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

Ugo Szachnowski[1], Sara Andus[1], Dominika Foretek[1], Antonin Morillon[1,2] and Maxime Wery[1,2].

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

[2] Corresponding authors

Running title: aslncRNAs in RNAi-capable budding yeast

Keywords: antisense lncRNA/Xrn1/Rrp6/Dicer/*N. castellii*

ABSTRACT

Antisense (as)lncRNAs are extensively degraded by the nuclear exosome and the cytoplasmic exoribonuclease Xrn1 in the budding yeast *Saccharomyces cerevisiae*, lacking RNA interference (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 lncRNAs (XUTs) levels in Xrn1-deficient cells. The *dcr1* and *xrn1* mutants display synergic growth defects, indicating that Dicer becomes critical in absence of Xrn1. Small RNA sequencing showed that Dicer processes aslncRNAs into small RNAs, with a preference for asXUTs. Consistently, Dicer localizes into the cytoplasm. Finally, we observed an expansion of the exosome-sensitive antisense transcriptome in *N. castellii* compared to *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.

INTRODUCTION

Initially considered as by-products of the pervasive transcription of eukaryotic genomes, long non-coding (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 and 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 and Chang 2016; Renganathan and Felley-Bosco 2017; Saha et al. 2017; Schmitt and 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 and Steinmetz 2013) and have attracted a lot of attention given their potential to regulate gene expression (Kopp and 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; Uhler et al. 2007; Berretta et al. 2008; Houseley et al. 2008; Camblong et al. 2009; 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 and 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 non-essential 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, the majority of CUTs and XUTs are antisense to protein-coding genes, at least partially.

This classification into CUTs or XUTs is interesting 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 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 RNA interference (RNAi) system during evolution, so it lacks the ribonuclease III Dicer that can process double-stranded (ds)RNA structures into small interfering (si)RNAs (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; Drinnenberg et al. 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 and 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 and 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 and 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 to *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 of 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, immuno-fluorescence 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 (Figure 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; Wery et al. 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 and Tsukiyama 2016). From all the ≥200 nt segments not overlapping ORF, 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 (Figure 1A; see details in Methods), we identified 146 stable unannotated transcripts (SUTs), 10 DUTs,1021 XUTs and 1280 CUTs (Figure 1B; see also Figure S1A-C).

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

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

Most of the transcripts we identified are novel (Figure 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 (Figure S1G). Interestingly, we observed that the solo lncRNAs (*ie* 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 (Figure 1F), but also for XUTs and CUTs in *xrn1Δ* and *rrp6Δ* cells, respectively (Figure S1H-I).

Overall, we annotated 2247 lncRNAs in *N. castellii*, 1583 of which are antisense to protein-coding genes. The vast majority of them are unstable and are primarily degraded by the nuclear exosome (CUTs) and/or Xrn1 (XUTs), with almost no effect of Dcr1. Figure S2A-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 Figure 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 of asXUTs levels compared to the single *xrn1Δ* mutant (Figure 2A; $P = 1.77e^{-5}$, Wilcoxon rank-sum test; see illustrative examples in Figure S2A-D), with no effect on the solo XUTs (Figure 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 (Figure 2B; $P = 0.513$ and 0.991, respectively; Wilcoxon rank-sum test).

The marginal effect of Dcr1 inactivation on the coding and non-coding transcriptomes (Figure 1B-C; see also Figure S2E) is consistent with the normal growth of the *dcr1Δ* mutant, which is undistinguishable from the WT strain (Figure 2C; see also Figure 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 (Figure 2C; see also Figure S2F). This effect is even stronger at higher (32°C) or lower (18°C) temperatures, or when cells are grown on synthetic medium (Figure 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 (Figure S3A), which corresponds to the expected size 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 RNAs densities are higher for the antisense SUTs, CUTs and XUTs in comparison to their solo counterparts, especially in the *xrn1Δ* context (Figure 3A; see also Figure 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, 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 (Figure 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 of small RNAs densities in *xrn1Δ* (Figure 3B; see also Figure S3E). In contrast, for the *CUT0672/C05770* and *CUT0275/A12440* pairs, the levels of 22-23 nt small RNAs in WT cells remain low (Figure 3B; see also Figure 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. Since asXUTs (*ie* 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 and Houseley 2014).

We constructed a Dcr1-GFP strain in *N. castellii*. Upon direct visualization in living cells, Dcr1-GFP appeared as individual discrete foci (Figure S4A). When detected using GFP nanobody by immuno-fluorescence in fixed cells, these foci were found in the cytoplasm (Figure 4). Importantly, small RNA sequencing showed that the Dcr1-GFP fusion remains functional for the production of 22-23 nt small RNAs (Figure S4B).

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 and 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 to the set of aslncRNAs in *S. cerevisiae* (Alcid and Tsukiyama 2016). As we considerably extended the repertoire of aslncRNAs in *N. castelli*, most of them being unstable due to 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 than 200 nt (the threshold commonly used to define lncRNAs). We decided to remove all these <200 nt CUTs from our analysis, in order 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 to *S. cerevisiae* (Figure 5A; median = 444 nt and 465 nt, respectively; $P = 9.328e^{-3}$, Wilcoxon rank-sum test). The size of asXUTs is also reduced in *N. castellii* (Figure 5B; median = 670 nt *vs* 709 nt in *S. cerevisiae*), but the difference is not significant ($P = 0.3473$, Wilcoxon rank-sum test). Surprisingly, the aslncRNAs annotated in this work are globally larger than the 170 previously annotated aslncRNAs (see Figure S5A). As a possible explanation of this discrepancy, we observe that 54/170 (32%) of the previously annotated aslncRNAs are shorter than the commonly used 200 nt threshold (Figure S5).

Independently of the size of the aslncRNA, the degree of overlap with the paired-sense mRNA is likely to be 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 (median length of the overlap = 357 bp and 370 bp, respectively; see Figure 5C). In contrast, the overlap between asXUTs and their paired-sense genes is significantly reduced in *N. castellii* (median = 400 bp, *vs* 462 bp in *S. cerevisiae*; see Figure 5D).

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, it is reduced in *N. castellii* in comparison to *S. cerevisiae* (8.1% *vs* 12.9%, respectively). 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, while in *S. cerevisiae*, it is mainly covered by asXUTs (Figure 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.

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), while 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 and 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*, *ie* 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 2247 lncRNAs, including 1583 aslncRNAs. Most of them are unstable and primarily degraded by the nuclear exosome (1280 CUTs) and/or Xrn1 (1021 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 (Figure S1D). This is marginal in comparison to the 1392 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. Firstly, asXUTs global levels significantly increase in the *xrn1Δ dcr1Δ* mutant, compared to the *xrn1Δ* single mutant (Figure 2A). Secondly, the number of Dcr1-sensitive protein-coding genes is larger in the *xrn1Δ* context, in comparison to WT and *rrp6Δ* (Figure S2E). Thirdly, the *dcr1Δ* and *xrn1Δ* mutants display synergic growth defects (Figure 2C). This indicates that the presence of Dcr1 becomes important for the cell viability in the absence of Xrn1, *ie* when aslncRNAs accumulate in the cytoplasm, presumably forming dsRNA structures with the paired-sense mRNAs. In contrast, *DCR1* deletion partially suppresses the growth defect of the *rrp6Δ* mutant (Alcid and 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 to date. Additional analyses will be 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 (Figure 4), which is consistent with previous observations made upon expression of a Dcr1-GFP fusion in *S. cerevisiae* (Cruz and Houseley 2014). Moreover, among the different classes of aslncRNAs, the asXUTs constitute the preferred target for small RNAs production (Figure 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, for example at the level of translation regulation, 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 (Figure 3A-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.

It has been recently proposed that the loss of RNAi might have allowed an expansion of the aslncRNAs transcriptome in *S. cerevisiae* (Alcid and 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 and 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 –deficient species are more subtle than initially proposed. In fact, we show that more than 1500 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 two-fold difference compared to *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 (Figure 5C-D). Moreover, we observed that globally, the coding regions are mainly overlapped by asCUTs in the RNAi-capable species, while 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 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 and 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 and Kim 2014; Borges and Martienssen 2015), will be of particular interest.

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 and Wolfe 2005). The orthology was confirmed by CLUSTALO alignments (Figure 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 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 (Figure 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-kanMX6*), 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 (1047 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, giving the YAM2826 strain.

N. castellii strains were grown at 25°C in rich Yeast Extract-Peptone-Dextrose (YPD) medium to mid-log phase (OD_{600} 0,5). For the microscopy analyses, YAM2478 (WT) and YAM2826 (Dcr1-GFP) cells were grown in Complete Synthetic Medium (CSM), in the same conditions.

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), then hybridized overnight at 65°C in PerfectHyb™ Plus Hybridization Buffer (Sigma).

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 x 50 nt) was performed on a HiSeq2500 system (Illumina).

The *N. castellii* reference genome was retrieved from version 7 of the Yeast Gene Order Browser (Byrne and Wolfe 2005); snoRNAs were annotated using the *S. cerevisiae* snoRNAs as queries for blastn alignements (E value cutoff e^{-8}). Reads were mapped using version 2.0.9 of TopHat (Kim et al. 2013), with a tolerance of 3 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, 12000 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 >2-fold enrichment in the *xrn1Δ* and *dcr1Δ* mutant, respectively, compared to 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 stable unannotated transcripts (SUTs).

For the annotation of CUTs, we used previously published RNA-Seq data from biological duplicates of *rrp6Δ* cells (Alcid and 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 being normalized on the total number of reads uniquely mapped on ORFs. We identified 1280 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, 1021 XUTs and 1280 CUTs. A 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. Libraries were constructed using the NEBNext Multiplex Small RNA Library Preparation Set for Illumina (New England Biolabs), starting from <100 nt small RNAs purified on 15% TBE-Urea polyacrylamide gels. 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* reference genome using the version 1.0.0 of Bowtie (Langmead et al. 2009), with a tolerance of 3 mismatches. Subsequent analysis used 22-23 nt uniquely mapped reads. Densities were normalized on tRNAs signals, which are not affected in *dcr1Δ* (Figure 1C) or *xrn1Δ* (Figure 1D).

For the control of Dcr1-GFP functionality, one library was prepared per strain (YAM2478, YAM2795 and YAM2826). Small RNA purification, libraries construction, sequencing and bioinformatics analyses were as above.

Microscopy

YAM2478 (WT) and YAM2826 (Dcr1-GFP) cells were grown to mid-log phase (OD_{600} 0.5) in CSM medium, at 25°C. For the live-cells 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 100x/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 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 immuno-FISH 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 minutes in phosphate-buffered saline (PBS) containing 4% paraformaldehyde and 2% of sucrose. After two washes with in PBS + 50 mM NH_4Cl, the fixed cells were blocked and permeabilized for 30 minutes in blocking/hybridization buffer (0.25% Triton X-100, 5% BSA, 0.004% NaN_3 in PBS), under gentle shaking. The cells were then labeled for 90 minutes with 100 µl of nanobody solution (10 µM ATTO

488 nanobody in blocking/hybridization buffer). Finally, the labeled cells were washed for 5 minutes in PBS, a drop of Vectashield mounting medium with DAPI (Vectorlabs) 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. A genome browser for visualization of processed RNA-Seq data is accessible at http://vm-gb.curie.fr/castellii.

Code accessibility

All codes are available upon request to the corresponding authors.

Availability of materials

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

Acknowledgments

We thank David Bartel for providing the WT and *dcr1Δ* strains of *N. castellii*. We also would like to thank Sylvain Baulande, Virginie Raynal and Patricia Legoix-Né from the NGS 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; 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 Canceropôle Ile-de-France. A.M.'s lab is supported by the ANR "DNA-Life", ERC "EpincRNA" starting and ERC "DARK" consolidator grants.

Author contributions

M.W. and A.M. designed the project; M.W., S.A. and D.F performed the experiments; U.S. performed all bioinformatics analyses and M.W. analyzed the data; M.W. wrote the manuscript with input from all authors.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

Figure legends

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 and Tsukiyama 2016) were segmented in parallel using the same tool. Among the ≥ 200 nt segments not overlapping 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 1280 CUTs (signal in *rrp6Δ* ≥1 FPKM; *rrp6Δ*/WT ratio >2, *P* <0.05).

B. Heatmap of the expression fold-change (ratio of tag densities, $\log_2$ scale) for SUTs (146), CUTs (1280), XUTs (1021) and DUTs (10) in the *dcr1Δ*, *xrn1Δ*, and *rrp6Δ* mutants, relative to the corresponding WT strain. For each class of lncRNA, the number of antisense and solo (*ie* 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. Ns = not significant. Outliers: not shown.

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 and 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 complete synthetic medium (CSM), then incubated at the indicated temperatures for 2-3 days.

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 aslncRNA and the paired-sense mRNA. The snapshot was produced using VING (Descrimes et al. 2015).

Figure 4. Subcellular localization of Dcr1 in *N. castellii*.

YAM2478 (WT) and YAM2826 (Dcr1-GFP) cells were grown to mid-log phase in complete synthetic medium, at 25°C. After fixation of cells, the subcellular localization of Dcr1-GFP was performed using immune-FISH using GFP nanobody (see details in Methods). DAPI staining was used to visualize DNA. Scale bars: 1 µm.

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 = 1152).

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 = 1252).

E. Cumulative coverage of the coding regions by asCUTs and asXUTs in *N. castellii* (grey bars) and *S. cerevisiae* (black bars).

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 and 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 and 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 and 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*, *rrp6Δ* and *rrp6Δ dcr1Δ* strains. The raw data for *rrp6Δ* and *rrp6Δ dcr1Δ* have been previously published (Alcid and Tsukiyama 2016). Signal for the + and - strands are visualized as heatmaps in the upper and lower panels, respectively, using the VING software (Descrimes et al. 2015). 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 Figure 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 Figure 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 and 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 small RNAs produced in WT (YAM2478), *xrn1Δ* (YAM2479), *dcr1Δ* (YAM2795) and *xrn1Δ dcr1Δ* (YAM2796) cells. Libraries were constructed using purified small RNAs.

B. Box-plot of 22-23 nt small RNAs densities (tag/nt, $\log_2$ scale) for the antisense (light grey) and solo (dark grey) SUTs in the WT, *xrn1Δ*, *dcr1Δ* and *xrn1Δ dcr1Δ* strains. Outliers: not shown.

C. Same as above for the antisense (light grey) and solo (dark grey) XUTs.

D. Same as above for the antisense (light grey) and solo (dark grey) CUTs.

E. Snapshot of small RNAs for 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 highlights the region of overlap between the aslncRNA and the paired-sense mRNA. The snapshot was produced using VING (Descrimes et al. 2015).

Figure S4. Dcr1-GFP localizes in the cytoplasm

A. Visualization of Dcr1-GFP in living cells. YAM2478 (WT) and YAM2826 (Dcr1-GFP) cells were grown to mid-log phase in complete synthetic medium at 25°C, harvested and then washed in sterile water before direct observation.

B. Size and first base distribution of small RNAs produced in the WT (YAM2478), *dcr1Δ* (YAM2795) and Dcr1-GFP (YAM2826) strains.

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 and 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 ^{32}P-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).

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.

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.

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.

Borges F, Martienssen RA. 2015. The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* **16**: 727-741.

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.

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.

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.

Cliften PF, Fulton RS, Wilson RK, Johnston M. 2006. After the duplication: gene loss and adaptation in Saccharomyces genomes. *Genetics* **172**: 863-872.

Cruz C, Houseley J. 2014. Endogenous RNA interference is driven by copy number. *Elife* **3**: e01581.

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.

Didion JP, Martin M, Collins FS. 2017. Atropos: specific, sensitive, and speedy trimming of sequencing reads. *PeerJ* **5**: e3720.

Drinnenberg IA, Fink GR, Bartel DP. 2011. Compatibility with killer explains the rise of RNAi-deficient fungi. *Science* **333**: 1592.

Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP. 2009. RNAi in Budding Yeast. *Science* **326**: 544-550.

Ha M, Kim VN. 2014. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**: 509-524.

Houseley J, LaCava J, Tollervey D. 2006. RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* **7**: 529-539.

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.

Jarroux J, Morillon A, Pinskaya M. 2017. History, Discovery, and Classification of lncRNAs. *Adv Exp Med Biol* **1008**: 1-46.

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.

Kopp F, Mendell JT. 2018. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172**: 393-407.

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

Lee JT, Lu N. 1999. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* **99**: 47-57.

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.

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.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Marquardt S, Hazelbaker DZ, Buratowski S. 2011. Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription* **2**: 145-154.

Mercer TR, Dinger ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**: 155-159.

Nagarajan VK, Jones CI, Newbury SF, Green PJ. 2013. XRN 5'-->3' exoribonucleases: structure, mechanisms and functions. *Biochim Biophys Acta* **1829**: 590-603.

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.

Pelechano V, Steinmetz LM. 2013. Gene regulation by antisense transcription. *Nat Rev Genet* **14**: 880-893.

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.

Renganathan A, Felley-Bosco E. 2017. Long Noncoding RNAs in Cancer and Therapeutic Potential. *Adv Exp Med Biol* **1008**: 199-222.

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.

Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81**: 145-166.

Saha P, Verma S, Pathak RU, Mishra RK. 2017. Long Noncoding RNAs in Mammalian Development and Diseases. *Adv Exp Med Biol* **1008**: 155-198.

Schmitt AM, Chang HY. 2016. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **29**: 452-463.

-. 2017. Long Noncoding RNAs: At the Intersection of Cancer and Chromatin Biology. *Cold Spring Harb Perspect Med* **7**: 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.

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.

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.

Tisseur M, Kwapisz M, Morillon A. 2011. Pervasive transcription - Lessons from yeast. *Biochimie* **93**: 1889-1896.

Tudek A, Candelli T, Libri D. 2015. Non-coding transcription by RNA polymerase II in yeast: Hasard or necessite? *Biochimie* **117**: 28-36.

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.

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.

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.

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.

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.

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.

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.

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.

Wery M, Kwapisz M, Morillon A. 2011. Noncoding RNAs in gene regulation. *Wiley Interdiscip Rev Syst Biol Med* **3**: 728-738.

Woolcock KJ, Buhler M. 2013. Nuclear organisation and RNAi in fission yeast. *Curr Opin Cell Biol* **25**: 372-377.

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.

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.

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.

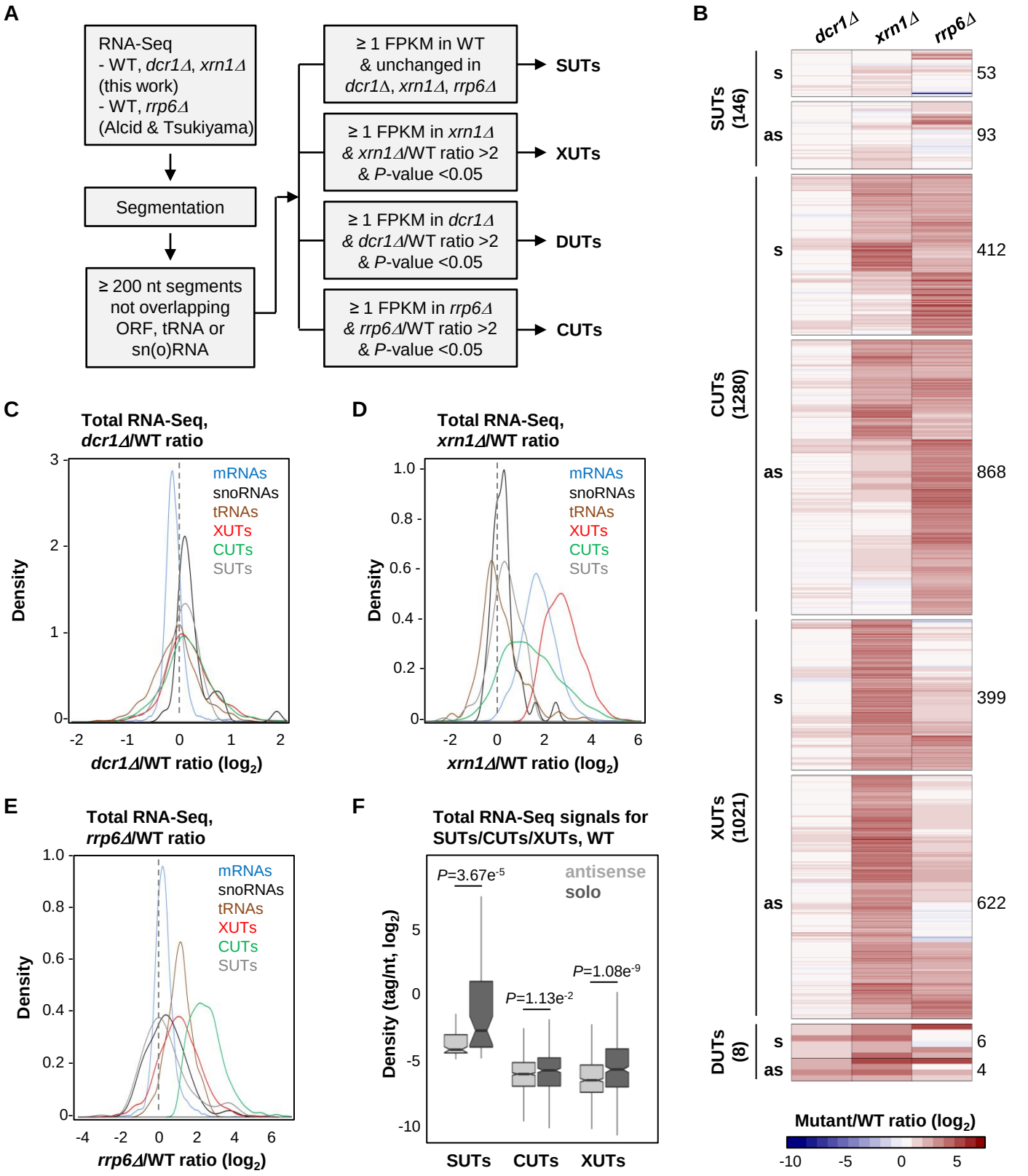
A
RNA-Seq
- WT, dcr1Δ, xrn1Δ
(this work)
- WT, rrp6Δ
(Alcid & Tsukiyama)
Segmentation
≥ 200 nt segments not overlapping ORF, tRNA or sn(o)RNA
≥ 1 FPKM in WT & unchanged in dcr1Δ, xrn1Δ, rrp6Δ
SUTs
≥ 1 FPKM in xrn1Δ & xrn1Δ/WT ratio >2 & P-value <0.05
XUTs
≥ 1 FPKM in dcr1Δ & dcr1Δ/WT ratio >2 & P-value <0.05
DUTs
≥ 1 FPKM in rrp6Δ & rrp6Δ/WT ratio >2 & P-value <0.05
CUTs
B
dcr1Δ
xrn1Δ
rrp6Δ
SUTs (146)
s
53
as
93
CUTs (1280)
s
412
as
868
XUTs (1021)
s
399
as
622
DUTs (8)
s
6
as
4
Mutant/WT ratio (log2)
-10
-5
0
5
C
Total RNA-Seq, dcr1Δ/WT ratio
Density
mRNAs
snoRNAs
tRNAs
XUTs
CUTs
SUTs
dcr1Δ/WT ratio (log2)
D
Total RNA-Seq, xrn1Δ/WT ratio
Density
mRNAs
snoRNAs
tRNAs
XUTs
CUTs
SUTs
xrn1Δ/WT ratio (log2)
E
Total RNA-Seq, rrp6Δ/WT ratio
Density
mRNAs
snoRNAs
tRNAs
XUTs
CUTs
SUTs
rrp6Δ/WT ratio (log2)
F
Total RNA-Seq signals for SUTs/CUTs/XUTs, WT
Density (tag/nt, log2)
P=3.67e-5
antisense
solo
P=1.08e-9
P=1.13e-2
SUTs
CUTs
XUTs

Figure 1

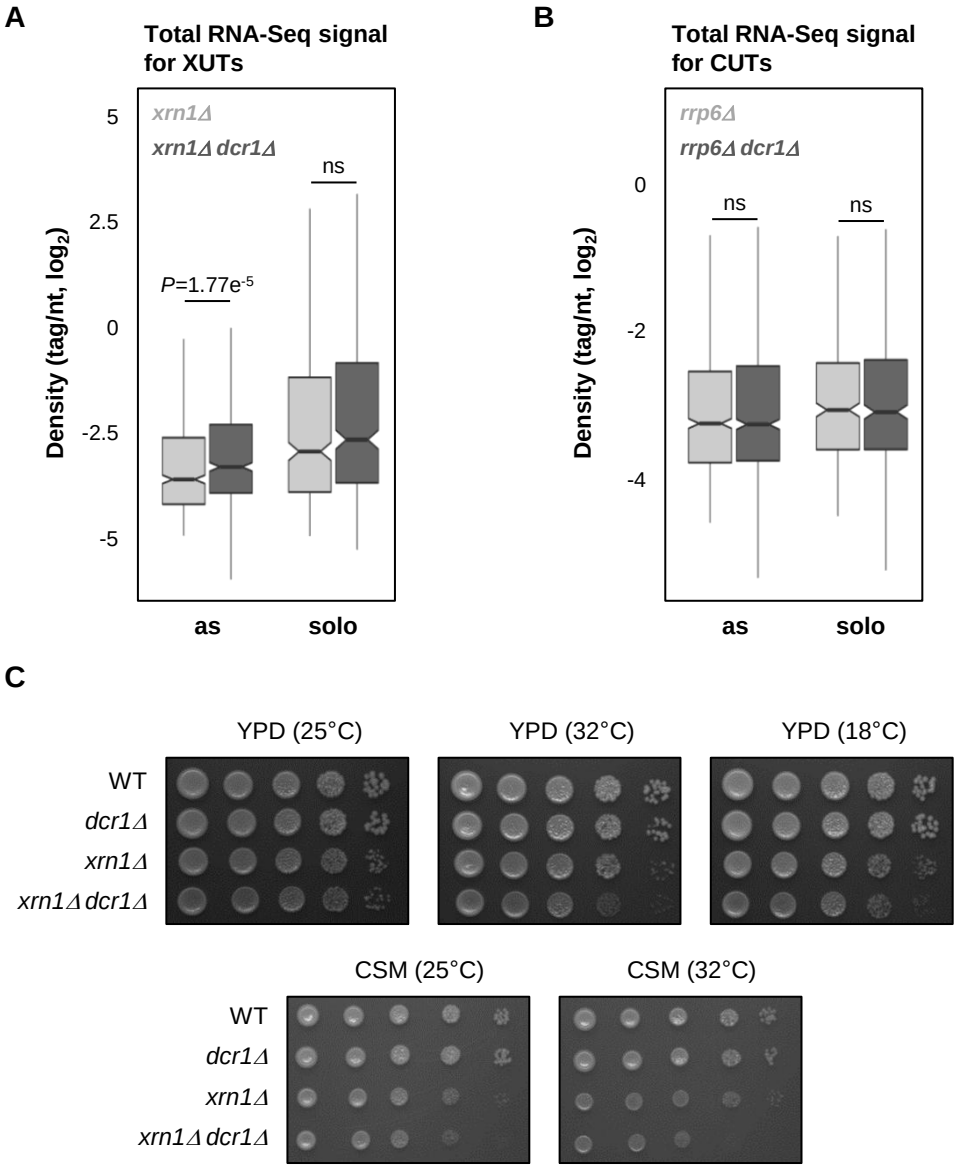
A
Total RNA-Seq signal for XUTs
xrn1Δ
xrn1Δ dcr1Δ
ns
P=1.77e-5
Density (tag/nt, log2)
5
2.5
0
-2.5
-5
as
solo
B
Total RNA-Seq signal for CUTs
rrp6Δ
rrp6Δ dcr1Δ
ns
ns
Density (tag/nt, log2)
0
-2
-4
as
solo
C
YPD (25°C)
YPD (32°C)
YPD (18°C)
WT
dcr1Δ
xrn1Δ
xrn1Δ dcr1Δ
CSM (25°C)
CSM (32°C)
WT
dcr1Δ
xrn1Δ
xrn1Δ dcr1Δ

Figure 2

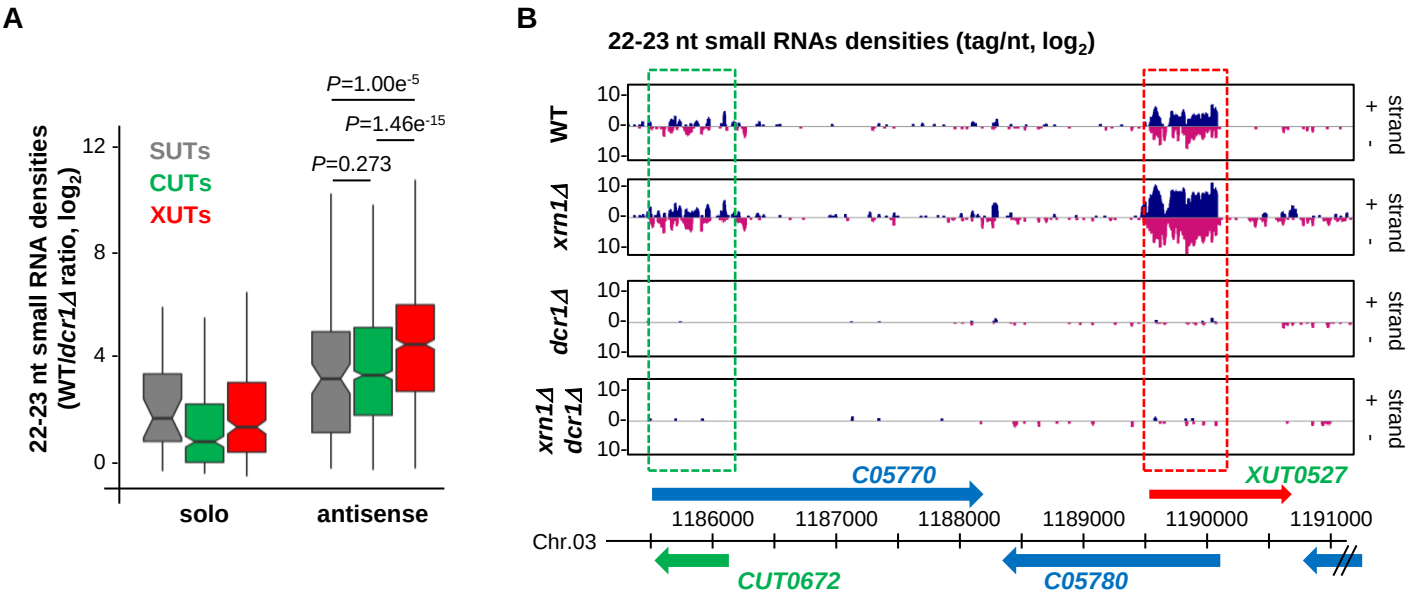
A
22-23 nt small RNA densities (WT/dcr1Δ ratio, log2)
SUTs
CUTs
XUTs
P=1.00e-5
P=1.46e-15
P=0.273
12
8
4
0
solo
antisense
B
22-23 nt small RNAs densities (tag/nt, log2)
WT
xrn1Δ
dcr1Δ
xrn1Δ dcr1Δ
10
0
10
+
-
strand
C05770
XUT0527
1186000
1187000
1188000
1189000
1190000
1191000
Chr.03
CUT0672
C05780

Figure 3

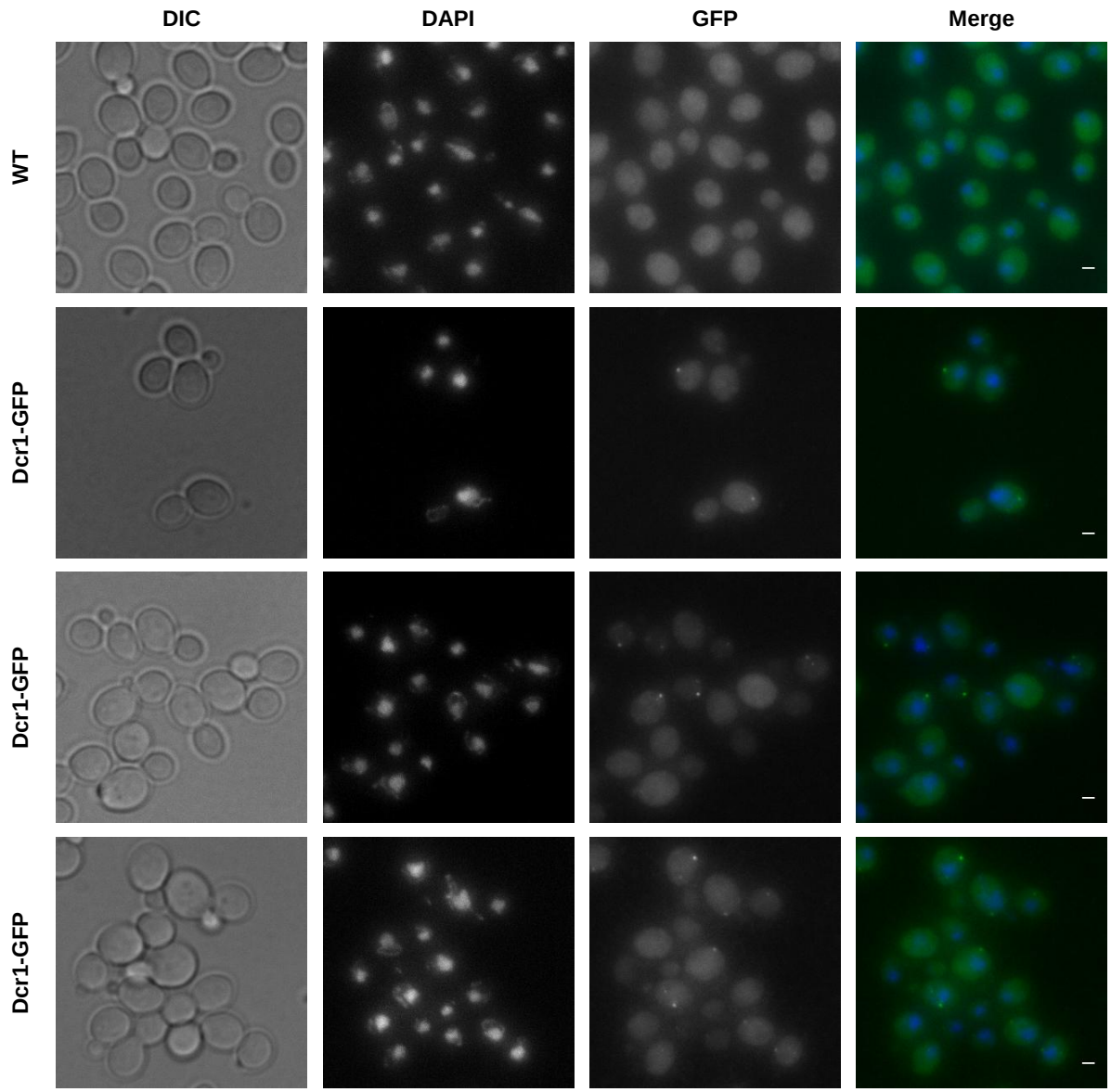

Figure 4

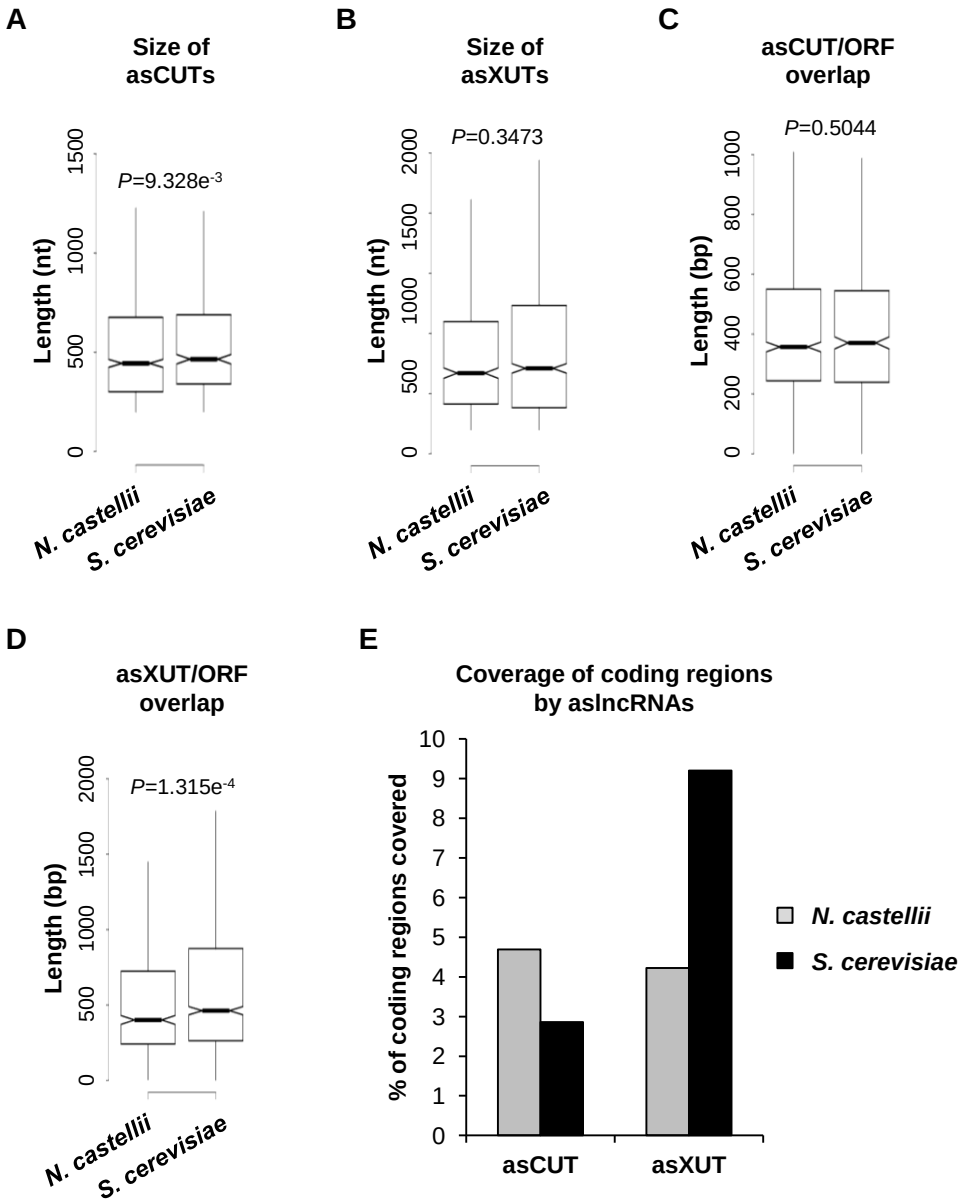
A
Size of asCUTs
P=9.328e-3
Length (nt)
0
500
1000
1500
N. castellii
S. cerevisiae
B
Size of asXUTs
P=0.3473
Length (nt)
0
500
1000
1500
2000
N. castellii
S. cerevisiae
C
asCUT/ORF overlap
P=0.5044
Length (bp)
0
200
400
600
800
1000
N. castellii
S. cerevisiae
D
asXUT/ORF overlap
P=1.315e-4
Length (bp)
0
500
1000
1500
2000
N. castellii
S. cerevisiae
E
Coverage of coding regions by asIncRNAs
% of coding regions covered
0
1
2
3
4
5
6
7
8
9
10
asCUT
asXUT
N. castellii
S. cerevisiae

Figure 5