The Regulation and Evolution of long, noncoding RNA Transcription

Eric Alcid

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Toshio Tsukiyama, Chair

Marshall Horwitz

Robert K. Bradley

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

The Regulation and Evolution of long, noncoding RNA Transcription

Eric Alcid

Chair of the Supervisory Committee:

Toshio Tsukiyama, Member

Division of Basic Sciences, Fred Hutchinson Cancer Research Center

Abstract Advances in genomics techniques, such as high-resolution tiled microarrays and deep sequencing, have revealed that transcription is highly pervasive. This suggests that transcription occurs not only at protein coding genes (open reading frames, or ORFs), but also at unannotated portions of the genome, leading to the production of thousands of long, noncoding RNAs (lncRNAs, > 200 basepairs), which are distinct from more traditional noncoding RNAs such as transfer RNAs or ribosomal RNAs. This striking finding is observed across all eukaryotes, from humans to budding yeast. It was once thought that these lncRNAs were largely the product of stochastic transcription. However, through detailed functional analyses, it is clear that some lncRNAs play key regulatory roles. For instance Xist is a lncRNA necessary for X-chromosome inactivation in higher eukaryotes. In budding yeast, lncRNAs produced at the *GAL10* and *PHO84* locus are needed for attenuating transcription of the overlapping mRNA. Though these examples illustrate that we understand the functions of some lncRNAs, the vast majority of lncRNAs have as-of-yet unidentified biological roles. Moreover, even less is known about how lncRNAs are regulated or the conditions underlying their evolution. This

dissertation describes the use of budding yeast as a model to elucidate fundamental principles of lncRNA regulation and evolution. I first describe the development of a high-throughput genetic screen in the budding yeast, *Saccharomyces cerevisiae,* that takes advantage of synthetic growth defects of mutants when RNA interference is reconstituted to identify 408 putative repressors of lncRNAs. Among these putative hits were four highly-conserved chromatin remodeling factors: Swr1, Isw2, Rsc, Ino80. I then use strand-specific, high-throughput RNA sequencing (ssRNAseq) to identify the lncRNA targets of these complexes. Further, I go on to show that these factors largely regulate distinct populations of lncRNAs genome-wide, that a subset of these lncRNAs are directly regulated by these remodeling factors (termed chromatin remodeling regulated antisense transcripts, CRRATs), and that some CRRATs might function to regulate the mRNA that they overlap with. Next, I describe the use of comparative genomics in five species of budding yeast to show that, since the loss of RNAi in the budding yeast lineage, levels of antisense, long noncoding RNAs (ASlncRNAs) have gradually risen genome-wide. I then identified a subset of ASlncRNAs that are highly conserved at the level of expression among the *sensu stricto* lineage of budding yeasts, and I assign putative biological roles for these ASlncRNAs using gene ontology. Finally, I show, using *S. cerevisiae* and *S. castellii* as models, that RNAi likely attenuates ASlncRNAs across the genome. Using genetic data, I suggest that this is likely due to deleterious effects when ASlncRNAs are elevated in the presence of an active RNAi pathway. lncRNAs have been shown to play key regulatory roles across many different eukaryotes. Understanding how lncRNAs are regulated and how they might have evolved will provide insights into their biological functions and their roles in disease states such as cancer.

Acknowledgements

There are a number of people that I must thank, without whom I likely would never have been able to finish my thesis work at all, much less in a relatively timely manner.

My first PI, Pradip Masharak, and my direct mentor in college, Mike Rose, were instrumental in developing my initial interest in laboratory science. They gave me projects that would not overwhelm me, and would help my confidence grow, and they wholeheartedly supported my long-term plans. Mike was especially great for restoring my confidence after I would make completely boneheaded mistakes.

Melissa Jurica, for whom I was a technician, is the single biggest reason I decided to pursue science. She took a chance on me after graduation despite no experience in structural biology research, she taught me how to develop proper controls, and, I think most importantly, she showed me how to push my project to completion when everything seems to be working and take advantage of a “hot streak”. There is no doubt that I learned how to work smart and work hard in the lab because of you. Furthermore, she was kind and generous with advice, and still serves as a great role model for balancing family with science.

My thesis advisor, Toshi Tsukiyama, was great throughout my graduate school career. He was incredibly hands-on in the beginning, which I definitely needed as I had no experience working in budding yeast. Once I grasped the techniques, he pushed me hard to establish my independence, and to take both of my projects in the directions that I want. He also taught me how to write in simple, elegant prose that concisely conveys exactly what I want, which might be the second most important skill to gain in graduate school after the execution of experiments. However, I think his mentorship was at its

best when I struggled. His excitement and enthusiasm for my projects never wavered, and he has an uncanny ability to motivate me and excite me after a prolonged losing streak. I have no doubt in my mind that I made the right decision to join his lab.

I must thank my family, especially my mom. My mom worked two jobs while I was growing up so that my brother and I could go to college. She's always supported my goals in life, and taught me that passion is more important than salary as long as the salary is somewhat livable. Thank you for all of the sacrifices that you have made for Alex and me.

And my wife Janine is simply incredible. She makes sure that I eat well, that I sleep well, that my powerpoint presentations do not look sloppy, and that I keep my enthusiasm for medicine and science even as I enter my 30s still a student. She does this all while being a new mother and finding a way to thrive in her own scientific career. Without you, I never would have been able to finish my PhD.

Table of Contents

Chapter 1: Introduction **9**
- **Regulatory lncRNAs in eukaryotes** **9**
- **Characteristics, expression level, degradation of lncRNAs** **12**
- **Chromatin introduction** **13**
- **Regulation of lncRNAs by Chromatin Remodeling** **14**
  - Isw2 represses lncRNAs originating at the 3' ends of genes 14
  - Chromatin remodeling represses lncRNAs in a stem cell-specific context 15
  - The histone variant H2A.Z represses antisense lncRNAs at convergent genes 16
  - Histone modifying enzymes and gene looping are associated with lncRNA repression 17
- **Development of Genetic Screens to identify novel lncRNA repressors- an introduction** **18**
  - The identification of suppressors of intragenic cryptic transcription 18
  - The identification of genes that promote promoter directionality 20
- **Evolutionary studies of lncRNA expression** **21**
- **Research Aims** **22**

Chapter 2: ATP-dependent Chromatin-remodeling Shapes the Long-Noncoding RNA Landscape **25**
- **Introduction** **26**
- **RESULTS** **29**
  - Genetic interactions between RNAi and mutations that elevate lncRNA levels 29
  - RNAi as a tool for systematic identification of lncRNA repressors 31
  - Four highly conserved ATP-dependent chromatin-remodeling complexes repress ASlncRNAs 34
  - ATP-dependent chromatin-remodeling factors repress unique sets of ASlncRNAs 36
  - Identification of ASlncRNAs directly repressed by chromatin-remodeling factors 38
  - Identification of ASlncRNAs with regulatory roles 43
- **Discussion** **45**
  - A novel genetic screen for ASlncRNA repressors 46
  - A large number of putative lncRNA repressors 47
  - Identification of lncRNAs with regulatory functions 48
- **Materials and Methods** **50**

Chapter 3: RNA interference constrains expansion of the anti-sense long, non-coding RNA transcriptome **56**
- **Abstract** **57**
- **Introduction** **57**
- **Results and Discussion** **59**
  - RNAi status is anti-correlated with expression of ASlncRNAs at GAL10 and PHO84 loci 59
  - Evolution of ASlncRNA transcriptomes across the budding yeast phylogeny 60
  - Expansion of ASlncRNA transcriptomes since divergence from S. castellii 62
  - Identification of conserved ASlncRNAs reveals species-specific bursts in expression and putative biological roles 65

Elevation of cryptic ASlncRNA transcriptomes after the loss of RNAi 67
Differential impact of RNAi on ASlncRNA levels in S. cerevisiae and S. castellii 69
Elevation of ASlncRNAs in the presence of RNAi leads to fitness defects in S. castellii. 70
Concluding remarks 71
Materials and Methods 72

Chapter 4: Perspectives and Conclusions 77
Chromatin structure maintenance and lncRNA regulation 77
Dedicated cellular resources to repress lncRNA genome-wide 78
Chromatin remodeling factors likely regulate functional lncRNAs 79
RNAi represses ASlncRNAs across evolutionary timescales 80
Rapid evolution of the ASlncRNA transcriptome 80
Cryptic ASlncRNA transcriptomes are constrained by RNAi 81
Reconciling RNAi and ASlncRNAs 82
Evidence of conservation of many ASlncRNAs at the level of expression 83

Chapter 5: Future Directions 85
Biological Functions of CRRATs 85
lncRNA mechanisms of gene regulation 86
Concluding Remarks 88

Appendix A: Supplemental Material for Chapter 2 90

Appendix B: Supplemental Material for Chapter 3 99

Chapter 1: Introduction

Abstract Long noncoding RNAs (lncRNAs) were discovered in eukaryotes more than thirty years ago (Vanhe´e-Brossollet 1998). However, advances in genomics have led to the discovery that lncRNAs are transcribed pervasively across the genome (Neil et al. 2009; Xu et al. 2009; Djebali et al. 2012; Pelechano and Steinmetz 2013). In depth functional studies have revealed that lncRNAs are capable of robust gene expression; however, the regulatory mechanisms and evolutionary pressures underlying lncRNA transcription are poorly understood. Here, we briefly review examples of functional lncRNA transcription events in humans and the budding yeast, *Saccharomyces cerevisiae* as well as how lncRNA steady-state levels are dictated, and the role that chromatin plays in lncRNA expression. This introduction then focuses on high-throughput genetic screens, conducted in budding yeast, whose aims are to identify repressors of lncRNAs, as well as comparative genomics studies of lncRNA evolution

Regulatory lncRNAs in eukaryotes

Recent advancement in technologies such as next-generation deep sequencing and tiled microarrays has enabled genome-wide analyses of the eukaryotic transcriptome . Such techniques have not only allowed for quantitative measurements of protein coding (mRNA) transcripts, but they have also revealed, surprisingly, that regions outside of open reading frames are also widely transcribed, producing long non-coding RNA (lncRNA). Because of the base-pair resolution afforded by tiled microarrays and high throughput sequencing, it also became appreciated that lncRNAs can arise under a variety of genomic contexts: intergenically or intragenically, and in the sense or antisense orientations (Figure 1). Even more surprising was the observation that

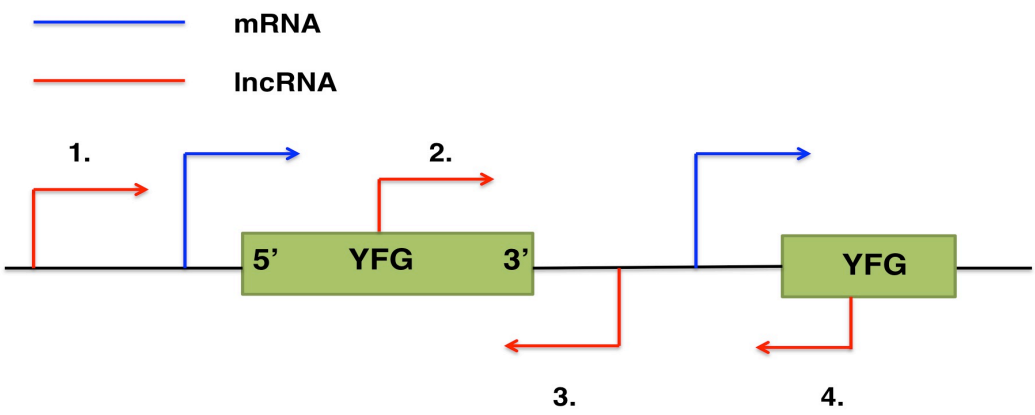


Figure 1 lncRNA arise from various genomic contexts. A schematic representation of the genomic environments wherein lncRNAs can initate transcription, shown in red. mRNAs are shown in blue. YFG= "your favorite gene". 1) sense/intergenic 2) sense/intragenic 3) antisense/intergenic 4) antisense intragenic

lncRNAs are transcribed in a pervasive manner across the genome, regardless of the organism (Neil et al. 2009; Xu et al. 2009; Graveley et al. 2011; Djebali et al. 2012), leading to the discovery of thousands of lncRNA transcripts that vastly outnumber the mRNAs with respect to the number of transcription units across the genome. After this surprising discovery, the natural next step was to elucidate the biological functions, if any, of this novel class of molecules. Initially thought to be products of stochastic transcription, it has become clear that lncRNAs are capable of possessing regulatory roles in gene expression.

There have been several examples in which transcription of antisense ncRNA leads to down-regulation of its cognate mRNA, and the underlying mechanisms have been reported for many of these cases. For instance, the lncRNA transcript can recruit histone-modifying enzymes to specific genomic loci, thereby creating repressive transcriptional environments. In mammals, the lncRNAs HOTAIR and Xist work in this manner to down-regulate the HOX genes and to inactivate one of the X-chromosomes,

respectively (Rinn et al. 2007; Gupta et al. 2010; Tsai et al. 2010; Lee and Bartolomei 2013). HOTAIR works *in trans,* being transcribed on chromosome 12 at the HOXC locus, while repressing transcriptional activity on chromosome 2, at the HOXD locus (Rinn et al. 2007; Tsai et al. 2010). In contrast, Xist works *in cis*, being transcribed on the X-chromosome and nucleating along the length of the chromosome to attenuate transcription on the inactivated X-chromosome (Lee and Bartolomei 2013). Interestingly, a recent study implicated HOTAIR in promoting tumor metastasis in mice. In the budding yeast, *Saccharomyces cerevisiae*, transcription of the antisense, lncRNA at the *PHO84* gene is coincident with recruitment of the histone deacetylase (HDAC) Hda1, which suppresses *PHO84* mRNA transcription, showing potential evolutionary conservation of lncRNA-mediated gene regulation mechanisms (Camblong et al. 2007; Camblong et al. 2009). Whether the lncRNA at *PHO84* directly recruits Hda1 is still not clear (Castelnuovo et al. 2013).

These examples in both humans and yeast show that lncRNAs are expressed from an array of contexts across all eukaryotes, and can work through various mechanisms to regulate gene expression, potentially underlying disease pathophysiology. Despite this, functional roles have still not been assigned to the vast majority of lncRNAs. Even less is known about any regulatory mechanisms and evolutionary constraints that might underlie lncRNA expression. This is in contrast to mRNA expression, whose regulatory and evolutionary principles have been extensively studied.

Characteristics, expression level, degradation of lncRNAs

As a class, lncRNA transcripts are present at roughly 10-fold lower levels than mRNAs. (Pelechano and Steinmetz 2013) Conceptually, this is consistent with the predominant functions of these two classes of transcripts. mRNAs, by definition, are intended to be translated by the ribosome in the cytoplasm, leading to the production of proteins, the main workhorse of cellular enzymatic activity. Therefore, though exceptions exist, mRNA abundance correlates relatively well with protein abundance (Nicholas T. Ingolia 2009). In contrast, lncRNAs tend to accumulate in the nucleus, and have largely been shown to associate with chromatin to regulate gene expression *in situ* (Pelechano and Steinmetz 2013). Therefore, past a certain steady-state threshold, an elevation lncRNA expression does not necessarily equate to increased lncRNA activity.

Broadly speaking, the steady-state levels of lncRNAs genome-wide can be regulated at the level of transcription, or at the level of degradation. Much of the work so far has focused on the latter. In eukaryotes, there exist two canonical pathways that are responsible for degrading lncRNAs: the exosome-mediated and Xrn1-mediated degradation. Upon transcription of the nascent lncRNA, the lncRNA transcript is targeted by targeted for degradation by the addition of poly-A nucleotides to the 3' end by the TRAMP complex (Wyers et al. 2005; Davis and Ares 2006). This 3'-end signature then leads to subsequent degradation by the exosome, a highly processive 3' to 5' nuclear exonuclease. Likewise, degradation of lncRNAs can also proceed down the Xrn1 pathway. Decapping of lncRNA transcripts signals Xrn1-mediated degradation (van Dijk et al. 2011; Geisler et al. 2012). Notably, in part due to its mainly cytoplasmic localization, Xrn1 is also largely responsible for mRNA destruction as well (Roy M Long

2003). Whether this is indicative of many lncRNAs being shuttled out of the nucleus and into the cytoplasm has not been investigated. Nevertheless, genetic perturbation of these pathways combined with genomics led to the identification of novel classes of lncRNAs: exosomal mutation led to the discovery of cryptic unstable transcripts (CUTs) while mutation of Xrn1 led to the identification of Xrn1-mediated unstable transcripts (XUTs) (Xu et al. 2009; van Dijk et al. 2011). Furthermore, these studies aided immensely in the discovery of pervasive transcription across the genome, and also provided suggestive evidence that the local chromatin environment might play roles in lncRNA expression.

Chromatin introduction

In eukaryotes, the genome is tightly packed into chromatin. The basic unit of chromatin is the nucleosome, which consists of 147 basepairs of DNA tightly wrapped around a histone octamer (Luger et al. 1997). As a result of direct steric hindrance, DNA contained within nucleosomes is less accessible to cellular machinery, such that DNA-dependent processes, like replication and transcription, must navigate this physical obstacle in order to proceed correctly (Teves et al. 2014; Weber and Henikoff 2014). This notion suggests that the positions of nucleosomes must be precisely tuned for proper regulation of lncRNA expression. If this is the case, one would expect that that the chromatin environment surrounding putative lncRNA start sites would be more "open", in order to allow the binding of factors that might facilitate transcription to the underlying DNA. Indeed, it has been shown that CUTs (see above) largely initiate from bidirectional promoters that are shared with mRNA transcripts. Furthermore, the chromatin environment surrounding bidirectional promoters is depleted of nucleosomes

(NDRs, nucleosome-depleted regions) relative to the rest of the genome (Xu et al. 2009). From this observation alone, it is difficult to assess whether the chromatin environment enables transcription of lncRNAs, or vice versa. However, it has been demonstrated that many CUTs have dedicated transcriptional pre-initiation complexes (PICs) directly adjacent to a +1 nucleosome, as determined by ChIP-exo (Rhee and Pugh 2012). This argues against lncRNA simply being the byproducts of stochastic transcription. Additionally, it further suggests that the positions of +1 nucleosomes might be regulated in order for proper stepwise assembly of PIC complexes assigned to lncRNAs. Together, these studies implicate a direct relationship between chromatin structure and lncRNA expression.

Regulation of lncRNAs by Chromatin Remodeling

Isw2 represses lncRNAs originating at the 3' ends of genes

Improvements in genomics techniques allowed not only the discovery of lncRNAs as a novel class of regulatory molecules, but they also allowed for the determination, at high resolution, of nucleosome positions across the genome (Segal et al. 2006; Whitehouse et al. 2007). As stated above, nucleosomes are generally inhibitory towards DNA-dependent processes, like transcription. One way the cell can regulate DNA accessibility is through the use of ATP-dependent chromatin remodelers, which can work through various mechanisms to mobilize nucleosomes such that the underlying DNA can be contacted (Bartholomew 2014). One such remodeler is Isw2 (Imitation Switch 2) in *S. cerevisiae*. Our lab has shown that Isw2 uses the energy of ATP to slide nucleosomes along the length of DNA (Fazzio and Tsukiyama 2003; Whitehouse and Tsukiyama 2006). Genome-wide analysis of Isw2 binding as well as nucleosome positions in Isw2 mutant revealed that Isw2 functions *in vivo* to slide nucleosomes away from open

reading frame (ORF) boundaries and into NDRs at the 5' and 3' ends of genes, leading to NDR shortening (Whitehouse et al. 2007). While the function at the 5' ends of genes is presumably to attenuate mRNA transcription, Isw2 also binds at the 3' ends of genes to repress antisense lncRNA transcription (Whitehouse et al. 2007; Yadon et al. 2010). To our knowledge, this was the first example of an ATP-dependent chromatin-remodeling factor playing any role in regulating the expression of lncRNAs. Further work out of our lab demonstrated that Isw2 might regulate a class of lncRNAs whose increase is coincident with a decrease in the levels of the overlapping mRNA, and these putative regulatory evens occur across the entire genome (Yadon et al. 2010; Alcid and Tsukiyama 2014). Evidence of nucleosome remodeling around putative lncRNA start sites substantiated direct regulation by Isw2 (Alcid and Tsukiyama 2014). This is suggestive of a novel function of Isw2: the regulation of mRNA expression through modulation of overlapping lncRNA levels.

Chromatin remodeling represses lncRNAs in a stem cell-specific context

A role for repression of lncRNAs by chromatin remodeling factors appears to extend to humans as well. esBAF is a chromatin remodeling factor that is required to maintain embryonic stem cell (ESCs) renewal and pluripotency (Ho et al. 2009b). It is a member of the SWI/SNF family of remodelers and is known to have roles in regulating mRNA expression (Ho et al. 2009a; Ho et al. 2009b). To address the role of esBAF in potentially regulating lncRNAs genome-wide, strand-specific RNAseq was conducted in ESCs, and revealed widespread de-repression of lncRNA expression when Smarca4, an ATP-dependent helicase subunit of esBAF, is depleted. esBAF accomplishes this function by localizing to DNase 1 hypersensitive sites (DHSs) and cryptic promoters to stabilize nucleosomes and maintaining a well-definied NDR. When esBAF is depleted,

nucleosomes flanking NDRs decrease in occupancy, leading to less well-defined chromatin structure, ultimately promoting aberrant transcription of lncRNAs, some of which are antisense to coding regions (Hainer et al. 2015).

The histone variant H2A.Z represses antisense lncRNAs at convergent genes

Another mechanism by which cells can alleviate histone-DNA contacts is through the use of histone variants, which can be substituted in lieu of one of the core histones. One such variant is H2A.Z (Htz1 in *S. cerevisiae* and *S. pombe*)., which is deposited and removed from chromatin by the ATP-dependent chromatin remodeling factors, Swr1 and Ino80, respectively (Kobor et al. 2004; Zofall et al. 2009; Papamichos-Chronakis et al. 2011; Yen et al. 2013). H2A.Z has been implicated in both activating and repressive roles in gene expression across many eukaryotes (Weber and Henikoff 2014). It has been suggested that H2A.Z fulfills this dual role by making DNA more accessible to transcriptional activators and repressors, presumably by decreasing nucleosome occupancy at factor binding sites (Weber and Henikoff 2014). Therefore, the gene regulatory function of H2A.Z *in vivo* is likely context and locus specific. In its canonical role, it localizes to the 5' ends of genes, at the +1 nucleosome, where it exerts its regulatory functions. However, it has been shown in fission yeast that H2A.Z plays a role in regulating lncRNA expression originating from read-through transcription at the downstream convergent gene. H2A.Z cooperates with the methyltransferase ClrC and RNAi machinery to stimulate degradation of these antisense transcripts (Zofall et al. 2009). Because of the role of Swr1 in the deposition of H2A.Z into chromatin, Swr1 is also implicated in a general role for the repression of lncRNAs. Indeed, two independently conducted genetic screens identified Swr1 as a putative lncRNA repressor in budding yeast (Alcid and Tsukiyama 2014; Marquardt et al. 2014). The

mechanism by which Swr1, and consequently H2A.Z, might repress lncRNAs is not well understood. The current consensus is that H2A.Z works synergistically with RNAi machinery (RITS complex) to induce a repressive chromatin environment and with RNA degradation machinery to process antisense transcripts.

Histone modifying enzymes and gene looping are associated with lncRNA repression

In addition to chromatin-remodeling factors, another means of regulating histone-DNA contacts is through the addition of histone modifications, which might also play a role in the genome-wide regulation of lncRNA transcription. One way to modify histones is by the covalent addition of an acetyl moiety onto specific lysine residues on histone tails. Indeed, it has been shown that levels of antisense lncRNAs highly correlate with histone H4 acetylation. Therefore, repression of lncRNA expression should be associated with the removal of acetyl groups on H4 by HDACs (Histone Deacetylase Complexes). The HDAC responsible for the removal of acetyl groups at H3K36 is Rpd3, which exists in the cell as three different subcomplexes: Rpd3L, Rpd3S, and Rpd3Mu. It has been shown that, as RNA polymerase II transcribes through gene boundaries to make pre-mRNA, the underlying nucleosomes become methylated at Histone H3 by the Set2 methyltransferase, and this serves as a signal to recruit Rpd3S to induce subsequent deacetylation of the underlying chromatin (Carrozza et al. 2005; Lickwar et al. 2009; Churchman and Weissman 2011). Failure to recruit Rpd3S, either by mutation of Rpd3 or Set2, leads to the initiation of transcription from cryptic promoters intragenically and at bidirectional, divergent promoters (Lickwar et al. 2009; Churchman and Weissman 2011). Thus, lncRNA expression is directly inhibited by the deacetylation of nucleosomes by Rpd3, and proper activity of Rpd3 is necessary for productive transcription in the coding

direction. Interestingly H4 deacetylation is also associated with genes in the genomes whose 5' and 3' ends are in proximity to each other, so-called "looped" genes. Disruption of the looping interaction through mutation of the phosphatase Ssu72 leads to increased antisense transcription at genes arranged in tandem (Tan-Wong et al. 2012). Because it is possible for this result to be confounded by coincided increase in H4 acetylation levels, the degree to which looping itself is at all responsible for the repression of lncRNAs still has to be further investigated. It is speculative to suggest that perhaps looping leads to local enrichment of transcription machinery at the coding regions of bidirectional promoters, resulting in mostly productive transcription of mRNAs.

Development of Genetic Screens to identify novel lncRNA repressors- an introduction

The examples discussed above suggest that regulation of chromatin structure and the expression of lncRNAs are directly linked. These examples also represent only a small proportion of all studies demonstrating a connection between chromatin and lncRNAs. However, the development of systematic methods to identify genes that regulate lncRNAs genome-wide in a high-throughput manner has been largely neglected. The genetic tractability of budding yeast makes it the ideal model for such studies. Indeed, two recent reports have conducted synthetic genetic array (SGA) analysis using the yeast deletion mutant library collection to identify putative lncRNA repressors (Tong et al. 2001).

The identification of suppressors of intragenic cryptic transcription

Work out of Fred Winston's laboratory sought to identify genes that repress intragenic transcription genome-wide (Cheung et al. 2008). To accomplish this, they constructed a

marker placing the *HIS3* open reading frame at a cryptic promoter from within the *FLO8* gene. Thus, in environments lacking histidine, only genes whose mutation leads to transcription from the cryptic promoter from within the construct leads to growth. Using this marker as the basis for their screen, they identified numerous genes involved in the regulation of chromatin remodeling and histone modifications. Importantly their screen identified the Rpd3 complex, substantiating the validity of their results. However, they also found genes that likely repress intragenic lncRNAs, but whose mutation is not coincident with defective H3K36 methylation, suggesting that many intragenic lncRNAs are repressed through mechanisms that are independent of Set2 and, potentially, Rpd3. One of the genes they identified was *SPT6*, which is involved in nucleosome assembly and disassembly coincident with the passage of RNA Pol II (Bortvin and Winston 1996; Hartzog et al. 1998). Microarray analysis of temperature-sensitive *SPT6* mutant revealed the de-repression of ~1000 cryptic intragenic transcripts. Notably, subsequent experiments showed that many of these cryptic lncRNAs are translated into short peptides. This further suggests that many of this class of lncRNAs are shuttled into the cytoplasm, alluded to above. Because many of these cryptic RNAs were translated into protein, this begged the question as to what the function of these lncRNAs, as well as the short peptides originating from intragenic lncRNAs, might be. To support that these intragenic lncRNAs might be functional, they found that shifting cells from rich medium to minimal medium led to increases levels of several lncRNAs even in the presence of wildtype Spt6. Whether the short peptides originating from “noncoding” cryptic transcripts have any functions have yet to be elucidated. Perhaps one function of these translated lncRNAs might be to increase proteasomal activity and provide a signal that the cell is in distress.

The identification of genes that promote promoter directionality

Recently, a genome-wide screen that identified genes that promote productive transcription in the coding direction at bidirectional promoters was developed (Marquardt et al. 2014). Similar to the screen out of the Winston lab, they constructed a marker using a promoter sequence known to undergo divergent transcription. At the coding end, they fused the mCherry fluorescent marker, and at the noncoding end, they fused YFP. Therefore, if noncoding transcription were to increase relative to coding, then the signal of YFP should rise relative to mCherry. Using this fluorescent reporter, they then performed SGA analysis to identify genes that increase the ratio of YFP to mCherry relative to wildtype. By taking the ratio, they are able to exclude genes that globally increase transcription with no specific bias for promoter directionality. Consistent with the screen described above, they identified many genes involved in chromatin remodeling and chromatin assembly. Notably, subunits of Swr1, Rsc, Ino80, and Isw2 chromatin remodeling factors were identified. They also identified three subunits of the Chromatin Assembly Factors 1 (CAF-1) complex: *CAC1, CAC2, CAC3.* NET-seq (Churchman and Weissman 2011) analysis on CAF-1 mutants revealed a loss of directionality at divergent promoters that occurs genome-wide. Given this notion, is it possible to enrich for functional lncRNA regulatory events using this strategy? Under the conditions tested, CAF-1 would predominantly be responsible for the deposition of histones following DNA replication or repair (Zabaronick and Tyler 2005; Kim and Haber 2009). CAF-1 likely regulates chromatin structure across the genome, with no bias for repressing lncRNAs involved with particular metabolic or developmental pathways. It is likely that mutation of CAF-1 simply leads to defective maintenance of chromatin structure, leading to increased stochastic tendencies of transcription. If this explains the

cause of the increase in divergent transcription genome-wide, then it becomes difficult to imagine many of the lncRNAs that they identified to be functional.

Evolutionary studies of lncRNA expression

Current evolutionary studies of mRNA transcriptomes, promoter sequences, nucleosome positions, and enhancer sequences have utilized comparative genomics to gain insights into how each of these genomic features have evolved over time (Tsankov et al. 2010; Brawand et al. 2011; Necsulea et al. 2014; Villar et al. 2015). These studies were aided immensely by the ability to sequence the genomes of organisms in a cheap and expedited manner. Examining how each feature changes at evolutionary transitions can reveal how organisms might have co-opted and shaped the respective feature landscape to adapt to its natural habitat, and can illuminate forces that act to influence its evolution. Comparative genomics on the lncRNA transcriptome have been performed in humans (Necsulea et al. 2014) and yeast (Yassour et al. 2010). In humans, it was shown that, in comparison to mRNA transcriptomes, lncRNAs evolve much more rapidly, and that they might function in embryonic development. Additionally, like mRNA transcriptomes, lncRNA transcripts show a high degree of tissue specificity that persists across species (Necsulea et al. 2014). In budding yeast, similar to mammalian studies, it was shown that lncRNAs can be conserved across the *sensu stricto* clade of species (Yassour et al. 2010). However, this analysis was not performed genome-wide for any of the species except for *S. cerevisiae.* Furthermore, in both mammals as well as budding yeast, no insights were provided as to the underlying cause for the evolutionary trends that were gleaned for lncRNA transcriptomes. Because of this, a more global view of lncRNA evolution is needed, as well as mechanistic insight as to what might be shaping lncRNA transcriptomes across evolutionary timescales. lncRNA studies are especially useful in

budding yeast, as RNAi was lost early in *sensu stricto* divergence from *Saccharomyces castellii.* Because RNAi can directly affect antisense lncRNA levels biochemically (Weinberg et al. 2011; Nakanishi et al. 2012), using budding yeast as a model might give insights into how the loss of RNAi might have affected lncRNA evolution.

Research Aims

lncRNAs are capable of key regulatory roles in gene expression across eukaryotes. However, the underlying regulatory mechanisms of lncRNA expression have not been elucidated. Many studies have strongly suggested a correlative relationship between the chromatin structure and lncRNA steady-state levels. However, the cause versus consequence relationship has not been established, and, furthermore, studies aimed at identifying novel lncRNA regulators is lacking. The screens described above were largely successful, and revealed how global maintenance of chromatin following transcription is necessary to prevent stochastic transcription events that produce lncRNAs. Despite this notion, both screens share limitations and caveats. The rationales of both screen take advantage of using one reporter construct as the output of their assays: Cheung et al. use a metabolic marker, while Marquardt et al. use a fluorescent marker. Given this notion, both screens make the assumption that the locus that they are testing is representative of all other intragenic cryptic and divergent events, respectively. However, this is likely not true, and because of this, it is also likely that some of the genes that they identified are locus-specific, and not general, repressors of lncRNA transcription events. In addition, both screens chose to follow up on the effect of chromatin assembly factors and their role in lncRNA regulation. It is not completely surprising that global disruption of chromatin structure leads to noisier transcription. Therefore, how functional the lncRNAs that both studies identified is difficult to assess.

In chapter 2, I aim to elucidate the fundamental principles underlying lncRNA transcriptional regulation. Specifically, I explore the role of chromatin remodeling factors in regulating lncRNA transcription events that might play functional roles in regulating mRNA transcription. Briefly, we found that reconstitution of RNAi and mutation of lncRNA repressor genes leads to synthetic growth phenotypes. Taking advantage of this, I use reconstituted RNAi to develop an unbiased, genome-wide screen. This screen identified many genes involved in transcription, RNA processing, chromatin assembly, and chromatin remodeling. Specifically, I identified subunits of four different, highly-conserved chromatin-remodeling factors: Swr1, Isw2, Rsc, and Ino80. Strand-specific, high-throughput RNA sequencing in mutants of these complexes revealed distinct subpopulations of lncRNAs regulated by each complex, and that many of these lncRNAs are likely directly regulated. I then identified a class of mRNAs whose levels are anti-correlated with a remodeling-factor regulated lncRNA. This strategy likely enriched for functional lncRNA transcription events as many enriched gene ontology terms were identified, revealing pathways that utilize remodeling-regulated lncRNAs as key regulatory players. In summary, I reveal a novel biological role for chromatin remodeling factors: the regulation of mRNA transcription through the use of lncRNAs.

In chapter 3, I build on a key observation from chapter 2 that increasing lncRNA levels genome-wide is detrimental in the presence of reconstituted RNAi in *S. cerevisiae.* Taking this a step further, I perform comparative genomics in budding yeast to assess how the antisense, lncRNA transcriptomes (ASlncRNA) have changed over the course of evolution. Strikingly, and in contrast to mRNA levels, we show that the steady-state levels of ASlncRNAs have risen globally since divergence from *S. castellii,* the budding yeast most related to *S. cerevisiae* to still possess active RNAi. We then show that this

had a greater effect on ASlncRNAs originating from genes arranged in tandem compared to those arranged convergently. A comparison of cryptic ASlncRNA transcriptomes from *S. cerevisiae* and *S. castellii* further shows that the amount of cryptic ASlncRNAs has also risen since divergence from *S. castellii.* We hypothesized that the increase that we observed is due, in part, to the loss of RNAi in *sensu stricto* budding yeasts. Measurement of ASlncRNA levels in *S. cerevisiae* with reconstituted RNAi was lower than in wild type *S. cerevisiae.* Importantly, abrogation of RNAi in *S. castellii* had minimal affect on its ASlncRNA transcriptome, and this is likely due to the co-evolution of the ASlncRNA transcriptome in the presence of RNAi. Finally, we further show that abrogation of RNAi partially rescues the growth of a mutation that increases steady-state levels of ASlncRNAs. Together, this data supports a model where the loss of RNAi in budding yeast led to the genome-wide increase in steady-state levels of ASlncRNAs.

Chapter 2: ATP-dependent Chromatin-remodeling Shapes the Long-Noncoding RNA Landscape

***Modified from an article of the same title from Genes and Development**

Eric A. Alcid[1,2] and Toshio Tsukiyama[1]*

[1]Divison of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, U.S.A.

[2]Molecular and Cellular Biology Program, University of Washington and Fred Hutchinson Cancer Research Center, Seattle, WA 98195, U.S.A.

*Correspondence to: ttsukiya@fhcrc.org

Running title: chromatin-remodeling factors repress lncRNAs

Keywords: chromatin, chromatin remodeling, lncRNAs, antisense RNAs

Abstract

Long non-coding RNAs (lncRNAs) are pervasively transcribed across eukaryotic genomes, but functions of only a very small subset of them have been demonstrated. This has led to active debate about whether many of them have any biological functions. In addition, very few regulators of lncRNAs have been identified. We have developed a novel genetic screen using reconstituted RNAi in *S. cerevisiae,* and systematically identified a large number of putative lncRNA repressors. Among them, we found that four highly conserved chromatin-remodeling factors are global lncRNA repressors that play major roles in shaping the eukaryotic lncRNA transcriptome. Importantly, we identified more than 250 anti-sense lncRNAs (CRRAT, chromatin-remodeling-repressed anti-sense transcripts) whose repression by these chromatin-remodeling factors is required for the maintenance of normal levels of overlapping mRNA transcripts. Our results strongly suggest that regulation of mRNA through repression of anti-sense lncRNAs is far more broadly used than previously appreciated.

Introduction

Recent advancement in genome-wide detection of RNA resulted in the unexpected finding that long non-coding RNAs (lncRNAs) are pervasively transcribed across the eukaryotic genomes (Djebali et al. 2012). It has been estimated that about 75 % of the human genome is transcribed (Djebali et al. 2012), despite the fact that about 1.2% of the genome encodes protein coding exons (Venter et al. 2001). Similarly, 75 to 85 % of the genome of the budding yeast *S. cerevisiae* is estimated to be transcribed (Nagalakshmi et al. 2008). Although lncRNAs are generally transcribed at lower levels than mRNAs, these new findings showed that lncRNAs exceed mRNAs in terms of the

number of transcription units on higher eukaryotic genomes. Some of these lncRNAs have critical regulatory roles. For example, *Xist* RNA is essential for mammalian X chromosome inactivation, and its expression is precisely regulated by other lncRNAs transcribed from neighboring regions (Lee and Bartolomei 2013). *HOTAIR* is expressed from human *HOXC* locus and regulates *HOXD* gene expression in *trans* by recruiting histone modifying enzymes PRC2 and CoRest (Rinn et al. 2007; Gupta et al. 2010; Tsai et al. 2010). Indeed, an increasing number of lncRNAs have been shown to play critical roles in transcriptional regulation, cell type specification and human diseases (Batista and Chang 2013; Flynn and Chang 2014). In budding yeast, transcription of several lncRNAs affects the abundance of the overlapping mRNA transcripts (Hongay et al. 2006; Camblong et al. 2007; Houseley et al. 2008; Camblong et al. 2009; Castelnuovo et al. 2013). Therefore, lncRNA products or the act of lncRNA transcription itself can play regulatory roles, both of which will be called "functional lncRNA transcription" in this manuscript for simplicity. However, given the large number of lncRNAs transcribed from eukaryotic genomes, the functions of the vast majority of lncRNA transcription events remain unknown. This led to an active debate as to whether most instances of lncRNA transcription lack any functional roles (Louro et al. 2009; Doolittle 2013). However, it was recently shown in budding yeast (Rhee and Pugh 2012) and humans (Venters and Pugh 2013) that a large fraction of lncRNA transcription initiates from the sites where pre-initiation complexes (PICs) are formed only for lncRNAs. This argues against the possibility that most lncRNAs are the unintended byproducts of stochastic initiation from mRNA start sites. Given that multiple basal transcription factors need to be targeted and assembled in an ordered fashion to form functional PICs, the fact that many PICs are dedicated for lncRNA transcripts is consistent with the possibility that eukaryotic cells intend to transcribe many of them. If this is the case, it is important to develop methods

to systematically identify or enrich for lncRNAs or lncRNA transcription events that play important biological roles.

If functional lncRNA transcription is prevalent, another significant challenge is to elucidate regulatory mechanisms of lncRNA transcription. Although transcription factors and chromatin regulators that control mRNA transcription have been extensively studied, much less is known about regulators of lncRNA transcription. Our lab has previously found that a highly conserved ATP-dependent chromatin-remodeling factor, Isw2, functions at 3' end of genes to repress anti-sense lncRNA (ASlncRNA) transcription in budding yeast (Whitehouse et al. 2007; Yadon et al. 2010). Recent genetic screens in yeast using reporter genes have identified histone deacetylases and histone chaperones as repressors of lncRNAs (Cheung et al. 2008; Marquardt et al. 2014). However, these factors affect only a small fraction of lncRNAs, and additional efforts are needed to identify regulators of lncRNA transcription.

In this report, we demonstrate that reconstitution of functional RNA interference (RNAi) in *S. cerevisiae* alone does not cause strong growth defects, but results in reduction of the abundance of mRNAs that have high levels of overlapping ASlncRNAs. We further found that global elevation of ASlncRNA levels in the presence of reconstituted RNAi in *S. cerevisiae* causes growth defects, likely due to destabilization of a large number of mRNAs and ASlncRNAs that overlap. Taking advantage of this finding, we systematically identified putative repressors of ASlncRNAs, including ATP-dependent chromatin-remodeling factors, histone modifying enzymes, the mediator complex subunits and transcription factors. We demonstrate that the four ATP-dependent chromatin-remodeling factors function as global ASlncRNA repressors, and provide evidence that they play major roles in shaping the eukaryotic lncRNA

transcriptome under a physiologically relevant condition (in the absence of RNAi). We further discovered that chromatin-remodeling factors are targeted to initiation sites of a large fraction of these ASlncRNA and subsequently alter chromatin structure. Finally, we identified more than 250 AslncRNAs whose repression by chromatin-remodeling factors is required for the maintenance of the normal level of overlapping mRNAs.

RESULTS

Genetic interactions between RNAi and mutations that elevate lncRNA levels

It has been shown, in both budding yeast (Neil et al. 2009; Xu et al. 2009) and mammalian cells (Core et al. 2008), that long non-coding RNAs (lncRNAs) frequently initiate from bidirectional promoters, where mRNA and lncRNA initiate in opposite directions. Because lncRNAs are highly prevalent in both mammals (Djebali et al. 2012) and budding yeast (Xu et al. 2009), this means that, at gene-dense regions of mammalian genomes (Teif et al. 2012) and at many loci in yeast, lncRNAs often overlap with protein coding genes in the anti-sense direction (Fig. 1A). This led us to hypothesize that global elevation of anti-sense lncRNA (ASlncRNA) levels in the presence of RNA interference (RNAi) would cause significant growth defects due to destabilization of a large number of mRNAs and lncRNAs that are processed by RNAi (Fig. 1A).

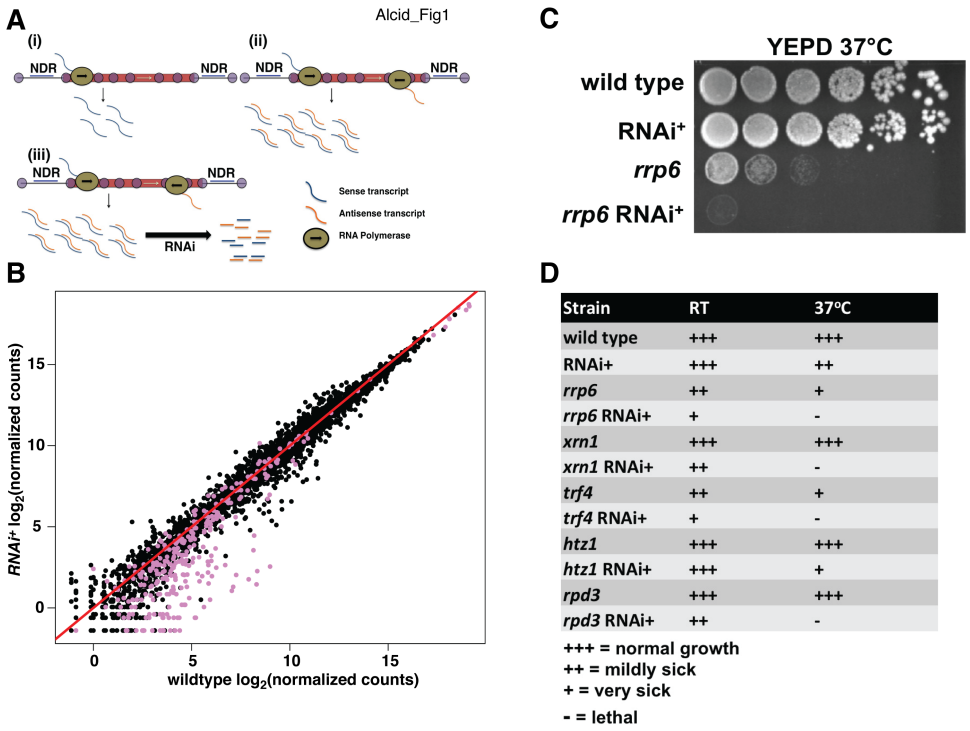


Figure 1. Genetic interactions between RNAi and mutations that elevate anti-sense lncRNA levels (A) The basis for the genetic interactions. If only the sense strand is transcribed (i), or even if both sense and anti-sense strands are transcribed in the absence of RNAi (ii), cells are fine as the double stranded RNAs (dsRNAs) are not processed. In contrast, if the levels of anti-sense lncRNAs (ASlncRNAs) are globally elevated in the presence of RNAi (iii), both mRNAs and ASlncRNAs are destabilized due to processing of dsRNAs by RNAi, causing growth defects. NDR stands for nucleosome depleted region. (B) Scatter plot of mRNA levels in wild type and RNAi+ strains. mRNAs with high endogenous antisense levels are marked by violet dots. (C) An example of genetic interactions between RNAi and a mutation that stabilizes lncRNAs (*rrp6*). A 5-fold dilution of saturated culture was spotted on YEPD and incubated at 37 °C. The genotypes of the spotted strains are shown on the left. (D) The result of the proof of concept genetic screen.

To test our hypothesis, we reconstituted RNAi in *S. cerevisiae* by expressing Argonaute (*AGO1)* and Dicer (*DCR1)* genes of *Saccharomyces castellii* (Drinnenberg et al. 2009; Drinnenberg et al. 2011). This caused a dramatic decrease in the abundance of Ty1 retrotransposon mRNA as reported (Drinnenberg et al. 2009) (Supplemental Fig. S1), showing that functional RNAi was reconstituted in our hands. Extensive phenotypic characterization by the Bartel lab (Drinnenberg et al. 2011) and our laboratory (data not shown) revealed that RNAi reconstitution in *S. cerevisiae* does not cause any detectable phenotypes other than a slight sensitivity to high temperature, as well as lithium

sensitivity, demonstrating that RNAi reconstitution alone does not significantly affect normal cell physiology. To test whether RNAi reconstitution leads to processing of endogenous double-strand RNA, we performed strand-specific RNA deep sequencing (ssRNA-seq) analysis of the *S. cerevisiae* strain with functional RNAi (hereafter RNAi+ strain) (Figure 1B). Consistent with the modest phenotype, the RNA profiles of the RNAi+ and wild type (WT) strains were overall quite similar (Spearman's rho = 0.986). Exceptions to this general trend are the genes with the highest levels of endogenous ASlncRNA (top 5th percentile shown in violet), which exhibited significant decrease in the mRNA abundance. This result demonstrates that the reconstituted RNAi can indeed process dsRNAs of the endogenous genes. We then crossed the RNAi+ strain with mutants that are known to globally stabilize lncRNAs, *xrn1* (van Dijk et al. 2011), *rrp6* (Davis and Ares 2006; Neil et al. 2009; Xu et al. 2009) and *trf4* (Wyers et al. 2005), as well as *rpd3,* which de-repress lncRNAs (Carrozza et al. 2005; Eden et al. 2009; Lickwar et al. 2009; Churchman and Weissman 2011). We also tested a histone variant mutant *htz1*, which causes global increase in lncRNAs due to failed transcriptional termination in *S. pombe* (Zofall et al. 2009). We observed strong synthetic growth defects between all of these mutations and the RNAi+ background (Fig. 1C, D). These results collectively show that global elevation of lncRNA levels can cause strong growth defects specifically in the presence of functional RNAi. Importantly, our results revealed that this synthetic RNAi system could be used as a tool to systematically identify genes that either repress or degrade ASlncRNAs.

RNAi as a tool for systematic identification of lncRNA repressors

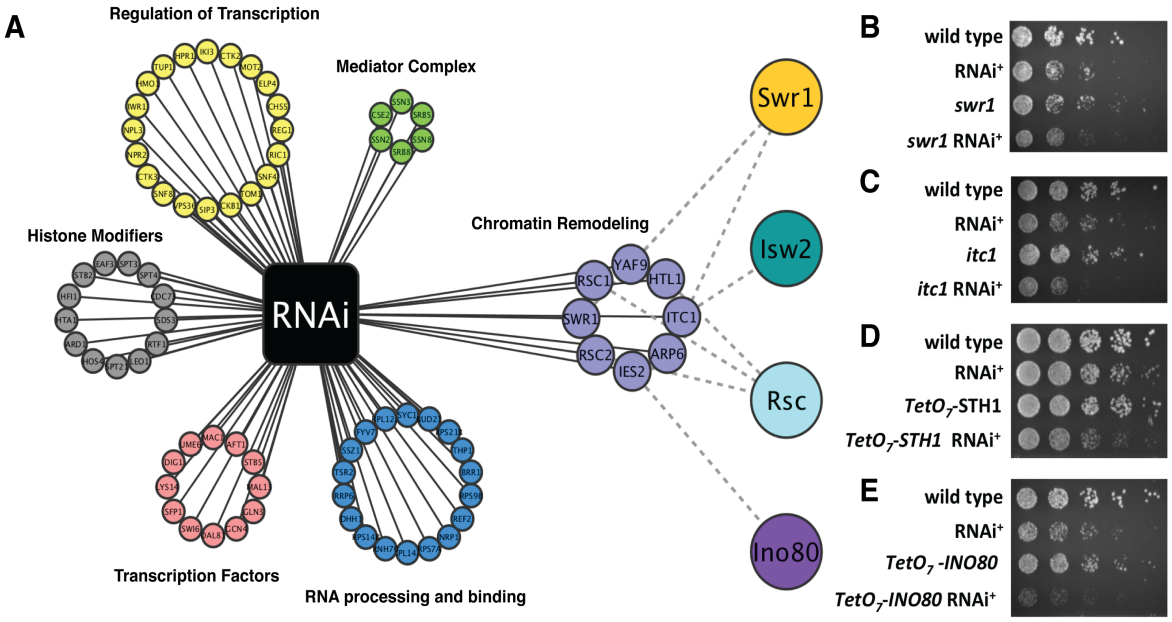


Figure 2. The results of the SGA screen using RNAi (A) *Cytoscape* network plot of the genes that interact with RNAi. Nodes represent genes and black edges represent a synthetic growth defect with RNAi. Nodes are grouped by statistically enriched gene ontology terms. Gray dashed edges connecting genes to nodes on right depict gene subunits of that chromatin-remodeling complex. (B), (C) Genetic interactions between RNAi and Δ*swr1* (swr1 complex) (B) and Δ*itc1* (Isw2 complex) (C) null mutants. (D), (E) Genetic interactions between RNAi and *tet-STH1* (Rsc complex) (D) and *tet-INO80* (Ino80 complex).

To systematically identify repressors of ASlncRNAs, we conducted a genetic screen to identify genes whose mutations cause synthetic growth defects in the RNAi+ background. To this end, we performed Synthetic Genetic Array (SGA) analysis (Tong et al. 2001) at three different conditions: 25 degrees, 37 degrees, and in the presence of 100 mM LiCl, because growth defects in the latter two conditions were common in the phenotype tests shown in Figures 1B and C. Elevated lithium inhibits the activity of Xrn1 and stabilizes lncRNAs (van Dijk et al. 2011). Enriched ($p < 0.001$) gene ontology (GO) terms among the putative gene hits include “RNA processing and binding”, “noncoding RNA metabolic process”, as well as “noncoding RNA processing” (Fig. 2A and data not shown). These genes are likely involved in destabilization of ASlncRNAs. We also found a large number of genes involved in regulation of mRNA transcription, including many

transcription repressors, as well as subunits of the mediator complex. Several genes involved in histone-modification were also among the hits, including *SDS3,* a subunit of the histone-deacetylase complex Rpd3L, which has been shown to repress lncRNAs (Lickwar et al. 2009). Given the role of transcription termination factors in the attenuation of ASlncRNAs genome-wide (Schulz et al. 2013), it was initially surprising to not identify any genes involved in termination. However, further inspection revealed that almost all genes involved in transcription termination are essential, and are not represented in the yeast deletion mutant library. Furthermore, we found eight gene encoding subunits of four different ATP-dependent chromatin-remodeling complexes: Isw2 (*ITC1*), Swr1 (*SWR1*, *ARP6, YAF9*), Ino80 (*IES2)*, and Rsc (*RSC1, RSC2, HTL1)* that are important for normal growth in the RNAi+ background (Fig. 2A). We confirmed genetic interactions between RNAi and mutants for these remodeling factors by creating mutations and performing genetic crosses in an independent genetic background (examples shown in Fig. 2B, C). Deletion of *ITC1* and *SWR1* genes is expected to completely abolish functions of Isw2 and Swr1 complexes, respectively. The null mutations of the ATPase subunits for Rsc and Ino80 complexes, *sth1* and *ino80* respectively, cause cell death and extreme growth defects, and thus are not represented in the deletion mutant library. For Rsc and Ino80 complexes, we created tetracycline-repressive alleles of the ATPase subunit *STH1* and *INO80* (*tet-STH1* and *tet-INO80*), respectively, and confirmed strong genetic interactions of these alleles with RNAi (Fig. 2D, E, Supplemental Fig. S2). These alleles are used for the following analyses. It should be noted that, although RNAi is a highly useful tool to identify putative lncRNA repressors, it is not normally present in *S. cerevisiae*. Therefore, all subsequent analyses were done in the in the absence of RNAi.

Four highly conserved ATP-dependent chromatin-remodeling complexes repress ASlncRNAs

Of the ATP-dependent chromatin-remodeling complexes we identified, only Isw2 had been shown to repress ASlncRNA transcript levels in *S. cerevisiae* at select loci (Whitehouse et al. 2007; Yadon et al. 2010). Despite this, genome-wide, transcriptomic analysis of lncRNA has not been performed for an Isw2 complex mutant. We therefore investigated whether these four chromatin-remodeling factors repress ASlncRNAs by ssRNA-Seq (Parkhomchuk et al. 2009; Sultan et al. 2012). Because it is not well established how lncRNAs repressed by chromatin-remodeling factors are processed, we did not perform any selection of RNA except for rRNA depletion (see Materials and Methods). Our analyses of the lncRNA transcriptome in chromatin-remodeling factor mutants identified a total of 1,799 ASlncRNAs that are repressed, revealing that ATP-dependent chromatin-remodeling factors play major roles in shaping the budding yeast ASlncRNA transcriptome (Fig. 3). Non-coding RNAs longer than 200b are generally considered as lncRNAs (Rinn et al. 2007). However, we found that ASlncRNAs repressed by chromatin-remodeling factors tend to be much longer than 200bp (Fig. 3 E-H, Supplemental Fig. S3). Among the four chromatin-remodeling factors, Rsc (Fig. 3C, G; n=545) and Ino80 (Fig. 3D, H; n=1,155) repress the largest number of ASlncRNAs, uncovering a novel function for these complexes. Swr1 has the most modest effects on the ASlncRNA transcriptome (Fig. 3A, E; n=10), but its mutation does de-repress ASlncRNAs. Isw2 complex has intermediate effects (Fig. 3B, F; n=89). Importantly, 620 out of the 1,799 ASlncRNAs repressed by chromatin remodeling factors have dedicated PICs around their transcription start sites (TSSs) (Rhee and Pugh 2012), suggesting that they are discrete transcription units, rather than cryptic transcripts. Of the 620 PICs

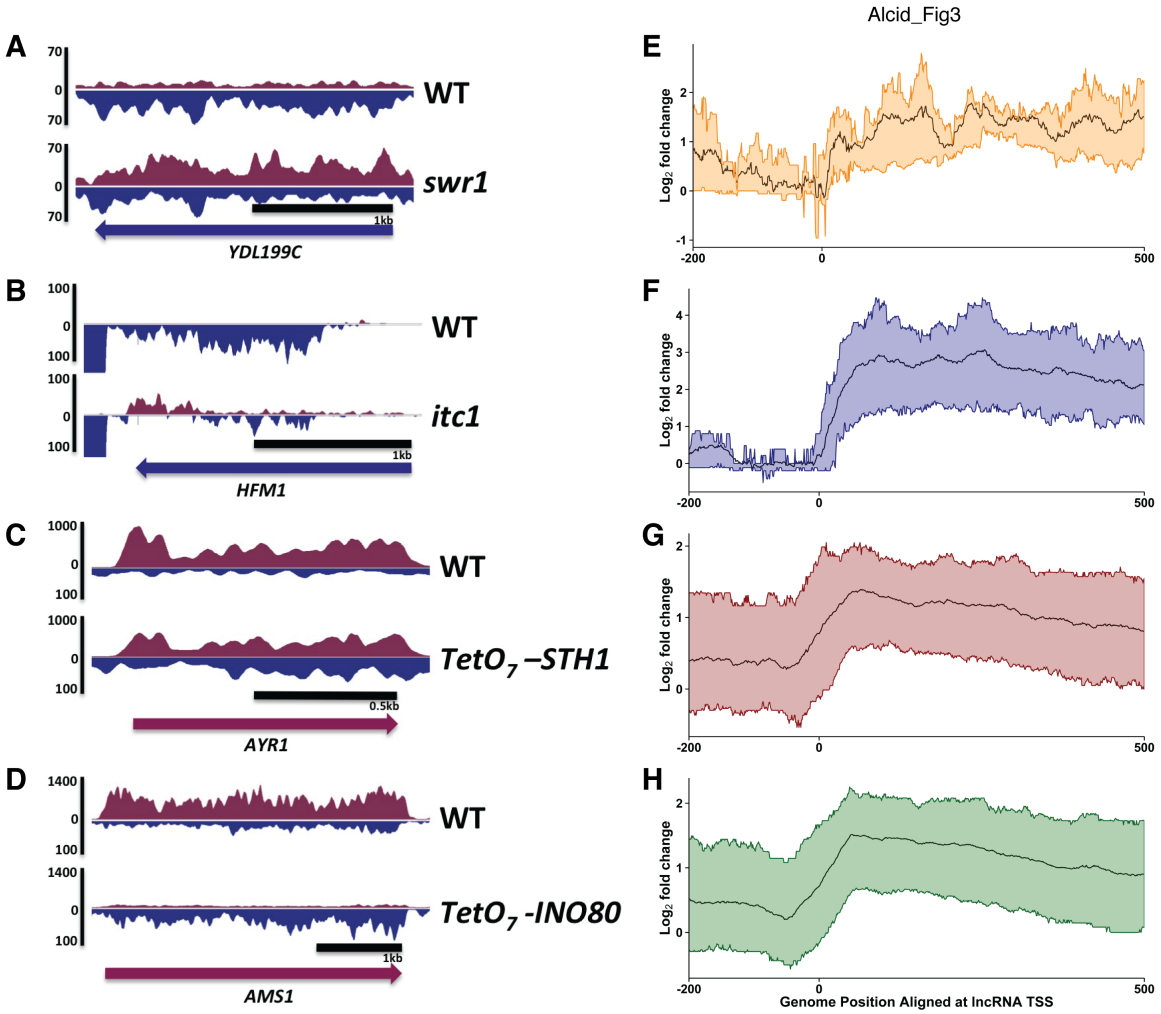


Figure 3. Chromatin-remodeling factors repress ASlncRNAs (A)-(D) Representative strand-specific RNA-seq data in D*swr1* (A), D*itc1* (B), *tet-STH1* (C), and *tet-INO80* (D) mutants. Blue and purple signals denote Watson and Crick strand transcripts, respectively. The direction of the coding gene transcription is shown at the bottom of each panel. (E)-(H) ASlncRNA meta analyses. The ratio of ASlncRNA levels in in D*swr1* (E), D*itc1* (F), *tet-STH1* (G), and *tet-INO80* (H) mutants relative to wild type cell levels is shown in $\log_2$-scale. The black lines represent the average signals, and the colored lines represent the RNA signals of the top 25% and bottom 25% changes in the mutants.

assigned to ASlncRNAs, 237 had been previously assigned to CUTs (cryptic unstable transcripts), 168 assigned to SUTs (stable unannotated transcripts), while 152 designated as “orphans”, and thus the lncRNAs controlled by these PICs have not been previously identified. Together, these results demonstrated that our genetic screen indeed identified ASlncRNS repressors as intended, and that four highly conserved ATP-

chromatin remodeling factors function as global repressors of ASlncRNAs under physiologically relevant conditions (in the absence of RNAi). Given that all four chromatin remodeling factors tested are ASlncRNA repressors, our results also suggest that many of the other genes we identified in the screen likely function as ASlncRNAs repressors.

ATP-dependent chromatin-remodeling factors repress unique sets of ASlncRNAs

The fact that Rsc, Isw2, Ino80 and Swr1 were identified as ASlncRNA repressors is intriguing, as they exhibit distinct biochemical activities. Rsc was previously reported to increase the size of NDRs (nucleosome depleted regions) at gene promoters (Hartley and Madhani 2009), whereas Isw2 decreases NDR size (Whitehouse et al. 2007; Yadon et al. 2010). On the other hand, Swr1 is required for replacement of canonical histone H2A with the histone variant Htz1 (Kobor et al. 2004), whereas Ino80 is implicated in replacing Htz1 with H2A (Papamichos-Chronakis et al. 2011). We therefore investigated whether these remodeling factors have distinct or overlapping sets of ASlncRNA targets. Systematic comparison of ASlncRNAs repressed by these remodeling factors revealed that there is little overlap among the ASlncRNAs that are repressed (Fig. 4A, Supplemental Fig. S4). This result suggest that there are multiple distinct ways by which chromatin regulation can repress ASlncRNAs, and that distinct sets of ASlncRNAs require different types of chromatin regulation for transcriptional repression.

A significant fraction of ncRNAs is degraded by RNA surveillance mechanisms *in vivo* (Neil et al. 2009; Xu et al. 2009; van Dijk et al. 2011; Schulz et al. 2013), and previous studies identified ncRNAs based on their sensitivity (cryptic unstable transcripts, CUTs and Xrn1-sensitive unannotated trancripts, XUTs) or insensitivity (Stable

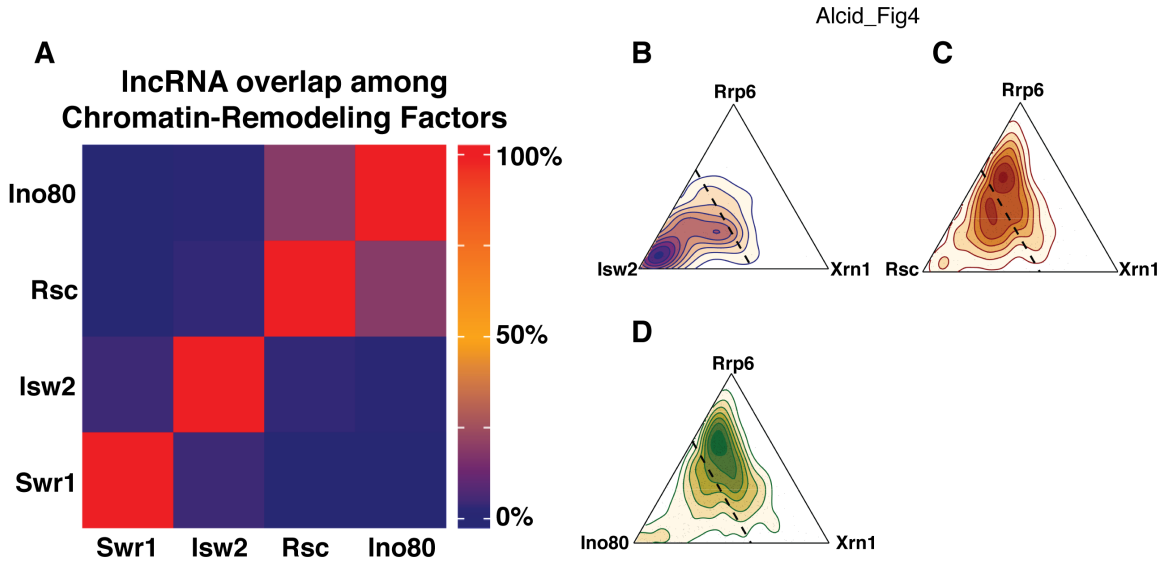


Figure 4. Chromatin-remodeling factors repress unique sets of targets (A) A heatmap representation of the degree of overlap of ASlncRNAs repressed by four ATP-dependent chromatin-remodeling factors by percentage. (B)-(D) Density ternary plots demonstrating a comparison of ASlncRNAs elevated in Isw2 (B), Rsc (C), and Ino80 (D) complexes versus in the exosome (*rrp6*) and *xrn1* mutants. The dotted lines represent the threshold where the remodeling mutant is responsible for at least 50% of the total normalized fold change.

Unannotated Transcripts, SUTs) to RNA surveillance mutation. As a result, ncRNAs have been mainly characterized in RNA surveillance mutant backgrounds. To determine whether chromatin remodeling factor-repressed ASlncRNAs have been identified in these studies, we next tested how much overlap there is among ASlncRNAs repressed by chromatin-remodeling factors and those that are degraded by the RNA surveillance mechanisms, as well as SUTs. Because of the large number of ASlncRNAs de-repressed in the mutants, we focused on Isw2, Rsc and Ino80 complexes in the following studies. There are two predominant nucleases responsible for the degradation of lncRNAs, the exosome and Xrn1. The exosome is a nuclease with 3'-5' exonuclease activity and is associated with the degradation of ~1000 lncRNAs throughout the genome (Neil et al. 2009; Xu et al. 2009). Likewise, Xrn1 is a 5'-3' cytoplasmic exonuclease associated destabilizing a large number of lncRNAs (van Dijk et al. 2011). To visualize

the extent to which the level of a particular lncRNA is dictated by chromatin remodeling, the exosome, and Xrn1, we constructed density ternary plots of lncRNAs regulated by each chromatin remodeler using the fold change of each ASlncRNA (see Material and Methods). For Isw2-repressed lncRNAs, a vast majority of the population is heavily biased towards the Isw2 vertex, suggesting that the levels of these lncRNAs are determined more by Isw2 than by degradation by the exosome and/or Xrn1. In contrast, for both Rsc and Ino80-regulated lncRNAs, large fractions of each population of ASlncRNAs are biased towards the exosome and Xrn1 vertices, suggesting that these ASlncRNAs are also degraded by the exosome and/or Xrn1. Despite this finding, we also found a significant portion of ASlncRNAs regulated by both Rsc and Ino80, whose abundance is dictated more by chromatin-remodeling factors than by exosome and/or Xrn1 (Fig. 4B-D, Supplemental Fig. S4). Unexpectedly, a very large fraction of ASlncRNAs that are de-repressed in chromatin-remodeling factor mutants exhibit a strong decrease in their abundance in *rrp6* and/or *xrn1* mutants (Fig. 4, Supplemental Fig. S4). We also found that, of the 1799 ASlncRNAs repressed by remodeling-factors, 400 overlap with previously identified SUTs (Xu et al. 2009). These results revealed that we have identified a large number of previously unidentified lncRNAs.

Identification of ASlncRNAs directly repressed by chromatin-remodeling factors

We next sought to identify ASlncRNAs that are likely directly repressed by chromatin-remodeling factors. Our criteria for these RNAs are that the ASlncRNA level significantly increases in a chromatin-remodeling factor mutant, and that the chromatin-remodeling factor is targeted to TSS of the ASlncRNA. To this end, we analyzed genome-wide ChIP

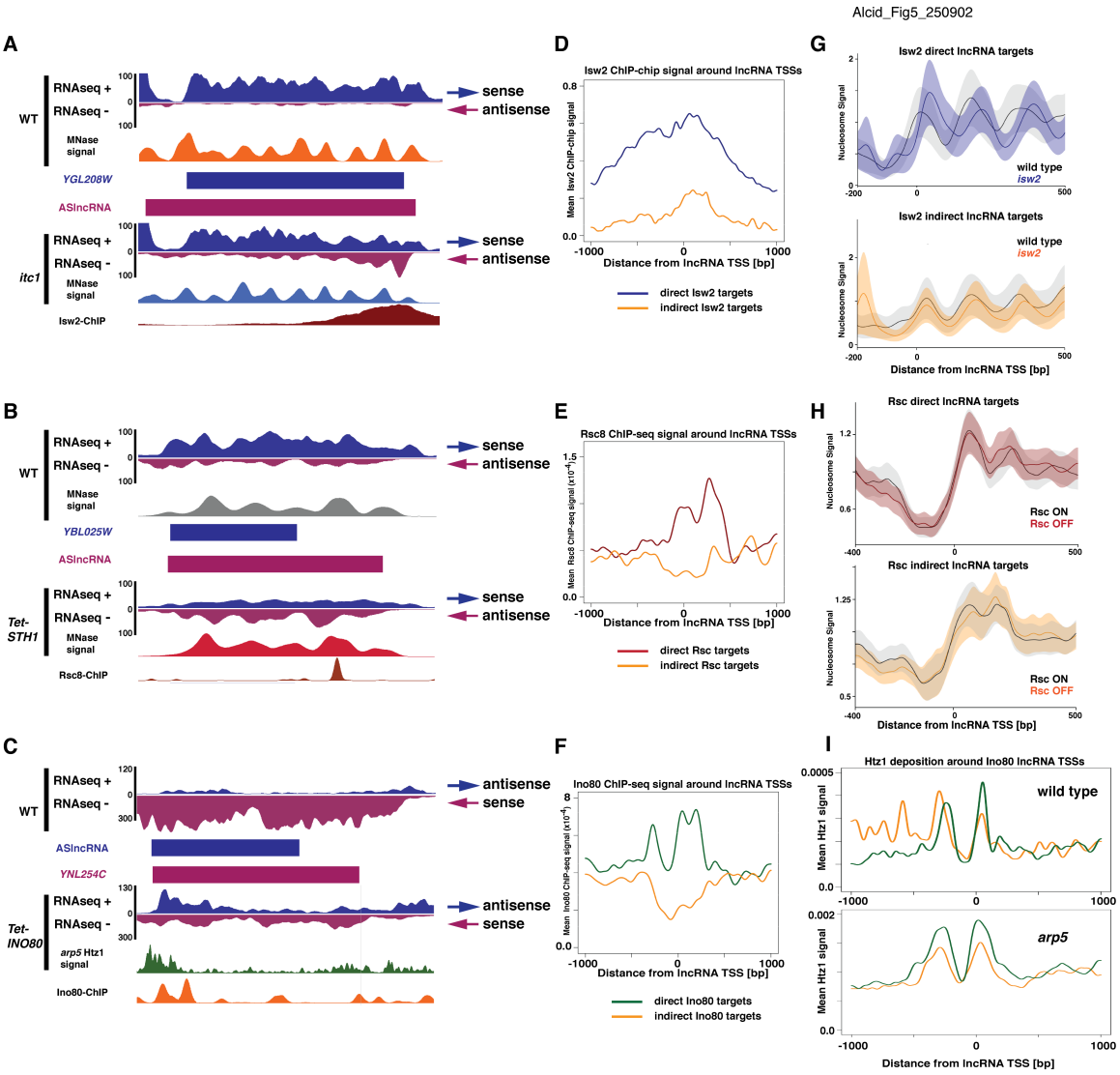


Figure 5. Targeting and chromatin changes at ASlncRNAs regulated by chromatin-remodeling factors (A) Isw2 targeting and Isw2-dependent chromatin changes at an Isw2-repressed ASlncRNA. RNA-seq reads and nucleosome signals (ref) (MNase signals) in wild type and *itc1* strains are shown. The orientation of sense and anti-sense transcripts are shown on the right. Isw2 ChIP signals (ref) are shown in the bottom. (B) Rsc targeting and Rsc-dependent chromatin changes at a Rsc-repressed ASlncRNA. RNA-seq reads and nucleosome signals (MNase signals) in wild type and *MET-STH1* strains (ref) are shown. The orientation of sense and anti-sense transcripts are shown on the right. Rsc8 ChIP signals (Yen et al. 2012) are shown in the bottom. (C) Ino80 targeting and Ino80-dependent chromatin changes at an Ino80-repressed ASlncRNA. RNA-seq reads and Htz1 signals in wild type and *arp5* strains are shown (Yen et al. 2013). The orientation of sense and anti-sense transcripts are shown on the right. Ino80 ChIP signals (Yen et al. 2012) are shown in the bottom. (D) Meta analysis of Isw2 ChIP signals around the TSSs (position 0) of direct (n=58, blue) and indirect (n=31, orange) Isw2 targets. (E) Meta analysis of Rsc8 ChIP signals around the TSSs (position 0) of direct (n=216, red) and indirect (n=329, orange) Rsc targets. (F) Meta analysis of INO80 ChIP signals around the TSSs (position 0) of direct (n=540, green) and indirect (n=615, orange) Ino80 targets. (G) Top: Ribbon plots of nucleosome signals around the TSSs (position 0) of direct Isw2 targets in wild type (black) and *isw2* mutant (blue). The lines represent the mean nucleosome signal, while the outer borders of the ribbon represent 1 standard-error of the mean away from the mean. Bottom: Ribbon plots (as above) of nucleosome signals around the TSSs (position 0) of indirect Isw2 targets in wild type (black) and *isw2* cells (orange). (H) Top: Ribbon plots (as in G) of nucleosome signals around the TSSs (position 0) of direct Rsc targets in wild type (black) and *MET-SHT1* cells (red). Bottom: Ribbon plots of nucleosome signals around the TSSs (position 0) of indirect Rsc targets in wild type (black) and *MET-SHT1* cells (orange). (I) Htz1 signals around the TSSs (position 0) of direct (green) and indirect (orange) Ino80 targets in wildtype cells. (J) Htz1 signals around the TSSs (position 0) of direct (green) and indirect (orange) Ino80 targets in *arp5* cells.

data of chromatin-remodeling factors (Whitehouse et al. 2007; Yen et al. 2012), and identified ASlncRNAs repressed by remodeling factors whose TSSs are also proximally located to remodeling factor-bound sites (see Materials and Methods). For Isw2, out of 89 repressed ASlncRNAs, 58 have Isw2 targeting around the ASlncRNS TSSs (Fig. 5A, D). For Rsc, we found 216 ASlncRNAs out of 545 exhibit Rsc targeting around TSSs (Fig. 5B, E). Finally, for Ino80, we found 540 ASlncRNAs out of 1,155 with proximal Ino80 targeting (Fig. 5C, F). Therefore, we found ~45% of ASlncRNAs repressed by chromatin-remodeling factors have the corresponding remodeling factor specifically targeting TSSs. Given that these transcripts represent unique sets of lncRNAs, we named the ASlncRNAs directly repressed by chromatin remodeling factors transcripts CRRATs (Chromatin-Remodeling Repressed Antisense Transcripts). Our *cis*-element search did not identify binding sites of any transcription factors whose binding sites are over-represented around the TSSs of CRRATs (data not shown). Therefore, the mechanisms by which chromatin remodeling factors are targeted to these sites are currently unknown.

We next examined how Isw2, Rsc and Ino80 alter chromatin around TSSs of CRRATs. Isw2-dependent repression of mRNA transcription is generally associated with reduction in the size of the nucleosome-depleted regions (NDRs) upstream of the mRNA initiation sites, which makes the TSSs covered by the upstream edges of the +1 nucleosomes (Whitehouse et al. 2007). Aligning all 58 Isw2-repressed lncRNAs at their putative TSSs in a wild type background (Yen et al. 2012) revealed that, like mRNA targets, Isw2-repressed ASlncRNAs are associated with an upstream NDR (Fig. 5G). In contrast to mRNA targets, however, nucleosomes more deeply occlude the TSSs of Isw2-repressed ASlncRNAs in wild type cells (Fig. 5G, upper panel). In the absence of

Isw2, nucleosomes shift away from NDRs, no longer occluding lncRNA TSSs. These results suggest that Isw2-repressed ASlncRNAs have strong intrinsic tendency to exclude nucleosomes at their TSSs, and that Isw2 represses them by sliding nucleosomes over their TSSs more deeply than at the mRNA TSSs. The lncRNAs that are de-repressed in *itc1* mutant but do not have Isw2 targeting around their TSSs, likely indirect targets of Isw2, are generally associated with much smaller Isw2-dependent changes in nucleosome positioning (Fig. 5G, lower panel), supporting our conclusion that they are indirectly affected by Isw2.

Rsc also functions around NDRs, but there have been reports on distinct chromatin regulation around NDRs for transcriptional regulation. Rsc was first reported to generally increase the size of NDRs to facilitate transcription (Hartley and Madhani 2009; Floer et al. 2010). However, it was recently reported that Rsc is also required for silencing of genes in subtelomeric regions, and that its activity leads to higher signals of +1 and -1 nucleosomes (Van de Vosse et al. 2013). These reports suggest that Rsc can exhibit different chromatin remodeling activities depending on the context. Analysis of chromatin structure at 216 Rsc-repressed ASlncRNAs revealed they also generally initiate from NDRs, and that Rsc depletion generally reduced signals of the -1 nucleosomes adjacent to lncRNA TSSs as well as modestly reducing the +1 nucleosome (Fig. 5H). Furthermore, Rsc depletion resulted in shifting of the +2, and +3 nucleosomes towards the 5' end of the ASlncRNA. This suggests that, like at subtelomeric mRNA targets, Rsc represses ASlncRNAs by increasing nucleosome occupancy around the TSSs. However, Rsc-repressed ASlncRNAs are generally not located within subtelomeric regions, revealing distinct location of mRNA and ASlncRNA targets of Rsc-dependent repression. In contrast, indirect Rsc target ASlncRNAs do not exhibit the

same degree of reduction of the -1 nucleosome signals, and instead showed the reduction of the +1 nucleosome signals (Fig. 5H), as well a considerably narrower NDR when compared to direct lncRNA targets of Rsc.

Ino80 acts at NDRs to remove the histone variant Htz1 at +1 nucleosomes of mRNA transcription units (Papamichos-Chronakis et al. 2011; Yen et al. 2013). *ARP5,* encoding a subunit of the Ino80 complex, is required for Ino80 biochemical activity, as well as DNA-binding activity of the complex. ChIP-exo analysis of Htz1 in an *arp5* background demonstrated that there is a substantial accumulation of Htz1 at the 5' ends of mRNA coding genes (Yen et al. 2013). To determine how Ino80 affects Htz1 levels for ASlncRNA repression, we examined the Htz1 level around the TSSs of 540 Ino80-repressed ASlncRNAs. In a wildtype background, levels of Htz1 around ASlncRNA TSSs were largely similar at indirect and direct ASlncRNA targets of Ino80 (Fig 5I) (Albert et al. 2007). Analysis of Htz1 levels in an *arp5* background at direct ASlncRNA targets of Ino80 revealed a substantial level of Htz1 enrichment directly overlapping the ASlncRNA TSSs that is much higher than neighboring regions (Fig. 5J). This result suggests that Ino80 represses ASlncRNA transcription by preventing the accumulation of abnormally high levels of Htz1 at the 3' ends of a very large number of genes. In contrast, the level of Htz1 was much lower around transcription initiation sites of indirect Ino80 target ASlncRNAs (Fig. 5I), again supporting the notion that their transcription is indirectly affected by Ino80.

These results collectively identified 814 CRRATs that are likely directly regulated by chromatin-remodeling factors through chromatin regulation around their TSSs. The fact that much weaker or different chromatin changes were found at TSSs of the indirect targets of the remodeling factors, despite the comparable level of ASlncRNA levels in the

mutants (Supplemental Fig. 5), argue against the possibility that these chromatin changes are the results of elevated ASlncRNA transcription. Therefore, these results provide further support for the direct regulation of CRRATs by chromatin-remodeling factors. Moreover, our results suggest that ASlncRNA can be repressed through diverse chromatin remodeling mechanisms.

Identification of ASlncRNAs with regulatory roles

Because lncRNAs are highly prevalent across eukaryotes, one important question is how many of them have functional roles *in vivo (Louro et al. 2009)*. This is a highly significant issue, as only a handful of lncRNAs transcription units with regulatory roles have been identified in *S. cerevisiae*, despite the fact that several thousand lncRNAs are transcribed across the genome (Hongay et al. 2006; Camblong et al. 2007; Houseley et al. 2008; Camblong et al. 2009; Castelnuovo et al. 2013). Similarly, although about 75% of the human genome is estimated to be transcribed by ncRNAs (Djebali et al. 2012), the functions of a very small fraction of them have been identified. Because known functional lncRNA transcription in *S. cerevisiae* is precisely regulated by environmental cues and cell-type specificities (Hongay et al. 2006; Camblong et al. 2007; Houseley et al. 2008; Camblong et al. 2009; Castelnuovo et al. 2013), we hypothesized that ASlncRNA transcription regulated by highly conserved chromatin-remodeling factors (CRRATs) may be enriched for those with regulatory functions. Most known regulatory lncRNA transcription events in *S. cerevisiae* repress overlapping mRNAs. Consistent with this, visual inspection of individual CRRATs revealed many cases in which de-repression of ASlncRNAs in chromatin remodeling factor mutants coincides with

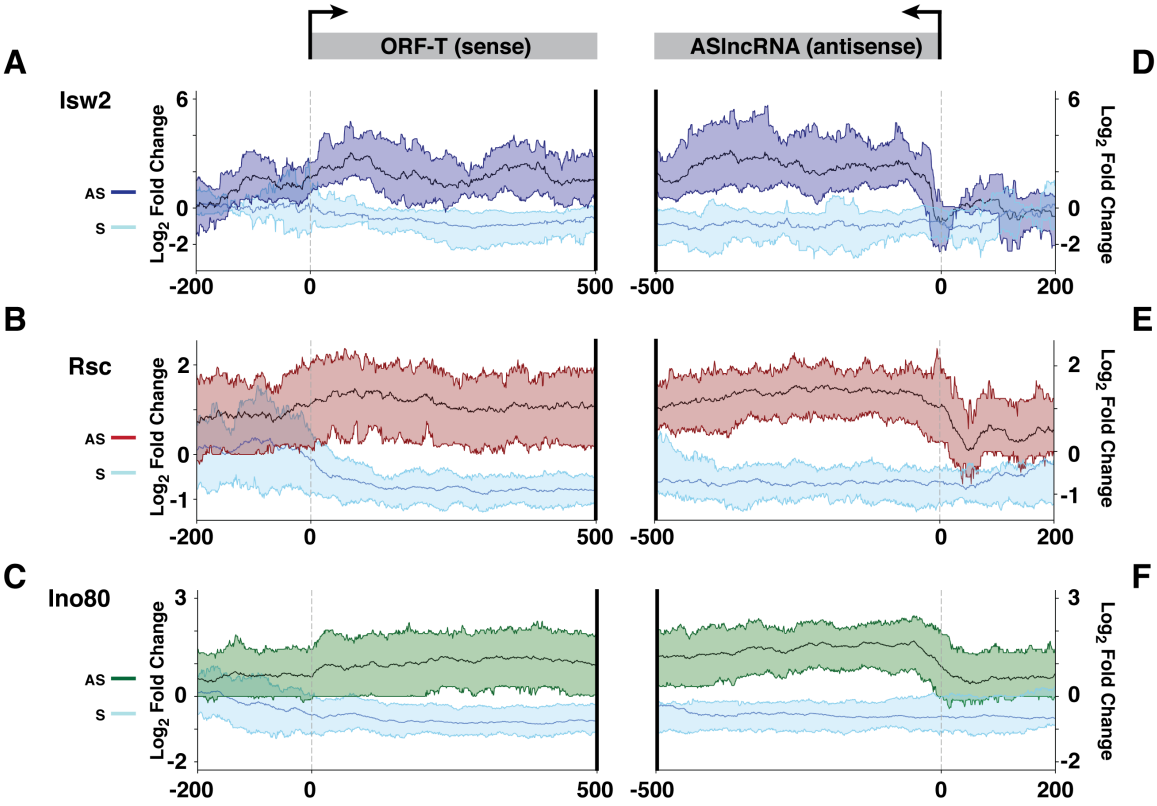


Figure 6. Identification of ASlncRNAs whose repression by chromatin-remodeling factors is required for the maintenance of the normal level of overlapping mRNAs (A) Changes in the levels of Isw2-repressed ASlncRNAs (purple) and overlapping mRNAs (light blue) in *itc1* cells. The solid lines denote the mean, and the colored ribbon show the RNA signals of the top 25% and bottom 25% values in the mutants. RNAs are aligned at the TSSs of mRNAs (left panel), and at the TSSs of ASlncRNAs (right panel). A dotted line is at the value "0". (B) Changes in the levels of Rsc-repressed ASlncRNAs (red) and overlapping mRNAs (light blue) in *tet-STH1* cells. The solid lines denote the mean, and the colored ribbon show the RNA signals of the top 25% and bottom 25% values in the mutants. RNAs are aligned at the TSSs of mRNAs (left panel), and at the TSSs of ASlncRNAs (right panel). A dotted line is at the value "0". (C) Changes in the levels of Ino80-repressed ASlncRNAs (green) and overlapping mRNAs (light blue) in *tet-INO80* cells. The solid lines denote the mean, and the colored ribbon show the RNA signals of the top 25% and bottom 25% values in the mutants. RNAs are aligned at the TSSs of mRNAs (left panel), and at the TSSs of ASlncRNAs (right panel). A dotted line is at the value "0".

decreased levels of overlapping mRNAs (Fig. 5A-C). We therefore systematically identified CRRATs whose de-repression in the mutant is associated with down regulation of the overlapping mRNA levels. We identified a total of 259 such CRRATs, out of 814, that fit to the criteria (Fig. 6A-C). Because these CRRATs are likely directly repressed by

chromatin remodeling factors, our results suggests that repression of these ASlncRNAs by chromatin-remodeling factors is required for the maintenance of the normal levels of overlapping mRNAs. We found that most of the CRRATs in this class transcribe through the entire open reading frame of the overlapping genes and the promoter regions (Fig. 6A-C), which is unusually long for lncRNAs. This result is consistent with the earlier conclusion that CRRATs represent a unique set of lncRNA transcripts, and suggests a possibility that the products of functional lncRNA transcription may have a tendency to be longer than the average ncRNAs. While changes in chromatin structure were the most modest for Rsc-CRRATs, measurement of nucleosome signal around putatively functional Rsc-CRRATs further accentuated the observed changes, suggesting that Rsc-CRRATs are enriched for direct targeting by Rsc (Supplementary Fig. 6). Furthermore, comparison of the fold-change of mRNAs targeted by CRRATs with all other mRNAs revealed a significant association with a reduction in levels when an overlapping CRRAT is present (Supplemental Fig. 7). Together, these results suggest that these 259 putatively functional CRRATs might play roles in regulating gene expression.

Discussion

Although lncRNAs are pervasively transcribed across eukaryotes, functions of the vast majority of them remain unknown. In addition, unlike mRNA transcription, very few regulators of lncRNAs have been identified. To address these crucial issues, we have developed a novel genetic screen to systematically identify repressors of ASlncRNAs. This allowed us to identify a large number of ASlncRNA repressors as well as ASlncRNAs whose transcription have regulatory roles.

A novel genetic screen for ASlncRNA repressors

Previously, Chung et al. identified genes that repress expression of cryptic transcripts from intragenic initiation sites using *HIS3* gene as a reporter (Cheung et al. 2008). More recently, Marquardt et al. developed a genetic screen to isolate genes involved in repression of lncRNA transcription from divergent promoters using YFP and mCherry reporters (Marquardt et al. 2014). Curiously, both of these screens identified genes involved in chromatin assembly as well as histone genes among the strongest hits. These results revealed that chromatin assembly defects lead to elevated transcription of both intragenic cryptic RNAs and lncRNAs from some bidirectional promoters. Whether these transcripts have biological roles remains unknown. One interesting possibility is that these cryptic transcripts may signal to chromatin surveillance systems the location of chromatin defects, resulting in their timely repair. Interestingly, the genes identified in our screen had only a few overlaps with those identified in these two screens. Our screen identified a large number of genes known for regulation of mRNA transcription. The difference in the results between our genetic screen and the other two is likely at least partly due to the fact that our screen did not use reporter genes and instead relied directly on genome-wide elevation of ASlncRNAs. What makes reporter assays particularly sensitive to chromatin assembly defects is an interesting question that needs to be addressed in the future. These results collectively demonstrate that different classes of lncRNAs regulated by diverse mechanisms exist, and that they likely have distinct biological functions.

A large number of putative lncRNA repressors

So far, only a very small number of regulators for lncRNA transcription have been identified. However, the fact that all four chromatin-remodeling factors identified in our screen are indeed ASlncRNA repressors suggest that many, if not all, other genes identified in our screen likely function as repressors of ASlncRNAs as well. This means a large amount of resources are used to control ASlncRNA transcription. Among the identified putative repressors, the mediator, Paf1 and Rpd3 complex subunits were also identified by the genetic screen for suppressors of intragenic cryptic transcripts by Chung et al., suggesting that they are likely involved in repression of multiple classes of lncRNAs. Most putative repressors identified in our screen also repress mRNA transcription, indicating that the dual roles of transcription repression on mRNA and lncRNAs are unexpectedly widespread. This means that the phenotypes of the mutants for these repressors need to be revisited, as they can be at least partly due to de-repression of lncRNAs.

Current estimation for the number of lncRNAs transcribed in *S. cerevisiae* is based on the number of RNAs detected in wild type cells, as well as in mutants that stabilize or prevent pre-mature termination of lncRNA transcripts, such as *trf4* (Wyers et al. 2005) and *rrp6* (Davis and Ares 2006) mutants. The estimated number of lncRNAs is already very large (~900 CUTs (Xu et al. 2009), 800 SUTs (Xu et al. 2009), 1600 XUTs (van Dijk et al. 2011), and 1500 NUTs (Schulz et al. 2013): Note that there are overlaps between the classes), so it was unexpected that a significant fraction of ASlncRNAs repressed by chromatin-remodeling factors are not detectably elevated in these mutants, and thus are novel lncRNAs. The fact that many uncharacterized putative ASlncRNA

repressors exist suggests a possibility that the number of lncRNAs is currently underestimated, and there are a large number of lncRNA to be identified

Identification of lncRNAs with regulatory functions

Compared to the enormous number of lncRNAs transcribed in eukaryotic cells, the number of functional lncRNA transcription is extremely small. This led to proposals that most lncRNAs transcribed in eukaryotes are non-functional ("noise" or "junk") (Doolittle 2013). However, proving the non-functionality of such a large number of transcripts is currently unfeasible. On the other hand, it has recently become clear that the small number of lncRNAs with regulatory functions play crucial roles in transcriptional control, cell type specification and human diseases (Batista and Chang 2013; Lee and Bartolomei 2013; Flynn and Chang 2014), raising the possibility that many lncRNAs with important functional roles are yet to be identified. These opposing views have resulted in active discussions in the field as to how many lncRNAs have any biological functions (Kellis 2014). One possibly way to address this debate is to systematically identify lncRNAs with regulatory functions. However, a systematic screen for functional lncRNAs has been difficult, and discovery of functional lncRNA has relied mostly on fortuitous events and locus-specific analyses. Based on the fact that transcription of lncRNAs with known regulatory functions are precisely regulated by environmental cues and/or cell type, we hypothesized that we would be able to enrich for ASlncRNAs with regulatory roles by focusing on those that are regulated by highly conserved ATP-dependent chromatin-remodeling factors. This makes an intuitive sense, as cellular events that play crucial roles are usually regulated, and there is little reason to regulate transcription of non-functional RNAs.

Our hypothesis was supported by the finding that, in chromatin-remodeling factor mutant, de-repression of ~32% (259 out of 814) of CRRATs is associated with a significant decrease in the level of mRNAs they overlap. This strongly suggests that chromatin-remodeling factors maintain the levels of these 259 mRNAs through repression of overlapping ASlncRNAs. This is most likely an underestimation of functional ASlncRNA transcription events, as only one growth condition (logarithmic growth in rich medium) was used for RNA analyses. Nonetheless, our results indicate that the mechanism to regulate mRNA levels through ASlncRNA control is far more widely used than currently appreciated. Moreover, given the many putative ASlncRNA repressors that we have identified, it is highly likely that a very large number of cases in which ASlncRNA control is used to regulate mRNA levels in a similar fashion are yet to be discovered.

Materials and Methods

Yeast strains

A list of all strains used in this study can be found in table S1. We carried out single-step gene deletions by standard lithium acetate transformation using KanMX, HygMX, and NatMX drug-resistance markers as described for *S. cerevisiae (Goldstein and McCusker 1999)*. Strains were also created using standard genetic crosses. For *S. cerevisiae,* genome sequences and annotations were downloaded from Ensembl or the Saccharomyces Genome Database.

Yeast growth conditions

Unless otherwise noted, strains were cultured at 30°C in either YPD or YC until OD600 = 0.4-0.7. For strains harboring Tet-repressible alleles, cells were cultured at 37°C until OD = 0.3, then doxycycline was added to YC-(-His) media at a final concentration of 20 ng/mL. Cells were grown for 3 hours before being harvested for RNA using standard hot acid phenol extraction.

Plasmid Construction

pRS406-(AGO1-DCR1) was created by inserting restriction digest fragments containing the coding sequences and associated promoters of *AGO1* and *DCR1* from pRS404-PTEF-AGO1 and pRS405-PTEF-DCR1(Drinnenberg et al. 2009) into pRS406. This plasmid was then used to integrate RNAi machinery into our lab strains.

Northern Analysis

10-20 μg of Total RNA was added to 1.5x sample buffer (75% formamide, 13.4% formaline, 1x MOPS) and loaded onto a 1% agarose gel prepared with running buffer (1% agarose, 1x MOPS, 5% formalin). The RNA was transferred onto a GeneScreen membrane (Perkin Elmer) in 10x SSC overnight. The blot was then UV cross-linked, and incubated with hybridization buffer (6x SSC, 0.1mg/ml salmon sperm DNA, 0.5% SDS) at 65°C for 1 hour. Radiolabeled probe to Ty1 or *ACT1* was then added and hybridization occurred overnight at 65°C. The membrane was then washed with 0.5x SSC for 30 minutes at 65°C. This wash was repeated 1x before exposure to film or phosphor screen.

Strand-specific library preparation and high-throughput RNA sequencing

RNA-seq was done in the presence and absence of reconstituted RNAi in Figure 1. For transcriptome analyses of chromatin remodeling factor mutants, all experiments were done using mutants that do not have RNAi. 3ug of Total RNA was depleted of ribosomal RNA species using Ribo-Zero magnetic rRNA removal kit (Human/Mouse/Rat) (Epicentre). Strand-specific libraries were then prepared using the dUTP method combined with TruSeq (Illumina) as previously described (Parkhomchuk et al. 2009; Sultan et al. 2012). 50 cycles of paired-end sequencing was performed on an Illumina HiSeq 2500 on either high-output mode or rapid run mode (FHCRC shared resources). All sequencing experiments were performed in biological duplicate.

RNA-seq analysis

Alignment

Reads were aligned to the *Saccharomyces cerevisiae* genome (Saccharomyces_cerevisiae.EF4.69.dna.toplevel.fa) (Flicek et al. 2014) using TopHat2 (Kim et al. 2013) with the following settings: tophat2 -p 4 -G <gene_annotation_file> -I 2000 --library-type=fr-firststrand -o <output_directory> <bowtie_index> <Read1.fastq> <Read2.fastq>. Reads were then trimmed of adapter sequences with a custom Python script using the Python module HTSeq (Anders et al. 2014).

Segmentation heuristic of RNA-seq data to identify putative transcript units

After reads were aligned, reads were filtered such that only properly aligned, uniquely mapped reads were kept using a custom Python script and pysam (Li et al. 2009). Because replicates were highly reproducible (data not shown), reads for each replicate were combined to make per-base, strand-specific pileup files using pysam. Using this pileup file, putative transcript units were segmented by defining a minimum expression threshold, defined below. tRNAs, and rRNAs were excluded for every step in analysis.

Defining a threshold level using empirically determined tag density

For a known open reading frame (ORF), expression was calculated by the following equation:

$$tag\,density = \frac{\frac{\sum_{i=start}^{end} count_i}{(end-start)}}{\overline{count}_{genome}} \qquad (1)$$

where *i* is the genomic position, *count* is the number of reads overlapping *i, end* is the last genomic position of the ORF, *start* is the beginning position of the ORF. This was repeated for every ORF in the genome. The threshold was defined by the bottom 5th

percentile expression value for transcripts longer than 250 bp (inclusive). For transcripts between 100bp and 249 bps (inclusive), the threshold was the bottom 25th percentile expression value.

Segmentation heuristic of pileup files

Using the threshold defined above, putative transcripts were identified by computing the tag density within a 100bp sliding-window using a 1bp step size. "Starts" and "Ends" of transcript units were defined by whether the tag density exceeded the defined threshold and were at least 100 bp in length. Segments closer than 50 bp, and were less than 2-fold different in tag density, were joined, which is commonly performed. See above for threshold differences based on length.

Identification of differentially expressed lncRNAs

Using the compiled putative transcript list, differentially expressed transcript units were defined by first enumerating the number of reads in each replicate that overlap with each transcript, then using a negative binomial distribution (R-package DESeq) (Anders and Huber 2010) to determine differential expression. For remodeling mutants, putative transcripts that were upregulated (>= 1.25 fold) and had a p-value < 0.05 were determined to be differentially expressed. Differentially expressed transcripts were then identified as ncRNAs by whether they overlap with annotated features of the genome in a strand-specific manner. This was performed using custom scripts written in Python, and BEDtools (Quinlan and Hall 2010). Fold-change, as well as absolute expression (in rpkm values) were determined using DESeq. Chromosome coordinates for all ASlncRNAs repressed by chromatin remodeling factors are listed in Table S2.

Construction of heatmaps and plots, and statistical analysis

Heatmaps, plots, and meta-gene plots were constructed in R using the packages "ggplot" or "ggtern." To create ternary plots, for each lncRNA, the fold-change was determined in the remodeling mutant, the exosome mutant, and the Xrn1 mutant. The fold changes were then normalized so that they summed to 1. These normalized fold-changes were then determined for each lncRNA, and the resulting matrix was used in ggtern to make the plots.

Identification of CRRATs

All ASlncRNA transcription start sites were extended 100bp in both the forward and reverse directions. The resulting 200 bp interval was then used to determine whether or not an ASlncRNA TSS was in proximity to a remodeler-bound nucleosome (Yen et al. 2012) using BEDtools (Quinlan and Hall 2010).

Gene Ontology Analysis

All gene ontology analysis was performed using GOrilla (Eden et al. 2009).

Author Contributions

E.A.A. contributed in planning and performing experiments, analyzing and interpreting data, and writing this manuscript. T.T. contributed in planning experiments, interpreting data and writing this manuscript.

Acknowledgements

We thank H. Malik, I.A. Drinnenberg, B.F. Pugh and the members of the Tsukiyama lab for helpful discussions; H. Malik, I.A. Drinnenberg for critical reading of the manuscript; F. Winston for yeast strains; A. Marty and FHCRC shared resources for deep sequencing. This work was supported by grant RO1 GM058465 (T.T.) and predoctoral fellowship F31 GM101944 (E.A.A.).

Chapter 3: RNA interference constrains expansion of the anti-sense long, non-coding RNA transcriptome

***A modified version of a submitted manuscript**

Eric A. Alcid[1,2] and Toshio Tsukiyama[1]*

[1]Divison of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, U.S.A.

[2]Molecular and Cellular Biology Program, University of Washington and Fred Hutchinson Cancer Research Center, Seattle, WA 98195, U.S.A.

*Correspondence to: ttsukiya@fhcrc.org

Running title: evolution of antisense lncRNAs

Keywords: RNAi, lncRNAs, antisense RNAs, evolution

Abstract

Antisense long, noncoding RNAs (ASlncRNAs) have been implicated in regulating gene expression in response to physiological cues. However, little is known about what drives the evolution of their expression. One pathway that can influence steady-state ASlncRNA levels is RNAi, which was lost during budding yeast evolution. Here, we present evidence that RNAi has imposed selective pressure to maintain the expression of ASlncRNA at low levels, and that the loss of RNAi has alleviated this evolutionary constrain. As a consequence, the loss of RNAi has allowed rapid genome-wide expansion of ASlncRNA transcriptomes among budding yeast species.

Introduction

Recent advancement in genome-wide detection of RNA has revealed that long noncoding RNAs (lncRNAs) are transcribed throughout eukaryotic genomes. One class of lncRNAs includes those that are antisense to open reading frames (ASlncRNAs). Many lncRNAs have been shown to play key regulatory roles in metazoans, such as *HOTAIR* and *Xist* (Rinn et al. 2007; Gupta et al. 2010; Tsai et al. 2010; Lee and Bartolomei 2013), as well as ASlncRNAs that overlap with the *GAL10* (Houseley et al. 2008) and *PHO84* genes (Camblong et al. 2007; Camblong et al. 2009; Castelnuovo et al. 2013) in the budding yeast *Saccharomyces cerevisiae*, which repress the expression of overlapping mRNAs in response to environmental cues. However, the biological functions and regulatory mechanisms of the vast majority of lncRNAs have not been determined. We have recently shown that global elevation of ASlncRNA levels is detrimental to *S. cerevisiae* in the presence of reconstituted RNA interference (RNAi), likely due to a large-scale destabilization of mRNAs and lncRNAs (Drinnenberg et al. 2009; Weinberg et al. 2011; Nakanishi et al. 2012; Alcid and Tsukiyama 2014). Taking

advantage of this finding, we used a reconstituted RNAi system in *S. cerevisiae* as a tool to systematically identify ASlncRNA repressors (Drinnenberg et al. 2009; Drinnenberg et al. 2011; Alcid and Tsukiyama 2014). One prediction that follows from the fitness cost caused by elevated ASlncRNA levels in the presence of RNAi is the global attenuation of ASlncRNA levels in the presence of RNAi during evolution. Thus, abrogation of RNAi could have led to genome-wide increases in ASlncRNA levels over the course of budding yeast evolution.

In this report, we investigate how the loss of RNAi in budding yeast has affected the evolution of the lncRNA transcriptome. Evolutionary analyses of lncRNAs have been performed in other organisms (Kutter et al. 2012; Necsulea et al. 2014). However, evolutionary analyses across the budding yeast lineage provide a unique opportunity to investigate how the loss of RNAi, which was likely caused by selection pressure imposed by killer, the double stranded RNA virus (Drinnenberg et al. 2011), has affected ASlncRNA transcription on a global level across evolution. Using an evolutionary genomics approach, we show that genome-wide, steady-state levels of ASlncRNAs have progressively increased since divergence from *Saccharomyces castelliii*, and that divergent, bi-directional promoters are a significant source of this increase. We then demonstrate that even cryptic ASlncRNAs, which are rapidly degraded by the exosome, are under the selective pressure to be maintained at low levels in the presence of RNAi. Finally, we provide evidence that the loss of RNAi in budding yeast is at least partially responsible for the observed ASlncRNA increases and that *S. castellii,* a natural host of RNAi, maintains low levels of ASlncRNAs to ensure optimal fitness.

Results and Discussion

RNAi status is anti-correlated with expression of ASlncRNAs at GAL10 and PHO84 loci

Because elevated levels of ASlncRNAs are deleterious to cells in the presence of RNAi, we hypothesized that the loss of RNAi enabled budding yeast species to expand their ASlncRNA transcriptome. We also hypothesized that ASlncRNAs with regulatory functions would be well-conserved at the level of expression among yeast species lacking RNAi machinery. To test these hypotheses, we first measured antisense expression at the *PHO84* (Camblong et al. 2007; Camblong et al. 2009; Castelnuovo et al. 2013) and *GAL10* (Houseley et al. 2008) genes, which are known to have regulatory functions in *S. cerevisiae.* Strand-specific reverse transcription followed by quantitative PCR across six species of budding yeast revealed that, at *PHO84,* antisense expression started at a very low level in *S. castellii,* and gradually increased across the *sensu stricto* phylogeny. (Fig. 1A) In contrast, antisense expression at the *GAL10* locus is very low in *S. castellii* and *S. bayanus,* but robustly expressed in all species of *sensu stricto* budding yeast that completely lack RNAi. Together, these results support the possibility that the presence of a functional RNAi pathway is anti-correlated with ASlncRNA transcription, and that the expression of functional ASlncRNAs is conserved throughout evolution.

Evolution of ASlncRNA transcriptomes across the budding yeast phylogeny

To globally identify ASlncRNAs conserved across budding yeast evolution, we next performed strand-specific, high-throughput RNA sequencing in haploid *S. castellii*, *S. bayanus, S. kudriavzevii, S. mikatae,* and *S. cerevisiae* to measure steady-state

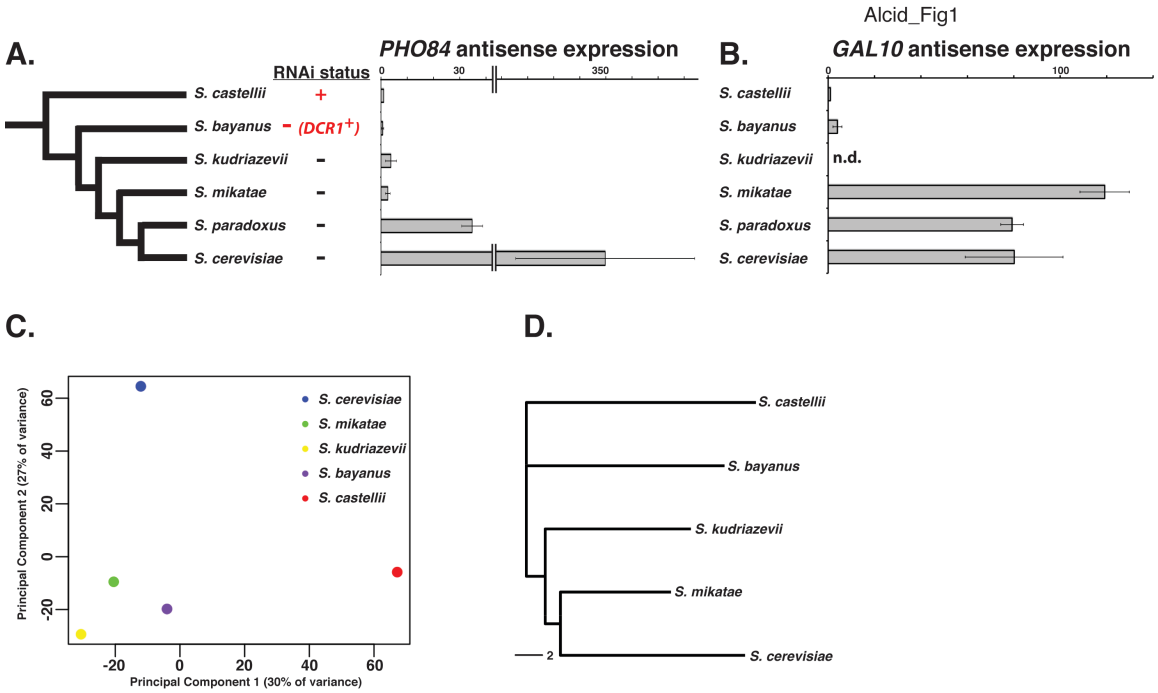


Figure 1 ASlncRNA expression patterns among budding yeast (A) From left to right: Cladogram of *S. castellii* and *sensu stricto* budding yeasts, functional RNAi status for each species, and expression levels of *PHO84* lncRNA for each species as determined by RT-qPCR. Expression level is relative to *S. castellii,* whose value was set to 1. (B) Expression levels of *GAL10* lncRNA for each species as determined by RT-qPCR. As in (A), the expression level is relative to *S. castellii,* which was set to 1. For (A) and (B), mean and standard error were determined using 3 replicates. (C) Principal Component Analysis (PCA) of ASlncRNA transcriptomes in budding yeast. (D) Neighbor-joining tree based on pairwise distance matrices (Jensen-Shannon distance metric) for *sensu stricto* budding yeasts and *S. castellii.*

ASlncRNA levels, using total RNA depleted of ribosomal RNA. We utilized the Yeast Gene Order Browser (K.P. and Wolfe 2005) and homology searches to identify a total of 5,031 orthologous genes for each species. Because it is possible that ASlncRNAs might overlap mRNA transcripts at Untranslated Regions (UTRs), we identified putative UTRs

using by finding local minima of sequencing read density within 300 basepairs of open reading frame (ORF) boundaries (see Materials and Methods). We then quantified and obtained normalized counts for antisense reads mapping to every orthologous gene for every species to obtain its ASlncRNA transcriptome (Anders and Huber 2010). To initially determine global ASlncRNA profile similarities among the species, we performed principal component analysis. Along the first principal component, the ASlncRNA transcriptomes of yeast species lacking RNAi clearly separated away from *S. castellii*, suggesting extensive rewiring of ASlncRNA transcriptomes since divergence from *S. castellii* (Fig 1C). This trend was not as pronounced for sense RNA transcriptomes (Supplemental Fig. 1A). Furthermore, ASlncRNA transcriptome similarity correlated with evolutionary distance (Supplemental Fig. 3). The striking difference in the ASlncRNA transcriptome of *S. castellii* as compared to the other yeast species is likely not simply due to RNAi system destabilizing ASlncRNAs globally, as *dcr1* deletion, which completely inactivates RNAi in *S. castellii* (Drinnenberg et al. 2009), had little effect on ASlncRNA transcriptome in this organism (see below).

To investigate the evolution of ASlncRNA transcriptomes further, we constructed distance matrices for each species using the Jensen-Shannon distance metric (Merkin 2012), and constructed ASlncRNA expression trees (Fig 1D). The resulting tree is highly congruent with the known budding yeast phylogeny, which is striking given the degree of relatedness between *sensu stricto* budding yeasts: *S. cerevisiae* and *S. bayanus* diverged from their common ancestor ~20-40 million years ago (Dujon 2006; D.R. et al. 2011). Interestingly, expression trees using sense RNA levels did not reflect the known budding yeast phylogeny (Supplemental Fig. 1B). This is likely due to mRNA transcriptomes being evolutionarily much more stable and highly conserved, making the

tree highly sensitive to even more subtle differences (Supplemental Fig. 2). These results suggest that changes in ASlncRNA transcriptomes occur across evolutionary transitions, and that they are much more divergent than mRNA transcriptomes.

Expansion of ASlncRNA transcriptomes since divergence from S. castellii

We next investigated how the loss of RNAi has affected the global levels of ASlncRNA transcripts among *Saccharomyces* species. When we measure the distribution of the transcript levels of all ASlncRNAs arising from 5031 orthologous ORFs, we find a clear, gradual increase in ASlncRNA levels across budding yeast evolution ($p << 2e\text{-}16$ for *S. cerevisiae* and *S. castellii,* 2-sided Kolmogorov-Smirnov test*,* Fig. 2A, left). Similar increases were not found when mRNA distributions for each species were assessed ($p = 0.9583$ for *S. cerevisiae* and *S. castellii,* 2-sided Kolmogorov-Smirnov test*,* Fig. 2A, right). This striking result suggests that many ASlncRNAs in *S. cerevisiae* started rapidly expanding immediately after the loss of RNAi (divergence from *S. castellii*)*,* and that the rewiring of the ASlncRNA transcriptomes continued as species progressively diverged from *S. castellii*. The increase in ASlncRNA expression since divergence from *S. castellii* could have largely come from two possible sources: transcription termination defects at convergent genes, or divergent promoters at nucleosome-depleted regions (NDRs) at genes arranged in tandem that overlap the upstream gene (Fig. 2B and 2C, bottom). To assess the contributions of termination defects and divergent promoters to the ASlncRNA transcriptome, we separated all ORFs into whether they are arranged convergently with their downstream gene, or in tandem. We then measured antisense tag density for each gene, and performed metagene analysis for each orientation category for each species. For every species, expression

of ASlncRNA from convergent genes is ~4-fold higher than ASlncRNA arising from tandem genes (Fig. 2B and 2C). At convergently oriented genes, *S. bayanus* exhibits modest, but significant increase in antisense read density compared to *S. castellii*. The same pattern in antisense read density is observed in the other yeast species, though

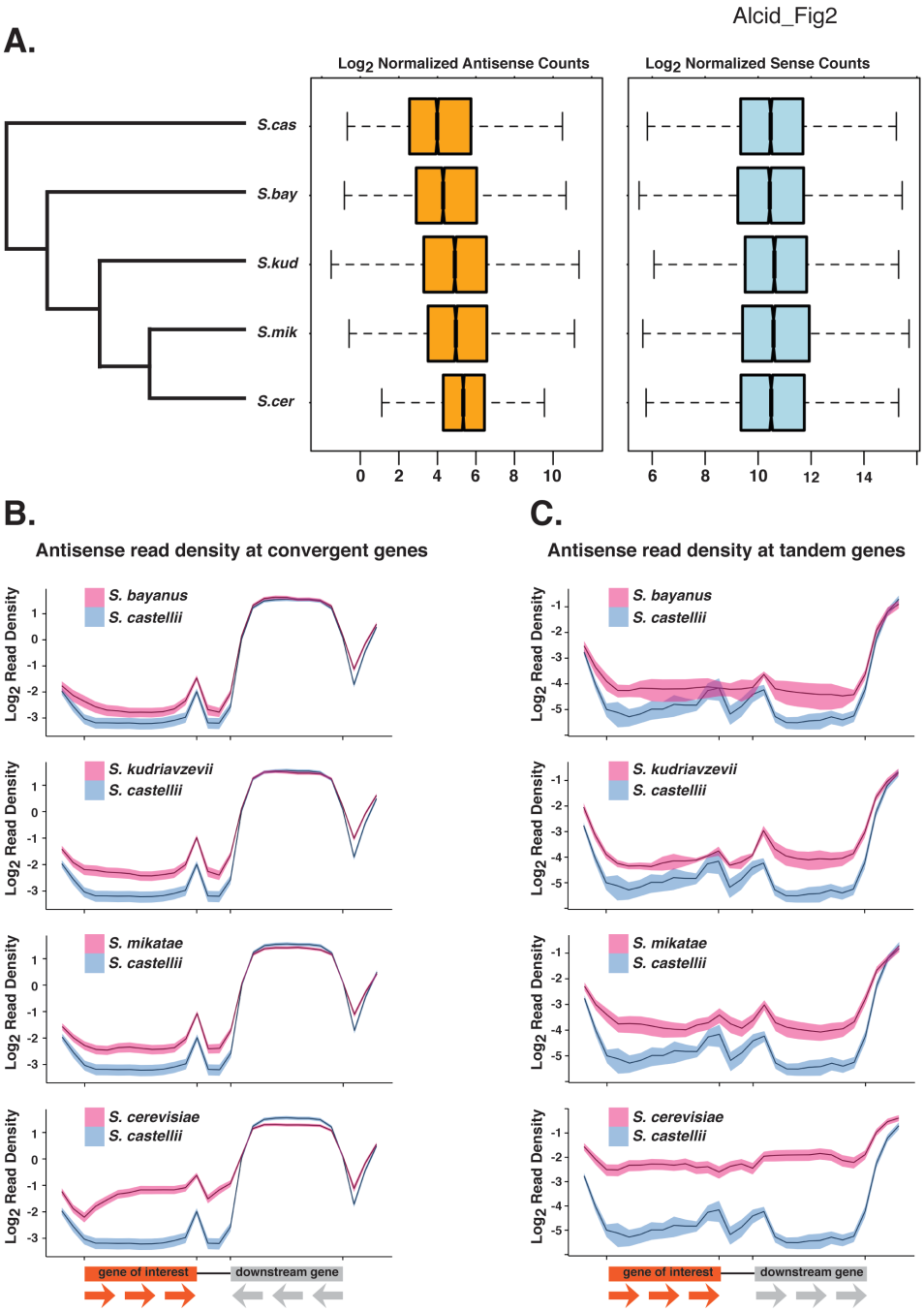


Figure 2 Elevation of ASlncRNA levels across budding yeast phylogeny (A) From left to right: Cladogram of *S. castellii* and *sensu stricto* budding yeast. Boxplots of distributions of normalized read counts (log2 scale) mapping antisense at 5031 orthologous open reading frames for budding yeasts. (B) As in (A), except reads mapping in the sense orientation. (C-D) ASlncRNA meta-analyses (C) Ribbon Plots of antisense read density in log2-scale at genes arranged in convergent orientation for (top to bottom) *S. bayanus, S. kudriavzevii, S. mikatae, S. cerevisiae*. *S. castellii* is represented in all the plots by the blue ribbon. The lines represent the antisense RNA-seq signal, while the outer borders of the ribbon represent 1 standard-error of the mean away from the mean. (D) As in (C), except for genes arranged in tandem orientation.

the relative increase compared to *S. castellii* expands across the phylogeny, with the difference between *S. cerevisiae* and *S. castellii* being the largest (Fig 2B). It should be noted that, for the convergent genes analyzed, the transcripts analyzed for the downstream genes are in the sense orientation, which shows striking similarities in abundance between yeast species. Similarly, antisense read density is low across the gene bodies for tandemly oriented genes in *S. castellii*. When we compare this to the other yeast species, we find a relative increase in the difference across they phylogeny, with the difference between *S. cerevisiae* and *S. castellii* again being the largest. Furthermore, the difference between these two species at tandem genes is even more pronounced than at convergent genes, (~8 vs ~4 fold, respectively) (Fig 2B and C, bottom panel). Similar analysis examining sense expression revealed minimal differences between all species (Supplemental Fig. 4). Taken together, this analysis revealed that after the divergence from *S. castellii* and the loss of RNAi, ASlncRNA levels have increased at both convergent and tandem genes, though more so at tandem genes, suggesting that increased divergent transcription is one of the driving forces underlying robust ASlncRNA transcription programs in *sensu stricto* species. Consistent with this possibility, 33% of ASlncRNAs in *S. cerevisiae* transcribed from divergent promoters have the preinitiation complex dedicated to them (not shared with mRNA transcripts, $p = 0.01$, hypergeometric test) (Rhee and Pugh 2012).

Identification of conserved ASlncRNAs reveals species-specific bursts in expression and putative biological roles

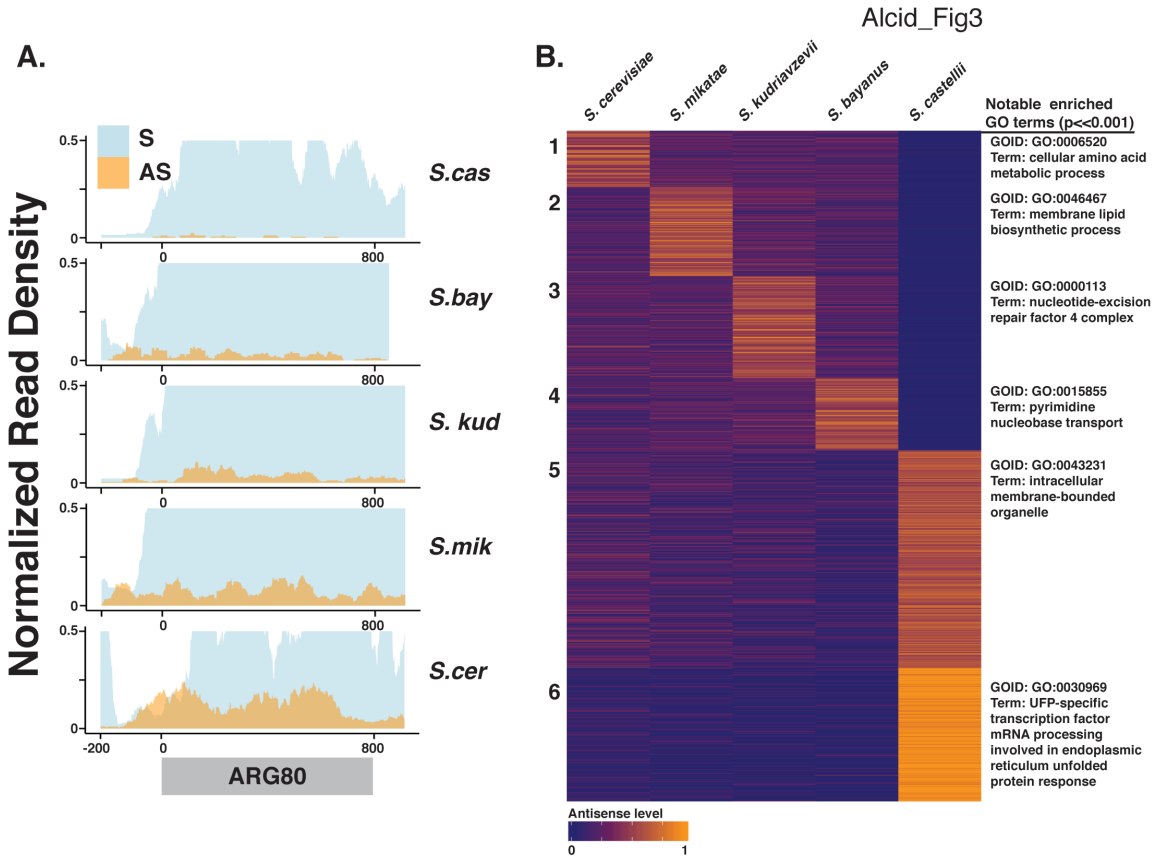


Figure 3 Identification of conserved ASlncRNAs in *sensu stricto* budding yeasts (A) A representative gene (*ARG80)* with conserved, normalized ASlncRNA expression in *sensu stricto* budding yeasts. Light blue represents sense read density (S), while orange denotes antisense read density (AS). The list of all genes with conserved antisense expression is in Table S1. (B) k-means clustering (number of clusters = 6) of all shared differentially expressed ASlncRNAs (n=1284) between *S. castellii* and *S. cerevisiae, S. mikatae, S. kudriavzevii, S. bayanus.* Right: enriched gene ontology terms within each cluster as determined by GOSeq.

We next sought to identify ASlncRNAs that arose after divergence from *S. castellii* and maintained expression across the *sensu stricto* phylogeny. To this end, we performed differential expression analysis of each *sensu stricto* species with *S. castellii,* and identified genes whose antisense expression is elevated relative to *S. castellii,* and present in all four *sensu stricto* species. In all, we identified 612 genes that have conserved antisense expression and that likely represent a lineage-specific shift in ASlncRNA expression (Fig 3A, Supplemental Table S1). As discussed above, although

thousands of lncRNAs are transcribed from the *S. cerevisiae* genome, only a handful of them have known regulatory roles. As such, outstanding questions concerning lncRNAs are whether we can enrich for lncRNAs with regulatory roles, and whether any specific biological processes might be regulated by lncRNAs. To address these questions, we analyzed ASlncRNAs that are expressed across *sensu stricto* species but are absent in *S. castellii*, as well as those that are expressed at significantly higher levels only in *S. castellii*. K-means clustering (n = 6) on these ASlncRNAs revealed a striking emergence of species-specific ASlncRNAs during budding yeast evolution (Fig. 3B). Gene Ontology analysis using the open reading frames that are overlapped by each ASlncRNA in each cluster lac (Young et al. 2010) revealed that genes involved in metabolism and growth are highly enriched among those conserved ASlncRNAs in *sensu stricto* species. On the other hand, genes involved in membrane-bound organelle and unfolded protein response are significantly enriched for *S. castellii*-specific ASlncRNAs. These results collectively suggest that the loss of RNAi alleviated the selective pressure to attenuate ASlncRNA levels in *sensu stricto* species, which allowed not only an expansion of the ASlncRNA repertoire, but a rewiring of regulatory circuits, presumably in a fashion most suitable to their natural environment.

Elevation of cryptic ASlncRNA transcriptomes after the loss of RNAi

The majority of lncRNAs are rapidly degraded by the exosome, a highly conserved exonuclease (Neil et al. 2009; Xu et al. 2009). We therefore investigated whether the stability of ASlncRNA transcripts is correlated with the degree of evolutionary constrain imposed on ASlncRNAs in the presence of RNAi. One possibility

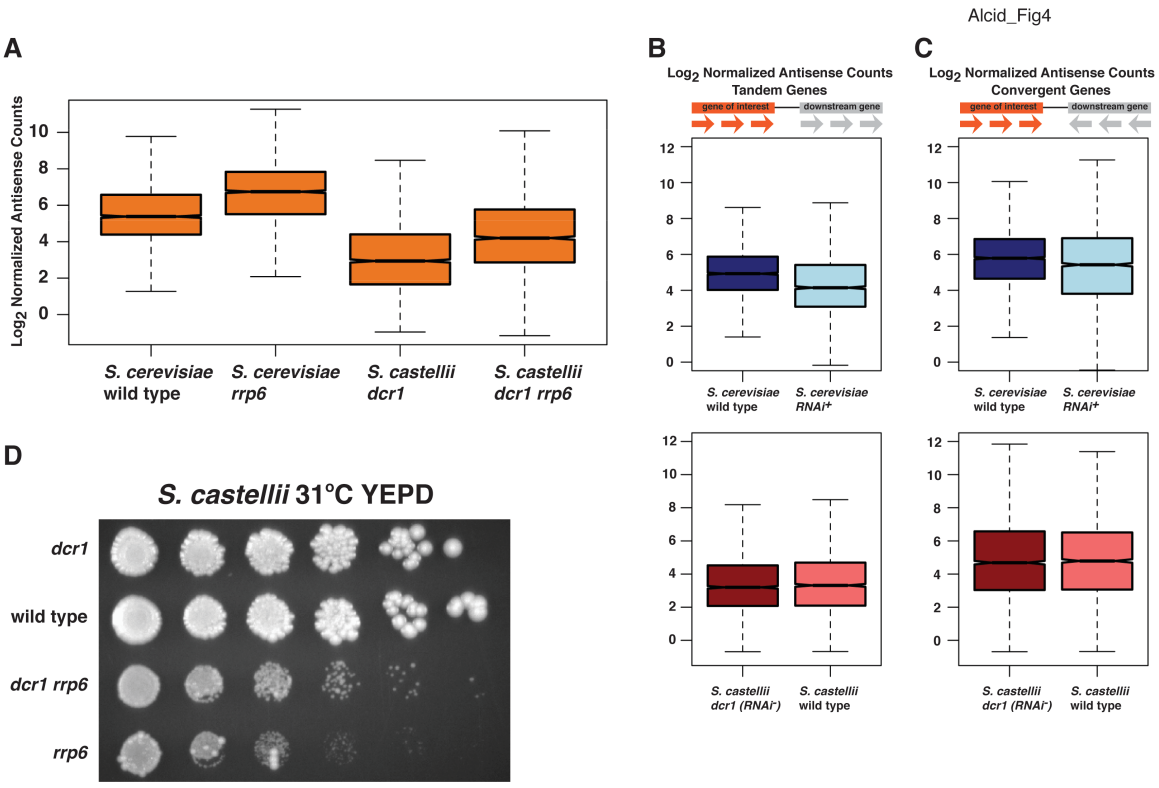


Figure 4 RNAi constrains ASlncRNA expression (A) Boxplots of distribution of normalized read counts at CUT-ASlncRNAs in control and *rrp6* strains of *S. cerevisiae* and *S. castellii.* (B) Boxplots of the distribution of normalized read counts of ASlncRNAs at tandem oriented genes for (Top) wild type and RNAi+ *S. cerevisiae* or (Bottom) wild type and *dcr1 S. castellii.* (C) As in (B), except at convergent oriented genes. (D) A spot test of 5-fold serial dilutions of *S. castellii dcr1,* wild type, *rrp6, dcr1 rrp6* strains on YEPD at an elevated temperature (*S. castellii* grows optimally at 25°C.)

is that Stable Unannotated Transcripts (SUTs) that are antisense (SUT-ASlncRNAs), whose levels persist even in the presence of the exosome, are more likely to form double-stranded RNA with mRNAs as compared to Cryptic Unstable Transcripts (CUT-ASlncRNAss) that are degraded by the exosome. If this is the case, SUT-ASlncRNAs would be under stronger constrain by RNAi. We would then expect the difference in the levels of SUT-ASlncRNAs between *S. cerevisiae* and *S. castellii* to be much larger than that of CUT-ASlncRNAs between the two species. Alternatively, the evolutionary constrain against transcription of ASlncRNAs in the presence of RNAi may be strong

enough to keep the abundance of both SUT- and CUT-ASlncRNAs low in *S. castellii*. To address these possibilities, we mutated *RRP6,* an exosome component, in *S. castellii* and globally compared its cryptic ASlncRNA transcriptomes to that of *S. cerevisiae's* (Alcid and Tsukiyama 2014). As expected, the abundance of ASlncRNAs strongly increases in *S. cerevisiae* (Figure 4A), due to stabilization of CUTs. In *S. castillii*, *DCR1* can also degrade ASlncRNAs, which can confound our analyses of CUT-ASlncRNAs. We therefore introduced null *RRP6* mutations in *S. castellii* in a *dcr1* background. In contrast to *S. cerevisiae*, our analyses revealed that mutation of *RRP6* in *S. castellii* revealed only a modest, though significant, increase in ASlncRNA expression when compared to the control strain (*dcr1* alone) (Fig 4A). As a result, the difference in the ASlncRNA levels between *S. cerevisiae rrp6* and *S. castillii dcr1 rrp6* mutants is comparable, if not larger, than that between wild type *S. cerevisiae* and *S. castillii dcr1* mutant (Supplemental Fig. 5A and 5B). Given that CUTs and SUTs are enriched in the transcriptome in *rrp6* mutant and cells with wild type exosome, respectively (Xu et al. 2009), our results strongly suggests that evolutionary constrain against transcription of ASlncRNAs is robust enough that expression of SUTs and CUTs has been similarly attenuated in the presence of RNAi.

Differential impact of RNAi on ASlncRNA levels in S. cerevisiae and S. castellii

Given that RNAi has likely constrained ASlncRNA transcription throughout evolution, we hypothesized that ASlncRNA transcription levels in *S. castellii* is low enough that the presence of RNAi may not affect ASlncRNA levels as strongly as in *S. cerevisiae.* To test this, we compared genome-wide levels of ASlncRNAs in our wild type *S. cerevisiae* strain, and an *S. cerevisiae* strain where RNAi is reconstituted

(Drinnenberg et al. 2011; Alcid and Tsukiyama 2014). This analysis shows that reconstitution of RNAi leads to a significant decrease of ASlncRNA levels at both convergently and tandemly oriented genes (Fig. 4B and 4C, top panels, $p << 2.2e-16$ for both orientations, 2-sided Kolmogorov-Smirnov test). In sharp contrast, we found that disabling RNAi in *S. castellii* by a *dcr1* mutation had no statistically significant effect on endogenous ASlncRNA levels at both convergent and tandem genes ($p = 0.52$ and $p = 0.08$, respectively, 2-sided Kolmogorov-Smirnov test) (Fig. 4B and 4C, bottom panels). These results support our model in which the loss of RNAi alleviated the necessity for *sensu stricto* species to keep the ASlncRNA levels low, which allowed them to evolve a large number of ASlncRNA transcripts.

Elevation of ASlncRNAs in the presence of RNAi leads to fitness defects in S. castellii

The results above support the notion that attenuation of ASlncRNA expression in the presence of RNAi ensures optimal cellular fitness. Indeed, it has been shown that elevation of ASlncRNAs in the presence of RNAi leads to synthetic slow growth/lethal phenotypes in *S. cerevisiae* (Alcid and Tsukiyama 2014). To examine whether this is the case in *S. castellii,* a natural host of RNAi (Drinnenberg et al. 2009), we performed a genetic test. Mutation of *RRP6* in *S. castellii* leads to a slow growth phenotype (Fig 4D). This suggests that, although the effects of this mutation on the abundance of ASlncRNAs are not as strong as in *S. cerevisiae* (Figure 4A), it does cause a fitness defect in *S. castillii*. If this fitness issue is at least partly due to RNAi globally destabilizing transcripts, deletion of *DCR1* is expected to rescue the slow growth phenotype. As shown in Figure 4D, this turns out to be the case, supporting our model

that processing of mRNA-ASlncRNA hybrids by RNAi is at least one way that RNAi has helped maintain low levels of ASlncRNA expression across budding yeast evolution.

Concluding remarks

We have shown that global ASlncRNA levels increase across the budding yeast phylogeny after the loss of an RNAi pathway in *sensu stricto* species of budding yeast. We further provided evidence that the loss of RNAi has alleviated the selective pressure to maintain the expression levels of ASlncRNA low, allowing steady expansion of ASlncRNA transcriptome in *sensu stricto* species. To our knowledge, this is the first report to demonstrate that RNAi profoundly affects the evolution of lncRNA transcriptomes, though it has been speculated before (Yassour et al. 2010). In this regard it is interesting to note that, despite possessing an active RNAi pathway, *S. castellii* still has detectable antisense expression at a large number of genes. This is analogous to many higher eukaryotes that keep RNAi while having abundant lncRNAs transcribed. As such, organisms that maintain both RNAi and ASlncRNAs likely possess currently unknown mechanisms that mitigate the deleterious effects of having both systems coexist. Given that elevation of ASlncRNAs in the presence of RNAi leads to a significant fitness cost to both *S. cerevisiae* and *S. castellii*, it will be very interesting to investigate whether the incompatibility between high levels of ASlncRNA across the genome and RNAi extends to metazoans.

Materials and Methods

Yeast strains

A list of all strains used in this study can be found in table S2. We carried out single-step gene deletions by standard lithium acetate transformation using NatMX drug-resistance markers as described for *S. cerevisiae* (Goldstein and McCusker 1999). For *S. castellii,* we performed gene deletions as previously described in (Krawchuck and Wahls 1999). Strains were also created using standard genetic crosses. For *S. cerevisiae,* genome sequences and annotations were downloaded from Ensembl (Cunningham et al. 2015) or the Saccharomyces Genome Database (Cherry et al. 2012). For all other yeast strains, genome sequences and annotations were downloaded from the Yeast Gene Order Browser (K.P. and Wolfe 2005).

Yeast growth conditions

Strains were cultured at 30°C or 25°C in YPD until OD600 = 0.4-0.7 before being harvested for RNA using standard hot acid phenol extraction.

Strand-specific library preparation and high-throughput RNA sequencing

For every strain, 3ug of Total RNA was depleted of ribosomal RNA species using Ribo-Zero magnetic rRNA removal kit (Human/Mouse/Rat) (Epicentre). Strand-specific libraries were then prepared using the dUTP method combined with TruSeq (Illumina) as previously described (Parkhomchuk et al. 2009; Sultan et al. 2012). 50 cycles of paired-end sequencing was performed on an Illumina HiSeq 2500 on either high-output mode or rapid run mode (FHCRC shared resources). All sequencing experiments were performed in biological duplicate.

Identification of orthologous genes among *S. cerevisiae, S. mikatae, S. kudriavzevii, S. bayanus, S. castellii* An initial set of orthologous genes was identified using the "Pillars.tab" file from YGOB, corresponding to 4894 orthologous genes. To identify additional orthologous genes, we aligned all open reading frame amino acid sequences for each species to all open reading frame amino acid sequences for *S. cerevisiae* using LAST (Kielbasa et al. 2011). We then identified the 20th percentile alignment score and set this as the minimum threshold. All remaining amino acid sequences not previously identified in "Pillars.tab" but having an alignment score at or above the minimum threshold were then identified as additional orthologs, resulting in a total of 5031 gene orthologs shared among the 5 yeast species..

RNA-seq analysis

Alignment

Reads were aligned to the species-specific genome using TopHat2 (Kim et al. 2013) with the following settings: tophat2 -p 4 -G <gene_annotation_file> -I 2000 --library-type=fr-firststrand -o <output_directory> <bowtie_index> <Read1.fastq> <Read2.fastq>. Reads were then trimmed of adapter sequences with a custom Python script using the Python module HTSeq (Anders et al. 2014).

Heuristic of RNA-seq data to identify putative untranslated regions (UTRs)

After reads were aligned, reads were filtered such that only properly aligned, uniquely mapped reads were kept using a custom Python script and pysam (Li et al. 2009). Because replicates were highly reproducible (data not shown), reads for each replicate were combined to make per-base, strand-specific pileup files using pysam. Using this pileup file, putative 5' and 3' UTRs were identified by starting at either the start codon

codon or stop codon coordinate, respectively, for each orthologous gene and extending away from the open reading frame boundary until a local minimum in the per/bp read density was encountered within 300 bp from the gene boundary. The coordinate where this is achieved served as the outer UTR coordinate. A custom python script was written that implements this (available upon request).

Identification of differentially expressed ASlncRNAs

Using the putative orthologous transcript list (with adjusted UTRs) for each species, differentially expressed ASlncRNA units were defined by first enumerating the number of reads in each replicate that overlap antisense to each transcript, then using a negative binomial distribution (R-package DESeq2) (Anders and Huber 2010) to determine differential expression. ASlncRNAs that had a p-adjusted value <= 0.2 were determined to be differentially expressed. Fold-change, as well as absolute expression (in normalized count values) were determined using DESeq. A list of all conserved ASlncRNAs is shown in Table S1.

Construction of CUT-ASlncRNA distributions

To identify CUT-ASlncRNAs, only those ASlncRNAs whose whose log2-fold change >= 0 were kept, as previously described (Xu et al. 2009), leading to 2420 and 2481 CUT-ASlncRNAs for *S. cerevisiae* and *S. castellii*, respectively. These populations were then used as distributions for boxplots and histograms.

Meta-analyses of RNA-seq data (Fig. 2, Supplemental Fig. 4)

To perform meta-analysis, we first normalized reads/per-base coverage files by the genome-wide average. Full-length transcripts (starts and ends adjusted by putative

UTRs) were then binned into 10 equally-sized bins, while upstream regions, downstream regions, and intergenic regions, were divided into 3 equally sized bines. Every binned region was then aligned by the putative transcription start-site, and the average of each aligned bin was found. This data was used to construct the ribbon plots (see below).

Construction of heatmaps, plots, statistical and phylogenetic analysis

Heatmaps, plots, and meta-gene plots were constructed in R (R Development Core Team 2013) using the packages ggplot (Wickham 2009). K-means clustering was performed using the "kmeans" function in R, with 6 clusters. Before clustering was performed, all genes (rows) were normalized such that expression summed to 1. Principal Component Analysis was performed on ASlncRNA transcriptomes and sense transcriptomes using the R function prcomp. Jensen-shannon distance metrics were calculated as previously described (Merkin 2012). Neighbor-joining trees were then created using the R-package "ape" (Paradis et al. 2004). 2-sided Kolmogorov-Smirnov tests were performed using the R function ks.test.

Strand-specific RT-PCR

Strand-specific RT-PCR was performed for *PHO84* and *GAL10* as previously described (Chatterjee et al. 2012). Calculation of relative expression was performed using the $\Delta\Delta C_t$ method, normalized to either *ACT1* or *IPP1.*

Gene Ontology Analysis

All gene ontology analysis was performed using GOSeq (Young et al. 2010)

Accession numbers

RNA-sequencing reads have been deposited to the Sequence Read Archive under BioProject PRJNA278671.

Author Contributions

E.A.A. contributed in planning and performing experiments, analyzing and interpreting data, and writing this manuscript. T.T. contributed in planning experiments, interpreting data and writing this manuscript.

Acknowledgements

We thank H. Malik, I.A. Drinnenberg, and the members of the Tsukiyama lab for helpful discussions; H. Malik, I.A. Drinnenberg for critical reading of the manuscript; M. Dunham and D. Gottschling for yeast strains; A. Marty and FHCRC shared resources for deep sequencing. This work was supported by grant RO1 GM058465 (T.T.) and predoctoral fellowship F31 GM101944 (E.A.A.).

Chapter 4: Perspectives and Conclusions

The incredible rate at which genomics technologies have improved has paved the way towards high-resolution studies of the transcriptomes of any organism, first with high-resolution tiled microarrays, followed by high-throughput RNA sequencing. Because these advances are relatively recent, we are only beginning to appreciate the complexity of transcription. In eukaryotes, these techniques have revealed that transcription is highly pervasive, such that number of lncRNA transcription units is vastly greater than the number of coding transcription units. Moreover, in contrast to coding RNAs, which have a relatively defined gene structure, lncRNAs are much more diverse in their genomic qualities, and can arise and terminate in any genomic boundary, and on both strands of DNA (Berretta and Morillon 2009). In this thesis, we have uncovered fundamental principles underlying lncRNA transcription and evolution. We have established chromatin-remodeling factors as having a general role in lncRNA repression. Additionally, we posit that the RNAi pathway constrains the levels of lncRNAs so as to ensure optimal cellular fitness. A major question in the lncRNA field concerns the function of the vast majority of identified lncRNAs, most of which are as of yet uncovered. We will now speculate on how the previous two chapters might address this question of fundamental importance.

Chromatin structure maintenance and lncRNA regulation

As mentioned in the introduction, there were two screens that served as precursors to the one presented in this thesis. Both took advantage of reporter constructs as their outputs to detect repressors of intragenic (Cheung et al. 2008) lncRNAs, and repressors of divergent lncRNAs (Marquardt et al. 2014). Both studies further established chromatin as playing a role in lncRNA regulation. However, the issue that both of these

studies raise concerns the functions of the lncRNAs subsequently identified. Both screens revealed roles for genes involved in chromatin maintenance and assembly. The screen described by Cheung, et al identified Spt6 as a major factor involved in the repression of intrangenic cryptic transcription. Spt6 is responsible for the remodeling of nucleosomes and maintenance of chromatin structure as RNA Pol II passes (Bortvin and Winston 1996; Hartzog et al. 1998). Microarray analysis revealed the de-repression of ~1000 lncRNAs when *SPT6* is mutated. Similarly, the screen described by Marquardt, et al identified CAF-1 as a repressor of noncoding, divergent transcription from bidirectional promoters. It should be mentioned that they also identified Swr1, Isw2, and Rsc in their screen, consistent with our study. Mutation of the *CAC1* subunit of CAF-1 also led to the de-repression of many lncRNAs across the genome initiating divergently from coding genes. These studies strongly suggest that defects in chromatin assembly leads to the de-repression of cryptic, intragenic lncRNAs and lncRNAs originating form bidirectional promoters. Whether these lncRNAs have any functions is still unknown, and needs to be followed up. A promising result from the study by Cheung, et al revealed that many of the lncRNAs that they identified are up-regulated upon nutrient depletion even in the presence of wild type Spt6 (Cheung et al. 2008).

Dedicated cellular resources to repress lncRNA genome-wide

In this thesis, we describe the development of an unbiased, genome-wide screen using reconstituted RNAi as a tool to identify repressors of lncRNA. Unlike the previously mentioned screens, the screen that we describe would theoretically identify repressors of lncRNAs that initiate both intragenically as well as intergenically. This is likely true, as our screen also identified subunits of the Rpd3 complex, which is known to repress cryptic intragenic transcripts (Carrozza et al. 2005; Houseley et al. 2008; Lickwar et al.

2009). Furthermore, our screen identified the chromatin remodeling factors Swr1, Isw2, Rsc, and Ino80 as direct repressors of lncRNAs. The fact that all four putative lncRNA repressors were confirmed as such by strand-specific RNA sequencing suggests that the screen that we developed was effective. Our screen, in total, identified 408 putative lncRNA repressors. This striking result suggests that the cell devotes numerous resources to repress lncRNAs genome-wide. One class of genes was those involved with mitochondrial maintenance. Whether intra-organellar communication between the mitochondria and the nucleus involving lncRNAs exists should be further investigated.

Chromatin remodeling factors likely regulate functional lncRNAs

Notably, these 4 remodeling factors have opposing biochemical activities. Isw2 is thought to decrease the size of NDRs by shifting the +1 and -1 nucleosomes into them (Fazzio and Tsukiyama 2003; Whitehouse et al. 2007; Yadon et al. 2010). Conversely, Rsc is thought to destabilize the +1 and -1 nucleosomes (Hartley and Madhani 2009; Floer et al. 2010; Van de Vosse et al. 2013). In a similar manner, Swr1 deposits the histone variant H2A.Z into +1 and -1 nucleosomes (Kobor et al. 2004; Yen et al. 2013), while Ino80 has been shown to remove H2A.Z from these nucleosomes (Papamichos-Chronakis et al. 2011). The +1 and -1 nucleosomes represent the major structural barrier to mRNA and divergent lncRNA transcription respectively (Teves et al. 2014; Weber and Henikoff 2014). These notions imply that, if the populations of lncRNAs regulated by each of these remodeling factors highly overlap, then, similar to the previously mentioned screens, simple disruption of chromatin structure is all that is needed for de-repression of lncRNAs. However, we found that there is minimal overlap between any of the lncRNA populations. Moreover, the highest degree of overlap is between lncRNAs regulated by Ino80 and Rsc, which, as we currently know, do not have

opposing biochemical mechanisms. This notion suggests that specific lncRNAs require precise biochemical activities for their proper repression, and is more suggestive of deliberate "regulation" by the cell. This is not surprising, as Swr1 has been found to be associated with depositing H2A.Z at "regulated" genes, as opposed to constitutively active genes (Li et al. 2005). Additionally, Isw2, Rsc, and Ino80 are all also associated with regulating mRNAs involved in specific pathways, from the regulation of meiotic genes, to the response to osmotic stress (Yukawa et al. 1999; Goldmark et al. 2000; Klopf et al. 2009). Therefore, whether these complexes also regulate lncRNAs that play roles in cellular responses to developmental and environmental cues is highly likely.

RNAi represses ASlncRNAs across evolutionary timescales

We performed comparative genomics using budding yeast as a model to explore the evolution of antisense, lncRNA (ASlncRA) expression. At the time of the writing of this thesis, we conducted the only study to investigate how ASlncRNAs have evolved over time. We found, since divergence from *S. castellii,* a gradual increase in the genome-wide steady-state levels of ASlncRNAs. We also find that, in comparison to mRNA transcriptomes, ASlncRNA transcriptomes evolve much more rapidly. Furthermore, we provide evidence that the rise in ASlncRNA levels is partly due to the loss of RNAi, which constrained endogenous and cryptic ASlncRNAs to a similar degree. We also suggest that the genome-wide elevation of ASlncRNAs in the presence of RNAi is detrimental to the cell.

Rapid evolution of the ASlncRNA transcriptome

mRNA transcriptomes are relatively stable, and the levels of each mRNA persist across species (Spearman's Rho for *S. cerevisiae* and *S. castellii* mRNA transcriptomes = 0.75).

Conversely, ASlncRNAs change rapidly from species to species, such that the correlation coefficients between even closely related species are very divergent. On its face, this notion alone might suggest that ASlncRNAs have modest biological importance; however, we know that ASlncRNAs are capable of regulating the overlapping mRNA and that lncRNAs rapidly evolve in mammalian species (Kutter et al. 2012; Necsulea et al. 2014). Therefore, it is likely that many of the changes in ASlncRNA populations are adaptive. The issue raised, then, is the identification of changes that are indeed adaptive. We do not think that the ASlncRNAs that we have identified represent noise, as only ~20% of ORFs have detectable ASlncRNA expression, many are long in length and proceed across the entire ORF boundary, and the levels of expression can be robust and comparable to that of mRNAs. We also show that, for many of the ASlncRNAs assigned to *S. cerevisiae,* a significant fraction of them have dedicated PICs. Furthermore, ASlncRNAs have minimal correlation with their adjacent mRNA neighbors, suggesting that their expression is independent of neighboring mRNA regulation (Pelechano and Steinmetz 2013).

Cryptic ASlncRNA transcriptomes are constrained by RNAi

One potential explanation for the elevated levels of ASlncRNAs in the *sensu stricto* clade relative to *S. castellii* is that exosomal activity in *S. castellii* is more efficient than in other species. We feel that this is unlikely for the following reasons. First, there is a high degree of sequence conservation between exosomal components among all species examined (data not shown). The catalytically active subunit of the exosome is encoded by the gene *DIS3.* Importantly, there is a high degree of amino acid conservation of this subunit. Additionally, none of the exosomal components are differentially expressed when the ssRNAseq data is analyzed. Together, this data argues against the possibility

of increased efficiency of ASlncRNA degradation in *S. castellii.* However, a more direct test of this is to examine the cryptic transcriptome of *S. castellii.*

We found that mutation of the exosome led to much higher cryptic ASlncRNA levels in *S. cerevisiae* as compared to *S. castellii.* This was initially surprising to us, as it is known that the exosome is a highly processive 3' to 5' exonuclease. Because of the efficiency of the exosome, we reasoned that cryptic ASlncRNAs would be under less constraint by RNAi than stable ASlncRNAs because the newly transcribed ASlncRNA would be degraded before forming a double-stranded intermediate with its cognate mRNA, and subsequent processing by Dcr1. Therefore, upon mutation of *RRP6,* we expected the distribution of cryptic ASlncRNAs to be at comparable levels in *S. castellii* and *S. cerevisiae.* However, we found that, like stable ASlncRNAs, cryptic ASlncRNAs are expressed to a much higher degree in *S. cerevisiae.* If it is assumed that transcription and degradation are uncoupled, then measurement of RNA levels in an exosome mutant is more indicative of total transcriptional output. Therefore, the data presented suggests that transcriptional activity of ASlncRNAs is much greater in *S. cerevisiae* than in *S. castellii,* and it is likely that the other species in the *sensu stricto* clade have elevated cryptic ASlncRNA transcriptomes relative to *S. castellii*.

Reconciling RNAi and ASlncRNAs

Despite *S. castellii* having the least amount of ASlncRNA transcription among all species examined, there were still many genes with elevated ASlncRNA expression. Indeed, most higher eukaryotes, with few exceptions (Ullu et al. 2004), possess RNAi while still having robust ASlncRNA levels across the genome. This raises the question as to how these organisms might reconcile these two systems, especially given that we have

shown that this comes at a detriment to the cell. It has been shown in higher eukaryotes that RNAi can be both positively and negatively regulated. One way to negatively regulate RNAi would be to outcompete Dcr1 for access the dsRNA intermediates. For instance, in *Caenorhabditis elegans*, the gene STAU-1 contains two double stranded RNA binding domains (dsRBDs), and has been shown to attenuate the RNAi response (LeGendre et al. 2013). However, STAU-1 homologs are not found in homology searches of the *S. castellii* and *S. cerevisiae* genomes. It is possible that other dsRBD containing genes in both genomes have current (in the case of *S. castellii)* or extant (in the case of *S. cerevisiae*) functions in RNAi regulation. Another way to modulate RNAi is by regulating Xrn1 activity. It has been shown that Xrn1 is needed for efficient target degradation following endonucleolytic cleavage by Ago1 (Orban and Izaurralde 2005). Indeed, amino acid alignments of Xrn1 across the budding yeast phylogeny reveals extended stretches that are deleted after divergence from *S. castellii.* It is therefore possible that *S. castellii* can modulate RNAi activity through Xrn1. Together, *S. castellii* , as well as other higher eukaroyotes, likely mitigate the deleterious effects of possessing both ASlncRNAs and RNAi by attenuating RNAi activity.

Evidence of conservation of many ASlncRNAs at the level of expression

We showed that 612 genes possess conserved ASlncRNAs at the level of expression across the *sensu stricto* lineage. Expression analyses of mRNAs demonstrate high conservation across evolutionary timescales (Brawand et al. 2011). Therefore, the fact that we identified a subpopulation of ASlncRNAs conserved at the level of expression suggests that this group is enriched for those with important biological functions. K-

means clustering on these ASlncRNAs revealed enriched GO terms that suggest that these ASlncRNAs might be involved in developmental pathways, and that this population of ASlncRNAs that we have identified might also be enriched for function. Furthermore, this analysis stratified these ASlncRNAs by species, suggesting that each species of yeast specifically co-opts some ASlncRNAs in order to adapt to their natural environment. Finally, RT-PCR analysis revealed conserved expression of ASlncRNAs at the *GAL10* locus across the budding yeast phylogeny. Because this ASlncRNA has a well-characterized function in the regulation of genes involved in galactose metabolism (Houseley et al. 2008), it is likely that identifying ASlncRNAs with conserved expression enriches for functional ASlncRNAs. Indeed, construction of co-expression networks for lncRNAs and mRNAs across evolution in mammals has identified potential functional relationships between both classes of these genes (Necsulea et al. 2014).

Chapter 5: Future Directions

The work described in this thesis has uncovered fundamental principles underlying lncRNA regulation and evolution. However, more, important questions have been raised that need to be addressed.

Biological Functions of CRRATs

We were able to identify a population of lncRNAs that are directly regulated by chromatin remodeling factors whose elevation is coincident with repression of the cognate mRNA, potentially uncovering a novel role for these complexes. Furthermore, gene ontology analysis uncovered enriched pathways for Rsc-CRRATs and Ino80-CRRATs. However, a caveat is that these lncRNAs were identified under nutrient-rich conditions. Null mutations of these factors will lead to severe selective disadvantages for yeast species in their natural habitat. Wild type budding yeast in their natural habitat regularly undergo periods of extended stress and/or rapid changes in conditions, such as nutrient deprivation. Therefore a major key to understanding lncRNA function is the identification of conditions where the cognate remodeler's gene is encoded in the genome, but is downregulated or has attenuated activity. One useful example to learn from is meiosis. It was shown that when cells undergo meiosis, exosomal activity is dramatically attenuated by post-translational modification, leading to the stabilization of many cryptic lncRNAs, which they term MUTs (Meiotic Unannotated Transcripts) (Lardenois et al. 2011). It was further shown that MUTs are likely critical for proper meiotic division and spore formation. It is conceivable that a similarly constructed experiment for CRRATs would prove meaningful. For instance, Ino80 undergoes a dramatic decrease in expression under conditions of high arsenic or osmotic stress (Klopf et al. 2009)Haugen

(Haugen et al. 2004). Performing strand-specific RNA sequencing of wild type yeast in these conditions would likely reveal an upregulation of many Ino80-CRRATs. To then show that Ino80-CRRATs are critical for the response to this stress, constitutive expression of the remodeler in the presence of stress would likely lead to a worse phenotype, though the effect on mRNAs cannot be discounted. It is likely that analogous conditions can be identified for Swr1, Isw2, and Rsc.

lncRNA mechanisms of gene regulation

In chapter 2, the 259 CRRATs that we identified as putatively functional can serve as candidates for further mechanistic studies. From the RNAseq data alone, it is difficult to assess whether the ASlncRNA upregulation is a cause of cognate mRNA downregulation, or a consequence. As a result, further experiments are needed to resolve the relationship.

*Determine that antisense ncRNA transcription down-regulates mRNA***.** Selected candidate genes will be tested on whether they are indeed negatively regulated by antisense ncRNA *in vivo*. To this end, in either mutant remodeler background, a transcription termination site will be inserted immediately downstream of the putative lncRNA start site, thus abolishing production of the lncRNA. mRNA levels will then be assayed. If transcriptional interference by the lncRNA is responsible for the down-regulation in mRNA as hypothesized, then blocking antisense lncRNA transcription by using a transcriptional terminator should result in an up-regulation of mRNA. If this is not observed, then it is likely that the particular mRNA is regulated by some other mechanism, and lncRNA upregulation is likely correlative.

Elucidate mechanisms of mRNA transcriptional interference by lncRNA. Down regulation of an mRNA at a given locus can be caused by at least three likely mechanisms: A) recruitment of histone-modifying enzymes by ncRNA. B) transcription-dependent nucleosome deposition. C) TF-displacement by antisense transcription. In model A, as described for HOTAIR and the antisense *PHO84* transcript, the lncRNA transcript serves as a landing pad that is targeted by histone-modifying enzymes, such as HDACs, to down-regulate mRNA transcription. In model B, antisense transcription through the promoter of the mRNA leads to deposition of nucleosomes that block transcription-factor binding in this region. Lastly, in model C, RNA polymerase travelling antisense, through the promoter of a mRNA encoding gene, displaces TFs necessary for mRNA transcription. To directly test these models, we can perform the following molecular genetic experiments, though this is certainly not an exhaustive list. The purpose here is to uncover whether the key regulatory factor is the lncRNA product, or the act of noncoding transcription.

(1) Abolition of ncRNA product: A hammerhead ribozyme sequence can be placed directly downstream of the putative antisense transcription start site in the antisense orientation. Doing so will obliterate the ncRNA product, while preserving the act of antisense transcription (Lacadie et al. 2006). mRNA levels will be assessed by Northern analysis or primer extension and compared to the unaltered sequence not containing the hammerhead ribozyme. If mRNA levels are unaffected, then the lncRNA is unlikely regulatory

(2) Ectopic expression of ncRNA: In addition to using the above-described hammerhead construct, ectopic expression of the lncRNA transcript can be induced by

placing it under the control of a Gal-inducible promoter at an ectopic locus, as previously described (Camblong et al. 2009). mRNA levels can be assessed as above.

(3) ncRNA transcript pull-down: The MS2 RNA sequence can be fused to the 3' end of the endogenous antisense transcript, which will be pulled down using the fusion protein MS2-MBP and an amylose column. Western blots to chromatin modifiers and/or proteomic analysis can be performed to identify factors that might be recruited by the lncRNA.

(4) Abolition of antisense transcription through the mRNA promoter: A transcription termination sequence will be inserted in the antisense orientation just downstream of the sense promoter, abruptly ending antisense transcription before it reaches the promoter region, but preserving sense transcription. Strand-specific Northern blot analysis will be used to assess mRNA levels, as described above.

The results of these tests will likely lead us to the mechanism of lncRNA-mediated regulation of mRNA transcription. A cautious interpretation of these results is required, however, as the underlying DNA sequence will be altered, potentially causing unexpected effects on transcription (Bassett et al. 2014). Whatever the case may be, it is clear that further mechanistic work is needed to better understand how lncRNAs function.

Concluding Remarks

As principles underlying the functions, regulation, and evolution of lncRNAs become elucidated, we can better appreciate the importance of lncRNAs in the biology of the cell

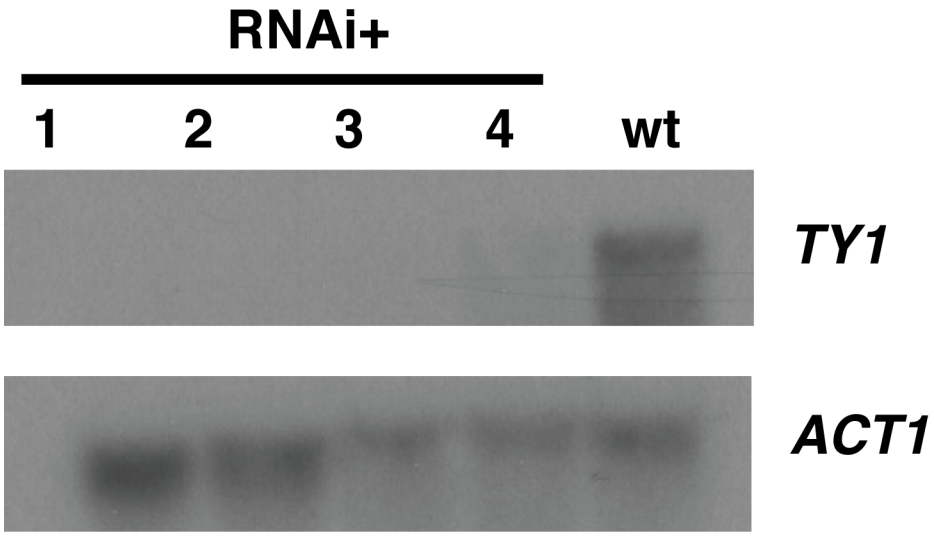
and the development of organisms. Indeed, it is likely that mRNAs and lncRNAs are subject to many of the same regulatory mechanisms: activation by site-specific transcription factors as well as the regulation of chromatin structure around their putative transcription start sites. Because of this, it is highly likely that mis-regulation of lncRNAs can be as devastating to the organism as mis-regulation of mRNAs. Indeed, misexpression of *HOTAIR,* one of the most well-characterized lncRNAs*,* is associated with breast cancers that are aggressively metastatic (Gupta et al. 2010). Therefore, understanding fundamental principles underlying lncRNA-mediated gene repression will allow us to understand normal cell physiology, as well as also disease pathogenesis, progression, and prognosis in the long term.

Appendix A: Supplemental Material for Chapter 2

ATP-dependent Chromatin-remodeling Shapes the Long-Noncoding RNA Landscape

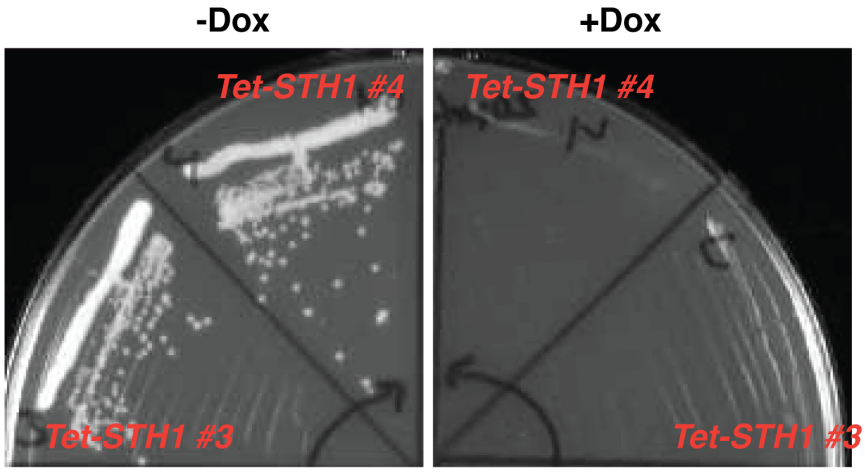
Eric A. Alcid and Toshio Tsukiyama

Figure S1 (related to main Figure 1). Reconstitution of RNAi in *S. cerevisiae.*
Figure S2 (related to main Figures 1 and 2). Growth of cells carrying tet-repressible alleles of *STH1* and *INO80.*
Figure S3 (related to main Figure 3). Length Distribution of Chromatin-Remodeling Regulated lncRNAs.
Figure S4 (related to main Figure 4). Chromatin Remodeling factors Regulate Unique Classes of lncRNAs.
Figure S5 (related to main Figure 5). Expression levels of direct ASlncRNA targets of chromatin-remodeling factors are similar to those of indirect ASlncRNA targets.
Figure S6 (related to main Figure 6). The levels of mRNAs overlapped with CRRATs are more strongly decreased than the rest of mRNAs in chromatin modeling factor mutants.
Table S1 (related to Materials and Methods). A table of all yeast strains used in this study.
Table 2 (related to main Figures 5 and 6). A table of all lncRNAs repressed by Swr1, Isw2, Rsc, and Ino80 chromatin remodeling complexes.

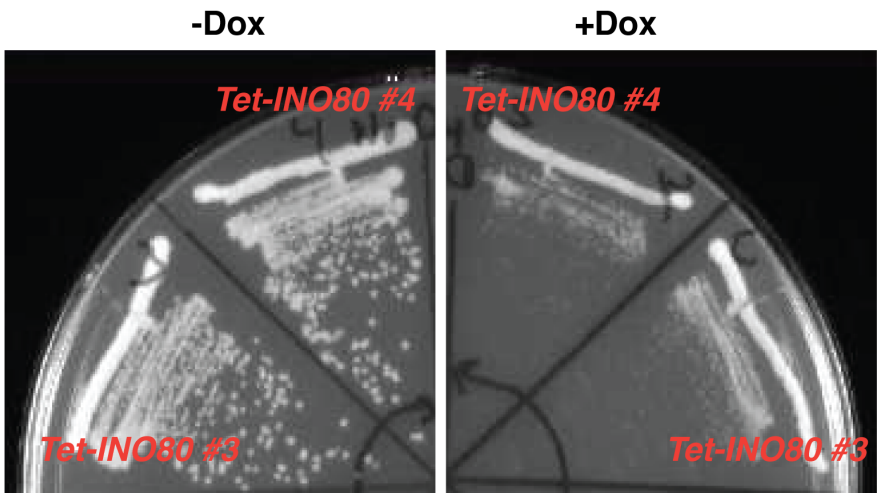


Supplemental Figure 1. Reconstitution of RNAi in *S. cerevisiae*
Northern blot probing Ty1 mRNA in 4 different isolates of RNAi+ *S. cerevisiae* and a lab wild type strain. *ACT1* was probed on the same blot as a loading control. All strains were grown to log phase before RNA was isolated.

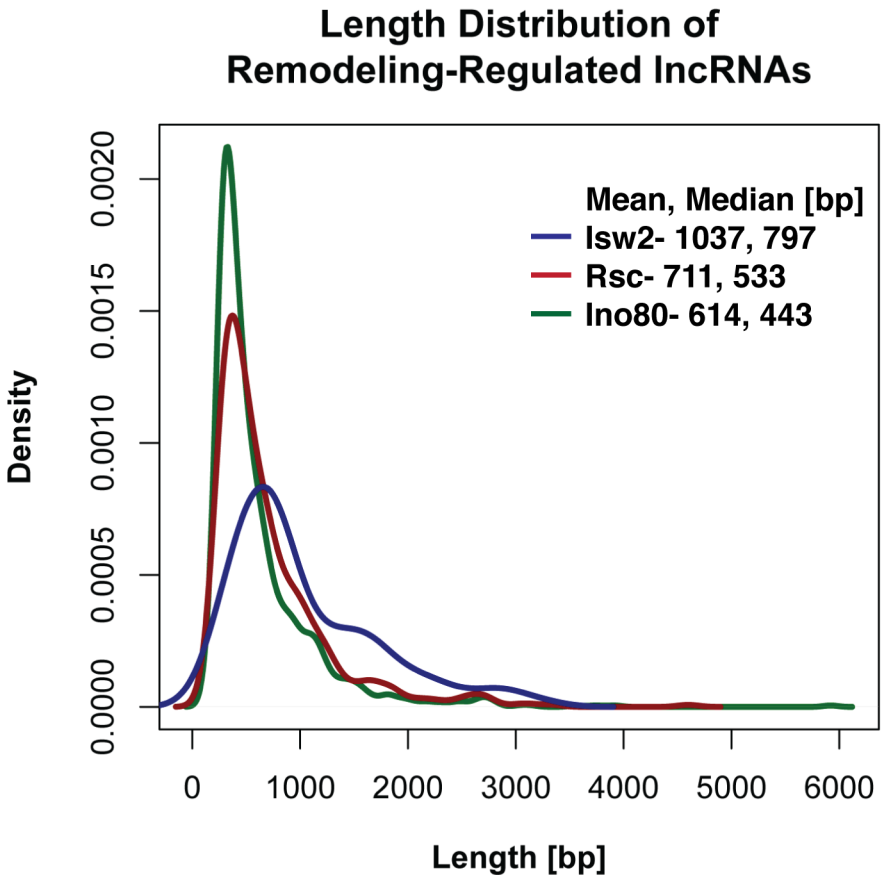
A



B

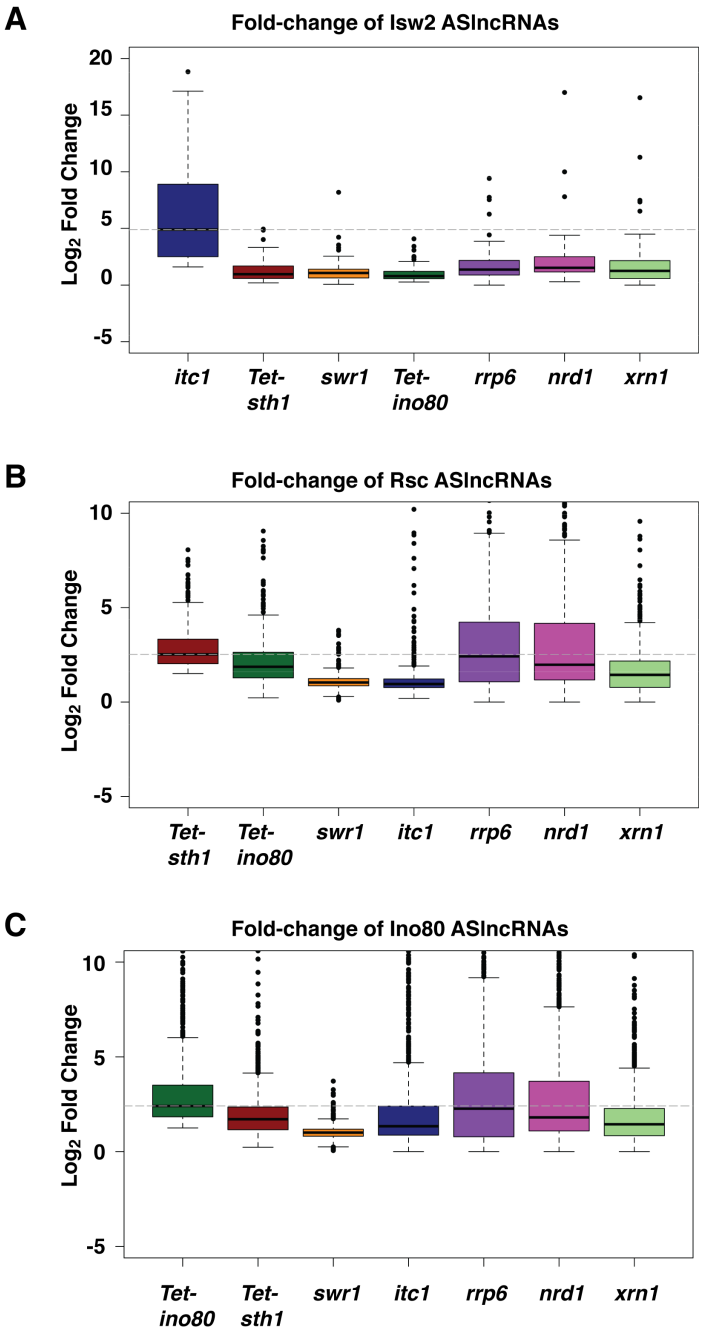


Supplemental Figure 2. Growth of cells carrying tet-repressible alleles of *STH1* and *INO80* (A) Growth on YC(-His) media of two isolates of *tet-STH1* with and without doxycycline. **(B)** As in (A), but for *tet-INO80.*
See Material and Methods for details on growth conditions.

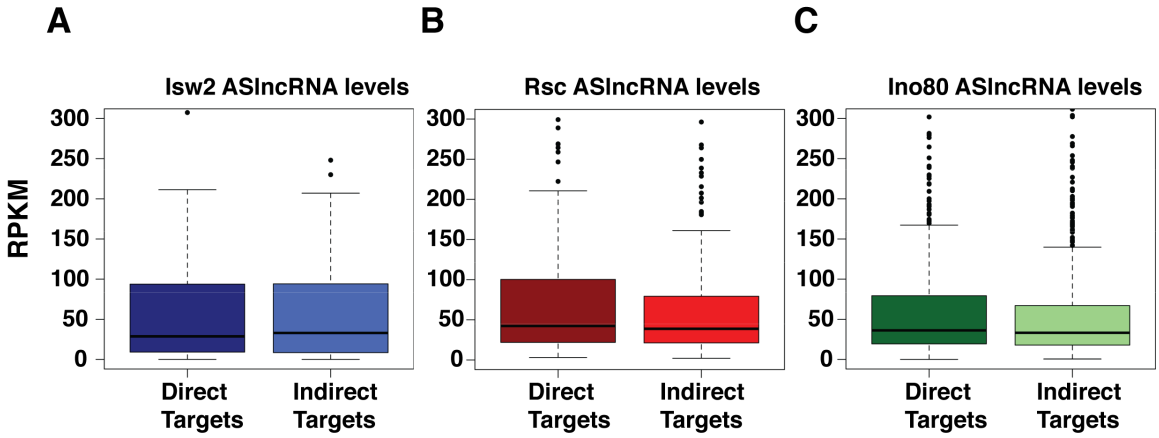


Supplemental Figure 3. Length Distribution of Chromatin-Remodeling Regulated lncRNAs

Kernal density estimates of length distribution of the ASlncRNAs repressed by each chromatin-remodeling factor. Median and Mean of each ASlncRNA class are provided in the legend.

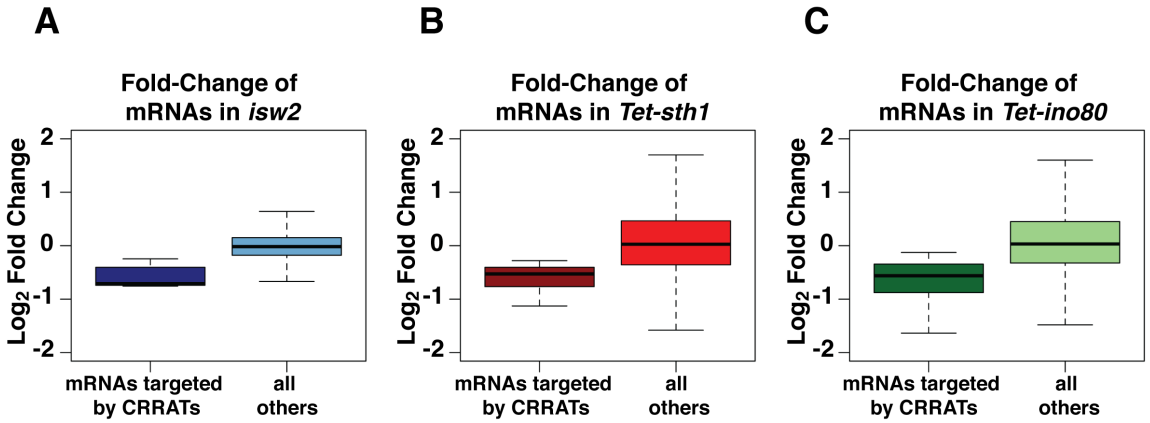


Supplemental Figure 4. Chromatin Remodeling factors Regulate Unique Classes of lncRNAs (A) Box plot showing the distribution of log2 fold-changes of Isw2-regulated lncRNAs in the indicated mutants as determined by strand-specific RNA-seq. The dashed gray line represents the median fold change of Isw2 regulated lncRNAs in the *itc1* background. **(B)** As in (A), but for Rsc regulated lncRNAs. The dashed gray line represents the median fold change of Rsc-regulated lncRNAs when Sth1 in depleted. **(C)** As in (A), but for Ino80-regulated lncRNAs. The dashed gray line represents the median fold change of Ino80-regulated lncRNAs when Ino80 is depleted.



Supplemental Figure 5. Expression Levels of Direct ASIncRNA Targets of Chromatin-Remodelers are Similar to those of Indirect ASIncRNA Targets

(A) Box plot showing the distribution of RPKM values of Isw2-direct lncRNA targets (left) and Isw2-indirect lncRNA targets (right).
(B) As in (A), but for Rsc-direct (left) and Rsc-indirect (right) lncRNA targets.
(C) As in (A), but for Ino80-direct (left) and Ino80-indirect (right) lncRNA targets.



Supplemental Figure 5. Expression Levels of Direct ASlncRNA Targets of Chromatin-Remodelers are Similar to those of Indirect ASlncRNA Targets

(A) Box plot showing the distribution of RPKM values of Isw2-direct lncRNA targets (left) and Isw2-indirect lncRNA targets (right).

(B) As in (A), but for Rsc-direct (left) and Rsc-indirect (right) lncRNA targets.

(C) As in (A), but for Ino80-direct (left) and Ino80-indirect (right) lncRNA targets.

Supplemental Table 1 Yeast strains used in this study

Figures	Genotype	Species	Reference and Strain Name
2A, 3A-B, 4A, 5A 6A	*MATa his3-1 leu2-0 met15-0 ura3-0*	*S. cerevisiae* s288c	Standard strain YTT0558
2B-E	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5+*	*S. cerevisiae* w303	Standard Strain YTT0166
SGA screen query strain Figure 2A	*MATα can1::STE2pr-LEU2 lyp1 leu2-0 met15-0 HIS3::pRS403(pRS406(AGO1-DCR1))*	*S. cerevisiae*	This study YTT5130
1B-C, 2B-E	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 URA3::pRS406(AGO1-DCR1) RAD5+*	*S. cerevisiae*	This study YTT5333
2C	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 URA3::pRS406(AGO1-DCR1) RAD5+ itc1::KanMX*	*S. cerevisiae*	This study YTT5935
2B	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 URA3::pRS406(AGO1-DCR1) RAD5+ swr1::HygMX*	*S. cerevisiae*	This study YTT6235
2D	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 URA3::pRS406(AGO1-DCR1) RAD5+ Tet-O7(HIS3)-STH1-FLAG(HygMX)*	*S. cerevisiae*	This study YTT6236
2E	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 URA3::pRS406(AGO1-DCR1) RAD5+ Tet-O7(HIS3)-INO80-FLAG(HygMX)*	*S. cerevisiae*	This study YTT6237
2C	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 RAD5+ itc1::KanMX*	*S. cerevisiae*	This study YTT0482
2B	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 RAD5+ swr1::HygMX*	*S. cerevisiae*	This study YTT2612
2D	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 RAD5+ Tet-O7(HIS3)-STH1-FLAG(HygMX)*	*S. cerevisiae*	This study YTT6238
2E	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 RAD5+ Tet-O7(HIS3)-INO80-FLAG(HygMX)*	*S. cerevisiae*	This study YTT6239

3, 4, 5, 6	*MATa his3-1 leu2-0 met15-0 ura3-0 itc1::NatMx*	*S. cerevisiae*	This study YTT6240
3	*MATa his3-1 leu2-0 met15-0 ura3-0 swr1::NatMx*	*S. cerevisiae*	This study YTT6241
3, 4, 5, 6	*MATa LEU2-0, MET15-0, URA3-0, Tet-O7(HIS3)-STH1-FLAG(HygMX)*	*S. cerevisiae*	This study YTT5869
3, 4, 5, 6	*MATa LEU2-0, MET15-0, URA3-0, Tet-O7(HIS3)-INO80-FLAG(HygMX)*	*S. cerevisiae*	This study YTT5792
4B-D	*MATa his3-1 leu2-0 met15-0 ura3-0 rrp6::NatMx*	*S. cerevisiae*	This study YTT3568
1B	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5+ rrp6::NatMx*	*S. cerevisiae*	This study YTT6242
1C	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5+ trf4::KanMx*	*S. cerevisiae*	This study YTT3819
1C	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5+ htz1::KanMx*	*S. cerevisiae*	This study YTT3122
3, 4, 5, 6, S3, S4, S5	*MATa leu2-0 met15-0 HIS3::pRS403*	*S. cerevisiae*	This study. Control strain for Tet-STH1 and Tet-INO80 strains.

Appendix B: Supplemental Material for Chapter 3

RNA interference constrains the antisense, long-noncoding RNA transcriptome

Eric A. Alcid and Toshio Tsukiyama

Figure S1 (related to main Figure 1). Sense RNA expression patterns among budding yeast.
Figure S2 (related to main Figures 1). Pairwise comparisons of sense RNA transcriptomes among budding yeast
Figure S3 (related to main Figure 1). Pairwise comparisons of ASlncRNA transcriptomes among budding yeast
Figure S4 (related to main Figure 2). Sense RNA expression meta-analyses at tandem and convergent genes
Figure S5 (related to main Figure 4). S5 SUT-ASlncRNA and CUT-ASlncRNA transcriptomes in *S. cerevisiae* and *S. castellii*
Table S1 (related Figure 3). A table of all genes with conserved ASlncRNAs
Table S2 (related to Materials and Methods). A table of all yeast strains used in this study
Table S3 (related to Materials and Methods) A table of all primers used for qRT-PCR analysis

Alcid_Supplemental Fig 1

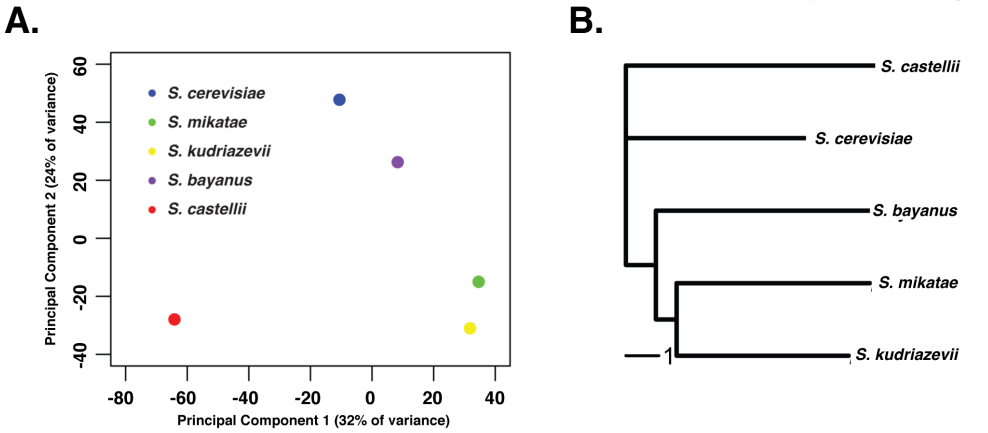


Figure S1- Sense RNA expression patterns among budding yeast (A) Principal Component Analysis (PCA) of sense RNA transcriptomes in budding yeast. (B) Neighbor-joining tree of sense RNA transcriptomes based on pairwise distance matrices (Jensen-Shannon distance metric) for *sensu stricto* budding yeasts and *S. castellii.*

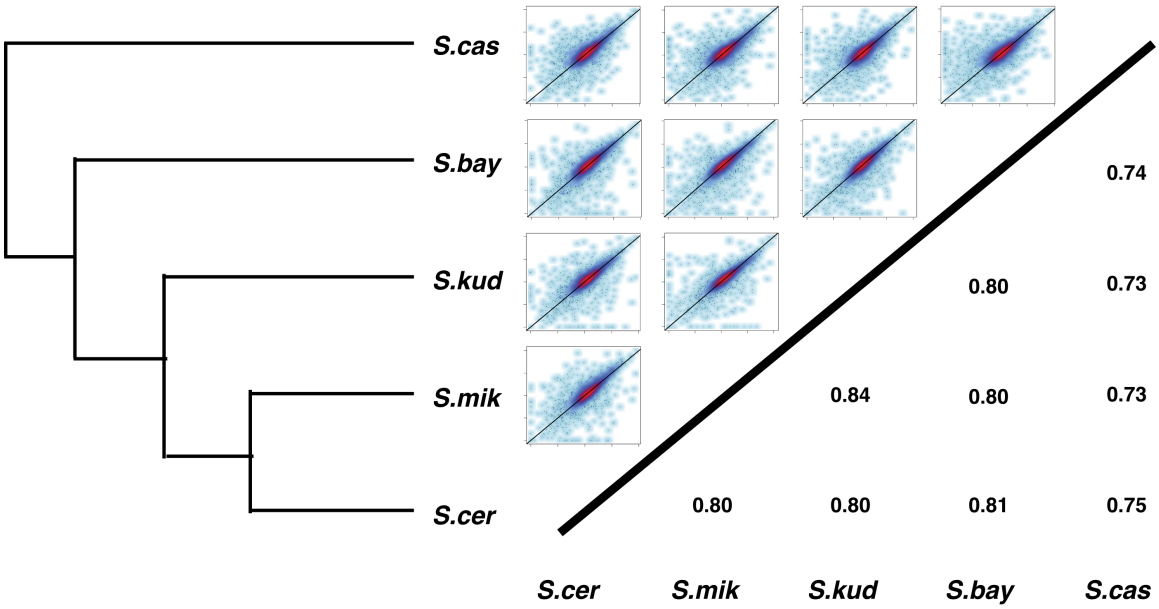


Figure S2 Pairwise comparisons of sense RNA transcriptomes among budding yeast (From left to right) Cladogram representing the phylogenetic relationships among budding yeasts; pairwise scatter-plots of sense RNA transcriptomes comprised of 5031 orthologous ORFs for all studied budding yeast species (top), and corresponding Spearman's rho values (bottom)

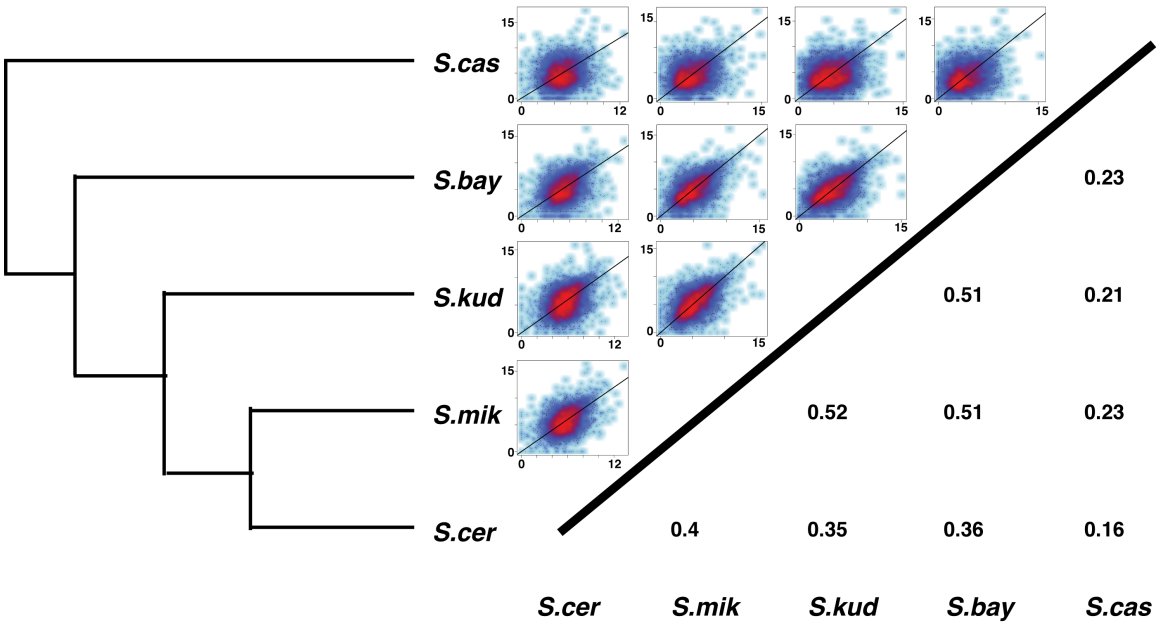


Figure S3 Pairwise comparisons of ASlncRNA transcriptomes among budding yeast (From left to right) Cladogram representing they phylogenetic relationship among budding yeasts; pairwise scatter-plots of antisense RNA transcriptomes comprised of 5031 orthologous ORFs for all studied budding yeast species (top), and corresponding Spearman's rho values (bottom)

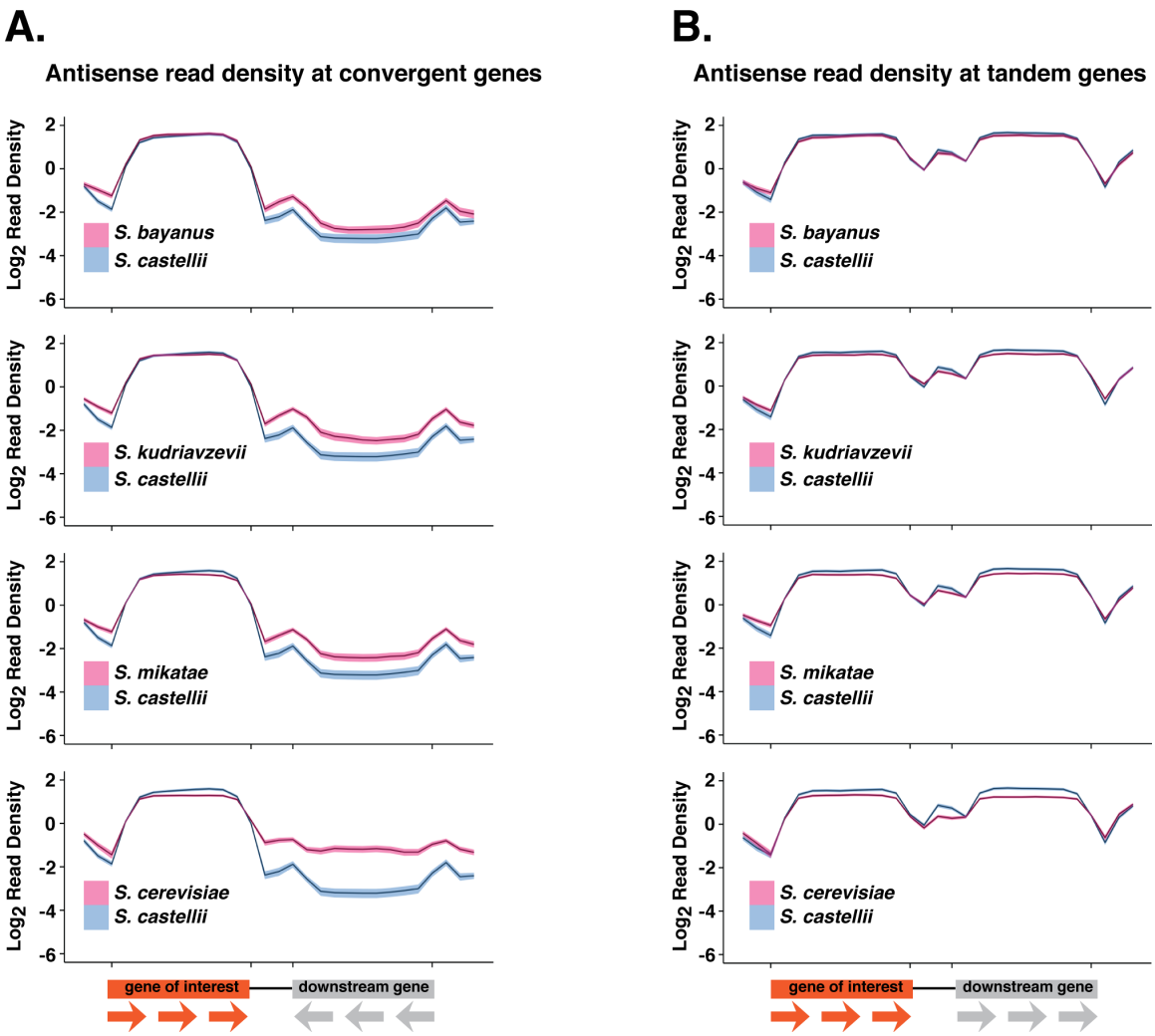


Figure S4 Sense RNA expression meta-analyses at tandem and convergent genes (A) Ribbon Plots of sense read density in log2-scale at genes arranged in convergent orientation for (top to bottom) *S. bayanus, S. kudriavzevii, S. mikatae, S. cerevisiae*. *S. castellii* is represented in all the plots by the blue ribbon. The lines represent the antisense RNA-seq signal, while the outer borders of the ribbon represent 1 standard-error of the mean away from the mean. (B) As in (A), except for genes arranged in tandem orientation.

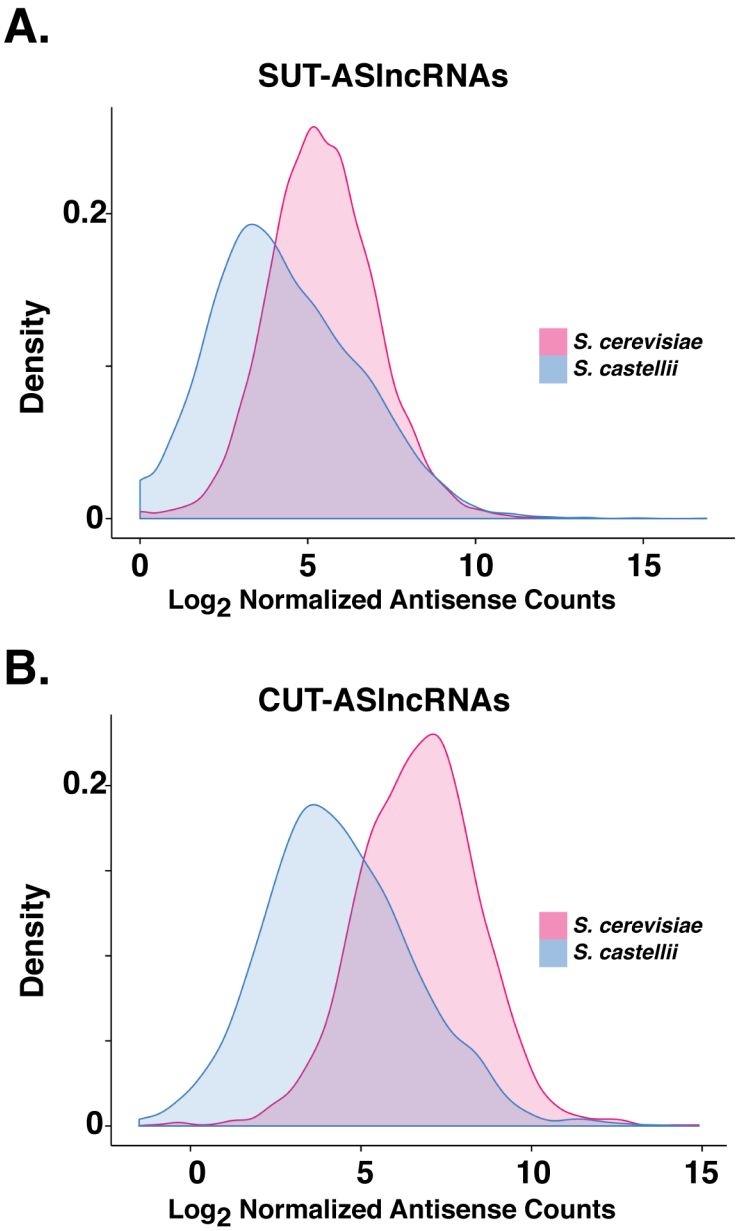


Figure S5 SUT-ASlncRNA and CUT-ASlncRNA transcriptomes in *S. cerevisiae* and *S. castellii* (A) Distribution of ASlncRNA levels in wild type *S. cerevisiae* and *S.castellii* (SUT-ASlncRNAs) (B) As in (A), but for CUT-ASlncRNAs. See Methods for identification of CUT-ASlncRNAs.

Table S1: Genes with conserved ASlncRNA expression

YPL076W	YPR025C	YOL111C	YLR243W	YLR341W	YLR310C
YLR130C	YBL027W	YHR025W	YGR096W	YBR033W	YEL002C
YDL214C	YJR052W	YKL151C	YML028W	YBR254C	YLR138W
YHR163W	YPR109W	YOR350C	YBL056W	YDL173W	YDL159W
YPL196W	YGR273C	YFR050C	YLR253W	YBR143C	YMR234W
YER172C	YNR047W	YKR045C	YAL039C	YHR072W-A	YPL217C
YJL185C	YBR117C	YOR163W	YNR018W	YPL072W	YJR014W
YLR272C	YGL105W	YOL058W	YDR476C	YDL192W	YKL209C
YER164W	YPR190C	YNL295W	YIL106W	YJL134W	YPR189W
YJL149W	YDL076C	YMR146C	YER003C	YMR252C	YGL018C
YLR244C	YDR083W	YHL024W	YDL231C	YHR046C	YOR261C
YKL059C	YPL221W	YOL156W	YGR092W	YPR183W	YKL175W
YGL092W	YLR066W	YNL042W	YGL122C	YJL126W	YGR187C
YLR024C	YDR514C	YPL010W	YLR355C	YCR054C	YFL046W
YLR089C	YOR253W	YPL084W	YPR063C	YLR059C	YER019C-A
YMR259C	YOL012C	YHR191C	YOR120W	YNL117W	YGR197C
YMR148W	YDL036C	YDR301W	YBR263W	YPL152W	YJR104C
YGR236C	YDL122W	YGL010W	YNL281W	YDR363W-A	YLR386W
YLR348C	YGR079W	YGR121C	YJL091C	YGR074W	YOL017W
YBR248C	YKL052C	YGR214W	YGR150C	YKL194C	YLR053C
YNR013C	YBR258C	YGR216C	YGR216C	YKR075C	YJL058C
YLL061W	YPR058W	YJR073C	YJR073C	YJL033W	YHR176W
YBR109C	YPR068C	YGL004C	YGL004C	YNR034W-A	YLR010C
YLR061W	YLR129W	YDL085C-A	YDL085C-A	YMR308C	YDR265W
YBR172C	YGL154C	YML107C	YML107C	YCR008W	YDR033W
YDR531W	YPR062W	YJL055W	YJL055W	YOL104C	YOL145C
YJL172W	YER163C	YEL018W	YEL018W	YAL056W	YFL004W
YDR151C	YMR289W	YHR193C	YHR193C	YNL312W	YDL128W
YJR040W	YER070W	YMR292W	YMR292W	YHR153C	YLR433C
YDL060W	YOR184W	YIL009C-A	YIL009C-A	YJR119C	YFR006W
YOR258W	YER060W	YOL122C	YOL122C	YJR117W	YBL068W
YDR421W	YDR374W-A	YNL024C	YNL024C	YDR169C	YGL011C
YGR186W	YAL015C	YLR073C	YLR073C	YLR098C	YMR033W
YDR099W	YKL045W	YER042W	YER042W	YLR110C	YDL112W
YDL142C	YMR210W	YNL316C	YNL316C	YIR038C	YNL323W

YJL062W	YNL074C	YBR157C	YBR157C	YDL134C	YPL007C
YNL277W	YLR033W	YNL197C	YNL197C	YNL240C	YJL198W
YOR346W	YLR009W	YCL036W	YCL036W	YBL003C	YML068W
YNR039C	YBR161W	YGR086C	YGR086C	YBR222C	YKL159C
YIL115C	YNL169C	YLR078C	YLR078C	YBR106W	YBL039C
YGR037C	YGL110C	YIL116W	YIL116W	YMR194W	YGR027C
YGR095C	YGL091C	YOR190W	YOR190W	YDL165W	YLR213C
YCR075C	YGR241C	YGL151W	YGL151W	YDL171C	YHR192W
YDR226W	YLR312C	YNL133C	YNL133C	YBR114W	YBL010C
YLR087C	YPL190C	YHR028C	YHR028C	YGR193C	YKR062W
YNL079C	YKL085W	YBR265W	YBR265W	YPL273W	YMR201C
YDR234W	YER182W	YML070W	YML070W	YER065C	YDR051C
YNL245C	YLR072W	YDR232W	YDR232W	YEL011W	YGL219C
YOR349W	YPL201C	YBR287W	YBR287W	YLR077W	YGR105W
YCL038C	YPL183C	YHR194W	YHR194W	YAL044W-A	YLR449W
YJR061W	YDR166C	YGR124W	YGR124W	YKL012W	YPL096C-A
YML079W	YDL092W	YBR017C	YBR017C	YDL164C	YOR131C
YBL004W	YOL089C	YOR295W	YOR295W	YNL125C	YBR071W
YLL029W	YPR118W	YMR260C	YMR260C	YFL022C	YNR038W
YIL138C	YGL251C	YNL244C	YNL244C	YPL031C	YNR022C
YDR464W	YJR016C	YAL032C	YAL032C	YNL115C	YDL054C
YER007W	YDR125C	YOR161C	YOR161C	YKL180W	YBL089W
YJL146W	YGL197W	YDR135C	YDR135C	YOR276W	YML123C
YBR244W	YOR095C	YHR137W	YHR137W	YPL036W	YMR087W
YMR246W	YGL108C	YBL049W	YBL049W	YNL305C	YOR155C
YEL003W	YCL014W	YDL208W	YDL208W	YJL004C	YMR253C
YOL135C	YNL280C	YML005W	YML005W	YDL093W	YDL072C
YLR027C	YLR090W	YML011C	YML011C	YLR298C	YMR208W
YLR351C	YIL073C	YER112W	YER112W	YER098W	YDL075W
YPR036W	YLR446W	YPL046C	YPL046C	YNR054C	YBR142W
YNL317W	YGL220W	YOR219C	YOR219C	YJR017C	YPR182W
YOR148C	YNL183C	YLR181C	YLR181C	YER147C	YNL200C
YLR058C	YBR233W	YKL206C	YKL206C	YJL016W	YNR004W
YKR081C	YAL027W	YPL246C	YPL246C	YCR053W	YKL187C
YGR229C	YMR029C	YNR011C	YNR011C	YCL058W-A	YGR075C
YKR035W-A	YNL147W	YBR108W	YBR108W	YBR291C	YNR012W
YLR001C	YJR076C	YLR063W	YLR063W	YAR015W	YKL207W
YDR142C	YMR163C	YGL111W	YGL111W	YHL004W	YLR405W
YJR053W	YMR313C	YMR060C	YMR060C	YHR201C	YCL028W

YFL044C	YBR283C	YHR139C	YHR139C	YKL171W	YHR129C
YDR025W	YKR002W	YMR319C	YMR319C	YPR187W	YDL069C
YGR195W	YGL137W	YGL156W	YGL156W	YLR007W	YHR045W
YOR222W	YBR003W	YLR380W	YLR380W	YNL082W	YLR299W
YOL061W	YML104C	YPR188C	YPR188C	YNL299W	YIL013C
YLR015W	YNL256W	YOR260W	YOR260W	YPL154C	YGL162W
YPL259C	YBR257W	YPL260W	YPL260W	YHR019C	YFL042C
YOL011W	YCR037C	YIL084C	YIL084C	YNL327W	YLR376C
YKR043C	YKR089C	YMR218C	YMR218C	YGR008C	YOL146W
YKL051W	YMR217W	YER093C	YER093C	YNL311C	YDR532C
YPR066W	YBR165W	YGR196C	YGR196C	YOR064C	YHR118C
YDR085C	YER024W	YLR064W	YLR064W	YML124C	YKR061W
YKR054C	YIL021W	YIR011C	YIR011C	YJL030W	YPL105C
YKL164C	YMR110C	YHR172W	YHR172W	YOR126C	YDR305C
YMR031C	YDR494W	YFL017C	YFL017C	YMR042W	YDR205W
YKL183W	YHR002W	YDL201W	YDL201W	YML066C	YER154W
YER030W	YKR051W	YIL134W	YIL134W	YIL007C	YDL168W
YIL095W	YJR110W	YDR332W	YDR332W	YIL037C	YDL179W
YMR138W	YNR009W	YLR263W	YLR263W	YER043C	YEL036C
YOL068C	YOR094W	YDR236C	YDR236C	YNL237W	YJL029C
YJR068W	YLR258W	YJL042W	YJL042W	YLR099W-A	YGL170C
YLR456W	YGL142C	YER037W	YER037W	YER057C	YFR043C
YGR156W	YNR015W	YOR061W	YOR061W	YNL243W	YAL029C
YBR163W	YBR127C	YOR356W	YOR356W	YGR157W	YIL008W
YOR118W	YFL024C	YOR311C	YOR311C	YPR008W	YCL005W-A
YER155C	YNR072W	YOL081W	YOL081W	YNL253W	YPL169C
YDR397C	YBR139W	YOR138C	YOR138C	YAL013W	YPL013C
YHL038C	YKL022C	YGL148W	YGL148W	YPL204W	YGR080W
YKL082C	YMR160W	YMR034C	YMR034C	YDR489W	YBL088C
YPR101W	YOR217W	YCL035C	YCL035C	YCR059C	YPL003W
YIL105C	YMR299C	YNL160W	YNL160W	YGR149W	YER019W
YER004W	YGL178W	YCR066W	YCR066W	YER130C	YOR027W
YDL232W	YFL023W	YJR112W-A	YJR112W-A	YGR130C	YER060W-A
YMR228W	YIR001C	YKL114C	YKL114C	YAL016W	YGR023W
YCL056C	YLR360W	YLR117C	YLR117C	YDR101C	YDR277C
YNR019W	YDR092W	YHR132C	YHR132C	YPR054W	YPR067W
YGR158C	YBR171W	YGL073W	YGL073W	YKR071C	YDL124W
YGL121C	YOR241W	YJR006W	YJR006W	YGL112C	YPR107C
YBR253W	YLR316C	YOR176W	YOR176W	YLR353W	
YPR172W	YLR425W	YDR158W	YDR158W	YIL047C	

YLR034C	YGL012W	YJL019W	YJL019W	YHR092C
YHR098C	YGL255W	YMR116C	YMR116C	YPR049C
YLR354C	YHR040W	YDL074C	YDL074C	YJL083W
YDR225W	YML108W	YOR164C	YOR164C	YPR026W
YNL134C	YML057W	YER029C	YER029C	YOR019W
YCR093W	YAR018C	YJL190C	YJL190C	YGR101W

Table S2: Yeast strains used in this study

Genotype	Species	Reference and Strain Name
MATa his3-1 leu2-0 met15-0 ura3-0	*S. cerevisiae* s288c	Standard strain YTT0558
MATα leu2-0 met15-0 HIS3::pRS403(pRS406(AGO1-DCR1))	*S. cerevisiae*	YTT5130
MATa his3-1 leu2-0 met15-0 ura3-0 rrp6::NatMx	*S. cerevisiae*	This study YTT3568
MATα ura3-1 hoΔ	*S. castellii*	DPB005
MATα ura3-1 hoΔ dcr1Δ::loxP-KanMX-loxP	*S. castellii*	DPB009
MATα ura3-1 hoΔ rrp6::HygMx	*S. castellii*	YTT
MATα ura3-1 hoΔ dcr1Δ::loxP-KanMX-loxP rrp6:: HygMx	*S. castellii*	
MATα/a prototrophic diploid	*S. bayanus* (CBS 7001)	From Maitreya Dunham
MATα/a prototrophic diploid	*S. kudriavzevii* (IFO 1802)	From Maitreya Dunham
MATα/a prototrophic diploid	*S. mikatae* (IFO 1815)	From Maitreya Dunham
MATα/a prototrophic diploid	*S. paradoxus*	From D. Gottschling
MATa ho::HYGb lys2-1 ura3::clonat	*S. bayanus Y2B5-113*	From Maitreya Dunham
MATa	*S. kudriavzevii FM527*	From Maitreya Dunham
*MAT**a** hoD::KanMX ura3D::HygMX*	*S. mikatae (FM356)*	From Maitreya Dunham

Table S3- Primers for RT-qPCR used in this study

Primer name	Sequence	Notes
Universal primer	CGGTCATGGTGGCGAATAA	Add with "Gene Specific Primers" (GSP) to amplify desired product from cDNA
Act1RT_cer	cggtcatggtggcgaataaATACCTGGGAACATGGTGG	Used for S. cerevisiae, S. paradoxus, S. mikatae, S. bayanus, S. castellii
Act1GSP_cer	CCAGAAGCTTTGTTCCATCC	Used for S. cerevisiae, S. paradoxus, S. mikatae, S. kudriavzevii, S. bayanus
Act1RT_kud	cggtcatggtggcgaataaGCGATACCTGGGAACATAGTG	
Act1GSP_cas	CTTTGTTCCACCCATCCATC	
IPP1RT_cer	cggtcatggtggcgaataaCCTCAACAGACCTGGGAAGT	
IPP1GSP_cer	GAAGGTGAGACCGATTGGAA	Used for S. cerevisiae, S. paradoxus
IPP1RT_par	cggtcatggtggcgaataaTCTGAACCATTCGTTGGTAGC	Used for S. paradoxus, S. mikatae
IPP1RT_kud	cggtcatggtggcgaataaAAACCATTCGTTGGTAGCTCTC	
IPP1GSP_kud	CTGGAAGGTTCTTGCCATTG	
IPP1RT_bay	cggtcatggtggcgaataaTTCTGAACCATTCGTTGGTG	
IPP1GSP_bay	TGAAGGTGAAACTGACTGGAAA	
IPP1RT_cas	cggtcatggtggcgaataaAACCATTCATTGGTGGCTCT	
IPP1GSP_cas	GAAGGTGAAACCGATTGGAA	
sPHO84RT_cer	cggtcatggtggcgaataaCTCCACATTTGGTCACAAGC	Used for S. cerevisiae, S. paradoxus, S.bayanus
sPHO84GSP_cer	AAATGGAGAGGTGCCATCAT	Used for S. cerevisiae, S. bayanus, S. castellii
sPHO84GSP_par	GCCATTATGGGTGCTGTCTT	
sPHO84RT_mik	cggtcatggtggcgaataaACAGGCCTTTTGACATCTGG	

sPHO84GSP_mik	GAGGTGCTATTATGGGTGCTG	
sPHO84RT_kud	cggtcatggtggcgaataaTCTCCACATTTGGTCACAGG	
sPHO84GSP_kud	GCCATTATGGGTGCTGTTTT	
sPHO84RT_cas	cggtcatggtggcgaataaCCACATTTGATCACAAGCTTTC	
asPHO84RT_cer	cggtcatggtggcgaataaAAATGGAGAGGTGCCATCAT	Used for S. cerevisiae, S. bayanus
asPHO84GSP_cer	CTCCACATTTGGTCACAAGC	Used for S. cerevisiae, S. paradoxus, S.bayanus
asPHO84RT_par	cggtcatggtggcgaataaGCCATTATGGGTGCTGTCTT	
asPHO84RT_mik	cggtcatggtggcgaataaGAGGTGCTATTATGGGTGCTG	
asPHO84GSP_mik	ACAGGCCTTTTGACATCTGG	
asPHO84RT_kud	cggtcatggtggcgaataaGCCATTATGGGTGCTGTTTT	
asPHOGSP_kud	TCTCCACATTTGGTCACAGG	
asPHO84RT_cas	cggtcatggtggcgaataaAAATGGAGAGGTGCCATCAT	
asPHO84GSP_cas	CCACATTTGATCACAAGCTTTC	
sGAL10RT_cer	cggtcatggtggcgaataaCGCAAGATAGCAAACTTCCAA	
sGAL10GSP_cer	TCCCAGAAGAATGTCCCTTAG	
sGAL10RT_par	cggtcatggtggcgaataaAATAGCGCAAGATGGCAAAC	
sGAL10GSP_par	TCCAGAAGAATGTCCCTTGG	
sGAL10RT_mik	cggtcatggtggcgaataaAACGCAAAATGGCAAACTTC	
sGAL10GSP_mik	TGTCCATTGGGACCTACTAACC	
sGAL10RT_bay	cggtcatggtggcgaataaTCGGATTGAAATAACGCAAA	
sGAL10GSP_bay	TAGGGCCCACCAATCCATA	
sGAL10RT_cas	cggtcatggtggcgaataaACGTAAGATGGCGAATTTCC	

sGAL10GSP_cas	TCCAGAAGAATGTCCCTTGG	
asGAL10RT_cer	cggtcatggtggcgaataaTCCCAGAAGAATGTCCCTTAG	
asGAL10GSP_cer	CGCAAGATAGCAAACTTCCAA	
asGAL10RT_par	cggtcatggtggcgaataaTCCAGAAGAATGTCCCTTGG	Used for S. paradoxus, S. castellii
asGAL10GSP_par	AATAGCGCAAGATGGCAAAC	
asGAL10RT_mik	cggtcatggtggcgaataaTGTCCATTGGGACCTACTAACC	
asGAL10GSP_mik	AACGCAAAATGGCAAACTTC	
asGAL10RT_bay	cggtcatggtggcgaataaTAGGGCCCACCAATCCATA	
asGAL10GSP_bay	TCGGATTGAAATAACGAAA	
asGAL10RT_cas	cggtcatggtggcgaataaTCCAGAAGAATGTCCCTTGG	
asGAL10GSP_cas	ACGTAAGATGGCGAATTTCC	

References:

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF. 2007. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature* **446**: 572-576.

Alcid EA, Tsukiyama T. 2014. ATP-dependent chromatin remodeling shapes the long noncoding RNA landscape. *Genes Dev* **28**: 2348-2360.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Anders S, Pyl PT, Huber W. 2014. HTSeq – A Python framework to work with high-throughput sequencing data. *bioRxiv*.

Bartholomew B. 2014. Regulating the Chromatin Landscape: Structural and Mechanistic Perspectives. *Annual Review of Biochemistry* **83**: 671-696.

Bassett AR, Akhtar A, Barlow DP, Bird AP, Brockdorff N, Duboule D, Ephrussi A, Ferguson-Smith AC, Gingeras TR, Haerty W et al. 2014. Considerations when investigating lncRNA function in vivo. *eLife* **3**.

Batista PJ, Chang HY. 2013. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**: 1298-1307.

Berretta J, Morillon A. 2009. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO reports* **10**: 973-982.

Bortvin A, Winston F. 1996. Evidence That Spt6p Controls Chromatin Structure by a Direct Interaction with Histones. *Science* **272**: 1473-1476.

Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* **478**: 343-348.

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.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP et al. 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.

Chatterjee SN, Devhare PB, Lole KS. 2012. Detection of negative-sense RNA in packaged hepatitis E virions by use of an improved strand-specific reverse transcription-PCR method. *J Clin Microbiol* **50**: 1467-1470.

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR et al. 2012. Saccharomyces

Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**: D700-705.
Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, Hughes TR, Winston F. 2008. Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. *PLoS Biol* **6**: e277.
Churchman LS, Weissman JS. 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**: 368-373.
Core LJ, Waterfall J, Lis J. 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* **322**: 1845-1848.
Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S et al. 2015. Ensembl 2015. *Nucleic Acids Res* **43**: D662-669.
D.R. S, Zill A.Z., Rokas A., Payen C., Dunham M.J., Eisen M.B., Rine J., Johnston M, C.T. H. 2011. The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3 (Genes, Genomics, Genetics)* **1**: 11-25.
Davis CA, Ares M, Jr. 2006. Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **103**: 3262-3267.
Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F et al. 2012. Landscape of transcription in human cells. *Nature* **489**: 101-108.
Doolittle WF. 2013. Is junk DNA bunk? A critique of ENCODE. *Proc Natl Acad Sci U S A* **110**: 5294-5300.
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.
Dujon B. 2006. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends in genetics : TIG* **22**: 375-387.
Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**: 48.
Fazzio TG, Tsukiyama T. 2003. Chromatin Remodeling In Vivo: Evidence for a Nucleosome Sliding Mechanism. *Mol Cell* **12**: 1333-1340.
Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S et al. 2014. Ensembl 2014. *Nucleic Acids Res* **42**: D749-755.
Floer M, Wang X, Prabhu V, Berrozpe G, Narayan S, Spagna D, Alvarez D, Kendall J, Krasnitz A, Stepansky A et al. 2010. A RSC/nucleosome

complex determines chromatin architecture and facilitates activator binding. *Cell* **141**: 407-418.
Flynn RA, Chang HY. 2014. Long Noncoding RNAs in Cell-Fate Programming and Reprogramming. *Cell Stem Cell* **14**: 752-761.
Geisler S, Lojek L, Khalil AM, Baker KE, Coller J. 2012. Decapping of long noncoding RNAs regulates inducible genes. *Mol Cell* **45**: 279-291.
Goldmark JP, Fazzio TG, Estep PW, Church GM, Tsukiyama T. 2000. The Isw2 Chromatin Remodeling Complex Represses Early Meiotic Genes upon Recruitment by Ume6p. *Cell* **103**: 423-433.
Goldstein G, McCusker J. 1999. Three New Dominant Drug Resistance Cassettes for Gene Disruption in Saccharomyces cerevisiae. *Yeast* **15**: 1541-1553.
Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW et al. 2011. The developmental transcriptome of Drosophila melanogaster. *Nature* **471**: 473-479.
Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL et al. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**: 1071-1076.
Hainer SJ, Gu W, Carone BR, Landry BD, Rando OJ, Mello CC, Fazzio TG. 2015. Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF. *Genes Dev* **29**: 362-378.
Hartley PD, Madhani HD. 2009. Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**: 445-458.
Hartzog GA, Wada T, Handa H, Winston F. 1998. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II inSaccharomyces cerevisiae. *Genes Dev* **12**: 357-369.
Haugen AC, Kelley R, Collins JB, Tucker CJ, Deng C, Afshari CA, Brown JM, Ideker T, Van Houten B. 2004. Integrating phenotypic and expression profiles to map arsenic-response networks. *Genome Biol* **5**: R95-R95.
Ho L, Jothi R, Ronan JL, Cui K, Zhao K, Crabtree GR. 2009a. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proceedings of the National Academy of Sciences* **106**: 5187-5191.
Ho L, Ronan JL, Wu J, Staahl BT, Chen L, Kuo A, Lessard J, Nesvizhskii AI, Ranish J, Crabtree GR. 2009b. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proceedings of the National Academy of Sciences* **106**: 5181-5186.
Hongay CF, Grisafi PL, Galitski T, Fink GR. 2006. Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell* **127**: 735-745.
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.

K.P. B, 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.

Kellis M, et. al. 2014. Defining functional DNA elements in the

human genome. *Proc Natl Acad Sci U S A* **111**: 6131-6138.

Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. *Genome Res* **21**: 487-493.

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.

Kim J-A, Haber JE. 2009. Chromatin assembly factors Asf1 and CAF-1 have overlapping roles in deactivating the DNA damage checkpoint when DNA repair is complete. *Proceedings of the National Academy of Sciences* **106**: 1151-1156.

Klopf E, Paskova L, Solé C, Mas G, Petryshyn A, Posas F, Wintersberger U, Ammerer G, Schüller C. 2009. Cooperation between the INO80 Complex and Histone Chaperones Determines Adaptation of Stress Gene Transcription in the Yeast Saccharomyces cerevisiae. *Mol Cell Biol* **29**: 4994-5007.

Kobor MS, Venkatasubrahmanyam S, Meneghini MD, Gin JW, Jennings JL, Link AJ, Madhani HD, Rine J. 2004. A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biol* **2**: E131.

Krawchuck M, Wahls W. 1999. High-efficiency Gene Targeting in Schizosaccharomyces pombe Using a Modular, PCR-based Approach with Long Tracts of Flanking Homology. *Yeast*: 1419-1427.

Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, Odom DT, Marques AC. 2012. Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet* **8**: e1002841.

Lacadie SA, Tardiff DF, Kadener S, Rosbash M. 2006. In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. *Genes Dev* **20**: 2055-2066.

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. *Proceedings of the National Academy of Sciences* **108**: 1058-1063.

Lee JT, Bartolomei MS. 2013. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**: 1308-1323.

LeGendre JB, Campbell ZT, Kroll-Conner P, Anderson P, Kimble J, Wickens M. 2013. RNA Targets and Specificity of Staufen, a Double-stranded RNA-binding Protein in Caenorhabditis elegans. *Journal of Biological Chemistry* **288**: 2532-2545.

Li B, Pattenden SG, Lee D, Gutiérrez J, Chen J, Seidel C, Gerton J, Workman JL. 2005. Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc Natl Acad Sci U S A* **102**: 18385-18390.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Lickwar CR, Rao B, Shabalin AA, Nobel AB, Strahl BD, Lieb JD. 2009. The Set2/Rpd3S pathway suppresses cryptic transcription without regard to gene length or transcription frequency. *PLoS One* **4**: e4886.

Louro R, Smirnova AS, Verjovski-Almeida S. 2009. Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* **93**: 291-298.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8[thinsp]A resolution. *Nature* **389**: 251-260.

Marquardt S, Escalante-Chong R, Pho N, Wang J, Churchman LS, Springer M, Buratowski S. 2014. A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* **157**: 1712-1723.

Merkin J, Russell, C., Chen, P., Burge, C. 2012. Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. *Science* **338**: 1593-1599.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**: 1344-1349.

Nakanishi K, Weinberg DE, Bartel DP, Patel DJ. 2012. Structure of yeast Argonaute with guide RNA. *Nature* **486**: 368-374.

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**: 635-640.

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.

Nicholas T. Ingolia SG, John R. S. Newman, Jonathan S. Weissman. 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324**: 218-223.

Orban TI, Izaurralde E. 2005. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**: 459-469.

Papamichos-Chronakis M, Watanabe S, Rando OJ, Peterson CL. 2011. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**: 200-213.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**: 289-290.

Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* **37**: e123.

Pelechano V, Steinmetz LM. 2013. Gene regulation by antisense transcription. *Nature reviews Genetics* **14**: 880-893.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rhee HS, Pugh BF. 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**: 295-301.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E et al. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**: 1311-1323.

Roy M Long MTM. 2003. mRNA Decay: X (XRN1) Marks the Spot. *Mol Cell* **11**: 1126-1128.

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.

Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang J-PZ, Widom J. 2006. A genomic code for nucleosome positioning. *Nature* **442**: 772-778.

Sultan M, Dokel S, Amstislavskiy V, Wuttig D, Sultmann H, Lehrach H, Yaspo ML. 2012. A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. *Biochem Biophys Res Commun* **422**: 643-646.

Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM, Proudfoot NJ. 2012. Gene Loops Enhance Transcriptional Directionality. *Science* **338**: 671-675.

Teif VB, Vainshtein Y, Caudron-Herger M, Mallm JP, Marth C, Hofer T, Rippe K. 2012. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol* **19**: 1185-1192.

Teves SS, Weber CM, Henikoff S. 2014. Transcribing through the nucleosome. *Trends in Biochemical Sciences* **39**: 577-586.

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**: 2364-2368.

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**: 689-693.

Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. 2010. The Role of Nucleosome Positioning in the Evolution of Gene Regulation. *PLoS Biol* **8**: e1000414.

Ullu E, Tschudi C, Chakraborty T. 2004. RNA interference in protozoan parasites. *Cellular Microbiology* **6**: 509-519.

Van de Vosse DW, Wan Y, Lapetina DL, Chen WM, Chiang JH, Aitchison JD, Wozniak RW. 2013. A role for the nucleoporin Nup170p in chromatin structure and gene silencing. *Cell* **152**: 969-983.

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

Vanhe´e-Brossollet C, Vaquero, C. 1998. Do natural antisense transcripts make sense in eukaryotes? *Gene* **211**: 1-9.

Venter JC Adams MD Myers EW Li PW Mural RJ Sutton GG Smith HO Yandell M Evans CA Holt RA et al. 2001. The sequence of the human genome. *Science* **291**: 1304-1351.

Venters BJ, Pugh BF. 2013. Genomic organization of human transcription initiation complexes. *Nature* **502**: 53-58.

Villar D, Berthelot C, Aldridge S, Rayner Tim F, Lukk M, Pignatelli M, Park Thomas J, Deaville R, Erichsen Jonathan T, Jasinska Anna J et al. 2015. Enhancer Evolution across 20 Mammalian Species. *Cell* **160**: 554-566.

Weber CM, Henikoff S. 2014. Histone variants: dynamic punctuation in transcription. *Genes Dev* **28**: 672-682.

Weinberg DE, Nakanishi K, Patel DJ, Bartel DP. 2011. The inside-out mechanism of Dicers from budding yeasts. *Cell* **146**: 262-276.

Whitehouse I, Rando OJ, Delrow J, Tsukiyama T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**: 1031-1035.

Whitehouse I, Tsukiyama T. 2006. Antagonistic forces that position nucleosomes in vivo. *Nat Struct Mol Biol* **13**: 633-640.

Wickham H. 2009. ggplot2: elegant graphics for data analysis. *Springer New York*.

Wyers F, Rougemaille M, Badis G, Rousselle J-C, Dufour M-E, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B. 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.

Yadon AN, Van de Mark D, Basom R, Delrow J, Whitehouse I, Tsukiyama T. 2010. Chromatin remodeling around nucleosome-free regions leads to repression of noncoding RNA transcription. *Mol Cell Biol* **30**: 5110-5122.

Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnirke A, Nusbaum C, Thompson DA, 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.

Yen K, Vinayachandran V, Batta K, Koerber RT, Pugh BF. 2012. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* **149**: 1461-1473.

Yen K, Vinayachandran V, Pugh BF. 2013. SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. *Cell* **154**: 1246-1256.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**: R14.

Yukawa M, Katoh S, Miyakawa T, Tsuchiya E. 1999. Nps1/Sth1p, a component of an essential chromatin-remodeling complex of Saccharomyces cerevisiae, is required for the maximal expression of early meiotic genes. *Genes to Cells* **4**: 99-110.

Zabaronick SR, Tyler JK. 2005. The Histone Chaperone Anti-Silencing Function 1 Is a Global Regulator of Transcription Independent of Passage through S Phase. *Mol Cell Biol* **25**: 652-660.

Zofall M, Fischer T, Zhang K, Zhou M, Cui B, Veenstra TD, Grewal SI. 2009. Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. *Nature* **461**: 419-422.