACCGTTTCATCATTGAACTTCGATCATTTCAATGCTGGCATGCGAAG
GAAAATGAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGATCCCGCGGTTCGCCGCGGTTTGT
TGGAGGATGCCGTACATAACGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGGGTACAACTCACAGGCGAGGGA
TGGTGTTCACTCGTGCgAGCAAACGCGGTGGGAGCAAAAAGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCAT
CTAAAGATGCTATATACcAATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTTATTTAGA
AGTGTCAACAACGTATCTACCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAATTTCTGGcAAACTGGAcAAGCCAACTACCT
TTATTGGAGACTTGACgAGACCTCTGGCAAAGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACA
ATGGAGATAGACTTGACAACGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAACCAAGACAACAGTACC
GTTCGCCCTACAGTATCTGGTAGAAGCTTCGATAGCGGCTTCGGAAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGG
TAGCCTTAACGACTGCGCTAACAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATTCT
TCCTTTCCTGGACCACCATCAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATTGAACAGCCAAAGAACCGACTACAAA
GACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGCTAGGATCCTAGACCACCAGCAGCAC
CAGAAATGGCCGCCCAGTGGCCTGCTCTCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGC
GCGACTTCAGCCAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGTATTCTTGGAAAGAACCGTC
GTGGGTGTAACCAGACAAGTCAGCGTGAGGACAGTTGGATTCGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGT
TCAACCATTTGATACCGGCGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCACCTTCG
TGACCACCAACTAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGTGCAAATCGGTGTGGCAGACACCAGAGTA
CTTGACGTTGATTAACAATTCGTTGGGCTTTGGCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGA
AGATAATGGCTTTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGATTGTATGCTTTTTG
TAGCTTGATATTCTATTTACCAAGAAGAAACAAGAAGTGATAAAAACAACAAGAGAGCAGTAGTAAGAGTATTTCGAGTGTGA
AAAAAGTCGCTACTGGCACTCTATTTATATGTGATAGGCATGCTATAGCTTTACCAAAAAGTGAACCCCATTTCTATGCTCTC
CTCTGCCTTTTTTTTTTTTTTTTTTTTCATTCTCTCAATCTGAAATTCTCTTATTTCTCCAACTTATAAGTTGGAGATGCCGG
C

```
CLUSTAL O(1.2.4) multiple sequence alignment

XUT0741         GAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTG  60
XUT0741-a       GAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTG  60
XUT0741-b       GAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTG  60
XUT0741-c       GAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTG  60
XUT0741-d       GAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTG  60
XUT0741-e       GAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTG  60
XUT0741-f       GAGAAATATCGAGGGAGACGATTCAGAGGAGCAGGACAAACTATAACCGACTGTTTGTTG  60
                ************************************************************

XUT0741         GAGGATGCCGTACATAACGAACACTGCTGAAGCTACCATGTCTACAGTTTAGAGGAATGG  120
XUT0741-a       GAGGATGCCGTACATAACGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGG  120
XUT0741-b       GAGGATGCCGTACATAACGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGG  120
XUT0741-c       GAGGATGCCGTACATAACGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGG  120
XUT0741-d       GAGGATGCCGTACATAACGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGG  120
XUT0741-e       GAGGATGCCGTACATAACGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGG  120
XUT0741-f       GAGGATGCCGTACATAACGAACACTGCTGAAGCTACCATGTCTACAGTTgAGAGGAATGG  120
                ************************************************ **********

XUT0741         GTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCTAGCAAACGCGGTGGGAGCAAAA  180
XUT0741-a       GTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAGCAAACGCGGTGGGAGCAAAA  180
XUT0741-b       GTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAGCAAACGCGGTGGGAGCAAAA  180
XUT0741-c       GTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAGCAAACGCGGTGGGAGCAAAA  180
XUT0741-d       GTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAGCAAACGCGGTGGGAGCAAAA  180
XUT0741-e       GTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAGCAAACGCGGTGGGAGCAAAA  180
XUT0741-f       GTACAACTCACAGGCGAGGGATGGTGTTCACTCGTGCgAGCAAACGCGGTGGGAGCAAAA  180
                ************************************ **********************

XUT0741         AGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATA  240
XUT0741-a       AGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATA  240
XUT0741-b       AGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATA  240
XUT0741-c       AGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATA  240
XUT0741-d       AGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATA  240
XUT0741-e       AGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATA  240
XUT0741-f       AGTAGAATATTATCTTTTATTCGTGAAACTTCGAACACTGTCATCTAAAGATGCTATATA  240
                ************************************************************

XUT0741         CTAATATAGGCATACTTGATAATGAAAACTATAAATCGTAAAGACATAAGAGATCCGCTT  300
XUT0741-a       CcAATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTT  300
XUT0741-b       CcAATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTT  300
XUT0741-c       CcAATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTT  300
XUT0741-d       CcAATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTT  300
XUT0741-e       CcAATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTT  300
XUT0741-f       CcAATATAGGCATACTgGAcAAgGAAAACTAcAAATCGTAAAGACAcAAGAGATCCGCTT  300
                * ************** ** ** ******** ************** *************

XUT0741         ATTTAGAAGTGTCAACAACGTATCTACCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAA  360
XUT0741-a       ATTTAGAAGTGTCAACAACGTATCTACCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAA  360
XUT0741-b       ATTTAGAAGTGTCAACAACGTATCTACCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAA  360
XUT0741-c       ATTTAGAAGTGTCAACAACGTATCTACCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAA  360
XUT0741-d       ATTTAGAAGTGTCAACAACGTATCTACCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAA  360
XUT0741-e       ATTTAGAAGTGTCAACAACGTATCTACCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAA  360
XUT0741-f       ATTTAGAAGTGTCAACAACGTATCTACCAGCAATTTGGCCCTTCTCCATCTTTTCGTAAA  360
                ************************************************************

XUT0741         TTTCTGGTAAACTGGATAAGCCAACTACCTTTATTGGAGACTTGACTAGACCTCTGGCAA  420
XUT0741-a       TTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAGACCTCTGGCAA  420
XUT0741-b       TTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAGACCTCTGGCAA  420
XUT0741-c       TTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAGACCTCTGGCAA  420
XUT0741-d       TTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAGACCTCTGGCAA  420
XUT0741-e       TTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAGACCTCTGGCAA  420
XUT0741-f       TTTCTGGcAAACTGGAcAAGCCAACTACCTTTATTGGAGACTTGACgAGACCTCTGGCAA  420
                ******* ******** **************************** *************
```

```
XUT0741         AGAAATCTAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGA  480
XUT0741-a       AGAAATCTAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGA  480
XUT0741-b       AGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGA  480
XUT0741-c       AGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGA  480
XUT0741-d       AGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGA  480
XUT0741-e       AGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGA  480
XUT0741-f       AGAAATCcAAGGCTTCTCTGGTATCAGCTCTGTTCCCCACGTAAGAGCCGACAATGGAGA  480
                ******* ************************************************** 

XUT0741         TAGACTTGACAACGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAA  540
XUT0741-a       TAGACTTGACAACGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAA  540
XUT0741-b       TAGACTTGACAACGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAA  540
XUT0741-c       TAGACTTGACAACGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAA  540
XUT0741-d       TAGACTTGACAACGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAA  540
XUT0741-e       TAGACTTGACAACGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAA  540
XUT0741-f       TAGACTTGACAACGTGGTTGAAGACATCAGAGGAGCACTTTGCACCGGCTGGCAAACCAA  540
                ************************************************************ 

XUT0741         CCAAGACAACAGTACCGTTCGCCCTACAGTATCTGGTAGAAGCTTCGATAGCGGCTTCGG  600
XUT0741-a       CCAAGACAACAGTACCGTTCGCCCTACAGTATCTGGTAGAAGCTTCGATAGCGGCTTCGG  600
XUT0741-b       CCAAGACAACAGTACCGTTCGCCCTACAGTATCTGGTAGAAGCTTCGATAGCGGCTTCGG  600
XUT0741-c       CCAAGACAACAGTACCGTTCGCCCTACAGTATCTGGTAGAAGCTTCGATAGCGGCTTCGG  600
XUT0741-d       CCAAGACAACAGTACCGTTCGCCCTACAGTATCTGGTAGAAGCTTCGATAGCGGCTTCGG  600
XUT0741-e       CCAAGACAACAGTACCGTTCGCCCTACAGTATCTGGTAGAAGCTTCGATAGCGGCTTCGG  600
XUT0741-f       CCAAGACAACAGTACCGTTCGCCCTACAGTATCTGGTAGAAGCTTCGATAGCGGCTTCGG  600
                ************************************************************ 

XUT0741         AAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCTAA  660
XUT0741-a       AAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCTAA  660
XUT0741-b       AAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCTAA  660
XUT0741-c       AAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCTAA  660
XUT0741-d       AAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCTAA  660
XUT0741-e       AAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCTAA  660
XUT0741-f       AAACGGAAACATTGATGATACCGTGGGCACCGCCGTTGGTAGCCTTAACGACTGCGCTAA  660
                ************************************************************ 

XUT0741         CAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATT  720
XUT0741-a       CAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATT  720
XUT0741-b       CAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATT  720
XUT0741-c       CAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATT  720
XUT0741-d       CAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATT  720
XUT0741-e       CAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATT  720
XUT0741-f       CAATGTCCTTCTCTTTGGTGAAGTCGATGAATACTTCACCACCGAGCGAGGTAAACAATT  720
                ************************************************************ 

XUT0741         CTTCCTTTCCTGGACCACCATCAATACCTAAGACTCTGTAACCCATCGCCTTAGCATATT  780
XUT0741-a       CTTCCTTTCCTGGACCACCATCAATACCTAAGACTCTGTAACCCATCGCCTTAGCATATT  780
XUT0741-b       CTTCCTTTCCTGGACCACCATCAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATT  780
XUT0741-c       CTTCCTTTCCTGGACCACCATCAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATT  780
XUT0741-d       CTTCCTTTCCTGGACCACCATCAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATT  780
XUT0741-e       CTTCCTTTCCTGGACCACCATCAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATT  780
XUT0741-f       CTTCCTTTCCTGGACCACCATCAATACCcAAGACTCTGTAACCCATCGCCTTAGCATATT  780
                ***************************** ****************************** 

XUT0741         GAACAGCCAAAGAACCTAGACCACCAGCAGCACCAGAAATGGCCGCCCAGTGGCCTGCTC  840
XUT0741-a       GAACAGCCAAAGAACCTAGACCACCAGCAGCACCAGAAATGGCCGCCCAGTGGCCTGCTC  840
XUT0741-b       GAACAGCCAAAGAACCTAGACCACCAGCAGCACCAGAAATGGCCGCCCAGTGGCCTGCTC  840
XUT0741-c       GAACAGCCAAAGAACCgAGACCACCAGCAGCACCAGAAATGGCCGCCCAGTGGCCTGCTC  840
XUT0741-d       GAACAGCCAAAGAACCgAGACCACCAGCAGCACCAGAAATGGCCGCCCAGTGGCCTGCTC  840
XUT0741-e       GAACAGCCAAAGAACCgAGACCACCAGCAGCACCAGAAATGGCCGCCCAGTGGCCTGCTC  840
XUT0741-f       GAACAGCCAAAGAACCgAGACCACCAGCAGCACCAGAAATGGCCGCCCAGTGGCCTGCTC  840
                *************** ****************************************** 
```

```
XUT0741         TCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGA  900
XUT0741-a       TCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGA  900
XUT0741-b       TCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGA  900
XUT0741-c       TCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGA  900
XUT0741-d       TCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGA  900
XUT0741-e       TCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGA  900
XUT0741-f       TCAAGTTGGCAGACTTCAAAGCCTTGTATACGGTGATACCAGCACACAAGATTGGCGCGA  900
                ************************************************************

XUT0741         CTTCAGCCAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGT  960
XUT0741-a       CTTCAGCCAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGT  960
XUT0741-b       CTTCAGCCAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGT  960
XUT0741-c       CTTCAGCtAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGT  960
XUT0741-d       CTTCAGCCAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGT  960
XUT0741-e       CTTCAGCCAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGT  960
XUT0741-f       CTTCAGCCAAGTCAGTACCTTGAGGAATGTGAGCGGCTTGAACAGCGTCAGCGGTAGCGT  960
                ******* **************************************************

XUT0741         ATTCTTGGAAAGAACCGTCGTGGGTGTAACCAGACAAGTCAGCGTGAGGACAGTTGGATT  1020
XUT0741-a       ATTCTTGGAAAGAACCGTCGTGGGTGTAACCAGACAAGTCAGCGTGAGGACAGTTGGATT  1020
XUT0741-b       ATTCTTGGAAAGAACCGTCGTGGGTGTAACCAGACAAGTCAGCGTGAGGACAGTTGGATT  1020
XUT0741-c       ATTCTTGGAAAGAACCGTCGTGGGTGTAACCAGACAAGTCAGCGTGAGGACAGTTGGATT  1020
XUT0741-d       ATTCTTGGAAAGAACCGTCGTGGGTGTAACCAGACAAGTCAGCGTGAGGACAGTTGGATT  1020
XUT0741-e       ATTCTTGGAAAGAACCGTCGTGGGTGTAACCAGACAAGTCAGCGTGAGGACAGTTGGATT  1020
XUT0741-f       ATTCTTGGAAAGAACCGTCGTGGGTGTAACCAGACAAGTCAGCGTGAGGACAGTTGGATT  1020
                ************************************************************

XUT0741         CGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTGATACCGG  1080
XUT0741-a       CGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTGATACCGG  1080
XUT0741-b       CGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTGATACCGG  1080
XUT0741-c       CGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTGATACCGG  1080
XUT0741-d       CGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTGATACCGG  1080
XUT0741-e       CGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTGATACCGG  1080
XUT0741-f       CGTTACCCAATTCACAGTATTCACAGGCCATACAAGAACCGTTCAACCATTTGATACCGG  1080
                ************************************************************

XUT0741         CGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCAC  1140
XUT0741-a       CGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCAC  1140
XUT0741-b       CGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCAC  1140
XUT0741-c       CGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCAC  1140
XUT0741-d       CGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCAC  1140
XUT0741-e       CGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCAC  1140
XUT0741-f       CGTAGTCACCGATCTTCCAGCCCTTAACGTTTTCACCCATGCCGACAACGACACCGGCAC  1140
                ************************************************************

XUT0741         CTTCGTGACCACCAACTAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGT  1200
XUT0741-a       CTTCGTGACCACCAACTAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGT  1200
XUT0741-b       CTTCGTGACCACCAACTAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGT  1200
XUT0741-c       CTTCGTGACCACCAACTAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGT  1200
XUT0741-d       CTTCGTGACCACCAACTAATGGTAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGT  1200
XUT0741-e       CTTCGTGACCACCAACcAATGGcAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGT  1200
XUT0741-f       CTTCGTGACCACCAACcAATGGcAACTTAGTTGGCAATGGCCAGTCACCATGCCAAGCGT  1200
                ********** ***** ************************************

XUT0741         GCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGATTAACAATTCGTTGGGCTTTG  1260
XUT0741-a       GCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGATTAACAATTCGTTGGGCTTTG  1260
XUT0741-b       GCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGATTAACAATTCGTTGGGCTTTG  1260
XUT0741-c       GCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGATTAACAATTCGTTGGGCTTTG  1260
XUT0741-d       GCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGATTAACAATTCGTTGGGCTTTG  1260
XUT0741-e       GCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGATTAACAATTCGTTGGGCTTTG  1260
XUT0741-f       GCAAATCGGTGTGGCAGACACCAGAGTACTTGACGTTGATcAACAATTCGTTGGGCTTTG  1260
                ************************************ ******************
```

```
XUT0741         GCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGG 1320
XUT0741-a       GCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGG 1320
XUT0741-b       GCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGG 1320
XUT0741-c       GCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGG 1320
XUT0741-d       GCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGG 1320
XUT0741-e       GCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGG 1320
XUT0741-f       GCTTTGGAACTGGGATATCCTTATGCTCCAACTTGCCGTTGGATTCGTAGAAGATAATGG 1320
                ************************************************************

XUT0741         CTTTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGA 1380
XUT0741-a       CTTTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGA 1380
XUT0741-b       CTTTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGA 1380
XUT0741-c       CTTTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGA 1380
XUT0741-d       CTTTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGA 1380
XUT0741-e       CTTTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGA 1380
XUT0741-f       CTTTTTGAGTTTCTGGAATAGACATTGTGTATTACGATATAGTTAATAGTTGATAGTTGA 1380
                ************************************************************

XUT0741         TTGTATGCTTTTTGTAGCTTGATATTCTATTTACCAAGAAGAAACAAGAAGTGATAAAAA 1440
XUT0741-a       TTGTATGCTTTTTGTAGCTTGATATTCTATTTACCAAGAAGAAACAAGAAGTGATAAAAA 1440
XUT0741-b       TTGTATGCTTTTTGTAGCTTGATATTCTATTTACCAAGAAGAAACAAGAAGTGATAAAAA 1440
XUT0741-c       TTGTATGCTTTTTGTAGCTTGATATTCTATTTACCAAGAAGAAACAAGAAGTGATAAAAA 1440
XUT0741-d       TTGTATGCTTTTTGTAGCTTGATATTCTATTTACCAAGAAGAAACAAGAAGTGATAAAAA 1440
XUT0741-e       TTGTATGCTTTTTGTAGCTTGATATTCTATTTACCAAGAAGAAACAAGAAGTGATAAAAA 1440
XUT0741-f       TTGTATGCTTTTTGTAGCTTGATATTCTATTTACCAAGAAGAAACAAGAAGTGATAAAAA 1440
                ************************************************************

XUT0741         CAACAAGAGAGCAGTAG  1457
XUT0741-a       CAACAAGAGAGCAGTAG  1457
XUT0741-b       CAACAAGAGAGCAGTAG  1457
XUT0741-c       CAACAAGAGAGCAGTAG  1457
XUT0741-d       CAACAAGAGAGCAGTAG  1457
XUT0741-e       CAACAAGAGAGCAGTAG  1457
XUT0741-f       CAACAAGAGAGCAGTAG  1457
                *****************
```

4. Supplementary Tables for Small and Single-Cell RNA-Seq.

Table S2. Yeast strains used for Small RNA-Seq.

Name	Background	Genotype	Source/Reference
YAM1725	W303	*MATα LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15*	(Drinnenberg et al., 2009)
YAM1730	W303	*MATα leu2-3,112 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15*	(Drinnenberg et al., 2009)
YAM1982	W303	*MATα LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15 xrn1::His3MX6*	(Wery et al., 2016)
YAM2271	W303	*MATα ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3::EGFP(S65T)-kanMX6can1-100 xrn1::His3MX6*	(Wery et al., 2016)
YAM2913	W303	*MATα ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3::EGFP(S65T)-kanMX6can1-100 upf1::His3MX6*	This work
YAM2915	W303	*MATα LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15 upf1::His3MX6*	This work

Table S5. Sense/Antisense Pairs in Common to CHX, *upf1Δ* and xrn1*Δ* cells from Single-Cell RNA-Seq.

Sense/Antisense Pair	Coordinates	Start	End	Strand	XUT as Type *
MAP2/XUT0020	chr02	46579	47370	+	type6
SEC17/XUT0953	chr02	126093	127562	-	type6
PDR3/XUT0961	chr02	218231	221115	-	type3
FLR1/XUT0034	chr02	252925	254972	+	type2
TAT1/XUT0051	chr02	376521	378723	+	type4
TIM12/XUT0056	chr02	427129	427701	+	type6
PHO3/XUT0056	chr02	427129	427701	+	type6
PHO3/XUT0057	chr02	427716	431776	+	type6
TAF5/XUT0071	chr02	615402	616253	+	type6
MRPS5/XUT0994	chr02	722288	723031	-	type6
DUT1/XUT0994	chr02	722288	723031	-	type6
SRD1/XUT0102	chr03	147892	148779	+	type3
RGT2/XUT1036	chr04	212631	217190	-	type6
ARF2/XUT1036	chr04	212631	217190	-	type6
RDI1/XUT0141	chr04	219148	219724	+	type6
PPH21/XUT0141	chr04	219148	219724	+	type6
HNT1/XUT0143	chr04	238906	239522	+	type3
PUS9/XUT0156	chr04	386276	390379	+	type6
GPR1/XUT0156	chr04	386276	390379	+	type6
LYS14/XUT0166	chr04	512001	513419	+	type4
SNF11/XUT1061	chr04	592666	592924	-	type2
RRP1/XUT0180	chr04	617523	618415	+	type4
BMH2/XUT1069	chr04	652632	653140	-	type4
AHA1/XUT1084	chr04	892883	893516	-	type2
EXG2/XUT0203	chr04	976491	979320	+	type6
CPR5/XUT0212	chr04	1071462	1072706	+	type6
MCM21/XUT1094	chr04	1102103	1103836	-	type6
ADE8/XUT0231	chr04	1288242	1289149	+	type4
HSP31/XUT0250	chr04	1500708	1502646	+	type5
YER062C/XUT0281	chr05	279882	280902	+	type4
GNA1/XUT0314	chr06	104097	104624	+	type6
SNF4/XUT1212	chr07	292833	293261	-	type6
HNM1/XUT0342	chr07	362122	364114	+	type6
YGL039W/XUT1228	chr07	423337	424708	-	type4
PTI1/XUT1261	chr07	800463	801860	-	type2
PBP1/XUT0387	chr07	850811	852577	+	type3
ECM14/XUT0453	chr08	368119	368535	+	type6
CHS7/XUT1308	chr08	383336	386979	-	type6
DSE2/XUT1308	chr08	383336	386979	-	type6
FMO1/XUT1311	chr08	454794	458503	-	type6
QDR2/XUT1340	chr09	132114	135272	-	type6
YRB2/XUT0490	chr09	243456	243984	+	type6
LAS21/XUT1369	chr10	319102	319790	-	type2
MCD4/XUT0574	chr11	136380	137933	+	type3
PMU1/XUT0580	chr11	200447	202741	+	type5
LAC1/XUT0595	chr11	427241	428746	+	type4
DYN1/XUT0602	chr11	533775	536472	+	type6

BET3/XUT0603	chr11	570435	570853	+	type2
GEX2/XUT1467	chr11	661427	662438	-	type4
TPO1/XUT1470	chr12	84514	85914	-	type4
MAS1/XUT0653	chr12	490883	492774	+	type3
HAP1/XUT1498	chr12	649048	650789	-	type2
BER1/XUT1520	chr12	948688	949949	-	type2
CTK3/XUT1536	chr13	45914	46562	-	type3
CPR3/XUT1537	chr13	110946	111706	-	type5
TCB3/XUT0689	chr13	127789	129735	+	type6
GAL80/XUT1539	chr13	171410	173537	-	type6
TAF4/XUT1553	chr13	275650	276306	-	type6
CCS1/XUT0703	chr13	347182	348501	+	type6
YMR178W/XUT1577	chr13	618238	619586	-	type5
HSC82/XUT1580	chr13	633022	633241	-	type6
MGL2/XUT1589	chr13	687386	688847	-	type4
YMR315W/XUT1600	chr13	902484	905066	-	type6
DIA1/XUT1600	chr13	902484	905066	-	type6
TCB2/XUT1625	chr14	462231	466625	-	type6
RNH201/XUT1628	chr14	490183	491247	-	type6
SUN4/XUT1630	chr14	500955	502870	-	type2
IDP3/XUT1646	chr14	614445	615668	-	type6
BDS1/XUT1658	chr15	5480	6898	-	type4
SMF1/XUT0797	chr15	90072	91426	+	type2
YSP3/XUT1683	chr15	331189	333373	-	type6
DNL4/XUT0811	chr15	335857	337470	+	type4
WTM1/XUT1711	chr15	770397	772231	-	type4
RDL1/XUT1717	chr15	848938	850081	-	type6
RRS1/XUT1718	chr15	868726	869597	-	type6
GYP5/XUT0860	chr16	76698	82620	+	type6
YPL216W/XUT1750	chr16	143391	147183	-	type5
EAF3/XUT0906	chr16	607534	609235	+	type3
NCE102/XUT1802	chr16	828394	830449	-	type4

* asXUTs (mRNA overlap $\geq$ 1 nt) are classified into full antisense (type 2), free 5' end (type 3), free 3' end (type 4), and free 5' and 3' ends (type 5) (Wery et al., 2016). Overlap with two mRNAs (type 6).

ANNEX II

non-coding RNA

MDPI

Review

From Yeast to Mammals, the Nonsense-Mediated mRNA Decay as a Master Regulator of Long Non-Coding RNAs Functional Trajectory

Sara Andjus [1], Antonin Morillon [2,*] and Maxime Wery [2,*]

[1] ncRNA, Epigenetic and Genome Fluidity, Institut Curie, PSL University, Sorbonne Université, CNRS UMR3244, 26 Rue d'Ulm, CEDEX 05, F-75248 Paris, France; sara.andus@curie.fr
[2] ncRNA, Epigenetic and Genome Fluidity, Institut Curie, Sorbonne Université, CNRS UMR3244, 26 Rue d'Ulm, CEDEX 05, F-75248 Paris, France
* Correspondence: antonin.morillon@curie.fr (A.M.); maxime.wery@curie.fr (M.W.)

Abstract: The Nonsense-Mediated mRNA Decay (NMD) has been classically viewed as a translation-dependent RNA surveillance pathway degrading aberrant mRNAs containing premature stop codons. However, it is now clear that mRNA quality control represents only one face of the multiple functions of NMD. Indeed, NMD also regulates the physiological expression of normal mRNAs, and more surprisingly, of long non-coding (lnc)RNAs. Here, we review the different mechanisms of NMD activation in yeast and mammals, and we discuss the molecular bases of the NMD sensitivity of lncRNAs, considering the functional roles of NMD and of translation in the metabolism of these transcripts. In this regard, we describe several examples of functional micropeptides produced from lncRNAs. We propose that translation and NMD provide potent means to regulate the expression of lncRNAs, which might be critical for the cell to respond to environmental changes.

Keywords: nonsense-mediated mRNA decay; Upf1; lncRNA; translation; micropeptide

Citation: Andjus, S.; Morillon, A.; Wery, M. From Yeast to Mammals, the Nonsense-Mediated mRNA Decay as a Master Regulator of Long Non-Coding RNAs Functional Trajectory. *Non-coding RNA* **2021**, *7*, 44. https://doi.org/10.3390/ncrna7030044

Academic Editor: George A. Calin

Received: 15 June 2021
Accepted: 25 July 2021
Published: 27 July 2021

1. Introduction

The accurate transmission of the genetic information is crucial for the cell, and several surveillance mechanisms have evolved to monitor the distinct steps of gene expression. RNA surveillance pathways are responsible for detecting and eliminating RNA intermediates that lack integrity or functionality [1–3]. Such transcripts can arise due to deleterious or genomic frameshift mutations or inappropriate processing, and the subsequent failure to produce functional proteins may result in disease.

If an mRNA is devoid of a stop codon (for instance, in the case of truncation or premature 3′-end cleavage and polyadenylation), it will cause the ribosome to progress to its 3′ extremity and stall. Such aberrant mRNAs are rapidly degraded through a process termed non-stop decay [4–6]. In contrast, the presence of stable structures or damaged nucleotides within an open reading frame (ORF) can impede ribosome progression, resulting into ribosome stalling upstream of the stop codon. In this case, the transcript is targeted to the degradation by the no-go decay pathway [7,8].

The Nonsense-Mediated mRNA Decay (NMD) is another quality control pathway targeting transcripts that terminate translation prematurely [9,10], such as mRNAs harboring a premature termination codon (PTC) within the ORF [11], as well as PTC-less mRNAs displaying long 3′ untranslated regions (UTRs) [12,13] or short upstream ORFs [13–16]. The NMD-targeted mRNAs are rapidly degraded [17,18], thus preventing the production of truncated, possibly deleterious proteins [19–21].

Here, we review the different mechanisms of NMD activation in yeast and mammals. We discuss the recent evidence showing that NMD also targets and regulates the expression of long non-coding (lnc)RNAs, including antisense (as)lncRNAs in yeast, indicating that

translation is part of the metabolism of transcripts initially thought to be devoid of coding potential. Supporting this idea, we describe several examples of functional micropeptides produced from small (sm)ORFs of lncRNAs. We propose that NMD and translation take part in the metabolism of lncRNAs, regulating their expression and providing the opportunity to produce micropeptides which might have a role in the cellular response to environmental changes.

2. Discovery, Conservation and Functions of NMD

NMD is a translation-dependent RNA decay pathway [22–24], which has been evolutionarily conserved [10,25]. It was originally discovered in the budding yeast *Saccharomyces cerevisiae* by Losson and Lacroute, when they observed that the presence of nonsense mutations reduces the level of a mutant mRNA without affecting its synthesis rate [26]. It was discovered afterwards in humans in the context of β^0-thalassemia, where it was observed that β-globin mRNAs levels dramatically decrease when carrying nonsense mutations [27,28].

Upstream frameshift proteins (Upfs) 1, 2 and 3 constitute the conserved core components of NMD [29] and were initially identified in *S. cerevisiae* [30–32].

Upf1 is a monomeric, highly regulated superfamily 1 helicase. Its ATPase and helicase activities are essential for NMD [33,34]. Upf1 has the ability to translocate slowly but with high processivity on nucleic acids and to unwind long double-stranded (ds)RNA structures [35]. Upf2 is the second core NMD factor and functions as a bridge between Upf1 and Upf3 [36–38]. Its interaction with Upf1 is a prerequisite for the phosphorylation of Upf1 [39]. However, NMD can be activated independently of Upf2 [40,41]. Upf3 is the least conserved of the three core NMD factors [42]. Vertebrates have two Upf3 paralogs, Upf3A and Upf3B [37]; in human cells, Upf3B seems to be the main contributor to NMD [43]. Like Upf2, Upf3 stimulates the ATPase and helicase activity of Upf1 in vitro [36]. In metazoans, NMD requires four additional factors: Smg1, Smg5, Smg6, and Smg7 [10,20,44,45]. Interestingly, there is a correlation between the organism complexity and the dependency on NMD; while Upf1 is essential in *Arabidopsis*, *Drosophila* and vertebrates [46–49], NMD-deficient mutants in yeast and *C. elegans* are viable [20,30,32,50].

At present, it has become clear that the mRNA quality control represents only one face of the multiple functions of NMD [51–55]. In yeast, almost half of protein-coding genes can generate NMD-sensitive mRNA isoforms, including truncated mRNAs for which transcription initiation occurs downstream of the canonical translation initiation site [56]. NMD also targets intron-containing pre-mRNAs that have escaped splicing and were exported to the cytoplasm [14]. In addition, NMD regulates 3–10% of physiological, non-mutated mRNAs in yeast, *Drosophila* and humans, including mRNAs with small upstream ORFs [13–16,57–61], long 3′ UTRs [12,13], as well as mRNAs displaying low translational efficiency and average codon optimality [14]; considered together, NMD provides a significant contribution to the post-transcriptional regulation of gene expression [55].

Numerous physiological processes rely on the capacity of the cell to adjust NMD activity at global and/or transcript specific levels. NMD factors are essential for embryonic development in vertebrates, as disrupted expression of core NMD factors confers lethality at an early embryonic stage [47,62]. NMD is also crucial for the maintenance of hematopoietic stem and progenitor cells [62], the maturation of T cells [62], as well as for liver development, function and regeneration in mice [63]. Furthermore, NMD is important for the response to multiple stresses [64–66], being itself regulated in response to stresses such as hypoxia [67] and amino acid deprivation [68]. In fact, many stress-related mRNAs are targeted by NMD under normal physiological conditions but are stabilized upon stress, due to the inhibition of NMD activity [69]. However, as Upf1 is also involved in diverse RNA decay pathways distinct from NMD, including staufen-mediated mRNA decay, replication-dependent histone mRNA decay, glucocorticoid receptor-mediated mRNA decay, regnase 1-mediated mRNA decay, and tudor-staphylococcal/micrococcal-like nuclease-mediated microRNA decay [70], it remains possible that some of the phenotypes associated with

mutants of Upf1 do not reflect the loss of NMD per se. Finally, seven NMD factors (Upf1, Upf2, Upf3B, Smg1, Smg5, Smg6, and Smg7) have been found to be NMD targets in mouse and human cells, revealing the existence of a regulatory feedback network between NMD factors, which is critical for the maintenance of physiological NMD activity and RNA homeostasis [71].

3. Molecular Bases of NMD Activation

In many organisms, NMD has been coupled to pre-mRNA splicing [24,72–77]. The Exon Junction complex (EJC) is deposited by the spliceosome at the level of the junction between two exons [78], and it is normally removed from the coding regions by the translating ribosomes [79]. The EJC is formed around four core components: the DEAD-box RNA helicase eIF4A3, MLN51, and the Magoh/Y14 heterodimer [80]. The presence of an EJC downstream of a stop codon is recognized as an abnormal situation and enhances the association and activity of Upf1 [81]. In the EJC-enhanced NMD model (Figure 1a), premature translation termination involves the SURF (Smg1–Upf1–eRF1–eRF3) complex, which consists of the Smg1 kinase, Upf1 and the eukaryotic release factors eRF1 and eRF3, and associates with the ribosome stalled at the PTC [82]. Upf2 and Upf3 are then recruited to SURF via the proximal EJC, leading to the formation of the DECID (DECay InDucing) complex [82]. The interaction with Upf2 induces a conformational change in Upf1, allowing its phosphorylation by Smg1 and its activation [82]. The activated Upf1 recruits the Smg6 endonuclease [83] and the Smg5–7 heterodimer [84], which in turn activates RNA deadenylation and decapping. In addition, phosphorylated Upf1 also prevents new translation initiation events by interacting with the translation initiation factor eIF3, inhibiting the formation of a competent translation initiation complex [85]. Finally, protein phosphatase 2 (PP2A) dephosphorylates Upf1, allowing it to return to its unphosphorylated state for another NMD cycle [84].

In addition to the EJC-enhanced NMD, examples of EJC-independent NMD have been described in human cells [86,87], as well as in fission yeast [88], *C. elegans* [89], *Drosophila* [90] and plants [72], all of which have orthologs of EJC factors. In contrast, in *S. cerevisiae*, not only is the proportion of intron-containing genes low (4%) [91], but EJC factors are absent, with the exception of eIF4A3 (Fal1), which acts in pre-rRNA processing in yeast [92].

The EJC-independent NMD targets RNAs with extended 3′ UTR but lacking EJC downstream of the translation termination codon [77,87,93–96]. Indeed, RNAs where long EJC-free sequences are inserted downstream of a stop codon show reduced levels due to accelerated degradation by NMD [93,95]. This EJC-independent NMD might be a vestige of an ancestral NMD mechanism associated with an abnormally long 3′ UTR, referred to as "faux 3′ UTR", which is still present in *S. cerevisiae* [97]. In this model, a compromised interaction between the polyadenylate-binding protein Pab1 and the prematurely terminating ribosome results in less efficient termination and enhanced interaction between Upf1 and eRF1/eRF3, triggering NMD (Figure 1b). In this context, a recent proteomics-based analysis in yeast characterized the composition of two distinct NMD complexes associated with Upf1 named Upf1-23 (Upf1, Upf2, Upf3) and Upf1-decapping [98]. The latter contained the decapping enzyme Dcp2 and its co-factor Dcp1, the decapping activator Ebs3, and two poorly characterized proteins, Nmd4 and Ebs1. The Upf1-23 complex is recruited and assembled on the RNA substrate, and then a complete re-organization leads to the replacement of the Upf2/3 heterodimer by Nmd4, Ebs1, Dcp2 and its co-factors (Figure 1b). Nmd4 and Ebs1 are accessory factors for NMD and could be functional homologues of human Smg6 and Smg5/7, respectively [98,99]. The discovery of these new factors in yeast suggests that NMD mechanisms could be more conserved than previously thought. However, how the switch from the "Upf1-23" to an "Upf1-decapping" complex occurs remains unclear.

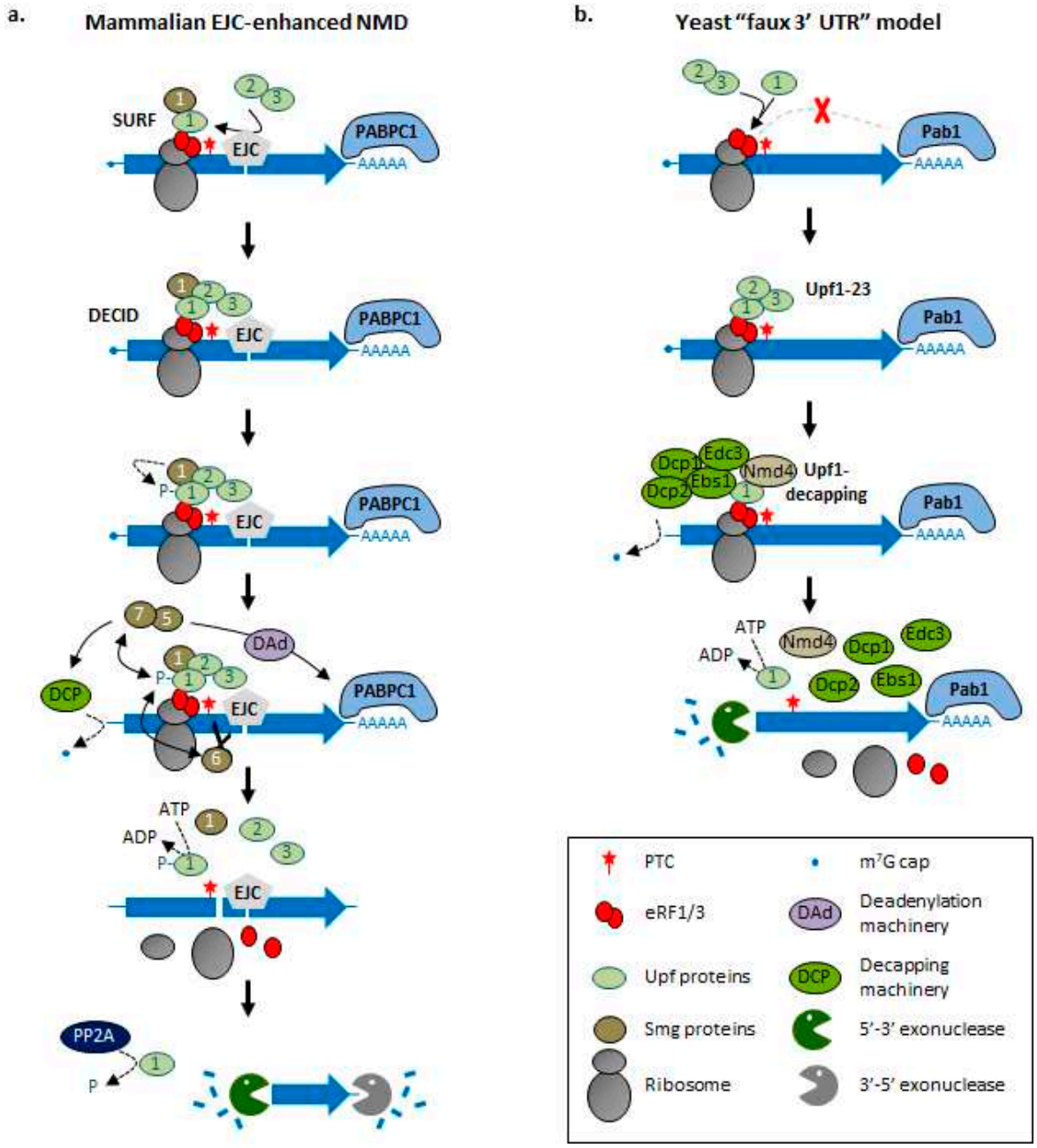


Figure 1. Models of NMD activation mechanisms in mammals and yeast. (**a**) Mammalian EJC-enhanced NMD. When an EJC remains bound to the RNA downstream of a termination codon, translation termination is inefficient as the EJC interferes with the interaction between PABPC1 and the eukaryotic release factors (eRF1/eRF3). Instead, a SURF complex (Smg1–Upf1–eRF1–eRF3) forms at the level of the PTC. Upf2 and Upf3 are then recruited by the downstream EJC and associate with SURF to form the decay-inducing (DECID) complex. Smg1 phosphorylates Upf1 (P), activating it. Phosphorylated Upf1 promotes RNA decay via Smg6-dependent endonucleolytic cleavage and the Smg5–Smg7-dependent triggering of mRNA deadenylation and decapping. ATP hydrolysis by Upf1 allows the dissociation of the termination complex and the release of the transcript, which can be degraded. Upf1 dephosphorylation by protein phosphatase 2A (PP2A) allows it to return to a dephosphorylated state. The coding region and the UTRs of the mRNA are represented as a large blue arrow and thin blue lines, respectively. See the key for the other symbols. (**b**) "Faux" 3′ UTR model in yeast. A long 3′ UTR results in inefficient translation termination and Upf1 interaction with eRF1/eRF3, promoting the formation of the Upf1-23 complex (Upf1, Upf2, Upf3) at the level of the terminating ribosome. The Upf2-Upf3 heterodimer is then replaced by Nmd4, Ebs1, the decapping enzyme Dcp2 and its co-factors Dcp1 and Edc3 in the Upf1-decapping complex, leading to RNA decapping. ATP hydrolysis by Upf1 promotes the disassembly of the mRNA/ribosome/Upf1-decapping complex, leading to the release of the transcript which can finally be degraded by Xrn1.

The polyadenylate-binding protein 1 (PABPC1 in mammals, Pab1 in yeast) is known to stimulate translation termination efficiency by recruiting the release factors to the ribosome [100]. A long distance between the PTC and Pab1/PABPC1 triggers NMD in all studied species [87,93,95,101,102], while tethering it close to the PTC suppresses the NMD sensitivity of the PTC-containing transcripts in yeast [97] and *Drosophila* cells [101]. Mechanistically, it has been proposed that the long 3′ UTR would act by impeding the efficient interaction between Pab1/PABPC1 and eRF1/eRF3, favoring the recruitment of Upf1 by the latter and the formation of a SURF complex at the level of the PTC.

Currently, several questions remain open regarding Upf1 recruitment to the target transcripts. Until recently, the classical view was that Upf1 is recruited at the level of the nonsense codon by the stalled ribosome through an interaction with eRF1/eRF3. However, it has been shown that substrate discrimination by NMD can occur independently of Pab1/PABPC1 or its interaction with eRF3 [103,104], indicating that other features contribute to RNA recognition by NMD. In addition, if Upf1 preferentially binds NMD-targeted transcripts [61,105–107], with a marked enrichment in the 3′ UTR [81,108–111], it is redistributed into the coding sequence upon translation inhibition [109–111]. This suggests that Upf1 can bind the RNA independently of translation as well as to NMD non-targets and is pushed away from the coding region by the elongating ribosomes (Figure 2). This means that NMD substrate selection occurs after Upf1 association with the RNA. In this regard, NMD substrate discrimination was shown to rely on a faster dissociation of Upf1 from non-target mRNAs, and this depends on its ATPase activity [106,112]. ATP hydrolysis by Upf1 is also required for ribosome release and recycling and efficient RNA degradation [113,114].

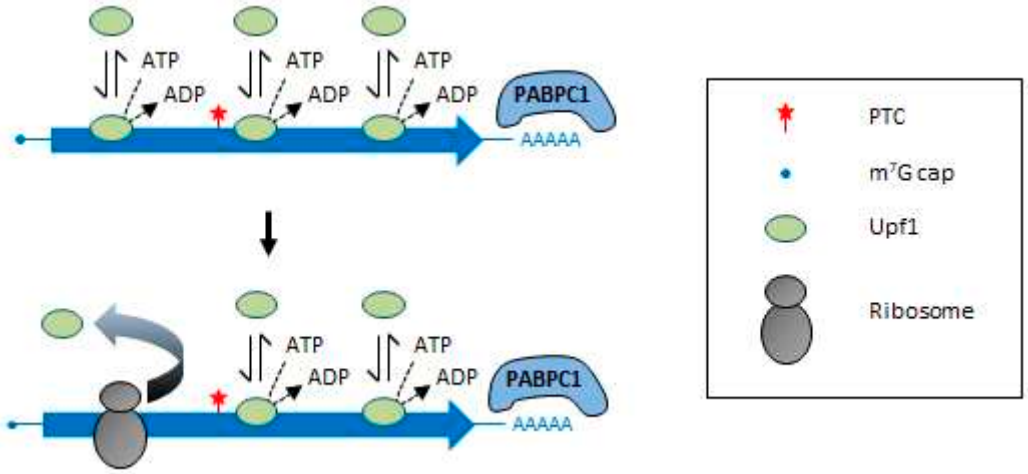


Figure 2. Model of translation-dependent displacement of Upf1 from the mRNA coding region. Upf1 binds promiscuously to any accessible RNA (including NMD non-targets), independently of translation. ATP hydrolysis promotes Upf1 dissociation from non-target RNA regions. Upf1 is also displaced from the coding region by the translating ribosome. This model implies that NMD substrate selection occurs after Upf1 associates with the RNA.

4. Long Non-Coding RNAs: An Unexpected Class of NMD Substrates

Unexpectedly, recent transcriptome-wide analyses of RNA binding sites of Upf1 in human and yeast cells revealed that, in addition to mRNAs, Upf1 can also bind lncRNAs [111,115,116].

LncRNAs are a prominent class of transcripts that play important roles in multiple cellular processes, including chromatin modification and regulation of gene expression [117–120]. They were a priori presumed to be devoid of coding potential [121]. However, this initial assumption has been challenged over recent years by a number of analyses showing that transcripts produced from non-coding regions of the genome, including intergenic regions and sequences antisense to protein-coding genes, associate

with the translation machinery in different models, including *S. cerevisiae* [122–126], fission yeast [127,128], plant [129], *Drosophila* [129,130], zebrafish [129,131,132], mouse [129,133] and human cells [129,131,134–136]. Thus, not only could the ribosome constitute a default destination for cytoplasmic lncRNAs [136], but the smORFs they carry are likely to be translated into micropeptides [129]. Furthermore, the observation that translation elongation inhibitors results in the stabilization of polysomal lncRNAs in human (K562) cells indicates that translation also determines the degradation of cytoplasmic lncRNAs [136].

In budding and fission yeasts, cytoplasmic lncRNAs are extensively degraded by the 5′-3′ exoribonuclease Xrn1/Exo2 [124,137–139]. Inactivation of Xrn1 leads to the stabilization of Xrn1-sensititve Unstable Transcripts (XUTs), the majority of which are antisense to protein-coding genes [124,138–140]. Strikingly, in *S. cerevisiae*, 70% of these XUTs are targeted to Xrn1 through NMD [56,124], indicating that most XUTs are translated and that translation constitutes a prerequisite for their degradation. In fact, NMD-sensitive XUTs display ribosome footprints restricted to their 5′ regions, followed by long downstream ribosome-free regions [124]. Conversely, antisense (as)XUTs were found to form dsRNA structures with their paired-sense mRNAs, thus modulating their sensitivity to NMD [124]. This suggests that unless blocked by dsRNA structures, ribosomes could rapidly bind smORFs in the 5′ region of cytoplasmic lncRNAs (Figure 3). The detection of a long 3′ UTR would trigger NMD, leading to the decapping of the transcript and its degradation by Xrn1. Alternatively, but not exclusively, dsRNA could also interfere with the recruitment of NMD factors to asXUTs. Given the current view of Upf1 binding to the RNA (Figure 2) and the observation that Upf1 physically interacts with yeast lncRNAs [115], we propose that Upf1 binds XUTs in a promiscuous manner, independently of translation, regardless of whether or not the transcript will be targeted to Xrn1 through NMD. As proposed for mRNAs, Upf1 would be displaced from the smORF of XUTs by the translating ribosomes and would accumulate on the 3′ UTR. Since NMD-sensitive XUTs are globally longer than NMD-insensitive ones [124], we speculate that longer XUTs carry longer 3′ UTRs, which will be more likely to impede the interaction between eRF1/eRF3 and Pab1. Instead, this situation would favor the Upf1–eRF1/eRF3 interaction, enclosing the XUT as an NMD target through a mechanism similar to the "faux 3′ UTR" model. Supporting this idea, XUTs are poly-adenylated [124,139], and their poly(A) tail is likely to be bound by Pab1.

We speculate that cytoplasmic smORFs-bearing lncRNAs, reminiscent of the yeast NMD-sensitive XUTs, could be targeted by NMD in other eukaryotic cells. Consistent with this idea, NMD inactivation in mouse embryonic stem cells results in the stabilization of a subset of lncRNAs [123]. Moreover, the levels of lncRNAs, including Natural Antisense Transcripts, are also modulated by NMD in *Arabidopsis* [141]. Further support comes from the observation that the *growth arrest-specific 5* (*GAS5*) lncRNA is targeted by NMD and accumulates in Upf1-depleted human cells [142].

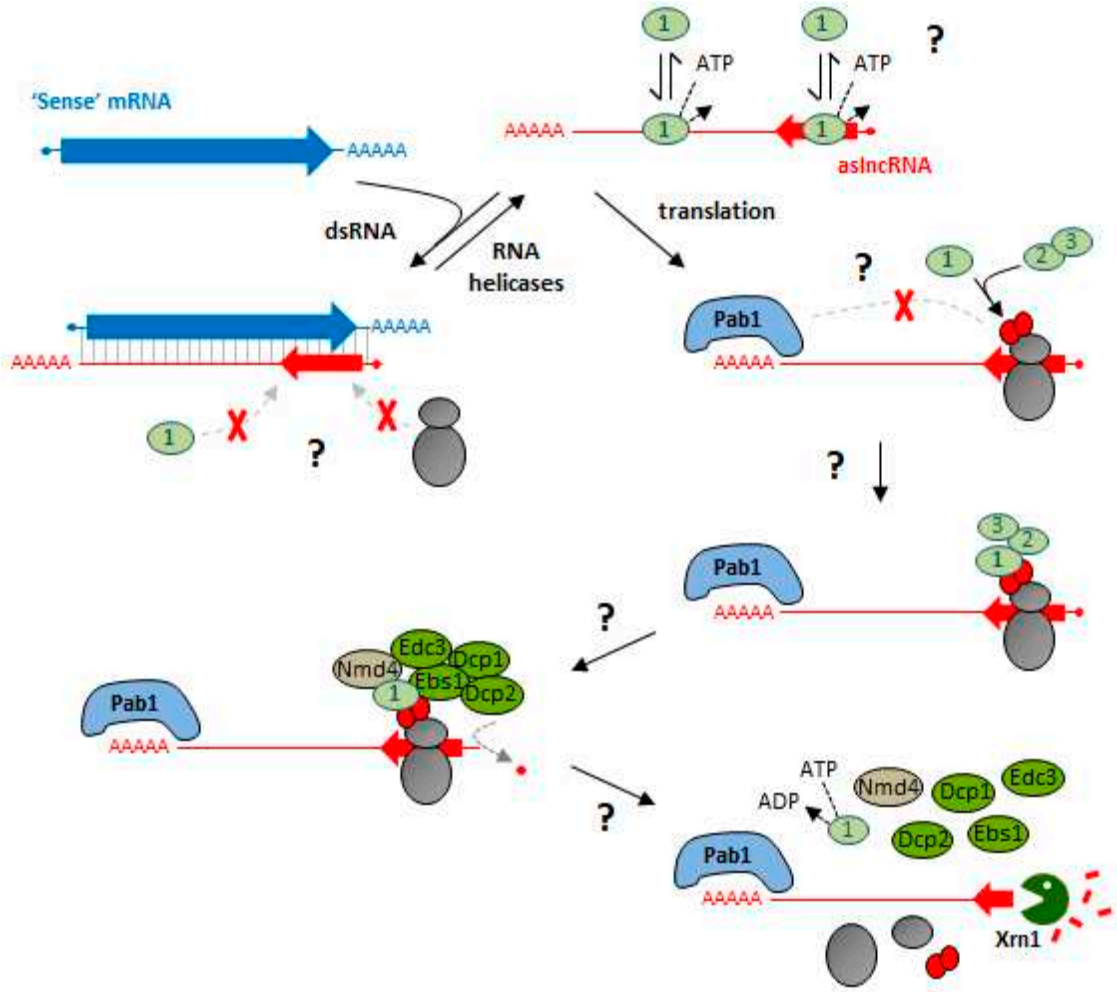


Figure 3. Model of yeast cytoplasmic aslncRNA degradation by NMD. Once in the cytoplasm, an aslncRNA (red) would rapidly be bound by ribosomes, unless in a dsRNA structure with its paired-sense mRNA (blue). This dsRNA might also interfere with Upf1 binding to the aslncRNAs and could be removed by the action of RNA helicases. The detection of a long 3′ UTR would trigger NMD by a mechanism similar to the "faux 3′ UTR". A Upf1-23 complex would form at the level of the termination codon thanks to the interaction between Upf1 and eRF1/eRF3. The subsequent formation of the Upf1-decapping complex would lead to the decapping of the aslncRNA by Dcp2. Upon ATP-dependent disassembly of the complex, the decapped aslncRNA is degraded by Xrn1. The mRNA and the aslncRNA are represented in blue and red, respectively. Large arrows and thin lines represent the coding regions and the UTRs, respectively. The ribosome and NMD/decapping factors are represented as in Figure 1.

5. Functional Importance of NMD and Translation in lncRNA Metabolism

NMD could be seen as an additional pathway contributing to the clearance of unproductive and potentially harmful spurious non-coding transcripts. However, we believe that this view is too reductive and that there might be more behind the involvement of NMD in the metabolism of lncRNAs (Figure 4). For example, as NMD is a cytoplasmic process, it could ensure that regulatory lncRNAs exhibit their functions exclusively in the nucleus by limiting their accumulation outside the nucleus. In addition, NMD could limit the accumulation of nonfunctional and potentially deleterious peptides from cytoplasmic lncRNAs during the de novo gene birth [125,126,143]. Additionally, the peptides produced from NMD-sensitive lncRNAs could be functional and important, despite their low levels. Even if this remains completely speculative for NMD-sensitive lncRNAs, we note that antigens of the MHC class I pathway are produced from PTC-containing mRNA [144], raising the question of how this process might be generalized for "cryptic" lncRNAs.

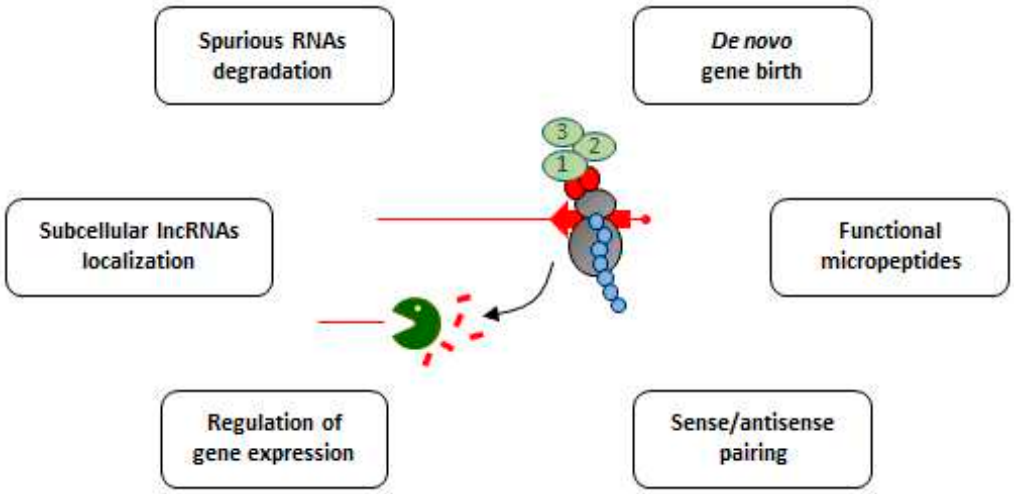


Figure 4. Possible roles of translation and NMD in the metabolism of (as)lncRNAs. Schematic representation of the functional importance of NMD and translation in lncRNAs metabolism (see main text for details). The chain of blue balls represents the micropeptide produced upon the translation of the smORF (red arrow) of the (as)lncRNA. The ribosome, the (as)lncRNA and the NMD/decay factors are represented as above.

NMD could also specifically modulate the levels of regulatory lncRNAs. The apoptotic lncRNA *GAS5* has been proposed to act in an NMD-based circuit, which is critical in response to serum starvation [142]. In normal conditions, NMD restricts the constitutive *GAS5* expression to low levels. However, in stress conditions associated with NMD inhibition (such as serum starvation), *GAS5* expression is up-regulated and binds the glucocorticoid receptor, perturbing its function as a transcription activator in the anti-apoptotic program [142].

More globally, by targeting regulatory cytoplasmic aslncRNAs, NMD could contribute to regulate gene expression. For instance, stabilization of subsets of Xrn1-sensitive aslncRNAs, most of which are NMD-sensitive [124,128], correlates with the transcriptional attenuation of the paired-sense genes, in budding and fission yeasts [137,139]. Interestingly, two independent studies in zebrafish embryos reported that NMD factors cycle to the nucleus to trigger transcriptional adaptation of genes with a sequence complementarity to the PTC-containing RNA in a mechanism called genetic compensation [145,146].

Finally, coupling translation to aslncRNA degradation via NMD could be important for cell recovery upon translation inhibitory stress. In such a condition, NMD-sensitive aslncRNAs are expected to be stabilized and form duplexes with their paired-sense mRNAs. By analogy with the protective effect on the aslncRNA [124], this interaction could also prevent the degradation of the mRNA partner, since local dsRNA formation correlates with higher mRNA stability [147]. After stress, the protected sense/as transcripts would be rapidly released upon the action of RNA helicases, thereby providing a pool of mRNAs in the cytoplasm that can be translated, while NMD-sensitive aslncRNAs would be rapidly degraded.

Together, the observations reported above support the idea that NMD is able to target lncRNAs, and that this might be important for the maintenance of RNA homeostasis, the regulation of gene expression, and for a robust response to several stress conditions. It also challenges the initial assumption that such transcripts are devoid of coding potential.

6. Insight into the Coding Potential of "Non-Coding" Transcripts

The accumulating evidence that cytoplasmic lncRNAs interact with the translation machinery raises the question of their coding potential. Numerous methods have been developed to assess this possibility [148–150]. Additionally, a growing body of experimental data indicate that "non-coding" RNAs can indeed be translated [151], including not only

lncRNAs but also circular RNAs [152–154] and primary microRNAs transcripts [155,156]; moreover, these translation events can produce functional peptides [134,151,157–159].

"Non-coding" RNAs contain one or more smORFs that can be translated into micropeptides (i.e., peptides not exceeding 100 amino acids in length.) Previously, such smORFs were ignored as the traditional gene annotation process filtered out ORFs shorter than 100 codons, considering them as noise or false positives. However, as ribosome profiling techniques and proteomics are growing in popularity and increasing in sensitivity, accuracy and efficiency, it is becoming clear that at least a fraction of short ribosome-bound sequences of (l)ncRNAs represent genuine smORFs.

Importantly, a recent work from Weissman's lab provided a catalog of smORFs and functional peptides derived from human lncRNAs, which included the identification of >800 novel lncRNA-associated smORFs and the observation that, for 91 of them, CRISPR-mediated knockout of the smORF resulted in a growth phenotype [134], indicating that the corresponding peptides are important for cell growth. Other studies previously showed that lncRNA-derived micropeptides are involved in the regulation of RNA decapping [160], in embryonic development [161,162], in muscle development [163–165], regeneration [166,167] or contraction [168–170], and in tumor development [154,171,172] (see Table 1).

Table 1. Examples of functional lncRNA-derived micropeptides.

Micropeptide	Species	Target	Function(s)	Ref.
NoBoDy	Human	mRNA decapping factors	Regulation of mRNA turnover and P-body numbers	[160]
CASIMO1	Human	Squalene epoxidase	Carcinogenesis; cell lipid homeostasis	[171]
PINT87aa	Human	Polymerase associated factor complex (PAF1c)	Oncogene transcriptional inhibition; tumor suppressive effect	[154]
HOXB-AS3	Human	hnRNP A1 splicing factor	Colon cancer growth suppression	[173]
RBRP	Human	m^6A reader IGF2BP1	Regulation of m^6A recognition by IGF2BP1 on c-Myc mRNA; tumorigenesis	[172]
Minion/Myomixer	Human, mouse	Unknown	Myoblast fusion; muscle formation and development	[163,164,167]
SPAR	Human, mouse	Lysosomal v-ATPase	Regulation of mTORC1 signaling pathway; muscle regeneration	[166]
TUG1-BOAT	Human, mouse	Unknown	Unknown; alters mitochondrial membrane potential when overexpressed	[174]
Mtln	Human, mouse	Cardiolipin	Increase of mitochondrial functions	[175]
DWORF	Mouse	SERCA	SERCA (sarcoplasmic reticulum Ca^{2+}-ATPase) activation	[168,169]
MLN	Mouse	SERCA	SERCA inhibition	[176]
Toddler	Zebrafish	Unknown	Promoting cell migration during embryogenesis	[174]
Pri	*Drosophila*	Ubr3 E3 ubiquitin ligase	Proteasome-dependent processing of the developmental Svb transcription factor	[162]
Scl	*Drosophila*	Ca-P60A SERCA	Calcium transport regulation	[170]

Mechanistically, if global information about the mode of action of lncRNA-derived peptides is still lacking, pioneer studies revealed that they can act by binding other proteins and regulate their activity [162,168,169,176], or as signaling pathway molecules [177]. We anticipate that future works will reveal additional modes of action.

In the light of the observations that micropeptides produced from lncRNAs can be biologically important, it is tempting to speculate that aberrant expression of endogenous lncRNA-derived peptides could be associated with diseases, including cancer [154,171]. In addition to providing a new perspective on pathogenicity, lncRNA-derived peptides could also constitute promising targets for targeted therapy [178], including tumor immunotherapy [179]. In this respect, it is interesting to note that a recent characterization of different murine cell lines and cancer patient samples showed that non-coding regions

constitute the major source of tumor specific neo-antigens [179], which could be pivotal for the development of future immunological treatments and cancer vaccines [180].

7. Conclusions

Today, NMD is extending far beyond its original definition assigning it only to the clearance of aberrant "nonsense" transcripts. The current research has revealed that it provides potent means to regulate the expression of many mRNAs and lncRNAs, as well as contributes to the establishment of suitable cellular responses to environmental changes, including adaptation, differentiation or apoptosis. The accumulating biochemical and transcriptomic evidence showing that NMD targets lncRNAs implores us to reconsider the idea that lncRNAs are devoid of coding potential, and challenges us to address how translation of smORFs could not only affect their stability, but also could be used to produce functional micropeptides. Revealing the possibility for a "dark peptidome" to arise from the "dark non-coding side of the genome" (i.e., "the dark side of the dark matter") is one of the challenges in the RNA field for the coming years and will open exciting perspectives regarding the roles of lncRNA-derived peptides.

Author Contributions: Conceptualization, S.A., A.M. and M.W.; writing—original draft preparation, S.A. and M.W.; writing—review and editing, S.A., A.M. and M.W.; funding acquisition, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: Research in Morillon's lab has been supported by the "DNA-life" (ANR-15-CE12-0007) grant from the ANR (Agence Nationale de la Recherche) and the "DARK" consolidator grant from the European Research Council (ERC).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to all members of the lab for discussions and to our colleague Michael Schertzer for his careful reading of this manuscript and English revisions. Sara Andjus is a recipient of a PhD fellowship from PSL university.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wolin, S.L.; Maquat, L.E. Cellular RNA surveillance in health and disease. *Science* **2019**, *366*, 822–827. [CrossRef] [PubMed]
2. Parker, R. RNA degradation in Saccharomyces cerevisae. *Genetics* **2012**, *191*, 671–702. [CrossRef]
3. Isken, O.; Maquat, L.E. Quality control of eukaryotic mRNA: Safeguarding cells from abnormal mRNA function. *Genes Dev.* **2007**, *21*, 1833–1856. [CrossRef] [PubMed]
4. Frischmeyer, P.A.; van Hoof, A.; O'Donnell, K.; Guerrerio, A.L.; Parker, R.; Dietz, H.C. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **2002**, *295*, 2258–2261. [CrossRef]
5. Van Hoof, A.; Frischmeyer, P.A.; Dietz, H.C.; Parker, R. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **2002**, *295*, 2262–2264. [CrossRef] [PubMed]
6. Klauer, A.A.; van Hoof, A. Degradation of mRNAs that lack a stop codon: A decade of nonstop progress. *Wiley Interdiscip. Rev. RNA* **2012**, *3*, 649–660. [CrossRef] [PubMed]
7. Doma, M.K.; Parker, R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **2006**, *440*, 561–564. [CrossRef]
8. Harigaya, Y.; Parker, R. No-go decay: A quality control mechanism for RNA in translation. *Wiley Interdiscip. Rev. RNA* **2010**, *1*, 132–141. [CrossRef] [PubMed]
9. Chang, Y.F.; Imam, J.S.; Wilkinson, M.F. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **2007**, *76*, 51–74. [CrossRef]
10. Behm-Ansmant, I.; Kashima, I.; Rehwinkel, J.; Sauliere, J.; Wittkopp, N.; Izaurralde, E. mRNA quality control: An ancient machinery recognizes and degrades mRNAs with nonsense codons. *FEBS Lett.* **2007**, *581*, 2845–2853. [CrossRef]
11. Muhlrad, D.; Parker, R. Premature translational termination triggers mRNA decapping. *Nature* **1994**, *370*, 578–581. [CrossRef]
12. Muhlrad, D.; Parker, R. Aberrant mRNAs with extended 3′ UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* **1999**, *5*, 1299–1307. [CrossRef] [PubMed]

13. Yepiskoposyan, H.; Aeschimann, F.; Nilsson, D.; Okoniewski, M.; Muhlemann, O. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* **2011**, *17*, 2108–2118. [CrossRef] [PubMed]
14. Celik, A.; Baker, R.; He, F.; Jacobson, A. High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. *RNA* **2017**, *23*, 735–748. [CrossRef]
15. Oliveira, C.C.; McCarthy, J.E. The relationship between eukaryotic translation and mRNA stability. A short upstream open reading frame strongly inhibits translational initiation and greatly accelerates mRNA degradation in the yeast Saccharomyces cerevisiae. *J. Biol. Chem.* **1995**, *270*, 8936–8943. [CrossRef]
16. Ruiz-Echevarria, M.J.; Peltz, S.W. The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. *Cell* **2000**, *101*, 741–751. [CrossRef]
17. Nicholson, P.; Muhlemann, O. Cutting the nonsense: The degradation of PTC-containing mRNAs. *BioChem. Soc. Trans.* **2010**, *38*, 1615–1620. [CrossRef]
18. Rebbapragada, I.; Lykke-Andersen, J. Execution of nonsense-mediated mRNA decay: What defines a substrate? *Curr. Opin. Cell Biol.* **2009**, *21*, 394–402. [CrossRef]
19. Bhuvanagiri, M.; Schlitter, A.M.; Hentze, M.W.; Kulozik, A.E. NMD: RNA biology meets human genetic medicine. *BioChem. J.* **2010**, *430*, 365–377. [CrossRef] [PubMed]
20. Pulak, R.; Anderson, P. mRNA surveillance by the Caenorhabditis elegans smg genes. *Genes Dev.* **1993**, *7*, 1885–1897. [CrossRef]
21. Hall, G.W.; Thein, S. Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: A mechanism for the phenotype of dominant beta-thalassemia. *Blood* **1994**, *83*, 2031–2037. [CrossRef] [PubMed]
22. Zhang, S.; Welch, E.M.; Hogan, K.; Brown, A.H.; Peltz, S.W.; Jacobson, A. Polysome-associated mRNAs are substrates for the nonsense-mediated mRNA decay pathway in Saccharomyces cerevisiae. *RNA* **1997**, *3*, 234–244.
23. Carter, M.S.; Doskow, J.; Morris, P.; Li, S.; Nhim, R.P.; Sandstedt, S.; Wilkinson, M.F. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. *J. Biol. Chem.* **1995**, *270*, 28995–29003. [CrossRef]
24. Thermann, R.; Neu-Yilik, G.; Deters, A.; Frede, U.; Wehr, K.; Hagemeier, C.; Hentze, M.W.; Kulozik, A.E. Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.* **1998**, *17*, 3484–3494. [CrossRef] [PubMed]
25. Causier, B.; Li, Z.; De Smet, R.; Lloyd, J.P.B.; Van de Peer, Y.; Davies, B. Conservation of Nonsense-Mediated mRNA Decay Complex Components Throughout Eukaryotic Evolution. *Sci. Rep.* **2017**, *7*, 16692. [CrossRef]
26. Losson, R.; Lacroute, F. Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 5134–5137. [CrossRef]
27. Maquat, L.E.; Kinniburgh, A.J.; Rachmilewitz, E.A.; Ross, J. Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia. *Cell* **1981**, *27*, 543–553. [CrossRef]
28. Baserga, S.J.; Benz, E.J., Jr. Nonsense mutations in the human beta-globin gene affect mRNA metabolism. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2056–2060. [CrossRef] [PubMed]
29. Culbertson, M.R.; Underbrink, K.M.; Fink, G.R. Frameshift suppression Saccharomyces cerevisiae. II. Genetic properties of group II suppressors. *Genetics* **1980**, *95*, 833–853. [CrossRef]
30. Leeds, P.; Peltz, S.W.; Jacobson, A.; Culbertson, M.R. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev.* **1991**, *5*, 2303–2314. [CrossRef]
31. Leeds, P.; Wood, J.M.; Lee, B.S.; Culbertson, M.R. Gene products that promote mRNA turnover in Saccharomyces cerevisiae. *Mol. Cell Biol.* **1992**, *12*, 2165–2177. [CrossRef] [PubMed]
32. Cui, Y.; Hagan, K.W.; Zhang, S.; Peltz, S.W. Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. *Genes Dev.* **1995**, *9*, 423–436. [CrossRef]
33. Franks, T.M.; Singh, G.; Lykke-Andersen, J. Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense-mediated mRNA decay. *Cell* **2010**, *143*, 938–950. [CrossRef]
34. Weng, Y.; Czaplinski, K.; Peltz, S.W. Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol. Cell Biol.* **1996**, *16*, 5477–5490. [CrossRef] [PubMed]
35. Fiorini, F.; Bagchi, D.; Le Hir, H.; Croquette, V. Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. *Nat. Commun.* **2015**, *6*, 7581. [CrossRef]
36. Chamieh, H.; Ballut, L.; Bonneau, F.; Le Hir, H. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.* **2008**, *15*, 85–93. [CrossRef]
37. Lykke-Andersen, J.; Shu, M.D.; Steitz, J.A. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **2000**, *103*, 1121–1131. [CrossRef]
38. Melero, R.; Buchwald, G.; Castano, R.; Raabe, M.; Gil, D.; Lazaro, M.; Urlaub, H.; Conti, E.; Llorca, O. The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3′ end. *Nat. Struct. Mol. Biol.* **2012**, *19*, 498–505, S491–S492. [CrossRef]
39. Chakrabarti, S.; Jayachandran, U.; Bonneau, F.; Fiorini, F.; Basquin, C.; Domcke, S.; Le Hir, H.; Conti, E. Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell* **2011**, *41*, 693–703. [CrossRef]
40. Aznarez, I.; Nomakuchi, T.T.; Tetenbaum-Novatt, J.; Rahman, M.A.; Fregoso, O.; Rees, H.; Krainer, A.R. Mechanism of Nonsense-Mediated mRNA Decay Stimulation by Splicing Factor SRSF1. *Cell Rep.* **2018**, *23*, 2186–2198. [CrossRef] [PubMed]

41. Gehring, N.H.; Kunz, J.B.; Neu-Yilik, G.; Breit, S.; Viegas, M.H.; Hentze, M.W.; Kulozik, A.E. Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol. Cell* **2005**, *20*, 65–75. [CrossRef] [PubMed]
42. Culbertson, M.R.; Leeds, P.F. Looking at mRNA decay pathways through the window of molecular evolution. *Curr. Opin. Genet. Dev.* **2003**, *13*, 207–214. [CrossRef]
43. Kunz, J.B.; Neu-Yilik, G.; Hentze, M.W.; Kulozik, A.E.; Gehring, N.H. Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. *RNA* **2006**, *12*, 1015–1022. [CrossRef]
44. Yamashita, A.; Ohnishi, T.; Kashima, I.; Taya, Y.; Ohno, S. Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev.* **2001**, *15*, 2215–2228. [CrossRef]
45. Cali, B.M.; Kuchma, S.L.; Latham, J.; Anderson, P. smg-7 is required for mRNA surveillance in Caenorhabditis elegans. *Genetics* **1999**, *151*, 605–616. [CrossRef]
46. Wittkopp, N.; Huntzinger, E.; Weiler, C.; Sauliere, J.; Schmidt, S.; Sonawane, M.; Izaurralde, E. Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. *Mol. Cell Biol.* **2009**, *29*, 3517–3528. [CrossRef] [PubMed]
47. Medghalchi, S.M.; Frischmeyer, P.A.; Mendell, J.T.; Kelly, A.G.; Lawler, A.M.; Dietz, H.C. Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* **2001**, *10*, 99–105. [CrossRef]
48. Metzstein, M.M.; Krasnow, M.A. Functions of the nonsense-mediated mRNA decay pathway in Drosophila development. *PLoS Genet.* **2006**, *2*, e180. [CrossRef]
49. Yoine, M.; Nishii, T.; Nakamura, K. Arabidopsis UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. *Plant. Cell Physiol.* **2006**, *47*, 572–580. [CrossRef] [PubMed]
50. Hodgkin, J.; Papp, A.; Pulak, R.; Ambros, V.; Anderson, P. A new kind of informational suppression in the nematode Caenorhabditis elegans. *Genetics* **1989**, *123*, 301–313. [CrossRef]
51. Nasif, S.; Contu, L.; Muhlemann, O. Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. *Semin. Cell Dev. Biol.* **2018**, *75*, 78–87. [CrossRef] [PubMed]
52. Lykke-Andersen, S.; Jensen, T.H. Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 665–677. [CrossRef]
53. Kurosaki, T.; Popp, M.W.; Maquat, L.E. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 406–420. [CrossRef] [PubMed]
54. Peccarelli, M.; Kebaara, B.W. Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. *Eukaryot. Cell* **2014**, *13*, 1126–1135. [CrossRef]
55. Smith, J.E.; Baker, K.E. Nonsense-mediated RNA decay—A switch and dial for regulating gene expression. *Bioessays* **2015**, *37*, 612–623. [CrossRef] [PubMed]
56. Malabat, C.; Feuerbach, F.; Ma, L.; Saveanu, C.; Jacquier, A. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife* **2015**, *4*, e06722. [CrossRef]
57. Gaba, A.; Jacobson, A.; Sachs, M.S. Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Mol. Cell* **2005**, *20*, 449–460. [CrossRef]
58. Arribere, J.A.; Gilbert, W.V. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res.* **2013**, *23*, 977–987. [CrossRef]
59. Guan, Q.; Zheng, W.; Tang, S.; Liu, X.; Zinkel, R.A.; Tsui, K.W.; Yandell, B.S.; Culbertson, M.R. Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. *PLoS Genet.* **2006**, *2*, e203. [CrossRef] [PubMed]
60. He, F.; Li, X.; Spatrick, P.; Casillo, R.; Dong, S.; Jacobson, A. Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5′ to 3′ mRNA decay pathways in yeast. *Mol. Cell* **2003**, *12*, 1439–1452. [CrossRef]
61. Johansson, M.J.; He, F.; Spatrick, P.; Li, C.; Jacobson, A. Association of yeast Upf1p with direct substrates of the NMD pathway. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20872–20877. [CrossRef]
62. Weischenfeldt, J.; Damgaard, I.; Bryder, D.; Theilgaard-Monch, K.; Thoren, L.A.; Nielsen, F.C.; Jacobsen, S.E.; Nerlov, C.; Porse, B.T. NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev.* **2008**, *22*, 1381–1396. [CrossRef]
63. Thoren, L.A.; Norgaard, G.A.; Weischenfeldt, J.; Waage, J.; Jakobsen, J.S.; Damgaard, I.; Bergstrom, F.C.; Blom, A.M.; Borup, R.; Bisgaard, H.C.; et al. UPF2 is a critical regulator of liver development, function and regeneration. *PLoS ONE* **2010**, *5*, e11650. [CrossRef]
64. Goetz, A.E.; Wilkinson, M. Stress and the nonsense-mediated RNA decay pathway. *Cell Mol. Life Sci.* **2017**, *74*, 3509–3531. [CrossRef]
65. Karam, R.; Lou, C.H.; Kroeger, H.; Huang, L.; Lin, J.H.; Wilkinson, M.F. The unfolded protein response is shaped by the NMD pathway. *EMBO Rep.* **2015**, *16*, 599–609. [CrossRef]
66. Rodriguez-Gabriel, M.A.; Watt, S.; Bahler, J.; Russell, P. Upf1, an RNA helicase required for nonsense-mediated mRNA decay, modulates the transcriptional response to oxidative stress in fission yeast. *Mol. Cell Biol.* **2006**, *26*, 6347–6356. [CrossRef] [PubMed]
67. Gardner, L.B. Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. *Mol. Cell Biol.* **2008**, *28*, 3729–3741. [CrossRef] [PubMed]

68. Mendell, J.T.; Sharifi, N.A.; Meyers, J.L.; Martinez-Murillo, F.; Dietz, H.C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* **2004**, *36*, 1073–1078. [CrossRef] [PubMed]
69. Usuki, F.; Yamashita, A.; Fujimura, M. Environmental stresses suppress nonsense-mediated mRNA decay (NMD) and affect cells by stabilizing NMD-targeted gene expression. *Sci. Rep.* **2019**, *9*, 1279. [CrossRef]
70. Kim, Y.K.; Maquat, L.E. UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. *RNA* **2019**, *25*, 407–422. [CrossRef] [PubMed]
71. Huang, L.; Lou, C.H.; Chan, W.; Shum, E.Y.; Shao, A.; Stone, E.; Karam, R.; Song, H.W.; Wilkinson, M.F. RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. *Mol. Cell* **2011**, *43*, 950–961. [CrossRef]
72. Kerenyi, Z.; Merai, Z.; Hiripi, L.; Benkovics, A.; Gyula, P.; Lacomme, C.; Barta, E.; Nagy, F.; Silhavy, D. Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. *EMBO J.* **2008**, *27*, 1585–1595. [CrossRef]
73. Woodward, L.A.; Mabin, J.W.; Gangras, P.; Singh, G. The exon junction complex: A lifelong guardian of mRNA fate. *Wiley Interdiscip. Rev. RNA* **2017**, *8*. [CrossRef]
74. Le Hir, H.; Gatfield, D.; Izaurralde, E.; Moore, M.J. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **2001**, *20*, 4987–4997. [CrossRef] [PubMed]
75. Sun, X.; Moriarty, P.M.; Maquat, L.E. Nonsense-mediated decay of glutathione peroxidase 1 mRNA in the cytoplasm depends on intron position. *EMBO J.* **2000**, *19*, 4734–4744. [CrossRef]
76. Sun, X.; Maquat, L.E. mRNA surveillance in mammalian cells: The relationship between introns and translation termination. *RNA* **2000**, *6*, 1–8. [CrossRef] [PubMed]
77. Zhang, J.; Sun, X.; Qian, Y.; Maquat, L.E. Intron function in the nonsense-mediated decay of beta-globin mRNA: Indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. *RNA* **1998**, *4*, 801–815. [CrossRef]
78. Le Hir, H.; Izaurralde, E.; Maquat, L.E.; Moore, M.J. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* **2000**, *19*, 6860–6869. [CrossRef]
79. Dostie, J.; Dreyfuss, G. Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr. Biol.* **2002**, *12*, 1060–1067. [CrossRef]
80. Tange, T.O.; Nott, A.; Moore, M.J. The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.* **2004**, *16*, 279–284. [CrossRef] [PubMed]
81. Kurosaki, T.; Maquat, L.E. Rules that govern UPF1 binding to mRNA 3′ UTRs. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 3357–3362. [CrossRef]
82. Kashima, I.; Yamashita, A.; Izumi, N.; Kataoka, N.; Morishita, R.; Hoshino, S.; Ohno, M.; Dreyfuss, G.; Ohno, S. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev.* **2006**, *20*, 355–367. [CrossRef]
83. Eberle, A.B.; Lykke-Andersen, S.; Muhlemann, O.; Jensen, T.H. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat. Struct. Mol. Biol.* **2009**, *16*, 49–55. [CrossRef]
84. Ohnishi, T.; Yamashita, A.; Kashima, I.; Schell, T.; Anders, K.R.; Grimson, A.; Hachiya, T.; Hentze, M.W.; Anderson, P.; Ohno, S. Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. *Mol. Cell* **2003**, *12*, 1187–1200. [CrossRef]
85. Isken, O.; Kim, Y.K.; Hosoda, N.; Mayeur, G.L.; Hershey, J.W.; Maquat, L.E. Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell* **2008**, *133*, 314–327. [CrossRef] [PubMed]
86. Metze, S.; Herzog, V.A.; Ruepp, M.D.; Muhlemann, O. Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. *RNA* **2013**, *19*, 1432–1448. [CrossRef] [PubMed]
87. Buhler, M.; Steiner, S.; Mohn, F.; Paillusson, A.; Muhlemann, O. EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3′ UTR length. *Nat. Struct. Mol. Biol.* **2006**, *13*, 462–464. [CrossRef] [PubMed]
88. Wen, J.; Brogna, S. Splicing-dependent NMD does not require the EJC in Schizosaccharomyces pombe. *EMBO J.* **2010**, *29*, 1537–1551. [CrossRef] [PubMed]
89. Longman, D.; Plasterk, R.H.; Johnstone, I.L.; Caceres, J.F. Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. *Genes Dev.* **2007**, *21*, 1075–1085. [CrossRef]
90. Gatfield, D.; Unterholzner, L.; Ciccarelli, F.D.; Bork, P.; Izaurralde, E. Nonsense-mediated mRNA decay in Drosophila: At the intersection of the yeast and mammalian pathways. *EMBO J.* **2003**, *22*, 3960–3970. [CrossRef] [PubMed]
91. Goffeau, A.; Barrell, B.G.; Bussey, H.; Davis, R.W.; Dujon, B.; Feldmann, H.; Galibert, F.; Hoheisel, J.D.; Jacq, C.; Johnston, M.; et al. Life with 6000 genes. *Science* **1996**, *274*, 546–567. [CrossRef]
92. Alexandrov, A.; Colognori, D.; Steitz, J.A. Human eIF4AIII interacts with an eIF4G-like partner, NOM1, revealing an evolutionarily conserved function outside the exon junction complex. *Genes Dev.* **2011**, *25*, 1078–1090. [CrossRef]
93. Eberle, A.B.; Stalder, L.; Mathys, H.; Orozco, R.Z.; Muhlemann, O. Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. *PLoS Biol.* **2008**, *6*, e92. [CrossRef]
94. Matsuda, D.; Hosoda, N.; Kim, Y.K.; Maquat, L.E. Failsafe nonsense-mediated mRNA decay does not detectably target eIF4E-bound mRNA. *Nat. Struct. Mol. Biol.* **2007**, *14*, 974–979. [CrossRef] [PubMed]
95. Singh, G.; Rebbapragada, I.; Lykke-Andersen, J. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol.* **2008**, *6*, e111. [CrossRef]

96. Wang, J.; Gudikote, J.P.; Olivas, O.R.; Wilkinson, M.F. Boundary-independent polar nonsense-mediated decay. *EMBO Rep.* **2002**, *3*, 274–279. [CrossRef]
97. Amrani, N.; Ganesan, R.; Kervestin, S.; Mangus, D.A.; Ghosh, S.; Jacobson, A. A faux 3′-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **2004**, *432*, 112–118. [CrossRef] [PubMed]
98. Dehecq, M.; Decourty, L.; Namane, A.; Proux, C.; Kanaan, J.; Le Hir, H.; Jacquier, A.; Saveanu, C. Nonsense-mediated mRNA decay involves two distinct Upf1-bound complexes. *EMBO J.* **2018**, *37*. [CrossRef] [PubMed]
99. Luke, B.; Azzalin, C.M.; Hug, N.; Deplazes, A.; Peter, M.; Lingner, J. Saccharomyces cerevisiae Ebs1p is a putative ortholog of human Smg7 and promotes nonsense-mediated mRNA decay. *Nucleic Acids Res.* **2007**, *35*, 7688–7697. [CrossRef]
100. Ivanov, A.; Mikhailova, T.; Eliseev, B.; Yeramala, L.; Sokolova, E.; Susorov, D.; Shuvalov, A.; Schaffitzel, C.; Alkalaeva, E. PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic Acids Res.* **2016**, *44*, 7766–7776. [CrossRef]
101. Behm-Ansmant, I.; Gatfield, D.; Rehwinkel, J.; Hilgers, V.; Izaurralde, E. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J.* **2007**, *26*, 1591–1601. [CrossRef]
102. Silva, A.L.; Ribeiro, P.; Inacio, A.; Liebhaber, S.A.; Romao, L. Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. *RNA* **2008**, *14*, 563–576. [CrossRef]
103. Meaux, S.; van Hoof, A.; Baker, K.E. Nonsense-mediated mRNA decay in yeast does not require PAB1 or a poly(A) tail. *Mol. Cell* **2008**, *29*, 134–140. [CrossRef]
104. Roque, S.; Cerciat, M.; Gaugue, I.; Mora, L.; Floch, A.G.; de Zamaroczy, M.; Heurgue-Hamard, V.; Kervestin, S. Interaction between the poly(A)-binding protein Pab1 and the eukaryotic release factor eRF3 regulates translation termination but not mRNA decay in Saccharomyces cerevisiae. *RNA* **2015**, *21*, 124–134. [CrossRef]
105. Johns, L.; Grimson, A.; Kuchma, S.L.; Newman, C.L.; Anderson, P. Caenorhabditis elegans SMG-2 selectively marks mRNAs containing premature translation termination codons. *Mol. Cell Biol.* **2007**, *27*, 5630–5638. [CrossRef]
106. Lee, S.R.; Pratt, G.A.; Martinez, F.J.; Yeo, G.W.; Lykke-Andersen, J. Target Discrimination in Nonsense-Mediated mRNA Decay Requires Upf1 ATPase Activity. *Mol. Cell* **2015**, *59*, 413–425. [CrossRef] [PubMed]
107. Hwang, J.; Sato, H.; Tang, Y.; Matsuda, D.; Maquat, L.E. UPF1 association with the cap-binding protein, CBP80, promotes nonsense-mediated mRNA decay at two distinct steps. *Mol. Cell* **2010**, *39*, 396–409. [CrossRef] [PubMed]
108. Hogg, J.R.; Goff, S.P. Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell* **2010**, *143*, 379–389. [CrossRef] [PubMed]
109. Gregersen, L.H.; Schueler, M.; Munschauer, M.; Mastrobuoni, G.; Chen, W.; Kempa, S.; Dieterich, C.; Landthaler, M. MOV10 Is a 5′ to 3′ RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3′ UTRs. *Mol. Cell* **2014**, *54*, 573–585. [CrossRef]
110. Hurt, J.A.; Robertson, A.D.; Burge, C.B. Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Res.* **2013**, *23*, 1636–1650. [CrossRef]
111. Zund, D.; Gruber, A.R.; Zavolan, M.; Muhlemann, O. Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3′ UTRs. *Nat. Struct. Mol. Biol.* **2013**, *20*, 936–943. [CrossRef]
112. Kurosaki, T.; Li, W.; Hoque, M.; Popp, M.W.; Ermolenko, D.N.; Tian, B.; Maquat, L.E. A post-translational regulatory switch on UPF1 controls targeted mRNA degradation. *Genes Dev.* **2014**, *28*, 1900–1916. [CrossRef]
113. Serdar, L.D.; Whiteside, D.L.; Baker, K.E. ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. *Nat. Commun.* **2016**, *7*, 14021. [CrossRef]
114. Serdar, L.D.; Whiteside, D.L.; Nock, S.L.; McGrath, D.; Baker, K.E. Inhibition of post-termination ribosome recycling at premature termination codons in UPF1 ATPase mutants. *eLife* **2020**, *9*. [CrossRef]
115. Sohrabi-Jahromi, S.; Hofmann, K.B.; Boltendahl, A.; Roth, C.; Gressel, S.; Baejen, C.; Soeding, J.; Cramer, P. Transcriptome maps of general eukaryotic RNA degradation factors. *eLife* **2019**, *8*. [CrossRef]
116. Colombo, M.; Karousis, E.D.; Bourquin, J.; Bruggmann, R.; Muhlemann, O. Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA* **2017**, *23*, 189–201. [CrossRef]
117. Jarroux, J.; Morillon, A.; Pinskaya, M. History, Discovery, and Classification of lncRNAs. *Adv. Exp. Med. Biol.* **2017**, *1008*, 1–46. [CrossRef]
118. Yao, R.W.; Wang, Y.; Chen, L.L. Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* **2019**, *21*, 542–551. [CrossRef] [PubMed]
119. Statello, L.; Guo, C.J.; Chen, L.L.; Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 96–118. [CrossRef] [PubMed]
120. Wery, M.; Kwapisz, M.; Morillon, A. Noncoding RNAs in gene regulation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2011**, *3*, 728–738. [CrossRef] [PubMed]
121. Verheggen, K.; Volders, P.J.; Mestdagh, P.; Menschaert, G.; Van Damme, P.; Gevaert, K.; Martens, L.; Vandesompele, J. Noncoding after All: Biases in Proteomics Data Do Not Explain Observed Absence of lncRNA Translation Products. *J. Proteome Res.* **2017**, *16*, 2508–2515. [CrossRef]
122. Brar, G.A.; Yassour, M.; Friedman, N.; Regev, A.; Ingolia, N.T.; Weissman, J.S. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **2012**, *335*, 552–557. [CrossRef]
123. Smith, J.E.; Alvarez-Dominguez, J.R.; Kline, N.; Huynh, N.J.; Geisler, S.; Hu, W.; Coller, J.; Baker, K.E. Translation of small open reading frames within unannotated RNA transcripts in Saccharomyces cerevisiae. *Cell Rep.* **2014**, *7*, 1858–1866. [CrossRef]

124. Wery, M.; Descrimes, M.; Vogt, N.; Dallongeville, A.S.; Gautheret, D.; Morillon, A. Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. *Mol. Cell* **2016**, *61*, 379–392. [CrossRef]
125. 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. Proto-genes and de novo gene birth. *Nature* **2012**, *487*, 370–374. [CrossRef] [PubMed]
126. Wilson, B.A.; Masel, J. Putatively noncoding transcripts show extensive association with ribosomes. *Genome Biol. Evol.* **2011**, *3*, 1245–1252. [CrossRef]
127. Duncan, C.D.; Mata, J. The translational landscape of fission-yeast meiosis and sporulation. *Nat. Struct. Mol. Biol.* **2014**, *21*, 641–647. [CrossRef]
128. Atkinson, S.R.; Marguerat, S.; Bitton, D.A.; Rodriguez-Lopez, M.; Rallis, C.; Lemay, J.F.; Cotobal, C.; Malecki, M.; Smialowski, P.; Mata, J.; et al. Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. *RNA* **2018**, *24*, 1195–1213. [CrossRef] [PubMed]
129. Ruiz-Orera, J.; Messeguer, X.; Subirana, J.A.; Alba, M.M. Long non-coding RNAs as a source of new peptides. *eLife* **2014**, *3*, e03523. [CrossRef] [PubMed]
130. Aspden, J.L.; Eyre-Walker, Y.C.; Phillips, R.J.; Amin, U.; Mumtaz, M.A.; Brocard, M.; Couso, J.P. Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *eLife* **2014**, *3*, e03528. [CrossRef] [PubMed]
131. Bazzini, A.A.; Johnstone, T.G.; Christiano, R.; Mackowiak, S.D.; Obermayer, B.; Fleming, E.S.; Vejnar, C.E.; Lee, M.T.; Rajewsky, N.; Walther, T.C.; et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* **2014**, *33*, 981–993. [CrossRef]
132. Chew, G.L.; Pauli, A.; Rinn, J.L.; Regev, A.; Schier, A.F.; Valen, E. Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. *Development* **2013**, *140*, 2828–2834. [CrossRef]
133. Ingolia, N.T.; Lareau, L.F.; Weissman, J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **2011**, *147*, 789–802. [CrossRef]
134. 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.; et al. Pervasive functional translation of noncanonical human open reading frames. *Science* **2020**, *367*, 1140–1146. [CrossRef]
135. Van Heesch, S.; van Iterson, M.; Jacobi, J.; Boymans, S.; Essers, P.B.; de Bruijn, E.; Hao, W.; MacInnes, A.W.; Cuppen, E.; Simonis, M. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* **2014**, *15*, R6. [CrossRef] [PubMed]
136. Carlevaro-Fita, J.; Rahim, A.; Guigo, R.; Vardy, L.A.; Johnson, R. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* **2016**, *22*, 867–882. [CrossRef] [PubMed]
137. Wery, M.; Gautier, C.; Descrimes, M.; Yoda, M.; Migeot, V.; Hermand, D.; Morillon, A. Bases of antisense lncRNA-associated regulation of gene expression in fission yeast. *PLoS Genet.* **2018**, *14*, e1007465. [CrossRef] [PubMed]
138. Szachnowski, U.; Andjus, S.; Foretek, D.; Morillon, A.; Wery, M. Endogenous RNAi pathway evolutionarily shapes the destiny of the antisense lncRNAs transcriptome. *Life Sci. Alliance* **2019**, *2*, e201900407. [CrossRef]
139. 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. XUTs are a class of Xrn1-sensitive antisense regulatory non coding RNA in yeast. *Nature* **2011**, *475*, 114–117. [CrossRef]
140. Wery, M.; Gautier, C.; Descrimes, M.; Yoda, M.; Vennin-Rendos, H.; Migeot, V.; Gautheret, D.; Hermand, D.; Morillon, A. Native elongating transcript sequencing reveals global anti-correlation between sense and antisense nascent transcription in fission yeast. *RNA* **2018**, *24*, 196–208. [CrossRef]
141. Kurihara, Y.; Matsui, A.; Hanada, K.; Kawashima, M.; Ishida, J.; Morosawa, T.; Tanaka, M.; Kaminuma, E.; Mochizuki, Y.; Matsushima, A.; et al. Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2453–2458. [CrossRef]
142. Tani, H.; Torimura, M.; Akimitsu, N. The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. *PLoS ONE* **2013**, *8*, e55684. [CrossRef]
143. Blevins, W.R.; Ruiz-Orera, J.; Messeguer, X.; Blasco-Moreno, B.; Villanueva-Canas, J.L.; Espinar, L.; Diez, J.; Carey, L.B.; Alba, M.M. Uncovering de novo gene birth in yeast using deep transcriptomics. *Nat. Commun.* **2021**, *12*, 604. [CrossRef] [PubMed]
144. Apcher, S.; Daskalogianni, C.; Lejeune, F.; Manoury, B.; Imhoos, G.; Heslop, L.; Fahraeus, R. Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11572–11577. [CrossRef] [PubMed]
145. Ma, Z.; Zhu, P.; Shi, H.; Guo, L.; Zhang, Q.; Chen, Y.; Chen, S.; Zhang, Z.; Peng, J.; Chen, J. PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature* **2019**, *568*, 259–263. [CrossRef]
146. El-Brolosy, M.A.; Kontarakis, Z.; Rossi, A.; Kuenne, C.; Gunther, S.; Fukuda, N.; Kikhi, K.; Boezio, G.L.M.; Takacs, C.M.; Lai, S.L.; et al. Genetic compensation triggered by mutant mRNA degradation. *Nature* **2019**, *568*, 193–197. [CrossRef]
147. Geisberg, J.V.; Moqtaderi, Z.; Fan, X.; Ozsolak, F.; Struhl, K. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* **2014**, *156*, 812–824. [CrossRef] [PubMed]
148. Mishra, S.K.; Wang, H. Computational Analysis Predicts Hundreds of Coding lncRNAs in Zebrafish. *Biology* **2021**, *10*, 371. [CrossRef]
149. Makarewich, C.A.; Olson, E.N. Mining for Micropeptides. *Trends Cell Biol.* **2017**, *27*, 685–696. [CrossRef]

150. Choi, S.W.; Kim, H.W.; Nam, J.W. The small peptide world in long noncoding RNAs. *Brief. Bioinform.* **2019**, *20*, 1853–1864. [CrossRef]
151. Wei, L.H.; Guo, J.U. Coding functions of "noncoding" RNAs. *Science* **2020**, *367*, 1074–1075. [CrossRef]
152. Legnini, I.; Di Timoteo, G.; Rossi, F.; Morlando, M.; Briganti, F.; Sthandier, O.; Fatica, A.; Santini, T.; Andronache, A.; Wade, M.; et al. Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* **2017**, *66*, 22–37.e9. [CrossRef]
153. Yang, Y.; Gao, X.; Zhang, M.; Yan, S.; Sun, C.; Xiao, F.; Huang, N.; Yang, X.; Zhao, K.; Zhou, H.; et al. Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J. Natl. Cancer Inst.* **2018**, *110*. [CrossRef] [PubMed]
154. Zhang, M.; Zhao, K.; Xu, X.; Yang, Y.; Yan, S.; Wei, P.; Liu, H.; Xu, J.; Xiao, F.; Zhou, H.; et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat. Commun.* **2018**, *9*, 4475. [CrossRef]
155. Montigny, A.; Tavormina, P.; Duboe, C.; San Clemente, H.; Aguilar, M.; Valenti, P.; Lauressergues, D.; Combier, J.P.; Plaza, S. Drosophila primary microRNA-8 encodes a microRNA-encoded peptide acting in parallel of miR-8. *Genome Biol.* **2021**, *22*, 118. [CrossRef] [PubMed]
156. Lauressergues, D.; Couzigou, J.M.; Clemente, H.S.; Martinez, Y.; Dunand, C.; Becard, G.; Combier, J.P. Primary transcripts of microRNAs encode regulatory peptides. *Nature* **2015**, *520*, 90–93. [CrossRef] [PubMed]
157. 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.; Saghatelian, A. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.* **2013**, *9*, 59–64. [CrossRef] [PubMed]
158. Matsumoto, A.; Nakayama, K.I. Hidden Peptides Encoded by Putative Noncoding RNAs. *Cell Struct. Funct.* **2018**, *43*, 75–83. [CrossRef] [PubMed]
159. 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. The Translational Landscape of the Human Heart. *Cell* **2019**, *178*, 242–260.e29. [CrossRef] [PubMed]
160. D'Lima, N.G.; Ma, J.; Winkler, L.; Chu, Q.; Loh, K.H.; Corpuz, E.O.; Budnik, B.A.; Lykke-Andersen, J.; Saghatelian, A.; Slavoff, S.A. A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* **2017**, *13*, 174–180. [CrossRef] [PubMed]
161. Kondo, T.; Plaza, S.; Zanet, J.; Benrabah, E.; Valenti, P.; Hashimoto, Y.; Kobayashi, S.; Payre, F.; Kageyama, Y. Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. *Science* **2010**, *329*, 336–339. [CrossRef]
162. Zanet, J.; Benrabah, E.; Li, T.; Pelissier-Monier, A.; Chanut-Delalande, H.; Ronsin, B.; Bellen, H.J.; Payre, F.; Plaza, S. Pri sORF peptides induce selective proteasome-mediated protein processing. *Science* **2015**, *349*, 1356–1358. [CrossRef] [PubMed]
163. Bi, P.; Ramirez-Martinez, A.; Li, H.; Cannavino, J.; McAnally, J.R.; Shelton, J.M.; Sanchez-Ortiz, E.; Bassel-Duby, R.; Olson, E.N. Control of muscle formation by the fusogenic micropeptide myomixer. *Science* **2017**, *356*, 323–327. [CrossRef] [PubMed]
164. Zhang, Q.; Vashisht, A.A.; O'Rourke, J.; Corbel, S.Y.; Moran, R.; Romero, A.; Miraglia, L.; Zhang, J.; Durrant, E.; Schmedt, C.; et al. The microprotein Minion controls cell fusion and muscle formation. *Nat. Commun.* **2017**, *8*, 15664. [CrossRef] [PubMed]
165. Lin, Y.F.; Xiao, M.H.; Chen, H.X.; Meng, Y.; Zhao, N.; Yang, L.; Tang, H.; Wang, J.L.; Liu, X.; Zhu, Y.; et al. A novel mitochondrial micropeptide MPM enhances mitochondrial respiratory activity and promotes myogenic differentiation. *Cell Death Dis.* **2019**, *10*, 528. [CrossRef]
166. Matsumoto, A.; Pasut, A.; Matsumoto, M.; Yamashita, R.; Fung, J.; Monteleone, E.; Saghatelian, A.; Nakayama, K.I.; Clohessy, J.G.; Pandolfi, P.P. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* **2017**, *541*, 228–232. [CrossRef]
167. Bi, P.; McAnally, J.R.; Shelton, J.M.; Sanchez-Ortiz, E.; Bassel-Duby, R.; Olson, E.N. Fusogenic micropeptide Myomixer is essential for satellite cell fusion and muscle regeneration. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3864–3869. [CrossRef] [PubMed]
168. Nelson, B.R.; Makarewich, C.A.; Anderson, D.M.; Winders, B.R.; Troupes, C.D.; Wu, F.; Reese, A.L.; McAnally, J.R.; Chen, X.; Kavalali, E.T.; et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* **2016**, *351*, 271–275. [CrossRef]
169. Makarewich, C.A.; Munir, A.Z.; Schiattarella, G.G.; Bezprozvannaya, S.; Raguimova, O.N.; Cho, E.E.; Vidal, A.H.; Robia, S.L.; Bassel-Duby, R.; Olson, E.N. The DWORF micropeptide enhances contractility and prevents heart failure in a mouse model of dilated cardiomyopathy. *eLife* **2018**, *7*. [CrossRef]
170. Magny, E.G.; Pueyo, J.I.; Pearl, F.M.; Cespedes, M.A.; Niven, J.E.; Bishop, S.A.; Couso, J.P. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **2013**, *341*, 1116–1120. [CrossRef]
171. Polycarpou-Schwarz, M.; Gross, M.; Mestdagh, P.; Schott, J.; Grund, S.E.; Hildenbrand, C.; Rom, J.; Aulmann, S.; Sinn, H.P.; Vandesompele, J.; et al. The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation. *Oncogene* **2018**, *37*, 4750–4768. [CrossRef]
172. Zhu, S.; Wang, J.Z.; Chen, D.; He, Y.T.; Meng, N.; Chen, M.; Lu, R.X.; Chen, X.H.; Zhang, X.L.; Yan, G.R. An oncopeptide regulates m(6)A recognition by the m(6)A reader IGF2BP1 and tumorigenesis. *Nat. Commun.* **2020**, *11*, 1685. [CrossRef]
173. Huang, J.Z.; Chen, M.; Chen, D.E.; Gao, X.C.; Zhu, S.; Huang, H.; Hu, M.; Zhu, H.; Yan, G.R. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol. Cell* **2017**, *68*, 171–184.e6. [CrossRef]
174. Lewandowski, J.P.; Dumbovic, G.; Watson, A.R.; Hwang, T.; Jacobs-Palmer, E.; Chang, N.; Much, C.; Turner, K.M.; Kirby, C.; Rubinstein, N.D.; et al. The Tug1 lncRNA locus is essential for male fertility. *Genome Biol.* **2020**, *21*, 237. [CrossRef] [PubMed]

175. Stein, C.S.; Jadiya, P.; Zhang, X.; McLendon, J.M.; Abouassaly, G.M.; Witmer, N.H.; Anderson, E.J.; Elrod, J.W.; Boudreau, R.L. Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. *Cell Rep.* **2018**, *23*, 3710–3720.e3718. [CrossRef]
176. Anderson, D.M.; Anderson, K.M.; Chang, C.L.; Makarewich, C.A.; Nelson, B.R.; McAnally, J.R.; Kasaragod, P.; Shelton, J.M.; Liou, J.; Bassel-Duby, R.; et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **2015**, *160*, 595–606. [CrossRef]
177. Pauli, A.; Norris, M.L.; Valen, E.; Chew, G.L.; Gagnon, J.A.; Zimmerman, S.; Mitchell, A.; Ma, J.; Dubrulle, J.; Reyon, D.; et al. Toddler: An embryonic signal that promotes cell movement via Apelin receptors. *Science* **2014**, *343*, 1248636. [CrossRef] [PubMed]
178. Zhu, S.; Wang, J.; He, Y.; Meng, N.; Yan, G.R. Peptides/Proteins Encoded by Non-coding RNA: A Novel Resource Bank for Drug Targets and Biomarkers. *Front. Pharm.* **2018**, *9*, 1295. [CrossRef]
179. Laumont, C.M.; Vincent, K.; Hesnard, L.; Audemard, E.; Bonneil, E.; Laverdure, J.P.; Gendron, P.; Courcelles, M.; Hardy, M.P.; Cote, C.; et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med.* **2018**, *10*. [CrossRef] [PubMed]
180. Lancaster, E.M.; Jablons, D.; Kratz, J.R. Applications of Next-Generation Sequencing in Neoantigen Prediction and Cancer Vaccine Development. *Genet. Test. Mol. Biomark.* **2020**, *24*, 59–66. [CrossRef] [PubMed]

RÉSUMÉ

La transcription des génomes eucaryotes ne se limite pas aux gènes codant pour des protéines. Au contraire, ces génomes sont transcrits de manière ‘pervasive’, ce qui est source de milliers de longs ARN non codants (nc). Certains sont anormalement exprimés dans certains cancers, ce qui en fait des candidats pour des biomarqueurs. Bien que définis comme ‘non codants’, ces ARN peuvent être traduits et donner lieu à des micropeptides pouvant affecter certaines étapes de la progression tumorale. Cependant, l'importance de la traduction de ces ARN est mal comprise. Chez la levure, ces ARN sont peu abondants car ciblés par une voie dégradation qui dépend de la traduction. En étudiant les mécanismes régulateurs de leur dégradation, nous démontrons que ces longs ARNnc peuvent être exploités pour fabriquer des peptides, alors qu’ils sont ciblés vers la dégradation.

MOTS CLÉS

ARNlnc, Xrn1, NMD, traduction, dsRNA, levure

ABSTRACT

Transcription in Eukaryotes is not limited to protein-coding genes. Rather, eukaryotic genomes are pervasively transcribed, producing thousands of long non-coding (lnc)RNAs. The expression of most mammalian lncRNAs in cell or tissue-specific. Abnormal lncRNA expression has been associated to cancer, making them excellent candidates for biomarkers and diagnostic tools. Human lncRNAs, despite defined as non-coding, encode peptides that can affect tumor phenotypes. However, the importance of lncRNA translation is poorly understood. In yeast, many lncRNAs are unstable due to translation-dependent degradation, suggesting they are translated. By investigating the molecular mechanisms of their decay, our data demonstrated that lncRNAs, prior to decay, can be exploited to make peptides, potentially functional.

KEYWORDS

lncRNA, Xrn1, NMD, translation, dsRNA, yeast