bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title

Antisense transcriptional interference mediates condition-specific gene repression in budding yeast

Alicia Nevers[1,3], Antonia Doyen[1], Christophe Malabat[1,4], Bertrand Néron[4], Thomas Kergrohen[1], Alain Jacquier*[,1,2] and Gwenael Badis*[,1,2].

[1]*Unité GIM, Institut Pasteur, Paris, France*
[2]*CNRS UMR3525, Paris, France*
[3]*Université Pierre et Marie Curie, Paris, France*
[4] *Bioinformatics and Biostatistics Hub, C3BI, Institut Pasteur, USR 3756 IP CNRS, Paris, France*
Corresponding authors:*
Gwenael Badis Tel: +33 140613331; Fax: +33 140613456; email: gbreard@pasteur.fr
Alain Jacquier Tel: +33 140613205; Fax: +33 140613456; email: jacquier@pasteur.fr

Corresponding Author upon submission.

Gwenael Badis Tel: +33 140613331; Fax: +33 140613456; E-mail: gbreard@pasteur.fr

Running title

Gene repression by antisense transcription

Keywords

Transcriptional interference/ antisense/ pervasive transcription/Quiescence

(55 165 characters).

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Pervasive transcription generates many unstable non-coding transcripts. Although a few examples of pervasive antisense have been shown to mediate gene regulation by transcriptional interference, whether pervasive transcription has a general functional role or merely represents transcriptional noise remains unclear. In a mutant context that stabilised pervasive transcripts, we characterized more than 800 antisense RNAs genome wide and analysed the corresponding sense mRNA behaviour. We observed that antisense non-coding transcripts were associated with genes tightly repressed during exponential growth compared to quiescence and with an opposite level of variation between the mRNA and the associated antisense. This suggested that antisense transcription might participate to gene repression during the exponential phase. We thus specifically interrupted the antisense transcription of a subset of genes, and found that it resulted in a de-repression of the corresponding mRNAs. Antisense-mediated repression involved a *cis*-acting mechanism and was dependent on several chromatin modification factors. Our data convey that transcriptional interference by pervasive antisense transcription is a general mechanism of gene repression between cellular states.

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

In steady state, the transcriptome reflects the equilibrium between RNA synthesis and degradation. Eukaryotes have developed sophisticated systems to control the turnover of mRNAs and ncRNAs necessary to the cell, undesired RNA species being rapidly eliminated by quality control mechanisms.

The development of genome-wide techniques to analyse transcriptomes, tilling arrays and cDNA next-generation sequencing, revealed that eukaryotic genomes are pervasively transcribed (Carninci & Hayashizaki, 2007). The genome of budding yeast is particularly compact and it has been hitherto conceded that more than 70% of it is composed of protein coding ORFs (Goffeau *et al*, 1996). Yet, this is only true if one does not distinguish the two DNA strands. If one takes into account sense and antisense sequences, non protein-coding genomic sequences represent up to 65% of the genome, leaving room to a large fraction of the genome for the generation of pervasive non-coding transcripts.

In yeast, pervasive transcription has been first reported more than a decade ago. If a fraction of it was uncovered in wild-type cells (David *et al*, 2006; Xu *et al*, 2009), a substantial part of the eukaryotic pervasive transcription is “hidden” as it generates very short-lived “cryptic” transcripts. These RNAs are difficult to detect unless they are stabilized by interfering with quality control mechanisms that normally eliminate them (Jensen *et al*, 2013). Pervasive transcripts detected in wild type yeast cells have been named “SUTs” for “Stable Unannotated Transcripts” (Xu *et al*, 2009), and different names have been given to cryptic transcripts depending on which factor was mutated in order to stabilise a particular class of RNAs. For example, CUTs, or Cryptic Unstable Transcripts were characterized upon removal of the exonuclease Rrp6, specific for the nuclear form of the exosome (Wyers *et al*, 2005; Neil *et al*, 2009; Xu *et al*, 2009), XUTs were revealed upon removal of the cytoplasmic exonuclease Xrn1 (van Dijk *et al*, 2011) and the NUTs correspond to the transcripts that accumulate when the nuclear termination factor Nrd1 is depleted (Schulz *et al*, 2013). Yet, there are in yeast only two main pathways responsible for the efficient elimination of pervasive transcripts, the nuclear Nrd1-Nab3-Sen1 (NNS) pathway, in which the early transcription termination of cryptic transcripts by the NNS complex is coupled to the degradation by the nuclear TRAMP-exosome complex (Arigo *et al*, 2006; Thiebaut *et al*, 2006; Schulz *et al*, 2013; Tudek *et al*, 2014) and the cytoplasmic non-sense mediated mRNA decay (NMD) pathway (Malabat *et al*, 2015; Wery *et al*, 2016). Many of pervasive transcripts require both pathways for their efficient and fast elimination (see Malabat *et al*, 2015).

Irrespective of which pathway predominates for their degradation, these transcripts all originate from nucleosome free regions (NFRs), which are essentially found 5’ and 3’ of mRNA coding sequences (Jiang & Pugh, 2009). When they originate from 5’ NFRs, they are most often transcribed divergently from mRNAs and result from an intrinsic low polarity of gene promoters (Neil *et al*, 2009; Xu *et al*, 2009). This divergent transcription has the potential to interfere with the expression of the neighbouring upstream gene. Likewise, when a non-coding transcript initiates from the 3’ NFR in antisense to the upstream gene, its transcription has the potential to interfere with the proper expression of the corresponding mRNA (van Dijk *et al*, 2011; Murray *et al*, 2012). Such transcription interference by pervasive transcription is largely prevented genome-wide by the NNS quality control pathway, which ensures the early transcription termination of these transcripts and prevent them to extend

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

into the promoter region of the corresponding antisense genes (Arigo *et al*, 2006; Thiebaut *et al*, 2006; Schulz *et al*, 2013; Castelnuovo *et al*, 2014).

Whether pervasive transcription has a general function is a matter of debate. The fact that highly efficient quality control mechanisms have been selected during evolution to eliminate most of these transcripts argue in favour of the idea that most of them are non functional, but pervasive transcription itself, more than it product, could be important. Yet, the existence of the NNS pathway, which, by terminating pervasive transcription early, is key in preserving pervasive transcription from interfering with the expression of many coding genes, also suggests that a large fraction of these events simply result from the low specificity of RNA polymerase II (PolII) transcription initiation.

There are a number of well-documented examples of individual coding gene regulation through the transcription of a non-coding RNA: *SER3* (Martens *et al*, 2004), *IME1* and *IME4* (van Werven *et al*, 2012), *GAL10/GAL1* (Houseley *et al*, 2008; Pinskaya *et al*, 2009), *PHO84* (Castelnuovo *et al*, 2013), *CDC28* (Nadal-Ribelles *et al*, 2014). In the vast majority of cases analysed in budding yeast, the synthesis of a non-coding transcript has only an effect in *cis*. The prevailing model is thus that repressive chromatin marks are deposited in the promoter regions of genes in the wake of PolII transcribing the associated non-coding RNAs (Castelnuovo & Stutz, 2015; Murray *et al*, 2015). It is thus the act of transcription rather that its product, which is important. Several distinct mechanisms can be at play, but the general theme is that methyltransferases carried on the carboxy-terminal domain (CTD) of the PolII large subunit deposit histone methylation marks that recruit repressive chromatin modifiers such as histone deacetylases or nucleosome remodelling complexes. In budding yeast, there are two such CTD associated histone methyl transferases, Set1, which methylates histone H3K4 at promoters and gene proximal regions of actively transcribed genes, and Set2, which methylates H3K36 within more distal gene regions. The role of Set1 is complex. It is responsible for both H3K4 di- and tri-methylation (H3K4me2 and H3K4me3). It has been proposed that H3K4me3 at the beginning of actively transcribed genes could enhance and help maintaining preinitiation complex assembly and an active acetylated chromatin state, thus playing a positive role on transcription. Conversely, Set1 generates H3K4me2 in the body of gene, which recruits the histone deacetylase complexes SET3 or RPD3L, which repress transcription initiation (see Venters & Pugh, 2009 for review). Set2 is responsible for the H3K36 methylation (H3K36me2) in the body of genes, which results in the recruiting the Rpd3S deacetylase complex that plays an essential role in preventing improper internal initiation (Carrozza *et al*, 2005; Keogh *et al*, 2005). Thus both Set1 and Set2 have the potential to mediate transcription interference and have been implicated in gene repression by non-coding RNA transcriptional interference (see Castelnuovo & Stutz, 2015 for review).

How widely the transcription of non-coding RNAs, in particular antisense RNAs (asRNAs), is associated to gene regulation? Is it only restricted to a few exceptions, pervasive transcription then being merely the result of transcriptional noise, or does it play a broader role in gene regulation?

Several large-scale studies attempted to find a role to this transcription. Genes with large expression variability (such as stress response and environment specific genes) often have antisense expression and this suggests a general regulatory effect of antisense gene expression (Xu *et al*, 2011). Others correlated antisense expression with chromatin marks, either in a wild type context or in a *rrp6* mutant condition (Murray *et al*, 2015; McKnight *et al*, 2015; Castelnuovo *et al*, 2014; Gu *et al*,

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2015; Kim *et al*, 2016) but no global anti-correlated trend was found between asRNA and mRNA expression.

Very recently, the question of the actual extent of the effect of asRNA transcription on gene regulation was addressed directly by measuring, at a large scale and under various conditions, the effect of specific antisense SUTs transcription interruption on the expression of the corresponding proteins fused to GFP (Huber *et al*, 2016). This study showed that, for 41 out of 152 (~27%) genes associated with an antisense SUT, a detectable antisense-dependent gene regulation could be observed under at least one condition. Although no specific biological pathway seemed enriched in the asRNA responsive genes, the analysis showed that repression by asRNA transcription interference helps reducing mRNA expression basal levels, especially for genes expressed at a low level, resulting in complete gene repression.

This analysis is restricted to SUTs, *i.e.* non-coding RNAs readily detectable in wild-type cells. Yet, we know that SUTs represent only a minority of the pervasive transcripts, most of which are too unstable to be detected in wild-type cells (Neil *et al*, 2009; Xu *et al*, 2009; van Dijk *et al*, 2011; Geisler *et al*, 2012). Whilst the nuclear NNS quality control pathway terminates early the transcription of many of the pervasive RNAs, to prevent them from interfering with mRNA expression (Vasiljeva *et al*, 2008; Schulz *et al*, 2013), many pervasive transcripts escape, at least in part, this first surveillance pathway and are extended up to cryptic cleavage and polyadenylation sites (polyA sites), potentially over the transcription start site (TSS) of their associated genes. This can lead to the export of long non-coding RNAs into the cytoplasm, where they are rapidly degraded by the NMD pathway (Malabat *et al*, 2015; Wery *et al*, 2016).

We analysed genome-wide the amount of asRNAs associated to each mRNA in a NMD mutant context (*upf1*Δ). In this mutant, hidden pervasive transcripts that escaped the nuclear NNS surveillance accumulate in the cytoplasm and can thus be quantified. An important fraction of the less expressed genes are associated with asRNAs, especially if the asRNAs overlapped the associated sense gene promoter. In addition, many of these genes with promoter-overlapping asRNAs were enriched for genes up-regulated in chromatin remodelling mutants such as *set2*Δ or *set1*Δ. Many of these mRNAs, poorly expressed during the exponential phase, corresponded to mRNAs found enriched during the stationary phase (G0). These observations strongly suggested that this particular class of genes is frequently subjected to asRNA transcription interference in order to be fully repressed during logarithmic growth, a prediction we validated for a subset of examples. This showed that antisense-mediated transcriptional interference is a common mechanism used when mRNA expression needs to be tightly repressed under specific conditions.

Results

Characterization of antisense transcription in upf1Δ cells

In order to reveal antisense transcription that escaped the nuclear surveillance by the NNS pathway and reached the proximal region of the genes, we quantified the asRNAs levels in the proximal region of the protein coding genes (over a +1 (mRNA TSS) to +200 nucleotides window) in a strain impaired for the cytoplasmic NMD surveillance pathway (*upf1*Δ mutant). Figure 1A shows these values (y axis) plotted against the average mRNA levels per gene (number of reads per nucleotide; x axis).

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The less expressed mRNAs appeared to be generally associated with high levels of asRNAs. In order to quantify this observation, we partitioned the genes according to their mRNAs levels in ten bins with an equal numbers of genes. The less expressed genes (bin 1) had significantly higher levels of asRNAs than the genes within higher mRNA expression categories (bins 2 to 10; see Figure 1B, Table EV1 and Dataset 1).

At least two non-exclusive phenomena could explain this observation. First, mRNA transcription itself could have a repressive effect on asRNA transcription initiation from their corresponding 3' NFRs (Xu *et al*, 2011). Hence, asRNAs initiating within NFRs situated downstream of non-expressed genes should be less subjected to such repression by mRNA transcription. Conversely, antisense-transcription from 3' NFRs could be a common mean to contribute to a tight gene repression. If the former explanation is correct, asRNAs associated with non-expressed genes should not show different termination characteristics than other asRNAs. In contrast, it was shown that repression by asRNAs correlates with mRNA TSS overlap (Huber *et al*, 2016). If asRNAs associated with the less expressed genes contribute to their tight repression, these asRNAs should overlap the mRNA TSSs more often than other asRNAs. We thus categorized genes depending on the occurrence of their associated asRNAs across TSSs by analysing a window between -50 nucleotides to +200 nucleotides relative to the mRNA TSS. We defined three types of genes. Genes without substantial asRNAs over the +1 to +200 nucleotide region (arbitrarily set below three reads per base over this window) defined class N (No antisense). Genes with asRNAs but with an average read number below three in the -50 to -1 nucleotide region, thus terminating before the mRNA TSS, defined class M (mRNA antisense). Conversely, genes with asRNAs with an average read number above three in the -50 to -1 nucleotide region defined genes with TSS overlapping asRNAs (class O - Overlapping antisense) (Figure 1C; Dataset 1). Figure 1D shows a heat map of the sense and antisense transcripts over a -200 to +200 nucleotides around the mRNA TSSs, ordered by these classes. Among the 5892 protein coding genes analysed, 5076 belong to class N, 259 to class M and 557 have an overlapping asRNA (class O). The higher proportion asRNAs in bin 1 mostly resulted from the over-representation of class O asRNAs (Figure 1E, Figure EV1 and Table EV2), strongly suggesting that many of them could be associated to antisense transcription mediated repression.

Genes up-regulated in the absence of RNA PolII associated chromatin regulators are enriched in the class of poorly expressed genes with TSS-overlapping asRNAs

If antisense transcription can affect sense transcription, one should expect that genes associated with asRNAs should be more up-regulated in chromatin modifier mutants implicated in transcriptional interference, in particular *set1* and *set2* mutants. Given that antisense transcriptional interference involves the extension of asRNA up to the promoter regions of the genes, Set2, which promotes H3K36me2 at late stages of PolII elongation, seemed a good candidate to mediate gene repression by asRNA transcription. *Set2* mutants are intrinsically difficult to analyse by RNAseq or tilling arrays since one of its main role is to suppress internal initiation both in sense and antisense of gene transcribed regions (Carrozza *et al*, 2005; Keogh *et al*, 2005; Venkatesh *et al*, 2012; Malabat *et al*, 2015). The cryptic initiation events observed in *set2*Δ mutants in the sense orientation can thus induce misleading global increase of

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the mRNA quantification by contributing to the overall sense RNAseq counts (Malabat *et al*, 2015). We thus took advantage of the analysis of individual TSSs in the Malabat *et al.* study, which allows the analysis of the specific mRNA TSSs, irrespective of internal transcription initiation. We considered a gene as up-regulated upon *SET2* deletion when its strongest mRNA-linked TSS cluster was induced at least two fold and with a p-value ≤0.05 (Supplementary file 3 in Malabat *et al*, 2015). Ninety-five of 5228 genes analysed in this dataset were up-regulated in *set2*Δ (see Dataset 1) Strikingly, genes with TSS-overlapping asRNAs (class O) showed the highest percentage of up regulation in *set2*Δ (9,1% compared to 1.8% for all genes; Figure 2A). Combining gene promoter classes with the mRNA expression level categories drastically increased this bias since class O of bin 1 showed the highest proportion of genes up-regulated in *set2*Δ cells (Figure 2B, right panel).

Direct measures of transcription levels by NETseq were also analysed in a *set2*Δ mutant (Churchman & Weissman, 2011). Although a higher number of genes were found up-regulated in absence of Set2, possibly due to internal initiation events not being filtered out, the same trend was observed (Figure EV2A). This prompted us to analyse the data for the *set1*Δ, as well as *rco1*Δ and *eaf3*Δ (two components of the Rpd3S deacetylase complex) mutants from the same dataset, as these factors have also been found involved in transcription interference. Genes up-regulated upon deletion of these factors were also clearly over represented in class O (Figures EV2B-D). Altogether these results suggest that repression by antisense transcriptional interference is frequent for poorly expressed genes, and is mediated by several, chromatin-modifying factors linked to elongating PolII. Figure EV2E reports the overlapping number of up-regulated genes in the different mutant strains.

Quiescence enriched genes are associated with high levels of asRNAs

We next questioned whether poorly expressed genes (bin 1) could belong to a particular category of regulated genes. An expected category of genes strongly repressed during exponential growth are genes found enriched is stationary phase and/or in quiescence (G0). We thus analysed a dataset reporting the time course of mRNA expression over a complete 9-days growth cycle (Radonjic *et al*, 2005). Figure 3A shows that the stationary phase enriched genes (SP-enriched in Radonjic *et al*, 2005) were most abundant in bin 1. The Radonjic dataset essentially documents mRNAs expression in wild-type cells. In addition, stationary phase cell populations are not homogeneous, being composed of dead, dying and quiescent cells (Allen *et al*, 2006; Aragon *et al*, 2008). We thus decided to perform a genome-wide RNAseq analysis of quiescent (G0) cells, which represents a much more homogeneous cell population, in wild type or *upf1*Δ cells in order to analyse both gene and pervasive transcription. To normalise the overall level of transcripts per genome, we spiked in the budding yeast cultures before RNA extraction with identical reference aliquots of a *Schizosaccharomyces pombe* exponential culture (see Materials and Methods for normalisation procedures). We defined quiescence-enriched (Q-enriched) mRNAs as being, after normalisation, five times more abundant in the G0 population compared to the cells in exponential growing phase (261 genes; see Table EV3). As anticipated, Q-enriched mRNAs were mostly found in bin 1 (Figure 3B and C). Accordingly, Figure 3D shows that, as for genes within bin 1, Q-enriched mRNAs were associated with higher asRNA levels than random (*** with $p = 1.5\ 10^{-4}$) and their distribution in the different asRNA associated genes classes (classes N, M & O) was similar to that of bin 1 (Figure EV3A). Moreover,

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

while 9.1% of all genes with TSS-overlapping asRNAs were up-regulated upon deletion of *SET2* (Figure 2A), this proportion raised to more than 35% when only considering Q-enriched genes (Figure EV3B). Breaking down these figures by bins and promoter classes showed that this strikingly high proportion was primarily contributed by class O genes, representing 13 out of 22 (59%) of the *set2*Δ up-regulated Q-enriched genes (Figure 3E). These observations strongly suggested that asRNA transcriptional interference could be a frequent mechanism of tight repression for this class of genes. In order to directly test this hypothesis, we chose for further analyses five representative examples of Q-enriched genes associated with an asRNA spanning the mRNA TSS: *PET10*, *SHH3*, *MOH1*, *CLD1* and *ARO10*. Among these genes, only *ARO10* was previously tested (Huber *et al*, 2016) for asRNA mediated transcription interference.

Time course of quiescence enriched mRNAs and corresponding asRNAs show an inverse expression pattern

In order to examine the relative behaviour of these mRNAs and their associated asRNAs during a growth cycle, we performed Northern-blots time course experiments starting (t0') with addition of rich medium to quiescence purified cells and using strand-specific RNA probes. The five selected Q-enriched mRNAs were not only accumulating during quiescence but were in fact strongly induced after about 48 hours of culture (Figure 4), which coincides with the post diauxic shift transition (Radonjic *et al*, 2005). The asRNAs started to accumulate between 5 and 30 minutes upon rich medium addition, to reach a peak of expression at ~24 hours, after which they rapidly disappeared. The mRNAs followed the inverse trend, although *ARO10* was not as substantially repressed during the exponential phase. These observations are compatible with the asRNA transcription contributing to mRNA repression. Conversely, they are also compatible with induction of the mRNA repressing the associated asRNAs.

Interruption of antisense transcription allows a de-repression of Q-enriched genes

One of the main effects of the NNS pathway is to prevent the expression of most pervasive transcription from interfering with the normal expression of genes genome-wide (Schulz *et al*, 2013). This mechanism is thus intrinsically optimized to terminate transcription early and in a strand specific manner. We thus chose to use it in order to specifically terminate asRNA transcription close to their transcription start by introducing in the TSS proximal region of the asRNAs a short (37 nucleotides) optimal NNS termination signal (NNS-ter; Porrua *et al*, 2012). In order to perturb as little as possible the corresponding mRNAs, we introduced this NNS-ter sequence seamlessly using a cloning-free method allowing chromosomal modifications without leaving selection markers (Erdeniz *et al*, 1997). This sequence was introduced in the proximal region of the asRNAs, corresponding to the terminal region of the mRNAs (see Figure EV4 and Table EV4). The introduction of the NNS-ter signal resulted in the proper elimination of all asRNAs and in a strong up-regulation of the corresponding mRNAs, except for *ARO10* (Figure 5). We note that *ARO10* is also, out of the five genes examined, the one that showed the weakest mRNA repression during the exponential phase (see Figure 4 and Discussion).

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The asRNA associated gene repression acts in cis

Although the majority of non-coding RNA associated gene regulation has been shown, when this was examined, to act only in *cis*, a *trans* effect of the asRNA itself has been invoked in a few cases (see Castelnuovo & Stutz, 2015 for a review). We directly addressed this question on *PET10* by comparing the mRNA and the asRNA expression in *cis* and in *trans*. In order to carry out this experiment, we built diploid strains where *PET10* sense and antisense transcripts were disrupted on one or two of the homologous chromosomes, allowing the expression of the asRNA either from the same chromosome as the mRNA (in *cis*), from the opposite chromosome (in *trans*) or no asRNA expression (see Figure 6A). RT-qPCR measurement showed that *PET10* mRNA is repressed only when its asRNA is expressed in *cis* (blue) but not in *trans* (green). In this case the mRNA level reached the same level as observed in the control strain without antisense (red). This is fully consistent with the hypothesis that the antisense transcription and not the asRNA itself, acts to repress mRNA by a transcriptional interference mechanism.

Several PolII elongation-associated chromatin modification factors cooperate to mediate antisense transcriptional interference

As described above, TSS-overlapping asRNA associated genes were more prone to be up-regulated upon deletion of chromatin modifiers such as *SET2* (Figure 2 and EV2) or *SET1*, *RCO1* and *EAF3* (Figure EV2, A-D) than the other categories of genes. Although the effects of *set2*Δ, *rco1*Δ and *eaf3*Δ are expected to be largely redundant as these factors act in the same chromatin modification pathway (Carrozza *et al*, 2005; Keogh *et al*, 2005; Churchman & Weissman, 2011), we also observed that more than half (80 out of the 155) of the genes that we computed in the Churchmann dataset (Churchman & Weissman, 2011) as the most up-regulated in *set2*Δ were also up-regulated in *set1*Δ (Figure EV2 E). This suggested that these different chromatin modifiers might cooperate to mediate asRNA transcriptional gene repression. The interpretation of such data are complicated by the fact that chromatin modifiers can affect both the mRNAs and their associated antisense (Murray *et al*, 2015; Castelnuovo & Stutz, 2015). To address this question, we directly measured the effects of *set1*Δ, *set2*Δ and *hda1*Δ on both the asRNA and the mRNA by strand-specific RT-qPCR (see Material and Methods). The analysis of the *PET10* locus indeed shows a complex picture. Not only the mRNAs were positively affected in different mutants, but the levels of asRNA was also impaired in all these mutants, making the evaluation of the antisense transcription interference on the mRNA difficult. In contrast, the *SHH3* asRNA was not repressed upon these deletions while the mRNA was significantly derepressed in all the mutants, although not at the level of the control strain in which the asRNA transcription elongation is restricted by the NNS-terminator (Figure 6B). This suggests that several chromatin modification pathways cooperate to mediate an efficient transcriptional interference to repress gene expression.

Sense and antisense transcription can mutually repress each other

As discussed above, the fact that the category of genes associated with asRNAs was enriched in the least expressed genes (bin 1 or quiescence enriched genes)

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

could result from two non-exclusive phenomena: either the asRNA transcription represses the mRNA or the absence of mRNA expression allows pervasive asRNA transcription from 3'-NFRs. We showed that, in four out of five genes we tested, asRNA transcription interruption led to an increase of sense mRNA levels, indicating a strong repressive effect of asRNA transcription on mRNA transcription. Indeed, as observed in Figure 4, the expression time courses of the mRNAs and asRNAs present inversed expression patterns, which is compatible with the mutual repression of sense and antisense transcription. We tested this hypothesis by interrupting the mRNA transcription of the *PET10*, *SHH3*, *ARO10* and *MOH1*. Figure 7A shows that in all cases but for *MOH1*, restricting mRNA transcription elongation led to an up-regulation of the corresponding asRNA at 48 hours (post-diauxic shift). The absence of effect observed for *MOH1* can be understood by the fact that its polyA site is the only of the four gene analysed, which is located upstream of its associated asRNA TSS (see Figure EV4). The mutual repression of sense and antisense transcription can thus be observed but will depend on *locus*-specific architecture.

As shown above, the mRNAs and their associated asRNAs exhibited an inverse pattern of expression between the exponential phase and quiescence, as expected if the expression of sense and antisense were mutually exclusive. Q-enriched genes were associated with more antisense transcription than average during the exponential phase, when these genes are repressed. Conversely, one could thus expect that they would be associated with less asRNA level than average in quiescence since they are the genes whose mRNAs are most abundant under this condition. This turned out not to be the case. Indeed, the Q-enriched genes remained associated with slightly higher asRNA levels than average even during quiescence (Figure 7B). This is consistent with the observation that there is no obligatory repression of asRNA transcription when sense transcription is induced and with the observation that the Q-enriched genes are, overall, more associated with asRNAs than other genes.

The analysis of sense and antisense expression in quiescence also revealed that, globally, if the mRNA levels strongly decreased in quiescence, as expected, the global level of asRNAs did not change (Figure EV5). A likely explanation is that mRNA transcription globally interferes with pervasive asRNA transcription during exponential phase. The global repression of transcription in quiescence (McKnight *et al*, 2015) could then be compensated for the asRNAs by the decrease of interference by mRNA transcription. This was verified at the *HIS1* locus, where the strong asRNA observed only in quiescence (Figure EV6A) could be revealed during the exponential phase by interrupting the *HIS1* mRNA transcription by a strand specific NNS terminator insertion (Figure EV6B).

Discussion

The discovery of an extensive pervasive transcription in yeast raised the question of whether it merely reflects transcription "noise" or whether it carries a widespread function. A large fraction of the yeast pervasive non-coding RNA transcription generates "cryptic" transcripts, which are rapidly degraded by the efficient NNS and NMD surveillance pathways. The NNS pathway restricts RNA PolII elongation of non-coding pervasive transcripts in order to prevent it from interfering with gene expression (Arigo *et al*, 2006; Thiebaut *et al*, 2006; Schulz *et al*, 2013). Yet, a very large number of pervasive transcription initiation events give rise to transcripts that

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

escape, at least in part, this first nuclear surveillance pathway. These transcripts are then terminated by the cleavage and polyadenylation machinery and exported to the cytoplasm as regular capped and polyadenylated RNAs. These RNAs, while non-coding in the sense of coding for functional proteins, all carry spurious ORFs and are thus engaged in translation. These short spurious ORFs are followed by long non-coding 3'-UTR like sequences. These transcripts thus exhibit all the characteristic features of efficient NMD targets and are therefore rapidly degraded by this cytoplasmic pathway (Malabat *et al*, 2015). Many of these transcripts are elongated from NFRs located at the 3'-end of genes. Those that are not fully terminated early by the NNS pathway can be elongated up to the promoter region of the corresponding gene, potentially interfering with their expression.

Targeted studies have previously described specific examples in which the transcription of a non-coding RNA was mediating gene regulation (see for reviews Donaldson & Saville, 2012; Castelnuovo & Stutz, 2015). To what extent this phenomenon results in a significant repression of the corresponding genes genome-wide is a question that has been recently addressed by Huber and colleagues, who performed a large-scale study focusing on the SUTs. This analysis concluded that less than 25% of the 162 mRNA/asRNA pairs analysed showed a weak suppressive effect of the transcription of antisense SUTs on the expression of the corresponding protein, in at least one condition (Huber *et al*, 2016). Antisense transcripts overlapping the mRNA TSS of weakly expressed genes were the most likely to exert such a suppressive effect.

This analysis, whatever comprehensive, is nevertheless restricted to a relatively small fraction of the naturally occurring antisense transcription since it focused only on SUTs. Analysing the transcriptome of NMD deficient cells (*upf1*Δ), we classified 816 genes as being associated with asRNAs (Figure 1). It was previously reported that the levels of antisense transcription were not correlated with that of the corresponding mRNAs (Murray *et al*, 2015). Figure 1A plots the respective levels of asRNAs and corresponding mRNAs in *upf1*Δ cells. While, similar to the Murray analysis, a linear regression analysis did not reveal any correlation between the two datasets (Pearson correlation coefficient R^2=0.07), we nevertheless observed that many of the least expressed genes were associated with high antisense levels. Dividing the 5892 genes of the dataset in ten bins of equal size, we found that the least expressed genes, corresponding to bin1, were significantly more associated with asRNAs than the other genes (Figure 1B). This could result in part from the fact that, in contrast to well-expressed genes, the asRNA transcription of the least expressed genes is not repressed by sense transcription. Yet, almost 78% (181/233) of these asRNAs overlapped the mRNA TSS of their corresponding genes in bin 1 and this percentage was very significantly higher than in other bins (Figures 1E, EV1 and Table EV2). Since this class of asRNAs has been shown to be more prone to suppress gene expression when compared to non TSS-overlapping asRNAs (Huber *et al*, 2016), it suggested that these least expressed genes could frequently be repressed by asRNA transcription interference. This idea was further strengthened by the observation that TSS overlapping asRNA associated genes are much more frequently up-regulated upon *SET2* deletion than average (Figure 2A) and this biais is for a large part contributed by genes in bin 1 (Figure 2B). Other chromatin modification factors, such as *SET1*, *RCO1* and *EAF3* showed the same trends (Figure EV2). Interestingly, while asRNA transcription mediated regulation was found to affect different genes of very diverse pathways, we uncovered that genes whose mRNAs accumulated in quiescence (Q-enriched genes) were particularly enriched in

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bin 1 and in the class of genes associated with TSS-overlapping asRNAs (Figures 3 and EV3).). It suggested that the full repression of this class of genes during the exponential phase often relies on interference by antisense transcription. We directly tested this hypothesis by analysing five of these Q-enriched genes associated with TSS-overlapping asRNA (*PET10*, *CLD1*, *MOH1*, *SHH3* and *ARO10*). For four out of these five genes, specifically interrupting asRNA transcription resulted in a strong induction of the corresponding mRNAs during the exponential phase (Figure 5). Interestingly, the only gene that did not respond was *ARO10* but it was also, out of the five genes analysed, the least repressed gene during the exponential phase (Figure 4). In addition, this was the only gene we analysed that was also analysed in the Huber study (Huber *et al*, 2016). Although they could not find a repressive effect of its associated asRNA in rich medium, they found it to be regulated by antisense transcription when the cells were grown in synthetic complete medium. It thus turns out that the TSS-overlapping asRNAs associated to all five Q-enriched genes we tested can have a repressive role on gene transcription.

Interestingly, we found the repression of the asRNA upon induction of the mRNA to be frequent, although not obligatory and depending on the fact that the induced mRNA transcription overlaps the asRNA TSS (Figure 7 and EV4). Overall, the asRNA levels remained high in quiescence, even slightly higher than average (Figure 7B, right panels). It suggests a model by which, in contrast to previously studied examples (see for examples Martens *et al*, 2004; van Werven *et al*, 2012), the asRNA expression is not regulated by specific transcription regulators. Rather these RNAs would be constitutively expressed, unless repressed by sense transcription when mRNAs are induced and overlap their TSSs. Their transcription would thus act “passively” as an amplifier of gene regulation, turning an non-induction into repression, as previously suggested for the *SUR7* gene (Xu *et al*, 2011). Consistent with this model, blocking asRNA transcription elongation during the exponential phase resulted in a 2.6 and 8.3 fold increase of *PET10* and *SHH3* mRNAs respectively (Figure 6B), which is markedly lower than the induction estimated by comparing the increase of their relative expression levels measured from the quiescence *versus* exponential phase transcriptome datasets (5.9 and 264.6 fold increase respectively in the *upf1*Δ background; dataset 1).

If, in a few instance, the asRNA itself was suggested to play a direct role in gene repression, in the majority of cases examined thus far this repressive effect was shown to be mediated in *cis* by antisense transcription interference, the asRNA being only a by-product of this process (see for review Donaldson & Saville, 2012). In the case where the asRNA itself is proposed to play a direct effector role, one of the main arguments in favour of this model is that the repression is observed only when the non-coding RNA is stabilised by mutating either the nuclear exosome (Berretta *et al*, 2008) or the cytoplasmic Xrn1 exonuclease (van Dijk *et al*, 2011). This is not the case for the examples we studied, the repressive effect being identical in a wild type or in a NMD deficient background, in which the asRNAs were stabilized. Using strand specific NNS terminators in a diploid strain, we directly confirmed, on the *PET10* locus, the prediction that the effects we observed act only in *cis*, which is a hallmark of the transcription interference mechanism (Figure 6A).

Interestingly, the presence of a *PET10* asRNA with an inversed expression profile compared to the mRNA was shown to be conserved in all five analysed *Saccharomyces* species, supporting its functional role (Yassour *et al*, 2010). More

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

generally, a phylogenetic conservation study of lncRNAs in budding yeasts has shown that, since the divergence with *N. castellii,* which has retained a functional RNAi machinery, the level of asRNAs and their extent has globally increased. It was thus suggested that the lack of RNAi favoured the development of asRNA transcription mediated gene regulation (Alcid & Tsukiyama, 2016).

Transcriptional interference was previously shown in yeast to rely on the chromatin modifications by factors bound to the elongating PolII (for review Castelnuovo & Stutz, 2015). The genes up-regulated in the absence of factors such as *SET2*, *SET1*, *RCO1* or *EAF3* were found more often associated to TSS-overlapping asRNA than average (Figures 2, EV2), and this was particularly true for Q-enriched genes (Figures 3E and EV3B). The specific study of the *SHH3* gene by RT-qPCR showed that the suppressive effect of the antisense transcription involved a combination of several factors, acting in different chromatin modification pathways (*SET2*, *SET1*, *HDA1*; Figure 6B). Taken together, these results strongly suggest that the observed asRNA-transcription mediated repression involves several redundant chromatin modification/remodelling pathways. This is reminiscent of previous observations showing that gene silencing is mediated by redundant mechanisms involving multiple histone modifiers (Verzijlbergen *et al*, 2009). In agreement with this idea, each individual mutation could only promote a partial de-repression of the *SHH3* gene when compared to the level observed in the positive control strain, in which the asRNA transcription is interrupted by the NNS terminator (AS NNS, Figure 6B). It suggests the possibility that any gene, associated to a TSS-overlapping asRNA, might be subjected to a regulation by a specific combination of chromatin modifiers.

In our study, we demonstrated that TSS-overlaping antisense-mediated transcriptional interference is a frequent mechanism used for full gene repression. This mechanism is often hidden since these antisense transcripts are rapidly degraded by the NMD pathway and therefore not detected in wild type conditions. We and others reported the existence of conditional asRNAs, such as for example the *HIS1* antisense RNA, specifically expressed during G0 (Figure EV6A), or Meiotic Unannotated transcripts (MUTs; Lardenois *et al*, 2011). In addition, asRNAs were shown to mediate protein expression regulation depending on various growth conditions (Huber *et al*, 2016). Widespread antisense transcription has thus the potential to repress the synthesis of sense RNA and participate to differential gene expression and adaptation to various environmental and growth conditions.

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and Methods

Detailed strains, cultures, libraries preparation, Northern blot and RT qPCR methods are detailed in Appendix supplementary Methods.

Data analysis

Illumina reads treatments

For RNAseq libraries, duplicated reads were first filtered out using fqduplicate (ftp://ftp.pasteur.fr/pub/gensoft/projects/fqtools/fqtools-1.1.tar.gz). Then sequencing error were corrected using Musket (Liu *et al*, 2013; version 1.1). Reads of bad quality were removed using fastq_qual_trimmer (https://github.com/ivars-silamikelis/fastq_qual_trimmer, version 1.0) with a threshold of 20. Illumina adpaters were finally removed using Flexbar (Dodt *et al*, 2012; version 2.7). After removal of the random sequence tag, resulting reads were mapped using bowtie (Langmead & Salzberg, 2012; version 2.2.3 with the following parameters: –N 1 –p 1 —no-unal –D 15 –R 2 –L 22 –I S,1,1.15) and a compilation of *S. cerevisiae* genome (S288C reference sequence, Release 64 obtained from the Saccharomyces Genome Database (SGD) [http://www.yeastgenome.org/]) and *S. pombe* genome (ASM294 reference sequence, v2.19 obtained from PomBase [http://www.pombase.org/]) as reference genomes.

For 3'Long SAGE libraries, duplicated reads were first filtered out using fqduplicate. Illumina adpaters were then removed using AlienTrimmer (Criscuolo & Brisse, 2013). Reads corresponding the 3' end of transcripts were identified by detection of a polyA sequence at the end of the reads with a minimal size of 6 nucleotides. After Poly A removal, the resulting reads were mapped using bowtie (same version and parameters that above) and the *S. cerevisiae* genome (S288C reference sequence, Release 64 obtained from the Saccharomyces Genome Database (SGD) [http://www.yeastgenome.org/]). False positive reads (*ie* reads identified by a ≥ 6nt encoded polyA sequence but not a true 3'end) were filtered out by matching with encoded PolyA sequence in the genome.

Mapped reads processing

For 3' Long SAGE libraries the 3′-end positions of the resulting mapped reads were used as TTS positions and extracted to wig files. For RNAseq libraries, reads corresponding to the whole transcripts and full read coverage were extracted to wig files.

Normalisation and differential expression

Transcript differential expressions were calculated using DESeq2 (4) within the SARTools pipeline (Varet *et al*, 2016; version 1.4.1).
Sample corresponding to cells in exponential phase were first treated together as a separated group, as well as samples corresponding to cells in G0. During the process, SARTools performed a normalisation step. Normalisation factors were extracted and used to produce normalized wig files.
G0 samples were normalized in a second time against exponential phase sample

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

using the spike-in of *S. pombe* transcripts. *S. pombe* transcripts median reads counts were determined for each sample after the first normalisation step. Then a global mean for *S. pombe* transcripts reads counts was calculated for quiescent and exponential phase samples. A ratio Exponential/Quiescent was calculated and applied to all G0 samples (wig files and transcripts reads counts).

Antisens / mRNA coverage was counted and visualized in a -50-+200 nucleotides windows using the Counter RNAseq window (CRAW) package version 0.9.0 (https://pypi.python.org/pypi/craw/0.9.0).

Accession number

The data reported here have been deposited in NCBI GEO under the accession number GSE101368.

Acknowledgements

We thank Frank Feuerbach for providing strains LMA2811 and LMA2819.
We thank Cosmin Saveanu and Micheline Fromont-Racine for discussions and critical reading of the manuscript.
This work was supported by the Pasteur Institute (AD, CM, TK and BN), the Centre National de la Recherche Scientifique (AJ and GB). AN received a Fellowship from the French Ministry of Research and the Fondation pour la Recherche Médicale. We acknowledge Jean Yves Coppee and Caroline Proux for the facilities and expertise of the Transcriptomic Platform (PF2) for RNAseq experiments.

Author contributions

AN, AJ and GB designed research. AN, AD and GB performed the experiments. AN, CM, TK, BN AJ and GB analysed the data. AN, AJ and GB interpreted the data and wrote the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References and Citations

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

Allen C, Büttner S, Aragon AD, Thomas JA, Meirelles O, Jaetao JE, Benn D, Ruby SW, Veenhuis M, Madeo F & Werner-Washburne M (2006) Isolation of quiescent and nonquiescent cells from yeast stationary-phase cultures. *J. Cell Biol.* **174** 89–100

Aragon AD, Rodriguez AL, Meirelles O, Roy S, Davidson GS, Tapia PH, Allen C, Joe R, Benn D & Werner-Washburne M (2008) Characterization of Differentiated Quiescent and Nonquiescent Cells in Yeast Stationary-Phase Cultures. *Mol. Biol. Cell* **19** 1271–1280

Arigo JT, Eyler DE, Carroll KL & Corden JL (2006) Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* **23** 841–851

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

Carninci P & Hayashizaki Y (2007) Noncoding RNA transcription beyond annotated genes. *Curr. Opin. Genet. Dev.* **17** 139–144

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia W-J, Anderson S, Yates J, Washburn MP & Workman JL (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123** 581–592

Castelnuovo M, Rahman S, Guffanti E, Infantino V, Stutz F & Zenklusen D (2013) Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat. Struct. Mol. Biol.* **20** 851–858

Castelnuovo M & Stutz F (2015) Role of chromatin, environmental changes and single cell heterogeneity in non-coding transcription and gene regulation. *Curr. Opin. Cell Biol.* **34** 16–22

Castelnuovo M, Zaugg JB, Guffanti E, Maffioletti A, Camblong J, Xu Z, Clauder-Münster S, Steinmetz LM, Luscombe NM & Stutz F (2014) Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res.* **42** 4348–4362

Churchman LS & Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469** 368–373

Criscuolo A & Brisse S (2013) AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* **102** 500–506

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW & Steinmetz LM (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U. S. A.* **103** 5320–5325

van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Né P, Loeillet S, Nicolas A, Thermes C & Morillon A (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475** 114–

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

117

Dodt M, Roehr JT, Ahmed R & Dieterich C (2012) FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology* **1** 895–905

Donaldson ME & Saville BJ (2012) Natural antisense transcripts in fungi. *Mol. Microbiol.* **85** 405–417

Erdeniz N, Mortensen UH & Rothstein R (1997) Cloning-free PCR-based allele replacement methods. *Genome Res.* **7** 1174–1183

Geisler S, Lojek L, Khalil AM, Baker KE & Coller J (2012) Decapping of long noncoding RNAs regulates inducible genes. *Mol. Cell* **45** 279–291

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H & Oliver SG (1996) Life with 6000 genes. *Science* **274** 546, 563–567

Gu M, Naiyachit Y, Wood TJ & Millar CB (2015) H2A.Z marks antisense promoters and has positive effects on antisense transcript levels in budding yeast. *BMC Genomics* **16** 99

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

Huber F, Bunina D, Gupta I, Khmelinskii A, Meurer M, Theer P, Steinmetz LM & Knop M (2016) Protein Abundance Control by Non-coding Antisense Transcription. *Cell Rep.* **15** 2625–2636

Jensen TH, Jacquier A & Libri D (2013) Dealing with pervasive transcription. *Mol. Cell* **52** 473–484

Jiang C & Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10** 161–172

Keogh M-C, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ, Boone C, Emili A, Weissman JS, Hughes TR, Strahl BD, Grunstein M, Greenblatt JF, Buratowski S & Krogan NJ (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123** 593–605

Kim JH, Lee BB, Oh YM, Zhu C, Steinmetz LM, Lee Y, Kim WK, Lee SB, Buratowski S & Kim T (2016) Modulation of mRNA and lncRNA expression dynamics by the Set2-Rpd3S pathway. *Nat. Commun.* **7** 13534

Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9** 357–359

Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, Chu A, Davis RW, Steinmetz LM & 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

Liu Y, Schröder J & Schmidt B (2013) Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. *Bioinforma. Oxf. Engl.* **29** 308–315

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

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Martens JA, Laprade L & Winston F (2004) Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature* **429** 571–574

McKnight JN, Boerma JW, Breeden LL & Tsukiyama T (2015) Global Promoter Targeting of a Conserved Lysine Deacetylase for Transcriptional Shutoff during Quiescence Entry. *Mol. Cell* **59** 732–743

Murray SC, Haenni S, Howe FS, Fischl H, Chocian K, Nair A & Mellor J (2015) Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Res.* **43** 7823–7837

Murray SC, Serra Barros A, Brown DA, Dudek P, Ayling J & Mellor J (2012) A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res.* **40** 2432–2444

Nadal-Ribelles M, Solé C, Xu Z, Steinmetz LM, de Nadal E & Posas F (2014) Control of Cdc28 CDK1 by a stress-induced lncRNA. *Mol. Cell* **53** 549–561

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

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

Porrua O, Hobor F, Boulay J, Kubicek K, D'Aubenton-Carafa Y, Gudipati RK, Stefl R & Libri D (2012) In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.* **31** 3935–3948

Radonjic M, Andrau J-C, Lijnzaad P, Kemmeren P, Kockelkorn TTJP, van Leenen D, van Berkum NL & Holstege FCP (2005) Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon S. cerevisiae stationary phase exit. *Mol. Cell* **18** 171–183

Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J & Cramer P (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* **155** 1075–1087

Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J & 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

Tudek A, Porrua O, Kabzinski T, Lidschreiber M, Kubicek K, Fortova A, Lacroute F, Vanacova S, Cramer P, Stefl R & Libri D (2014) Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell* **55** 467–481

Varet H, Brillet-Guéguen L, Coppée J-Y & Dillies M-A (2016) SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PloS One* **11** e0157022

Vasiljeva L, Kim M, Mutschler H, Buratowski S & Meinhart A (2008) The Nrd1–Nab3–Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **15** 795–804

Venkatesh S, Smolle M, Li H, Gogol MM, Saint M, Kumar S, Natarajan K & Workman JL (2012) Set2 methylation of histone H3 lysine 36 suppresses histone exchange on

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transcribed genes. *Nature* **489** 452–455

Venters BJ & Pugh BF (2009) How eukaryotic genes are transcribed. *Crit. Rev. Biochem. Mol. Biol.* **44** 117–141

Verzijlbergen KF, Faber AW, Stulemeijer IJ & van Leeuwen F (2009) Multiple histone modifications in euchromatin promote heterochromatin formation by redundant mechanisms in Saccharomyces cerevisiae. *BMC Mol. Biol.* **10** 76

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

Wery M, Descrimes M, Vogt N, Dallongeville A-S, 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

Wyers F, Rougemaille M, Badis G, Rousselle J-C, Dufour M-E, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, Libri D & Jacquier A (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, Clauder-Münster S, Smolik M, Huber W & Steinmetz LM (2011) Antisense expression increases gene expression variability and locus interdependency. *Mol. Syst. Biol.* **7** 468

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

Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnirke A, Nusbaum C, Thompson D-A, Friedman N & Regev A (2010) Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biol.* **11** R87

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends

Figure 1 - Antisense ncRNAs are over-represented in lowly expressed genes.

A Scatter plot representing the antisense level (ordinate) function of the corresponding mRNA level (abscissa) in log_{10} read/base. Gene read count was determined for 5892 genes, and divided into ten bins (grey strips) of equal length (N = 589 genes per bin for bin1 to bin9; N=591 genes for bin10). The Pearson correlation coefficient $R^2 = 0.07$.

B Comparison of the average antisense level distribution between bins. Boxplots shows the distribution of the average antisense levels within each bin. Brackets indicate the results of an Anova test on pairs of distributions, with *** for P < 0.001.

C Schematic of the gene associated promoter class categories depending of the presence and the characteristics of asRNA : N = No asRNA, M = asRNA within the mRNA, O = TSS-overlapping asRNA.

D Heatmap distribution of mRNA and antisense around the TSS of all genes, sorted by antisense and promoter class categories. Depending the class of promoter defined in C, a category N, M or O was assigned to each gene.

E Promoter class categories count per bin. The total number of genes that belong to each class of promoter is indicated (Class N : N=5076; Class M : N=259; Class O : N= 557). Bar charts represent the percentage of each class within the 10 bins defined in A (see also dataset 1). Brackets indicate the results of a statistical inference test on pairs of distributions between bin1 and each other bins, with *** for P < 0.001.

Figure 2 - Promoter overlapping antisenses are overrepresented in *set2*Δ targets.

A Gene distribution across promoter categories in *set2*Δ up-regulated genes (dataset from Malabat *et al*, 2015). Stacked histograms represent the proportion of *set2*Δ up-regulated genes across all genes (All), or depending the presence of a TSS-overlapping asRNA ("+" = class O) or not ("-" = classes N+M). Brackets indicate the results of a statistical inference test on pairs of distributions, with *** for P < 0.001, n.s. for "not significant" (P>0,05).

B *Set2*Δ up-regulated genes count depending the presence of a TSS-overlapping asRNA or not and per bin. Brackets indicate the results of a statistical inference test on pairs of distributions between bin1 and each other bins, with *** for P < 0.001

Figure 3 - Quiescent-enriched genes are associated with high antisense level.

A Bar plot of stationary phase-enriched genes count (SP-enriched) versus other genes count (not SP-enriched) within each bin (dataset from Radonjic *et al*, 2005).

B Distribution of Quiescence-enriched genes among the 5892 yeast genes. Scatter plot of the antisense level as a function of the corresponding mRNA level. 261 genes

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were found enriched at least 5 times between exponential and quiescence, defining the Quiescence-enriched genes (Q-enriched, green dots, see also Dataset 1)

C Bar chart of the 261 Q-enriched genes within the 10 bins.

D Distributions of antisense level for different gene categories in exponential phase. Boxplots show the mean antisense level of 261 corresponding Q-enriched genes (green) or "Random" (grey) genes. Random-1, -2, and -3 were defined by random sampling of 261 genes among all the 5892 genes. "Bin_1" (blue) or "All" categories (black) are the measures of all 589 genes from bin1 or all 5892 genes respectively. Brackets indicate the results of an Anova test on pairs, with * for $P < 0.05$, and *** for $P < 0.001$.

E *set2*Δ up-regulated genes count among Q-enriched genes per promoter class and bin. The bar charts represent the count of *set2*Δ up-regulated genes within each category of promoter and each bin (see also Dataset 1).

Figure 4 - Q-enriched genes mRNA and corresponding asRNAs are anti-regulated.

Northern-Blot expression time course of mRNA and antisense transcripts in a Δ*upf1* strain for five examples of Q-enriched genes : *PET10*, *SHH3*, *MOH1*, *CLD1* and *ARO10*. Time point 0' is the time at which quiescent-arrested cells are restarted in rich YPD medium. *SCR1* is used as a loading control. RNA probes are described in Figure EV4 and Table EV4.

Figure 5 - Antisense transcription interruption during the exponential growth relieves repression of quiescence-enriched genes.

Northern blot analysis of *PET10*, *CLD1*, *MOH1*, *SHH3* and *ARO10* mRNA and antisense RNAs in the WT and Δ*upf1* strains with (+) or without (-) the insertion of an antisense Nrd1-Nab3-Sen1 terminator (AS NNS). *SCR1* is used as a loading control. RNA probes and NNS insertion are described in Figure EV4 and Table EV4 (see also Material and Methods for strain construction and AS NNS-corresponding strains in Appendix table 1).

Figure 6 - Antisense repression is mediated by transcriptional interference mechanisms.

A Strand-specific RT-qPCR analysis of *PET10* mRNA and antisense RNA abundance in diploid strains. PET10 antisense is transcribed *in cis* (blue), in trans (green) or not produced (red).

B Strand-specific RT-qPCR analysis analysis of *PET10* (upper panel) and *SHH3* (lower panel) mRNAs and antisense abundances in a mutant strain where the deletion of *UPF1* (ref. strain) is either combined to an antisense NNS terminator insertion (AS NNS, purple; positive control), or to the deletion of a chromatin modification factor (*set1*Δ*, hda1*Δ and *set2*Δ*).*

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 7 - Gene expression is repressive for antisense non-coding transcription.

A Northern blot analysis of *PET10*, *MOH1*, *SHH3* and *ARO10* mRNA and antisense RNAs in Δ*upf1* strain, after 24h or 48h of growth in YPD and with (+) or without (-) the insertion of a sense Nrd1-Nab3-Sen1 terminator (NNS S). RNA probes and NNS insertion are described in Figure EV4 and Table EV4 (see also Material and Methods for strain construction) and S NNS-corresponding strains in Appendix table 1. *SCR1* is used as a loading control.

B Comparison of density plots between all (black lines) and Q-enriched genes (green lines) for mRNAs (left panels) or associated asRNA (right panels) from cultures harvested in exponential (upper panels) or G0 (lower panels) phases. Log_{10} RNA levels are plotted (abscissa) function of the frequency (ordinate).

Figure EV1 - Heatmap of asRNA containing genes categories sorted by bin.

Figure EV2 - Promoter overlapping asRNA associated genes are overrepresented in chromatin modifiers mutants up-regulated targets (datasets from Churchman & Weissman, 2011):

A, B, C, D : Gene count across promoter categories in chromatin modifier mutants up-regulated genes (Datasets from Churchman & Weissman, 2011). Stacked histograms represent the proportion of respectively *set2*Δ, *set1*Δ*, rco1*Δ and *eaf3*Δ up-regulated genes across all genes (All genes), or depending the presence of a TSS-overlapping asRNA ("+" ; class O) or not ("-" ; classes N+M). Brackets indicate the results of a statistical inference test on pairs of distributions, with *** for P < 0.001

E Overlap of up-regulated genes between the different mutants (respectively *set2*Δ, *set1*Δ*, rco1*Δ and *eaf3*Δ) in (Churchman & Weissman, 2011). The table represents the gene overlap between each chromatin modifiers mutants (see also Dataset 1).

Figure EV3 - Q-enriched genes are more enriched in TSS-overlapping asRNA containing genes.

A Comparison of promoter class categories count per bin between Q-enriched genes and all genes.

The total number of genes that belong to each category is indicated (Q-enriched : N=261, All genes : N=5892). The Bar charts represent the percentage of each class within either the 261 Q-enriched or the 10 bins shown Figure 1E (see also dataset 1). Brackets indicate the results of a statistical inference test on pairs of distributions between bin1 and each other bins, with "n.s." for not significant (P > 0.05), ** for P < 0.01, and *** for P < 0.001.

B Comparison of *set2*Δ up-regulated genes between Q-enriched genes and all genes Stacked histograms represent the proportion of *set2*Δ up-regulated genes across all genes (All), or depending the presence of a TSS-overlapping asRNA ("+" = class O) or not ("-" = classes N+M). Brackets indicate the results of a statistical inference test on pairs of distributions, with ** for P < 0.01 and *** for P < 0.001.

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure EV4 – Schematic organisation around *PET10, SHH3, MOH1, CLD1* and *ARO10 loci.*

Figure EV5 – mRNA and asRNAs global levels of expression in exponential phase and G0.

Comparison of density plots between for mRNAs (blue lines) or associated asRNA (orange lines) from cultures harvested in exponential (upper panels) or G0 (lower panels) phases. Log_{10} RNA levels are plotted (abscissa) function of the frequency (ordinate).

Figure EV6 – *HIS1* associated asRNA is induced in quiescence or when HIS1 mRNA transcription is interrupted.

A, B Northern blot analysis of *HIS1* mRNA and antisense RNAs in WT and Δ*upf1* strains in exponential phase or quiescence (A) or in exponential phase with (+) or without (-) the insertion of a sense Nrd1-Nab3-Sen1 terminator (NNS S ; B). RNA probes and NNS insertion are described in Figure EV4 and Table EV4 (see also Material and Methods for strain construction) and S NNS-corresponding strains in Appendix table 1. *SCR1* is used as a loading control.

bioRxiv preprint doi: https://doi.org/10.1101/169730; this version posted August 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 7

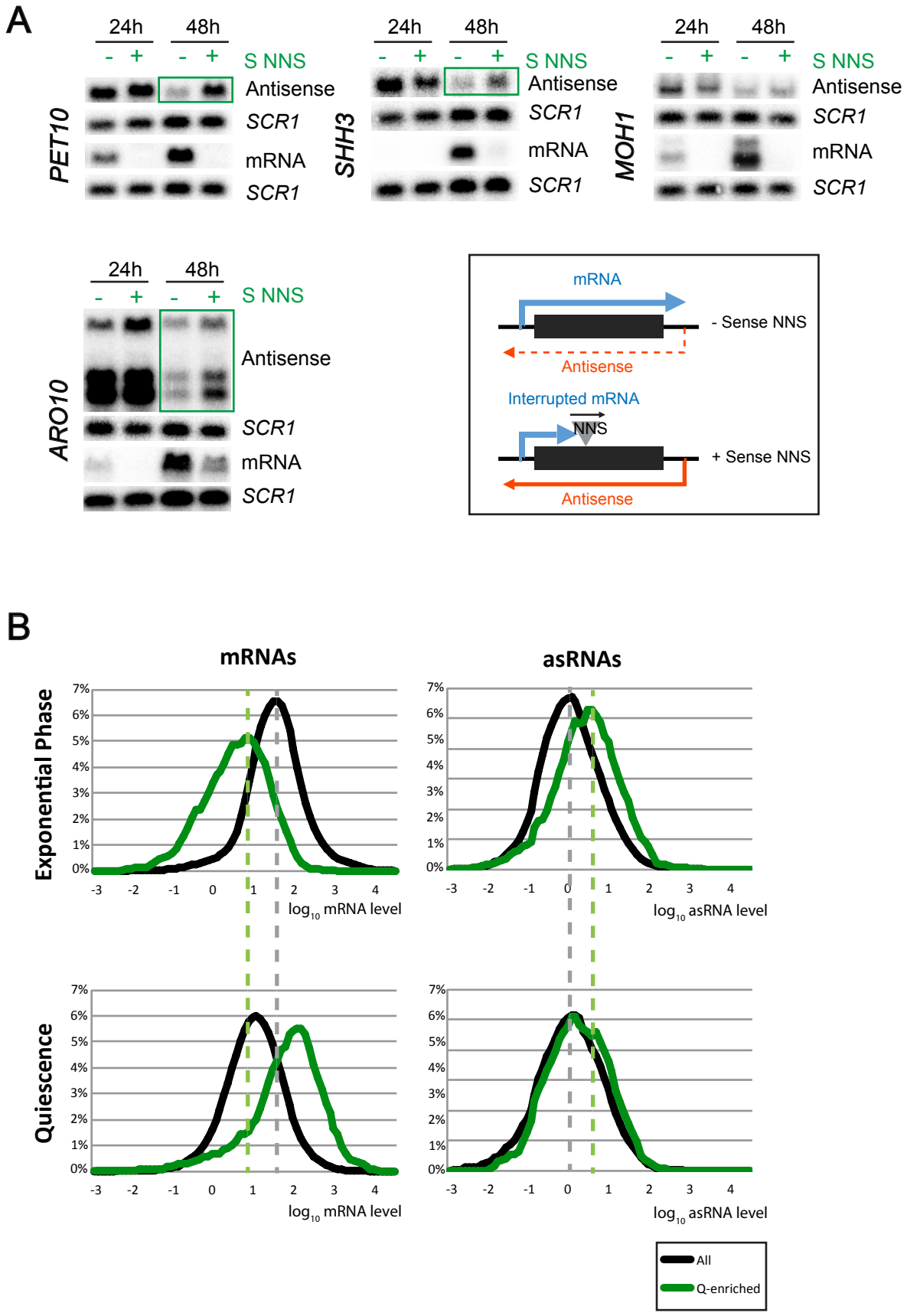
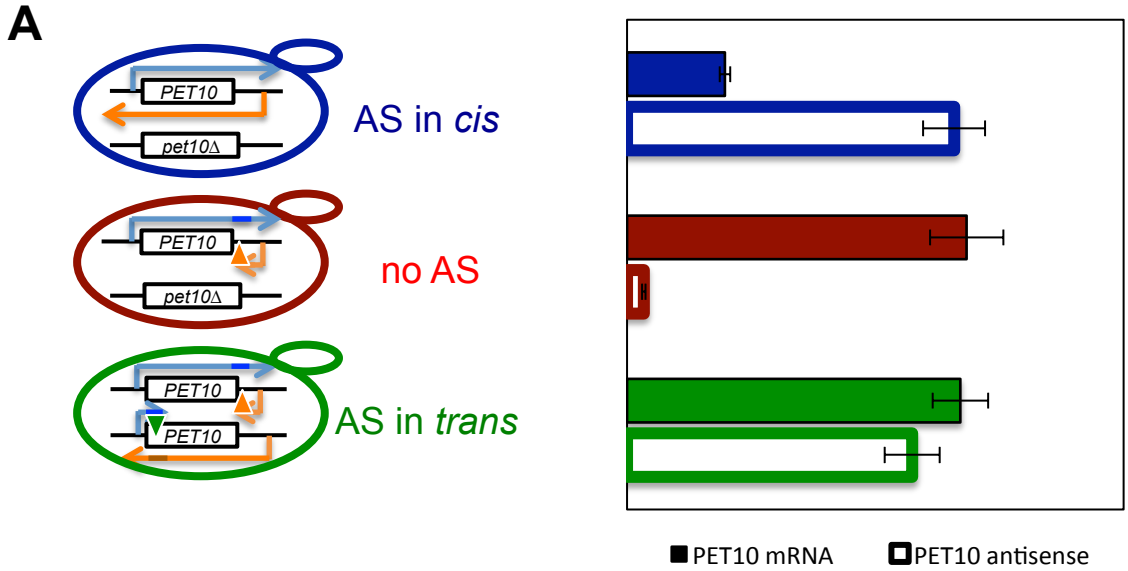


Figure 6


A
PET10
pet10Δ
AS in cis
PET10
pet10Δ
no AS
PET10
PET10
AS in trans
PET10 mRNA
PET10 antisense

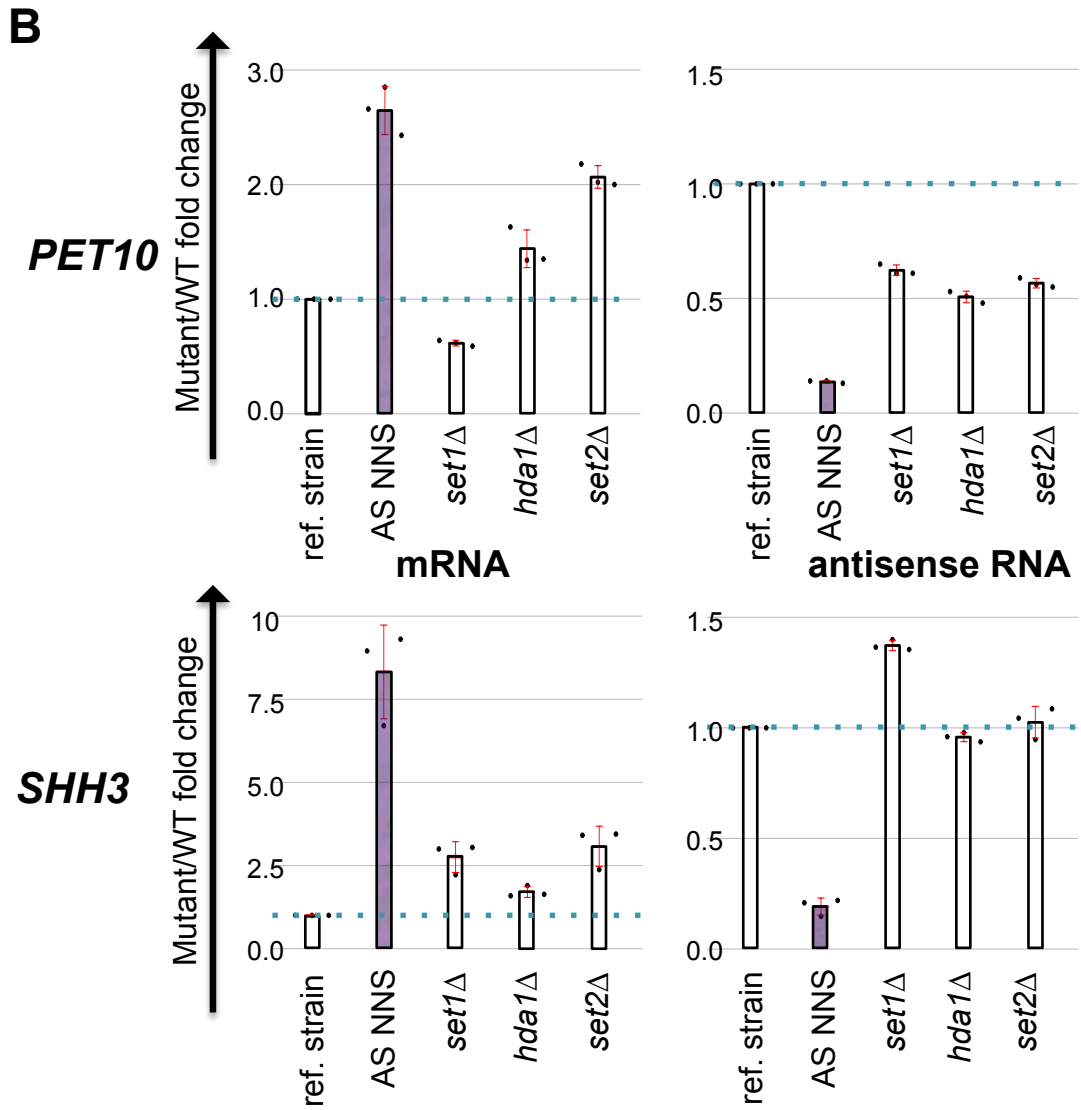

B
PET10
Mutant/WT fold change
3.0
2.0
1.0
0.0
ref. strain
AS NNS
set1Δ
hda1Δ
set2Δ
mRNA
1.5
1.0
0.5
0.0
antisense RNA
SHH3
10
7.5
5.0
2.5

Figure 5

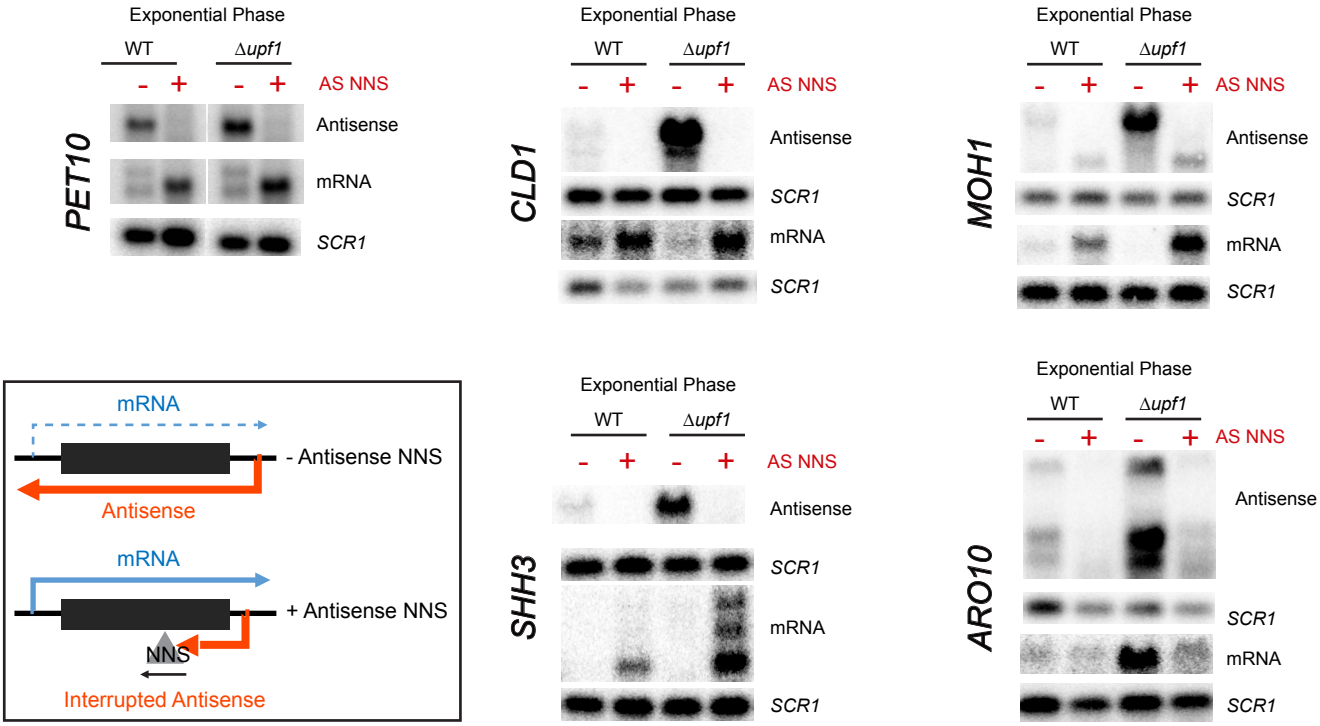

Exponential Phase
WT
Δupf1
- + - +
AS NNS
PET10
Antisense
mRNA
SCR1
CLD1
MOH1
SHH3
ARO10
mRNA
- Antisense NNS
Antisense
+ Antisense NNS
NNS
Interrupted Antisense

Figure 4

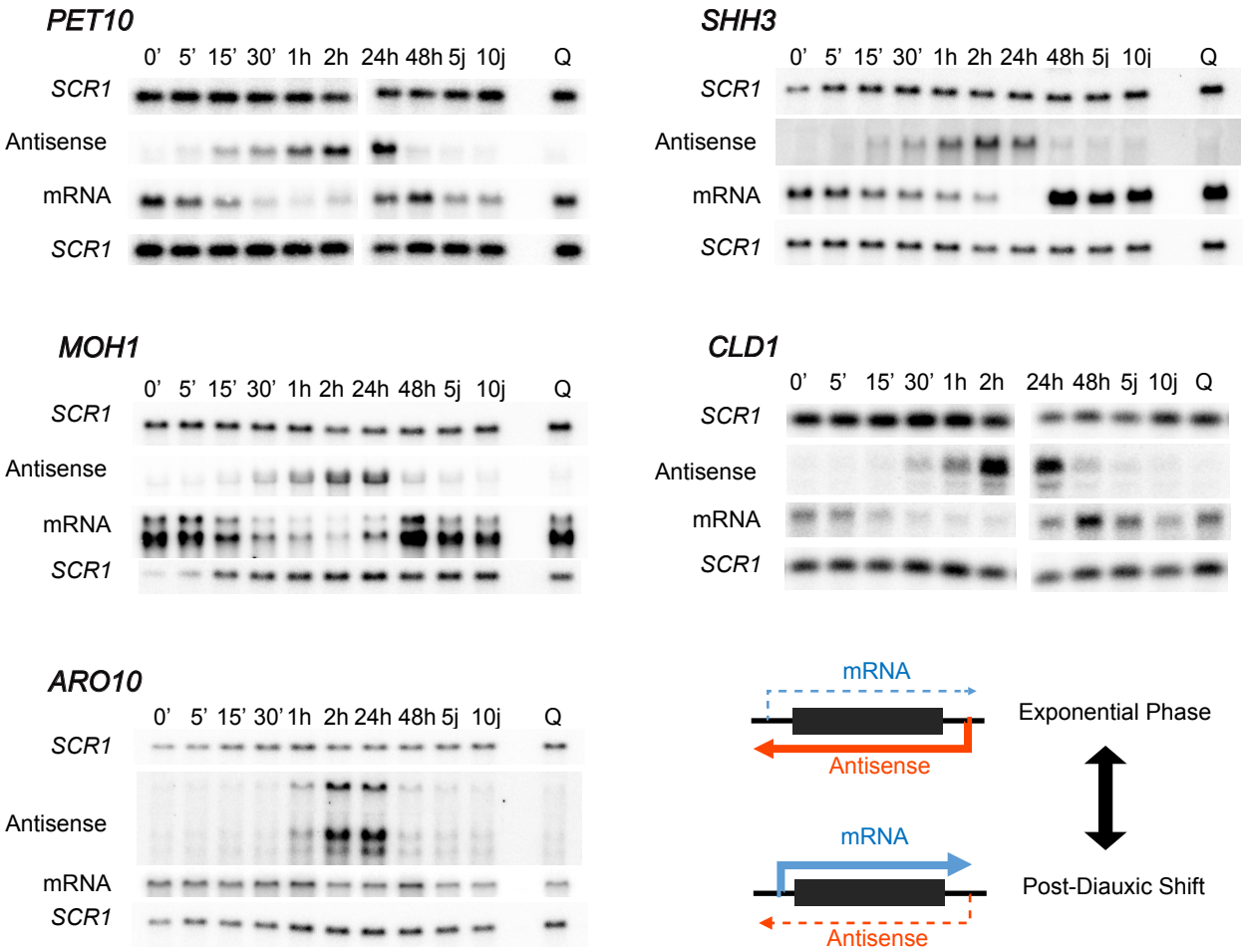

PET10
0' 5' 15' 30' 1h 2h 24h 48h 5j 10j Q
SCR1
Antisense
mRNA
SCR1
SHH3
0' 5' 15' 30' 1h 2h 24h 48h 5j 10j Q
SCR1
Antisense
mRNA
SCR1
MOH1
0' 5' 15' 30' 1h 2h 24h 48h 5j 10j Q
SCR1
Antisense
mRNA
SCR1
CLD1
0' 5' 15' 30' 1h 2h 24h 48h 5j 10j Q
SCR1
Antisense
mRNA
SCR1
ARO10
0' 5' 15' 30' 1h 2h 24h 48h 5j 10j Q
SCR1
Antisense
mRNA
SCR1
mRNA
Antisense
Exponential Phase
mRNA
Antisense
Post-Diauxic Shift

Figure 1

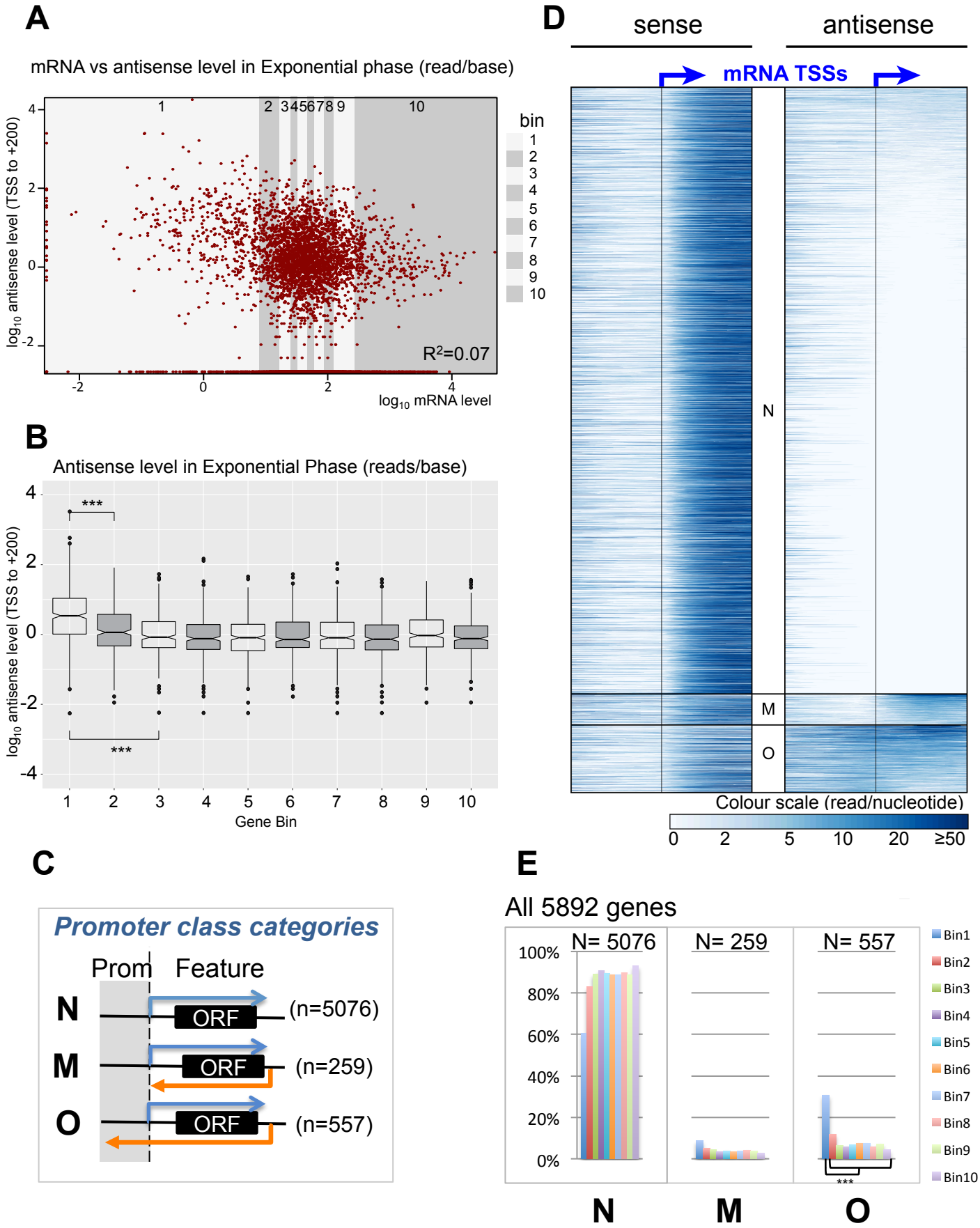

A
mRNA vs antisense level in Exponential phase (read/base)
log10 antisense level (TSS to +200)
log10 mRNA level
R2=0.07
bin
1
2
3
4
5
6
7
8
9
10
B
Antisense level in Exponential Phase (reads/base)
log10 antisense level (TSS to +200)
Gene Bin
***
***
C
Promoter class categories
Prom
Feature
N
M
O
ORF
(n=5076)
(n=259)
(n=557)
D
sense
antisense
mRNA TSSs
N
M
O
Colour scale (read/nucleotide)
0
2
5
10
20
≥50
E
All 5892 genes
N= 5076
N= 259
N= 557
100%
80%
60%
40%
20%
0%
Bin1
Bin2
Bin3
Bin4
Bin5
Bin6
Bin7
Bin8
Bin9
Bin10
***
N
M
O

Figure 2

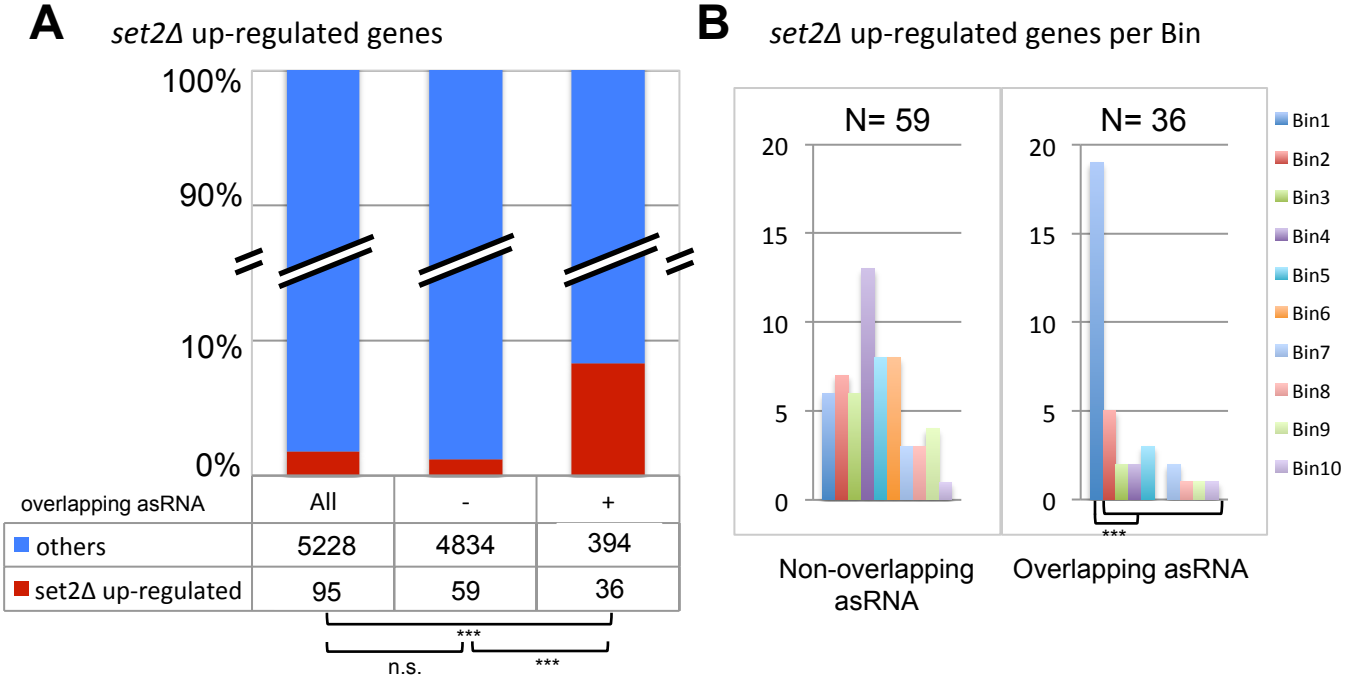
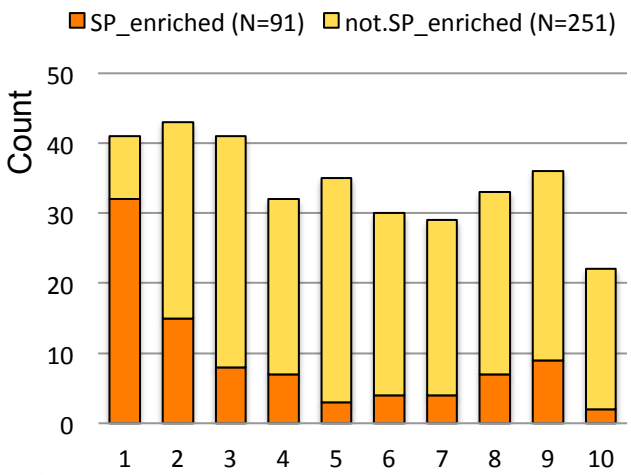
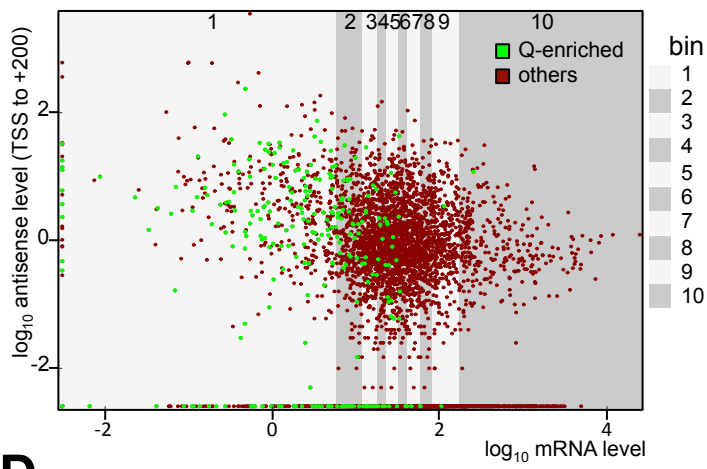

A
set2Δ up-regulated genes
100%
90%
10%
0%
overlapping asRNA
All
-
+
others
5228
4834
394
set2Δ up-regulated
95
59
36
***
n.s.
***
B
set2Δ up-regulated genes per Bin
N= 59
N= 36
20
15
10
5
0
***
Bin1
Bin2
Bin3
Bin4
Bin5
Bin6
Bin7
Bin8
Bin9
Bin10
Non-overlapping asRNA
Overlapping asRNA

Figure 3

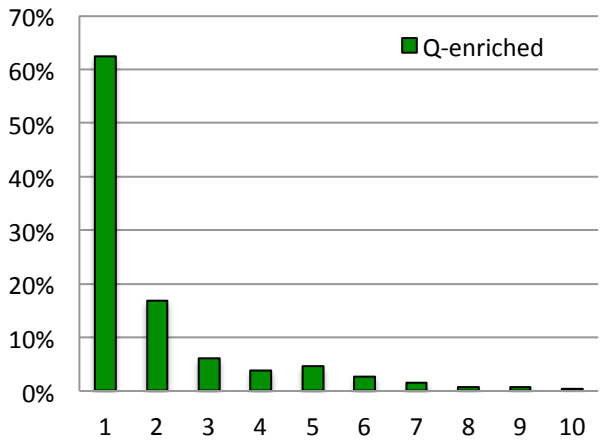
A Bin repartition (from Radonjic *et al.* Dataset)



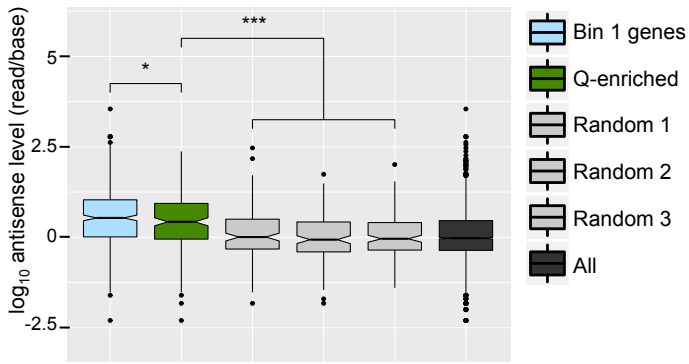
B mRNA vs antisense in Exponential phase (read/base)



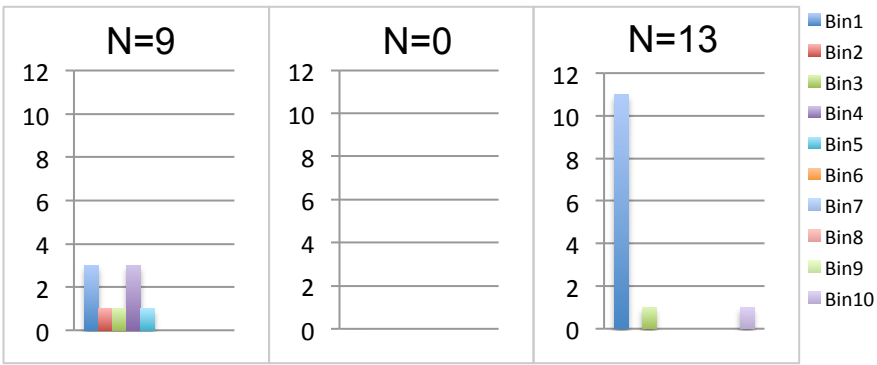
C Bin repartition of Q-enriched genes (N=261)



D Antisense level in exponential phase



E *Set2*Δ up-regulated genes among Q-enriched genes per bin and promoter class



N **M** **O**