THE UNIVERSITY
of EDINBURGH

Functional long non-coding RNA transcription in *Schizosaccharomyces pombe*

Ryan Ard

Thesis presented for the degree of

Doctor of Philosophy

The University of Edinburgh

2016

Declaration of originality

This thesis is the result of my own work; the research presented herein is my own unless otherwise indicated. Note, some of the findings presented in this thesis have previously been published in the following research paper:

Ard R, Tong P, and Allshire RC (2014) Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. *Nature Communications*, 5:5576. doi: 10.1038/ncomms6576. URL: http://www.nature.com/ncomms/2014/141127/ncomms6576/full/ncomms6576.html

Ryan Ard

2016

Lay summary

Only a small fraction of the human genome contains genes that code for protein. Instead, the vast majority of the human genome is made up of DNA that does not code for protein at all. In fact, such regions of “non-coding DNA” make up the bulk of essentially all plant and animal genomes studied to date. Despite the abundance of non-coding DNA in different genomes, it is still unclear what proportion serves genuine biological functions or might simply be inconsequential “junk DNA”. Recent studies have shown that non-coding regions are frequently transcribed into long non-coding RNA molecules and that such acts of non-coding transcription are sensitive to the cellular environment. Moreover, non-coding transcription appears to aid environmental responses to stress at the molecular level in cells. While attention has mostly been paid to potential functions for non-coding RNA products, a growing body of evidence suggests the mere process of transcribing non-coding regions of the genome can itself be a regulatory event. Here I present evidence for acts of non-coding transcription that regulate two distinct genes involved in controlling nutrient levels in the unicellular fission yeast *Schizosaccharomyces pombe*. These results underscore the importance of the act of non-coding transcription in controlling gene activity and provide important clues to better understand the role(s) of non-coding DNA in plants and animals.

Abstract

Eukaryotic genomes are pervasively transcribed and frequently generate long non-coding RNAs (lncRNAs). However, most lncRNAs remain uncharacterized. In this work, a set of positionally conserved intergenic lncRNAs in the fission yeast *Schizosaccharomyces pombe* genome are selected for further analysis. Deleting one of these lncRNA genes (*ncRNA.1343*) exhibited a clear phenotype: increased drug sensitivity. Further analyses revealed that deleting *ncRNA.1343* also disrupted a previously unannotated lncRNA, termed *nc-tgp1*, transcribed in the opposite orientation of the predicted *ncRNA.1343* gene and into the promoter of the phosphate-responsive permease gene *tgp1*$^{+}$. Detailed analyses revealed that the act of transcribing *nc-tgp1* into the *tgp1*$^{+}$ promoter increases nucleosome density and prevents transcription factor access. Decreased *nc-tgp1* transcription permits *tgp1*$^{+}$ expression upon phosphate starvation, while *nc-tgp1* loss induces *tgp1*$^{+}$ in repressive phosphate-rich conditions. Notably, drug sensitivity results directly from *tgp1*$^{+}$ expression in the absence of *nc-tgp1* transcription. Similarly, lncRNA transcription upstream of *pho1*$^{+}$, another phosphate-regulated gene, increases nucleosome density and prevents transcription factor binding to repress *pho1*$^{+}$ in phosphate-replete cells. Importantly, the regulation of *tgp1*$^{+}$ and *pho1*$^{+}$ by upstream lncRNA transcription occurs in the absence of RNAi and heterochromatin components. Instead, the regulation of *tgp1*$^{+}$ and *pho1*$^{+}$ by upstream lncRNA transcription resembles examples of transcriptional interference reported in other organisms. Thus, *tgp1*$^{+}$ and *pho1*$^{+}$ are the first documented examples of genes regulated by transcriptional interference in *S. pombe*.

Table of contents

List of figures

List of tables

List of abbreviations

ade	Adenine
arg	Arginine
bp	Base pair
CAF	Caffeine
cDNA	Complementary deoxyribonucleic acid
CloNAT	Nourseothricin
Cp	Crossing point
CTD	Carboxy-terminal domain
C-terminal	Carboxy-terminal
cs	Cold-sensitive
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleic triphosphate
EDTA	Ethylene di-amine tetra acetic acid
5-FOA	5-fluoro-orotic acid
G418	Geneticin
GFP	Green fluorescence protein
H3	Histone H3
H3K9me2	Histone H3 lysine 9 dimethylation
HA	Haemagglutinin
HDAC	Histone deacetylase
his	Histidine
HU	Hydroxyurea
HTP	6x Histidine, Tobacco Etch Virus site, Protein A
IgG	Immunoglobulin G

kb	Kilobase
kDa	Kilodalton
LacZ	β-galactosidase gene
LB	Lysogeny broth
leu	Leucine
lncRNA	long non-coding RNA
Mb	Megabase
ME	Malt extract
mRNA	Messenger RNA
ncRNA	non-coding RNA
N-terminal	Amino-terminal
nmt	No message in thiamine
nt	Nucleotide
ONPG	Ortho-nitrophenyl-β-D-galactopyranoside
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
PMG	Pombe minimal glutamate
RACE	Rapid amplification of cDNA ends
RDRC	RNA-directed RNA polymerase complex
RITS	RNA-mediated initiation of transcriptional silencing
RNA	Ribonucleic acid
RNAi	RNA interference
RNAPII	RNA polymerase II
RPM	Revolutions per minute

RT-qPCR	Quantitative reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
siRNA	Small interfering RNA
SSC	Saline-sodium citrate
TBE	Tris-borate EDTA
TBZ	Thiobendazole
Tgp1	Transporter for glycerophosphodiester 1
tRNA	Transfer RNA
ts	Temperature sensitive
TSS	Transcription start site
ura	Uracil
wt	Wild type
YES	Yeast extract with supplements

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Robin Allshire for the opportunity to work in his laboratory, the freedom and independence to explore research questions I found interesting, and his continued support, guidance, and encouragement. I would also like to thank my co-supervisors Prof. Elizabeth Bayne, Prof. Sander Granneman, and Prof. David Tollervey for their interest in my work and for sharing their expertise. The final copy of this thesis benefited enormously from the discussion with my internal and external examiners, Prof. Jean Beggs and Prof. Jürg Bähler. I am grateful to Prof. Tomoyasu Sugiyama and Prof. Lidia Vasilieva for generously sharing *S. pombe* strains. I am indebted to the current and past members of the Allshire lab. Many thanks to Dr. Tanya Auchynnikava, Dr. Pauline Audergon, Dr. Sandra Catania, Max Fitz-James, George Michaels-Hamilton, Dr. Matthew Miell, Dr. Alison Pidoux, Dr. Manu Shukla Claus, Puneet Singh, Dr. Nick Toda, Dr. Pin Tong, and Dr. Sharon White. You've helped to make my PhD an unusually enjoyable experience. Beyond the confines of the Allshire laboratory, the Wellcome Trust Centre for Cell Biology houses a considerable amount of shared and communal state-of-the-art facilities which my work has reaped many benefits. To the many new friends I've made here in Edinburgh and to all the old friends that have overcome long and sometimes arduous journeys to visit me here in Scotland: you've kept the loneliness at bay. I must acknowledge Rosie Priest in particular for keeping me well hydrated, fed, and relatively sane as I wrote this thesis. And above all, I need to thank my loving and supportive family, even if they might occasionally drive me insane.

Dedication

I dedicate this thesis to FBI Special Agent Dale Cooper and to all the damn fine coffee in the world.

“I have no idea where this will lead us.

But I have a definite feeling it will be a place both wonderful and strange.”

– Dale Cooper

CHAPTER 1

Introduction

1.1 General background

1.1.1 *The central dogma of molecular biology*

All living organisms and many viruses store heritable genetic information in deoxyribonucleic acid (DNA). DNA encodes this information in the arrangement of covalently linked nucleotide bases, which include purines adenine (A) and guanine (G) and pyrimidines thymine (T) and cytosine (C). The two helical strands of DNA are held together by hydrogen bonds that form between specific purine/pyrimidine pairs (A with T and G with C), providing the basic copying mechanism for the inheritance of genetic information as complementary strands unwind and serve as templates for the production of two identical new strands of DNA (Watson and Crick, 1953). Nucleotide pairing is equally important to copy DNA into ribonucleic acid (RNA), a related nucleic acid polymer that can be used as a template for protein synthesis. The "central dogma of molecular biology" provides a simplified framework for this linear flow of genetic information from DNA to functional units in the cell, whereby discrete sequences within DNA are transcribed into messenger RNA (mRNA) that is later translated into protein (Crick, 1970) (**Fig. 1.1**). While the translation step of mRNA into protein is unidirectional, this flow of genetic information is not actually linear. Genetic information stored in RNA can be copied into a complementary strand of RNA or reverse transcribed into a complementary DNA strand (Astier-Manifacier and Cornuet, 1971; Baltimore, 1970; Duda *et al.*, 1973; Temin and Mizutani, 1970). In addition, eukaryotic genomes produce many

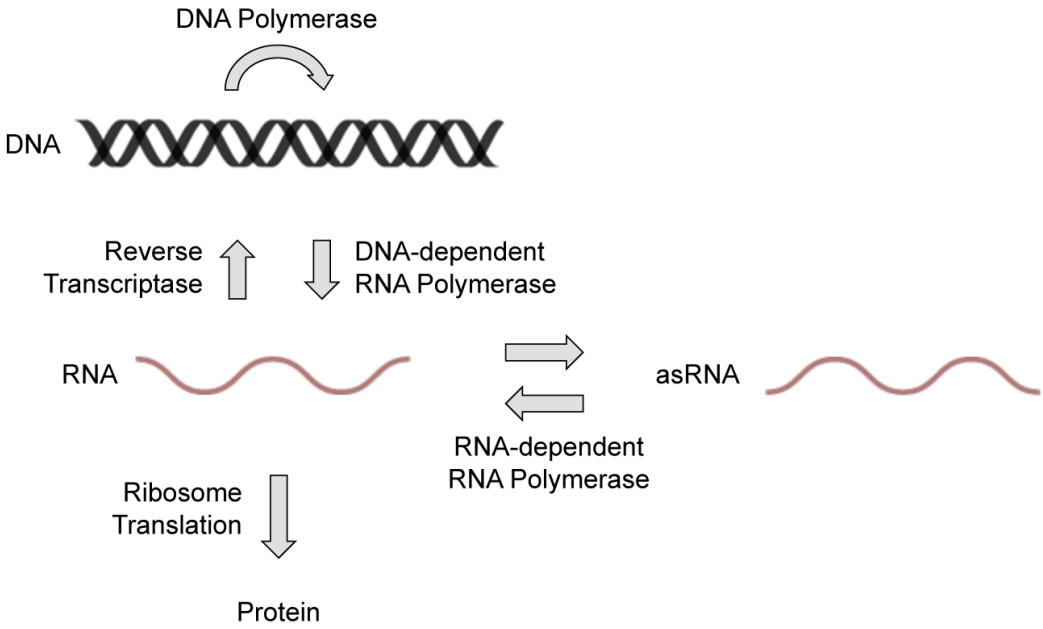

Figure 1.1. The central dogma of molecular biology. This diagram depicts the flow of genetic information from DNA to functional cellular units. Protein-coding genes are copied into a sense-stranded messenger RNA, the template required for protein synthesis. This flow of genetic information from DNA to RNA to protein is over simplified since many non-protein coding RNAs (ncRNAs) are also copied from DNA and can exert specific cellular functions akin to proteins. Moreover, specialized enzymes are capable of copying a strand of RNA into a complementary DNA or RNA strand.

non-protein coding RNAs (ncRNAs) with diverse cellular functions (Cech and Steitz, 2014).

Decoding an mRNA for protein translation is achieved by the ribosome, a large ribonucleoprotein (RNP) complex composed of a variety of proteins and specialized ncRNAs called ribosomal RNA (rRNA) (Schmeing and Ramakrishnan, 2009). The information required to assemble a given protein is stored in sequential three nucleotide codon units in the mRNA, whereby different nucleotide combinations within a codon specify distinct amino acids (Crick *et al.*, 1961). This genetic code evolved very early in the history of life on Earth and is nearly universal among all extant organisms (Koonin and Novozhilov, 2009). Mechanistically, the ribosome decodes this information by facilitating the binding of codons present in mRNA with complementary anticodon sequences in transfer RNA (tRNA), specialized ncRNA adaptors that carry amino acids (Ramakrishnan, 2002). Remarkably, the catalytic step that links these amino acids together to form a polypeptide is mediated by the ribozyme activity of rRNA, not by ribosome proteins (Nissen *et al.*, 2000). Once the linear polypeptide is synthesized by the ribosome, it physically folds into a functional three-dimensional protein structure (Dill and MacCallum, 2012).

Many of the processing events and/or chemical modifications required for the maturation of most mRNAs, rRNAs, and tRNAs are performed by RNP complexes that utilize other specialized RNA molecules, such as small nucleolar RNAs (snoRNAs) (Dieci *et al.*, 2009). The central role of mRNA and ncRNA in the flow of genetic information, in addition to the unique ability of RNA to both store genetic information (like DNA) and catalyze chemical reactions (like enzymatic proteins), are cited as evidence for the RNA world hypothesis, which proposes that all extant life on Earth descended from self-replicating RNA molecules (Gilbert, 1986). It is now

widely believed that early ribozymes catalyzing peptide linkages allowed the formation of polypeptides long enough and diverse enough to catalyze novel biological reactions and spur evolution (Zhang and Cech, 1997). This hypothesis posits that heritable genetic information later became stored in DNA, which is more stable than RNA, and proteins acquired the primary structural and catalytic functions of the cell. Despite this, RNA remains an important intermediate and integral regulator of this process. Moreover, ncRNAs have acquired new functions during the course of evolution, many of which are only presently being discovered.

1.1.2 *Eukaryotic genomes are pervasively transcribed*

An organism's genome contains all the genetic information required for it to grow, develop, and reproduce. While the relatively small genomes of multiple viruses were sequenced as early as the 1970s (Fiers *et al.*, 1978; Sanger *et al.*, 1977), it was not until the 1990s that developments in DNA sequencing technology permitted the assembly of the first bacterial genome (*Haemophilus influenza*e) (Fleischmann *et al.*, 1995). The first archaean (*Methanococcus jannaschii*) and eukaryotic (*Saccharomyces cerevisiae*) genomes were published shortly thereafter (Bult *et al.*, 1996; Goffeau *et al.*, 1996). At the turn of the millennium, rapid advances in sequencing technologies and computational strategies to manage large sequencing datasets culminated in the assembly of the draft human genome (Lander *et al.*, 2001; Venter *et al.*, 2001). Further technological improvements have since permitted more and more organisms to have their genomes sequenced in an increasingly time effective and cost effective manner.

Large-scale bioinformatic approaches now permit evolutionary and biomedical studies on an unprecedented genome-wide scale (Alföldi and Linblad-Toh, 2013). One of the most remarkable outcomes from these studies has been the discovery

that most eukaryotic genomes contain large swaths of DNA that do not actually code for protein. The amount of this non-coding DNA varies considerably between different eukaryotes, making up as little as ~3% of the carnivorous plant *Utricularia gibba* genome and as much as ~98% of mammalian genomes (Elgar and Vavouri, 2008; Ibarra-Laclette *et al.*, 2013). Once dismissed as "junk DNA", some regions of non-coding DNA serve important biological functions. Examples of functional non-coding elements in DNA include genes for ncRNAs, sequences involved in regulating the transcription and translation of protein coding genes, centromere sequences upon which the kinetochore attaches for segregating identical copies of chromosomal DNA to daughter cells during mitosis, repetitive telomere sequences at chromosome ends to protect chromosomes from deterioration and genomic instability, and sequences specifying DNA replication origins (Bell and Dutta, 2002; ENCODE Project Consortium, 2012; Lamb and Birchler, 2003; O'Sullivan and Karlseder, 2010). However, these few examples do not account for all non-coding DNA present in most eukaryotes. Recent estimates suggest that less than 10% of the human genome is constrained and that non-coding regions evolve very rapidly (Rands *et al.*, 2014). It is therefore still unclear how much non-coding DNA serves a real biological function.

A byproduct of high-throughput next-generation sequencing has been the advent of RNA sequencing (RNA-seq), which measures stable transcriptional activity genome-wide. While pre-existing hybridization-based approaches, such as genomic tiling microarrays, had already existed to measure genome-wide transcription patterns, these methods have limited resolution and poor dynamic range due to high background signals from non-specific hybridization and signal saturation. In addition, microarray probes often lack coverage over intergenic regions and regions antisense to protein-coding genes. RNA-seq bypasses these limitations by utilizing

deep sequencing platforms to profile all RNA transcripts present in cells at near single nucleotide resolution, while simultaneously providing information about their strand of origin and expression levels (Wang *et al.*, 2009). Studies using this powerful new tool have revealed that the bulk of non-coding DNA in eukaryotic genomes is actively transcribed, including genomic regions that were long thought to be transcriptionally silent (Jacquier, 2009). Many of these previously undetected transcripts are greater than 200 nt in length and resemble protein-coding mRNAs in many important ways but do not actually code for protein (Mercer *et al.*, 2009). In recent years an enormous amount of effort has been devoted to functionally characterizing these long non-coding RNAs (lncRNAs), which arguably represent the least understood products of eukaryotic genomes.

In this chapter, I will review the basic processes involved in regulating gene expression, starting with how chromatin controls the accessibility of DNA to permit transcription, mechanisms responsible for RNA synthesis and quality control, and present the emerging roles for lncRNAs as functional products of the eukaryotic genome. Where necessary, I will highlight contentious findings and shifting paradigms in this relatively new and rapidly growing discipline. Lastly, I will detail what is currently known about lncRNAs in the fission yeast *Schizosaccharomyces pombe*, a model system that is widely used to study eukaryotic chromatin and RNA biology and provides the basis for much of the original work presented in this thesis.

1.2 Chromatin

1.2.1 *Eukaryotic DNA is organized into chromatin*

Eukaryotic DNA is folded and compacted into a condensed macromolecular structure called chromatin, which consists of DNA, proteins, and RNA (Lilley and Pardon, 1979). Chromatin facilitates the packing of DNA into a much smaller volume, which is required to fit large eukaryotic genomes into the relatively small nucleus of cells. The functional consequences of this packaging include gene expression control, mitigating DNA damage, and permits chromosome segregation in mitosis and meiosis (Li and Reinberg, 2011).

The fundamental, repeating structural unit of chromatin is the nucleosome, an octameric protein core composed of two copies each of the histone proteins H2A, H2B, H3, and H4, that tightly wraps ~147 base pairs of DNA (Luger *et al.*, 1997). However, in some cases one or more of these canonical histone proteins can be substituted with a non-canonical histone variant, which provide nucleosomes with new functional properties (Weber and Henikoff, 2014). Repeating arrays of nucleosomes linked by short segments of DNA and linker histones are organized into higher-order structures that contribute to the condensed compaction of chromosomes (Tremethick, 2007) (**Fig. 1.2**). However, the organization of DNA into chromatin is far from uniform. This is an important feature of chromatin since DNA replication, DNA repair, and transcription all require specialized factors to access the DNA template.

Nucleosome structure and stability control accessibility to the underlying DNA. Changes in the ability of nucleosomes to package DNA are conferred in many ways,

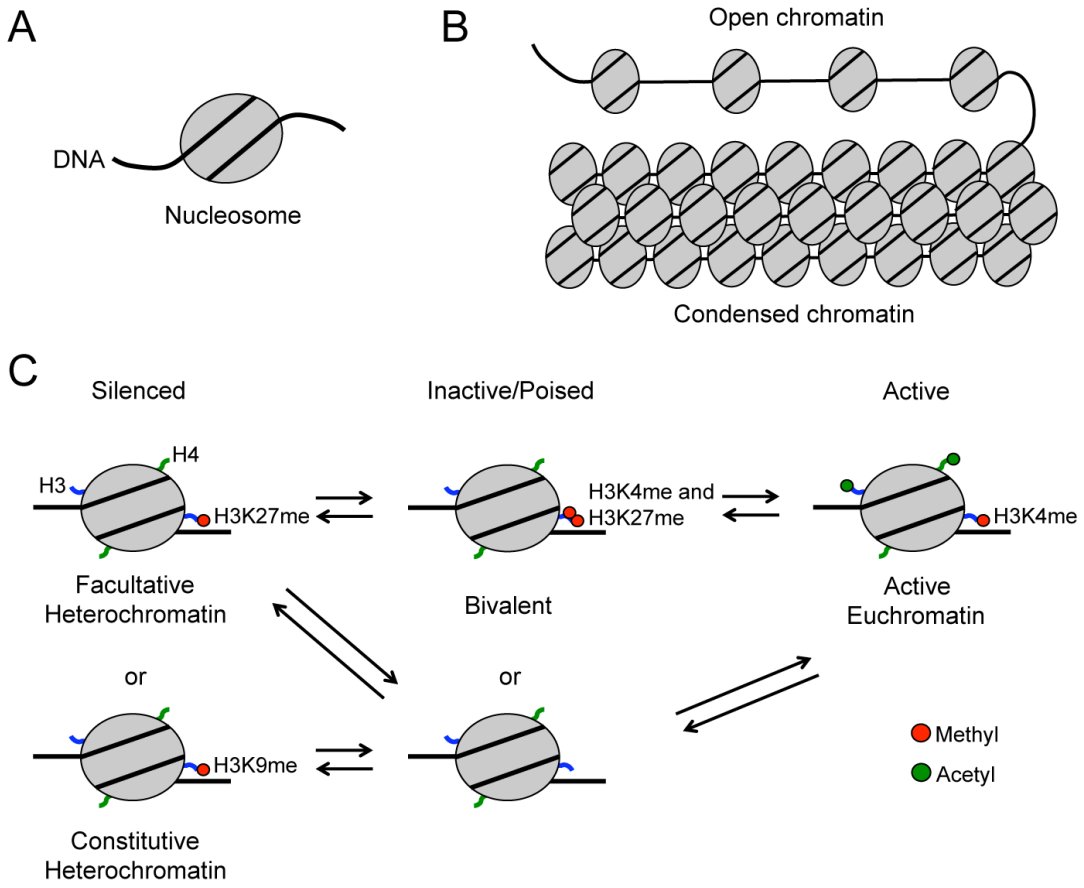

Figure 1.2. Chromatin states. (A) Eukaryotic DNA wrapped around a nucleosome, composed of two copies of each histone protein H2A, H2B, H3, and H4. (B) Actively transcribed genes reside in regions of open chromatin with acetylated nucleosomes that wrap DNA less tightly. Conversely, silenced genes and other condensed regions of repressed chromatin are tightly packaged. (C) Repressed chromatin features nucleosomes with modifications characteristic of facultative and constitutive heterochromatin such as H3K27 and H3K9 methylation, respectively. Inactive regions might contain a mixture of inactive marks (e.g. H3K27me) and active marks (e.g. H3K4me) present at active or poised promoters. Note that the methylation state (i.e. mono-, di-, or tri-methylation) of specific lysine residues on histone proteins often confers distinct properties to nucleosomes. Where necessary, these distinctions will be made clear.

some of these changes include reversible post-translational modifications to histones, the incorporation of variant histone proteins into nucleosomes, chemical modification to DNA, as well as by factors that recognize, maintain, and/or propagate a given chromatin state (Li and Reinberg, 2011). Together, many levels of complex regulation are required to establish and maintain chromatin status at local, and sometimes, chromosome-wide levels. Histone acetyltransferase (HAT) complexes, for example, are an important family of enzymes that transfer acetyl groups to lysine residues on histone tails that protrude from the nucleosome core (Lee and Workman, 2007). Acetylation neutralizes the positive charge of lysine, decreasing the affinity of nucleosomes for the net negative charge inherent to the DNA molecule (Hong *et al.*, 1993). Lysine residues can also be methylated up to three times by histone methyltransferases (HMTs). In the case of the lysine K4 residue on histone H3, tri-methylation (H3K4me3) is important to inhibit the binding of repressive complexes while simultaneously recruiting chromatin-remodeling factors such as HATs to bring about a more open chromatin structure termed euchromatin (Flanagan *et al.*, 2005; Li *et al.*, 2006; Nishioka *et al.*, 2002). Actively transcribed genes are generally concentrated in this type of lightly packed chromatin since DNA is much more accessible to the transcription machinery. Conversely, the ability to silence genes is carried out in part by the action of histone deacetylase complexes (HDACs) that remove acetyl groups from histones, increasing the affinity of nucleosomes for DNA and creating a much more compact chromatin state (Haberland *et al.*, 2009). Chromatin can be further packed into a structure much less permissive to transcription called heterochromatin.

Specific genomic loci in many organisms are dynamically regulated by repressive di- and tri-methylation on H3K27, which is deposited by Polycomb-group proteins and imposes transient “facultative heterochromatin” (Maison and Almouzni, 2004).

Elsewhere in the genome, large domains are enriched in H3K9 di- and tri-methylation, which recruits the heterochromatin protein-1 (HP1) to help establish "constitutive heterochromatin". This form of highly repressed chromatin is frequently observed next to telomeres, flanking centromeres, and at repetitive sequences. Heterochromatin at these sites controls many aspects of chromosome biology, such as ensuring faithful chromosome segregation, controlling the nuclear organization of chromatin, and preventing the spread of harmful transposable DNA elements (i.e. transposons) (Allshire *et al.*, 1995; Csink and Henikoff, 1996; Dernburg *et al.*, 1996; Slotkin and Martienssen, 2007). In special cases, entire chromosomes are silenced by heterochromatin. This is the case for "X-inactivation" in female mammals where one of the two X chromosomes found in each cell is packaged into a repressed heterochromatic structure called a Barr body to achieve dosage compensation between XX females and XY males (Heard and Disteche, 2006).

1.2.2 *DNA methylation influences chromatin status*

Chemical modifications to DNA play an important role regulating gene expression and chromatin status in many eukaryotes. The most well understood DNA modification is cytosine methylation. In mammals, this modification generally occurs on sequences that are unusually GC rich, called CpG islands, and are often associated with regulatory promoter sequences upstream of genes. Nucleosomes within CpG islands are inherently unstable (Ramirez-Carrozzi *et al.*, 2009), which facilitates transcription initiation and therefore likely accounts for their presence at mammalian promoters. DNA methyltransferases (DNMTs) reduce gene expression by depositing methyl groups on cytosine nucleotides in these regions (Saxonov *et al.*, 2006). Gene repression in this context is brought about by methyl-CpG-binding domain (MBD) containing proteins that recognize this modification and recruit histone-modifying activities that compact local chromatin structure (Lunyak *et al.*,

2002; Soppe *et al.*, 2002). The methylation of H3K9 and H3K36 can in turn direct DNMT activity to deposit cytosine methylation, creating a feedback loop that stabilizes repressive chromatin (Baubec *et al.*, 2015; Esteve *et al.*, 2006; Lehnertz *et al.*, 2003). In addition to regulating chromatin structure generally, there is also evidence that methylated cytosines preclude some transcription factors from recognizing DNA binding motifs and can therefore directly prohibit transcription initiation (Choy *et al*., 2010). Importantly, cytosine methylation is copied to new DNA strands during replication, meaning daughter cells are able to inherit the chromatin status of methylated loci following cell division (Bird, 2002).

1.2.3 *Epigenetic inheritance of chromatin states*

Changes in gene expression drive the emergence of different phenotypes from a single genotype. In multicellular organisms, the inherited memory of chromatin states is essential for imprinted allele-specific gene expression and to commit specialized cell types to the appropriate developmental lineage (Feng *et al.*, 2010). Mechanisms involved in propagating specific chromatin states independent of underlying DNA sequence are said to be epigenetic (i.e. the Greek prefix “epi-” meaning “above” genetics). Briefly introduced above, DNA methylation provides a heritable change in phenotype (i.e. gene expression control) without altering the genotype and therefore behaves in an epigenetic manner. However, the prevalence of this epigenetic mark differs greatly between eukaryotes. Cytosine methylation is abundant in plants and vertebrates, rarely present in fruit flies, present in some nematode worm species but not the well-studied *Caenorhabditis elegans*, and absent from all yeast species examined to date (Capuano *et al.*, 2014; Gao *et al.*, 2012). Organisms lacking cytosine methylation provide important systems for studying the ability of other factors, such as histone modifications, non-canonical histone variants, or even RNA, to behave epigenetically.

The specific chromosomal location of some histone variants can be inherited in an epigenetic manner. For example, the histone H3 variant CENP-A is present at centromeres in most eukaryotes and is predominantly maintained there by epigenetically regulated processes (Karpen and Allshire, 1997). In higher eukaryotes, a different histone H3 variant, termed H3.3, is present in nucleosomes that have been displaced by the transcription machinery and has been proposed to transmit a memory of transcriptional activity across cell divisions (Ng and Gurdon, 2008). There is also evidence that the histone H2A variant H2A.Z, which is often distributed near a transcription start site (TSS), can establish a memory of active transcription that poises recently repressed genes for rapid reactivation (Brickner *et al.*, 2007). It is therefore plausible that other context-dependent histone variants are also capable of acting in an epigenetic manner.

Many chromatin-modifying complexes associate with the transcription machinery and/or localize at DNA replication forks, raising the possibility that histone modifications left by these factors and/or the factors themselves might be retained and facilitate the reestablishment of chromatin states to pass epigenetic information to newly divided cells (Esteve *et al.*, 2006; Hansen *et al.*, 2008; Li *et al.*, 2011; Milutinovic *et al.*, 2002; Petruck *et al.*, 2012; Sarraf and Stancheva, 2004). However, the heritability of any given histone mark is limited not only by the presence of the modifying-complex that deposits it but also by the stability of the mark itself. The stability of different histone modifications varies greatly: acetylation and phosphorylation last only minutes, while histone methylation can persist for hours to days (Jackson *et al.*, 1975; Zee *et al.*, 2010). The position of the active methyl-H3K4 and repressive methyl-H3K27 marks have been shown to propagate across generations in the nematode worm *C. elegans* and the fruit fly *Drosophila*

melanogaster, respectively (Gaydos *et al.*, 2014; Greer *et al.*, 2014), while the epigenetic transmission of methyl-H3K9, the constitutive heterochromatin mark, has been demonstrated in the fission yeast *S. pombe* (Audergon *et al.*, 2015; Ragunathan *et al.*, 2015). While further studies are required to assess the capacity of other histone marks to behave in an epigenetic manner, work in diverse systems suggests that multiple methyl-marks are capable of transmitting epigenetic memory.

It is now evident that RNA plays a central role in epigenetics. Many small and long ncRNAs have been discovered to play important roles in diverse chromatin-modifying pathways that establish and/or maintain chromatin states (Bernstein and Allis, 2005). Beyond these findings, exciting new evidence suggests that the stable transfer of these specialized RNA molecules to new daughter cells provides an additional mechanism for establishing epigenetic memory (Holoch and Moazed, 2015). Moreover, the transmission of these regulatory RNAs during gametogenesis might contribute to epigenetic inheritance in higher eukaryotes (Liebers *et al.*, 2014). It is therefore possible that heritable RNA could allow generations of organisms to adapt to rapidly changing environments without the need for changes at the genetic level. Although this is an attractive idea, it remains to be determined to what extent regulatory RNAs are involved in the transmission of epigenetic memory. Future research will reveal how significant and widespread roles for RNA are in the epigenetic inheritance of phenotypes in organisms.

1.3 Gene expression

1.3.1 *Transcription initiation*

The first step of gene expression involves the transcription of RNA from DNA. DNA-dependent RNA polymerases (referred to as RNA polymerases) are a related family of multi-subunit enzymes that are responsible for catalyzing primary RNA synthesis from template DNA. Bacteria and archaea use a single RNA polymerase (RNAP) to synthesize both mRNAs and ncRNAs, while eukaryotic organisms have evolved multiple specialized RNA polymerases that are generally responsible for synthesizing distinct RNA classes (Werner and Grohmann, 2011). In eukaryotes, RNA polymerase I (RNAPI) transcribes rRNAs, RNA polymerase II (RNAPII) synthesizes mRNAs, lncRNAs, and many short regulatory ncRNAs, and RNA polymerase III (RNAPIII) mainly produces tRNAs and the 5S rRNA. Plants are unique in that they have acquired two additional RNA polymerase complexes, RNA polymerase IV (RNAPIV) and RNA polymerase V (RNAPV), which synthesize small interfering RNAs (siRNAs) involved in post-transcriptionally silencing transcripts that contain complementary nucleotide sequences (Haag and Pikaard, 2011). Despite these different functions and slight variations in molecular mechanisms and subunit composition, RNA polymerases are highly conserved from prokaryotes to eukaryotes and all originate from a common ancestor early in the history of life on Earth (Werner and Grohmann, 2011).

The initiation of transcription by RNA polymerase requires a core promoter sequence in DNA. In most bacteria, specialized proteins called sigma (σ) factors directly contact specific promoter DNA sequences and recruit RNAP to initiate transcription (Browning and Busby, 2004). Promoter regions in eukaryotes are much

more complex and require different transcription factors and co-activators to associate with promoters in order to facilitate RNA polymerase binding (Thomas and Chiang, 2006). Eukaryotic gene promoters are typically located upstream of a gene but can also have regulatory elements, such as enhancers or silencers, many kilobases (kb) or even mega bases (Mb) away from the actual TSS (Harmston and Lenhard, 2013). Specific DNA elements in the promoter direct the association of factors essential for initiating transcription. The TATA box, a short TATAAA sequence or a variant thereof, is the best-characterized proximal promoter element known in eukaryotes. Located roughly 30 base pairs (bps) upstream of the TSS (Wang *et al.*, 1996), the TATA-binding protein (TBP) associates with this motif and recruits TBP-associated transcription factors important for transcription initiation (Bushnell *et al.*, 2004; Miller and Hahn, 2006). For this reason, TBP binding is a tightly regulated step and flanking elements adjacent to the TATA box can recruit transcription factor II B (TFIIB) to stabilize the binding of TBP to DNA. It is important to note that the majority of eukaryotic gene promoters do not actually contain TATA box elements (Yang *et al.*, 2007). Instead, TATA-less promoters contain other DNA elements that function analogously by recruiting general transcription factors and later the transcription machinery (Anish *et al.*, 2009; Emami *et al.*, 1998; Seizl *et al.*, 2011; Somboonthum *et al*., 2005). Additional sequence-specific transcription factors and co-activators can vary from gene to gene, increasing the specificity and control of gene expression (Spitz and Furlong, 2012).

In a highly integrated series of steps, RNA polymerase, general and specific transcription factors, and the Mediator complex combine to form what is called the pre-initiation complex (Lewis and Reinberg, 2003). At this point, melting double stranded DNA is a prerequisite to the formation of an open complex between RNA polymerase and the DNA template. This essential step, carried out by the DNA

helicase activity of TFIIH, allows RNA polymerase to synthesize RNA by complimentary nucleotide base pairing with the template DNA strand (Kim *et al.*, 2000). The final RNA product is identical in sequence to the DNA coding strand, with two key exceptions: (1) T in DNA is replaced by the RNA-specific pyrimidine uracil (U) in the nascent transcript and (2) RNA nucleotides are composed of ribose (5-carbon) sugar-phosphate backbones instead of the deoxyribose sugar-phosphate backbones found in DNA. These chemical differences make RNA less stable than DNA but also provide it with many of the additional biochemical properties discussed earlier.

The organization of DNA into chromatin poses a significant physical challenge to eukaryotic transcription. Chromatin must be altered in order to allow transcription factors and RNAPII accessibility to the DNA template. Active eukaryotic promoters exhibit nucleosome-free regions immediately upstream of the TSS (Yuan *et al.*, 2005). Not surprisingly, this pattern is most frequently observed at highly expressed housekeeping genes. Conversely, increased nucleosome density is often found at stress-response gene promoters, which controls expression by masking key regulatory DNA sequences. Numerous chromatin remodelers, histone chaperones, and specific histone modifications grant RNA polymerase access to DNA and remodel nucleosomes to permit transcription into gene bodies (Li *et al.*, 2007). These factors are often targeted directly or indirectly by histone modifications on nearby nucleosomes and/or by specific post-translational modifications to the C-terminal domain (CTD) of Rpb1, the largest subunit of RNAPII (Eick and Geyer, 2013). The Rpb1 CTD is composed of tandem hepta-peptide repeats (YSPTSPS) and reversible post-translational modifications to this domain recruit factors involved in coupling RNAPII transcription to RNA processing and maturation events, in addition to recruiting chromatin-modifying activities that deliver important changes to

the status of chromatin that permit transcription initiation and elongation (Hsin and Manley, 2012; Komarnitsky *et al.*, 2000; Phatnani and Greenleaf, 2006). Although the function of this domain is highly conserved among eukaryotes, the actual number of YSPTSPS repeats differs widely from species to species: 26 repeats in budding yeast *S. cerevisiae* CTD domain, 29 in fission yeast *S. pombe*, 32 in nematode worm *C. elegans*, 34 in flowering plant *Arabidopsis thaliana*, 45 in fruit fly *Drosophila melanogaster*, and 52 in mammals.

In the context of the pre-initiation complex, the Rpb1 CTD is generally non-phosphorylated when RNAPII is first loaded onto to a promoter (Usheva *et al.*, 1992) (**Fig. 1.3**). The successful formation of the pre-initiation complex does not however guarantee productive transcription elongation. For most genes, the transition from initiation to elongation is regulated by a phenomenon referred to as promoter-proximal pausing whereby RNAPII is restrained ~20 - 60 nt downstream of the TSS (Levine, 2011). This provides a major rate-limiting step for transcription. Inhibitive protein complexes such as the DRB sensitivity inducing factor (DSIF), the negative elongation factor (NELF), and Pol II-associated factor 1 (PAF1) play a central role in promoter-proximal pausing and frequently stall RNAPII before it has left the promoter (Chen *et al.*, 2015; Wada *et al.*, 1998; Yamaguchi *et al.*, 1999). The acquisition of Ser-5 phosphorylation on the Rpb1 CTD is thought to dissociate initiation-specific factors and target the Set1 HMT to deposit the active H3K4me mark at promoters (Lee and Skalnik, 2008; Ng *et al.*, 2003; Svejstrup *et al.*, 1997). However, multiple cycles of aborted initiations usually occur, causing sequential RNAPII stalling at promoters before all inhibitive factors finally dissociate. Such events are generally characterized by the presence of both active H3K4me and repressive H3K27me marks on nucleosomes flanking the promoter (Bernstein *et al.*, 2006). Transcription from bivalent promoters such as this is inhibited but also poised

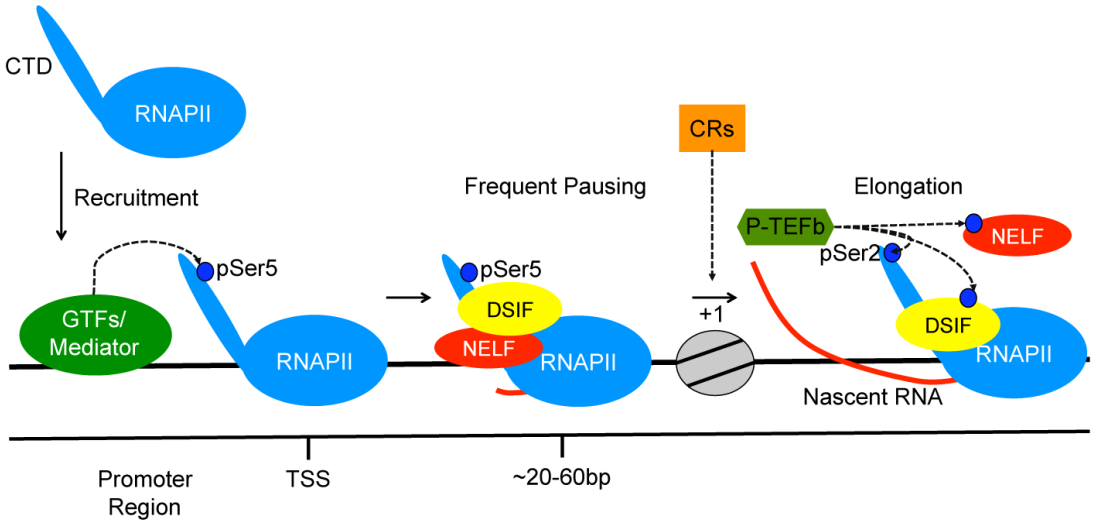

Figure 1.3. Transcription initiation and elongation. General transcription factors (GTFs) and Mediator cooperate to bring RNAPII to nucleosome-depleted promoters. The first nucleosome after the transcription start site (TSS), known as the +1 nucleosome, provides a physical barrier to transcription elongation that must be overcome. Inhibitive factors such as the DRB sensitivity inducing factor (DSIF) and the negative elongation factor (NELF) contribute to RNAPII stalling the promoter. The positive transcription elongation factor b (P-TEFb) phosphorylates inhibitory factors DSIF, NELF, and the Rpb1 CTD on Ser-2, while chromatin remodelers (CRs) disassemble nucleosomes ahead of RNAPII. Together these activities favour productive transcription elongation into the gene body.

for rapid transcription initiation following the removal of H3K27me and associated inhibitory factors. This is an important step for controlling the rate of transcription from a given promoter. It is also suggested that this level of regulation helps to control the proper directionality of transcription since most, if not all, eukaryotic promoters are capable of initiating transcription in either direction (Xu *et al.*, 2009; Wei *et al.*, 2011). Ultimately, the positive transcription elongation factor b (P-TEFb) phosphorylates inhibitory factors DSIF, NELF, and the Rpb1 CTD to favour productive transcription elongation from the promoter into the gene body (Bres *et al.*, 2008).

1.3.2 *Transcription elongation*

The Rpb1 CTD loses Ser-5 phosphorylation as RNAPII transcription travels away from the initiation site (Brodsky *et al.*, 2005). Thus, this modified form of RNAPII is predominantly confined to promoters and 5' regions of genes (**Fig. 1.4**). However, as RNAPII clears promoters, the CTD acquires Ser-2 phosphorylation, which is necessary for transcription elongation, termination, and 3'-end formation (Eick and Geyer, 2013; Ni *et al.*, 2008). The Set2 HMT interacts with this elongating form of RNAPII and deposits H3K36 methylation on nucleosomes over the gene body of actively transcribed genes (Li *et al.*, 2003; Xiao *et al.*, 2003). H3K36 tri-methylation impedes histone chaperones from incorporating acetylated nucleosomes (Venkatesh *et al.*, 2012) and serves as a docking site for HDACs (Carrozza *et al.*, 2005; Keogh *et al.*, 2005). In mammalian cells, H3K36me3 can also target DNA methylation to the body of actively transcribed genes (Baubec *et al.*, 2015). Together, these activities are thought to prevent aberrant transcription from initiating at cryptic promoters within gene bodies.

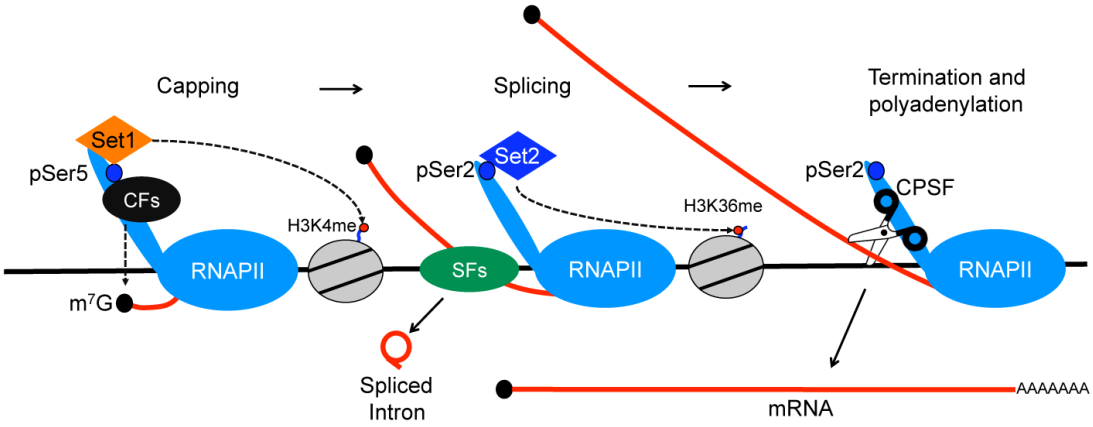

Figure 1.4. Co-transcriptional RNA processing and chromatin modifications. RNAPII-associated factors actively process the nascent transcript (shown in red) during transcription and modify chromatin. Capping factors (CFs) and 3′ end processing/termination factors such as the cleavage and polyadenylation specificity factor (CPSF) all bind directly to the CTD of RNAPII subunit Rpb1. Ser-5 phosphorylation on the Rpb1 CTD (pSer5) recruits CFs as well as the Set1 HMT, which is responsible for methylating H3K4 on nucleosomes adjacent to active gene promoters. Elongating RNAPII can also recruit splicing factors (SFs) to co-transcriptionally remove introns from the nascent transcript. During transcription elongation, the Rpb1 CTD loses pSer5 and acquires Ser-2 phosphorylation (pSer2), which recruits Set2 HMT allowing methylation of H3K36 on nucleosomes positioned over gene bodies. Thus, the pattern of Rpb1 CTD phosphorylation and histone H3 methylation change predictably during the course of transcription.

Importantly, RNA polymerase elongation is a highly discontinuous process, moving forward at variable rates, frequently pausing, and at times backtracking. Pausing is a natural feature of RNAPII transcription and an important regulatory step to control gene expression, as introduced above, but is also important to help facilitate RNA folding (Pan and Sosnick, 2006), to allow time for the co-ordination of co-transcriptional processing and termination (Alexander *et al.*, 2010; Gusarov and Nudler, 1999), as well as to permit quality control measures to take place (Thomas *et al.*, 1998). Pauses are often reversible and regulated by a myriad of factors (Jonkers and Lis, 2015). In cases where elongation factors are unable to overcome halted RNA synthesis, transcription will arrest. Arrested RNAPII can be cleared from chromatin and even targeted for destruction by proteasome-mediated degradation (Svejstrup, 2007).

1.3.3 *RNA processing and transcription termination*

Unlike bacteria, where transcription and translation are coupled (Robinson and van Oijen, 2013), eukaryotic mRNAs require extensive processing and export to the cytoplasm before translation can occur. Common modifications to pre-mRNA transcripts in eukaryotes occur simultaneously with transcription and include capping the 5'-end of the transcript, splicing, and modified 3'-ends. Evidence for co-transcriptional processing comes in part from the ability of the Rpb1 CTD to directly recruit factors that stimulate RNA processing of the nascent transcript (Hirose and Manley, 1998; Hirose *et al*., 1999; Ho and Shuman, 1999). In turn, many factors involved in transcription initiation and elongation also influence capping (Chiu *et al.*, 2002), RNA splicing (Ji and Fu, 2012), and 3'-end processing events (Rosonina *et al.*, 2003; Nagaike and Manley, 2011).

Capping involves the addition of a modified guanine nucleotide, 7-methylguanosine (m^7G), to the 5'-end of the growing RNAPII transcription product (Rasmussen and Lis, 1993). The m^7G cap improves RNA stability and recruits RNA splicing factors to excise sequences within genes that do not code for protein (introns) (Görnemann *et al.*, 2005). It also plays a role in directing mRNA export to the cytoplasm and helps to guide the ribosome to the mRNA for protein translation (Cheng *et al.*, 2006; Mitchell *et al.*, 2010; Preiss and Hentze, 1998).

Many eukaryotic genes are interrupted by non-coding intron sequences. Specific sequence motifs within introns direct the spliceosome, a large RNP complex composed of five small nuclear RNAs (snRNAs) and a range of associated proteins, to excise introns and ligate flanking coding regions called exons (Will and Luhrmann, 2011). The recognition of splice sites relies on many variables, including RNAPII kinetics and auxiliary factors that are predominantly recruited by the Rpb1 CTD (de la Mata and Kornblihtt, 2006; Fong *et al.*, 2003). Alternative splicing can include or exclude particular exons from the final RNA product (Matlin *et al.*, 2005). Sequence-specific RNA-binding protein factors, the packaging of nascent transcripts into heterogeneous nuclear RNP (or hnRNP) complexes that hide strong splice sites or expose weak splice sites, and RNA secondary structure all contribute to alternative splicing events (Caputi and Zahler, 2002; McManus and Graveley, 2011; Olson *et al*., 2007). The process is further complicated by the possibility that some elements within promoters influence the decision to alternatively splice exons in a gene (Cramer *et al.*, 1997). Ultimately, alternative splicing allows a single gene to encode multiple protein products (isoforms) and vastly increase the diversity of proteins encoded by eukaryotic genomes. An additional outcome of splicing can be the formation of circularized RNA (circRNAs) molecules. Although long dismissed as insignificant byproducts of splicing, an accumulating number of individual circRNAs

have been found to serve genuine biological functions in cells (Lasda and Parker, 2014). In these cases, splicing produces more than alternate isoforms of individual protein, but also provides the opportunity to generate functional circRNA products from a gene. Finally, splicing activity has been found to correlate with changes in histone modifications (Luco and Misteli, 2011), coupling co-transcriptional processing activities to changes in chromatin and vice versa.

Transcription terminates when a polyadenylation signal sequence in the nascent RNA is recognized by a cleavage and polyadenylation specificity factor (CPSF). The CPSF recruits additional factors to cleave the 3'-end of the transcript and add a long poly-adenine (poly-A) tail (Zhao *et al.*, 1999). The poly-A tail is important for mRNA transport to the cytoplasm and efficient protein translation (Huang and Carmichael, 1996; Preiss and Hentze, 1998). The poly-A tail also controls mRNA stability as poly-A tail shortening triggers RNA degradation (Laird-Offringa *et al.*, 1990). Similar in concept to different splice isoforms, the 3'-end of genes can have multiple polyadenylation signal sequences, leading to the possibility of alternative polyadenylation products (Di Giammartino *et al.*, 2011). This is important as the non-coding sequences between the translation stop codon and the poly-A tail of an mRNA, known as the 3'-untranslated region (3'-UTR), influences RNA localization, translation, and stability (Matoulkova *et al.*, 2012). Alternative polyadenylation can therefore modulate both translation efficiency and mRNA abundance. In addition, small regulatory RNAs have been found to originate from the cleavage of 3' regions in bacterial mRNAs (Miyakoshi *et al.*, 2015). It is currently unclear whether eukaryotic genes increase the output of single mRNAs by producing short functional ncRNAs in this manner.

1.3.4 *RNA-mediated translation control*

Unlike prokaryotic translation, which is a continuous process with transcription in the cytoplasm, eukaryotic mRNAs are produced from DNA in the nucleus and must be exported to the cytoplasm for translation into protein. Specialized export receptors utilize GTPase activity to deliver many of the small ncRNAs important for translation and translational control to the cytoplasm, while much longer transcripts, such as mRNAs or lncRNAs, require much more sophisticated mechanisms for recruiting/assembling exporter complexes (Köhler and Hurt, 2007). Once in the cytoplasm, mature mRNAs are transported to ribosomes for translation. However, an important level of translation control in many eukaryotes involves a class of small ncRNAs termed microRNAs (miRNAs). These short ncRNAs (~21-22 nt in length) often originate from their own genes and are transcribed by RNAPII (Lee *et al.*, 2004). miRNAs require a great deal of processing, both in the nucleus and following export to cytoplasm, and primarily function by imperfect base-pairing with complementary sequences in specific target mRNAs (Winter *et al.*, 2009). This binding generally occurs in the 3' regions of transcripts and inhibits protein synthesis by directly repressing translation initiation or by stimulating mRNA degradation, but some miRNAs have also been found to stimulate protein synthesis (Fabian *et al.*, 2010). There is emerging evidence that specific lncRNAs are also involved in regulating the translation of mRNAs (Carrieri *et al.*, 2012). It is therefore evident that diverse species of ncRNAs play important roles in all levels of gene expression, including RNA processing steps, protein translation control, and post-transcriptional gene regulation and even degradation.

1.3.5 *RNA degradation*

The final stage in the lifespan of RNA involves degradation by highly conserved RNA surveillance pathways. There are three main classes of RNA-degrading

enzymes, termed ribonucleases (RNases), common to all living organisms: 1) exonucleases that degrade RNA from the 5'-end, 2) exonucleases that degrade RNA from the 3'-end, and 3) endonucleases that make internal excisions in RNA (Houseley and Tollervey, 2009). Different RNA surveillance pathways are often redundant and many are also involved in different RNA processing steps (Doma and Parker, 2007). In fact, nearly every step in RNA biogenesis involves meticulous quality control measures performed by RNA surveillance pathways to detect errors in transcription, processing, and export. Therefore, it is thought that specificity for RNA processing and degradation activities is imparted by the interactions of transcripts with specific RNA-binding proteins and complexes (Bühler *et al.*, 2008; LaCava *et al.*, 2005; Wang *et al.*, 2005). Ultimately, these pathways play a critical role in the ability of cells to tightly regulate mRNA and ncRNA turnover as well as to provide an appropriate and timely response to environmental and development cues at the level of gene expression control.

Virtually all RNA molecules are processed and/or degraded by the exosome complex, a highly conserved multi-subunit protein complex with endonuclease and 3'→5' exonuclease activity that is present in all eukaryotes (Januszyk and Lima, 2014). The exosome is composed of a nine-subunit core that directly binds proteins that confer catalytic activity (**Fig. 1.5**). One essential catalytic subunit is Dis3/Rrp44, which aids substrate recognition and possesses both endonuclease and 3'→5' exonuclease activity (Lebreton *et al.*, 2008; Liu *et al.*, 2006; Schneider *et al.*, 2007). The nuclear exosome complex contains an additional catalytic factor called Rrp6 and has been shown to associate with actively transcribing genes to influence transcription itself (Allmang *et al.*, 1999; Castelnuovo *et al.*, 2013; Lemay *et al.*, 2014; Shah *et al.*, 2014; Wagschal *et al.*, 2012). Importantly, Dis3/Rrp44 and Rrp6 bind opposite ends of the exosome core and target distinct RNA substrates, thereby

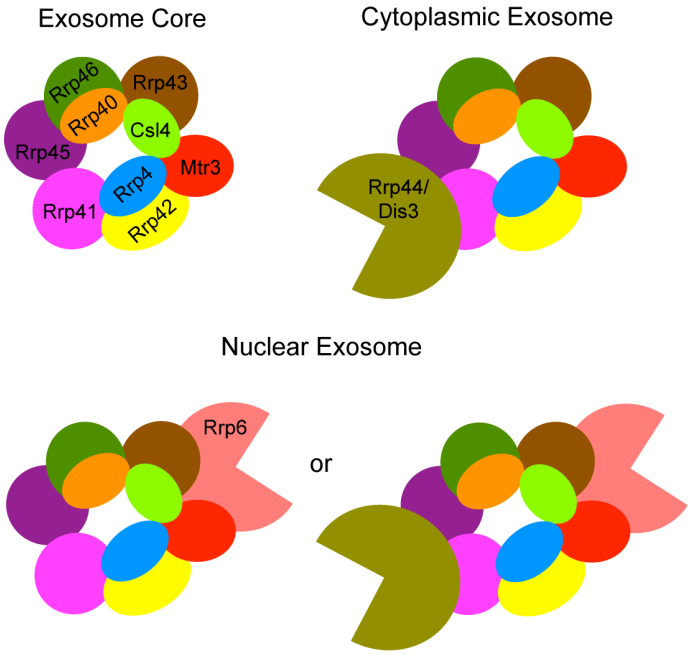

Figure 1.5. The exosome complex. The exosome complex is composed of a 9 sub-unit, catalytically inactive core. In the cytoplasm this exosome core associates with Rrp44/Dis3, an RNase with endonuclease and exonuclease activity. The nuclear exosome contains an additional subunit, Rrp6, which provides a secondary exonuclease activity. Degradation/processing by the exosome requires that RNA substrates enter the internal chamber of the exosome core, from either side, to reach the active sites of Rrp44/Dis3 or Rrp6, which are each positioned at opposite sides of this chamber. Importantly, Rrp44/Dis3 and Rrp6 often target a unique set of transcripts, increasing the specificity of the exosome complex.

providing specificity for exosome targets (Makino *et al.*, 2013; Kiss and Andrulis, 2010). Specificity is further conferred by interactions with auxiliary factors that direct exosome activities to distinct classes of RNAs (Houseley and Tollervey, 2009). These is also evidence that auxiliary factors mediate the association of the exosome complex with factors involved in heterochromatin formation (Bühler *et al.*, 2007; Murakami *et al.*, 2007; Reyes-Turcu *et al.*, 2011; Vasiljeva *et al.*, 2008; Zhang *et al.*, 2011), providing a possible link between co-transcriptional RNA surveillance mechanisms and changes in chromatin status.

Mature RNA products are often modified to prevent/postpone degradation. Eukaryotic mRNAs are primarily degraded from the 3'-end by the activity of the exosome complex. The poly-A tail therefore provides protection for the 3'-end of the transcript. Stepwise deadenylation of the poly-A tail by the exosome controls mRNA turnover (Tran *et al.*, 2004). Eukaryotic mRNAs are also protected from degradation at the 5'-end by the m^7G cap. The m^7G cap must first be removed by decapping enzymes before 5'→3' exonucleases can actively degrade the transcript (Coller and Parker, 2004). 5'→3' exonuclease activity is also important for preventing read-through transcription into neighbouring genes since polyadenylation site cleavage during 3'-end processing exposes a free, unmodified 5'-end on the nascent RNA, which allows 5'→3' exonuclease degradation by Xrn2 to chase the polymerase and terminate transcription (West *et al.*, 2004). Interestingly, several factors involved in transcriptional termination have also been implicated in promoter-proximal pausing (Gardini *et al.*, 2014; Stadelmayer *et al.*, 2014), revealing multi-layered gene expression control. In sum, many of the modifications made during RNA maturation are critical for cells to control both the stability and quality of transcripts produced from the genome.

1.3.6 *RNA interference*

In most eukaryotes, gene expression can be regulated post-transcriptionally by a process termed RNA interference (RNAi), which involves either miRNAs or siRNAs. While miRNAs are generally transcribed from their own genes, RNAi silencing by siRNA is initiated by the cleavage of double-strand RNA (dsRNA) molecules into short siRNA fragments (~20-24 nt long) by the RNA endonuclease Dicer (Bernstein *et al.*, 2001) **(Fig. 1.6)**. Dicer-derived siRNAs are then incorporated into the RNA-induced silencing complex (RISC) and used to recognize complementary RNA transcripts and target them for degradation by Argonaute, the catalytic subunit of RISC (Hannon, 2002). Beyond operating in this manner, elements of the RNAi pathway are also involved in the biogenesis of miRNAs (Bartel, 2004).

RNAi pathways provide an efficient mechanism for post-transcriptional control of gene expression and serve many important biological roles, such as defending cells against foreign genetic material from viruses and other parasites and preventing transposons from propagating through an organism's genome (Obbard *et al.*, 2009). In addition, elements of the RNAi pathway are important to establish repressive heterochromatin in many species (Volpe and Martienssen, 2011), and recent studies in *C. elegans* and *S. pombe* suggest that a memory of RNAi-mediated silencing activities can be passed trans-generationally (Buckley *et al.*, 2012; Kowalik *et al.*, 2015). Thus, RNAi is capable of facilitating the transmission of epigenetic states.

RNAi has become an incredibly powerful research tool to reduce target gene expression by introducing synthetic siRNAs into cells or whole organisms with sequences complementary to genes of interest (Mello and Conte, 2004). Large-

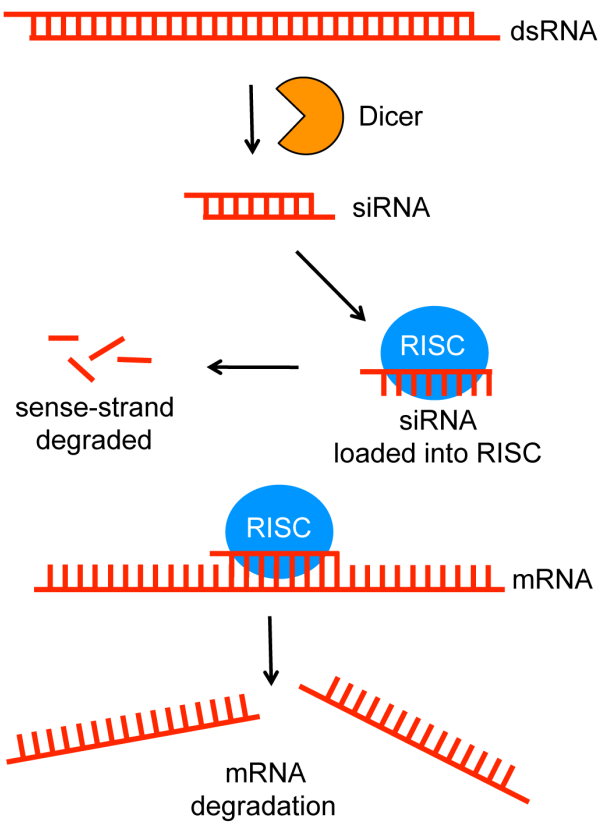

Figure 1.6. RNA interference (RNAi). The RNAi pathway processes dsRNA into siRNAs and silences target transcripts as depicted. Dicer cuts dsRNA into siRNAs. siRNAs are loaded into the RNA-induced silencing complex (RISC). The RNase activity of Argonaute, the catalytic subunit of RISC, cleaves the passenger sense strand. RISC targets homologous mRNAs by base-pairing complementarity and cleaves target mRNAs.

scale RNAi-directed knockdown approaches have permitted high-throughput, genome-wide screens to identify new genes and molecular pathways associated with specific phenotypes, accelerating functional genomics research (Mohr *et al*., 2010). However, RNAi based studies are now beginning to lose some of their appeal following the emergence of CRISPR-Cas9 technology, an RNP-based adaptive immune system in bacteria that has been engineered to allow efficient and rapid genome editing in eukaryotes (Sander and Young, 2014). Nonetheless, silencing genes of interest with RNAi remains a useful approach for examining the effects of reduced gene expression, especially in genome-wide screens and in the case of studying individual genes essential to survival.

1.4 Long non-coding RNAs

1.4.1 *Functional lncRNAs or transcriptional noise?*

Having largely escaped detection until recently, due to technological limitations, it is now widely accepted that most eukaryotic genomes generate an abundance of lncRNA transcripts, defined as RNAPII transcripts that lack protein-coding open reading frames (ORFs) and are greater than 200 nt in length. This arbitrary size threshold of 200 nt is a useful cutoff since experimental procedures can easily select RNAs that are larger than 200 nt in length from shorter transcripts, which represent better known classes of small regulatory RNAs introduced above. The fact that lncRNAs are defined by their size rather than on any common function is evidence that it is still unknown what roles, if any, many of these transcripts play in cells.

lncRNAs can be transcribed antisense to protein-coding genes, from within introns, or from intergenic regions of the genome (**Fig. 1.7**). Genome-wide profiling of

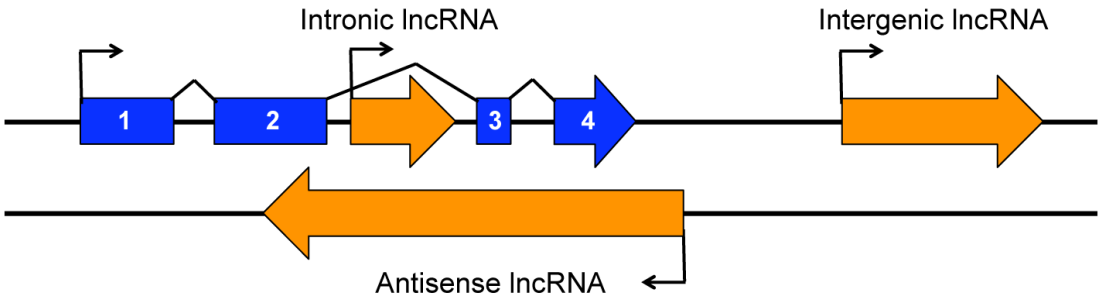

Figure 1.7. Origins of eukaryotic long non-coding RNAs. Eukaryotic genomes produce an abundance of long non-coding RNAs (lncRNAs). lncRNAs are synthesized by RNAPII and can originate from within protein-coding gene introns, be transcribed antisense to protein-coding gene ORFs, or from intergenic regions of the genome as illustrated.

RNAPII occupancy and histone modifications associated with RNAPII transcription initiation and elongation indicate that the same transcriptional machinery that generates mRNAs is responsible for lncRNA expression (Guttman *et al.*, 2009). Further consistent with lncRNAs being transcribed by RNAPII, many of these transcripts are co-transcriptionally processed in the same manner as mRNAs (e.g. m^7G capped, spliced, and polyadenylated). Despite these similarities, lncRNAs as a class are poorly conserved in primary nucleotide sequence when compared with mRNAs (Pang *et al.*, 2006). Furthermore, the fate of mRNAs and lncRNAs is notably different. In contrast to stable mature mRNAs that are exported to the cytoplasm for protein synthesis, lncRNAs remain predominantly nuclear and many are rapidly degraded by the exosome and/or other RNA decay pathways (Ponting *et al.*, 2009). Consequently, the majority of lncRNAs exhibit low steady-state levels compared to mRNAs. A series of elegant experiments performed in budding yeast *S. cerevisiae* show that RNA processing factors involved in 3'-end formation govern the fate of these transcripts (Tuck and Tollervey, 2013). Based on these findings, lncRNAs with 3' cleavage and polyadenylation motifs resembling those of mRNAs are generally more stable and more likely to be exported to the cytoplasm. Likewise, mRNAs with 3' cleavage and polyadenylation motifs resembling those of lncRNAs are less stable and generally represent mRNAs with lower abundance in cells.

Unlike other classes of small ncRNAs or rRNAs, which are relatively well characterized, a limited but growing number of lncRNAs have been characterized in detail. Circumstantial evidence for their functional significance originally came from genome-wide expression studies showing that many lncRNAs exhibit cell type-specific expression patterns and are regulated during development (Wilusz *et al.,*2009). Altered patterns of lncRNA expression have also been observed in human diseases and developmental disorders (Lee and Bartolomei, 2013), implicating

some of these transcripts in human health and disease. Whether such changes in lncRNA abundance are merely symptomatic of the disease state or actually drive important phenotypic changes associated with disease progression is still unclear and the focus of ongoing research. More generally, the question of whether the bulk of lncRNAs encoded by eukaryotic genomes serve genuine cellular functions or might simply result from inconsequential acts of “transcriptional noise” arising from low RNAPII fidelity casts a modicum of doubt on the biological significance of fluctuations in lncRNA expression catalogued by genome-wide approaches (Struhl, 2007).

A clear challenge for assigning function to lncRNAs has been the general absence of sequence conservation. Despite this drawback, the order of genes flanking the transcription units that encode lncRNAs can be preserved through evolution (i.e. conserved synteny) (Ulitsky *et al.*, 2011; Necsulea *et al.*, 2014), raising the possibility that such transcripts might represent functionally conserved lncRNAs whose primary sequences have diverged too greatly to retain detectable homology. Further evolutionary support for lncRNA function stems from the observation that lncRNA and mRNA promoters exhibit similar levels of sequence conservation (Derrien *et al.*, 2012), while splice motifs are also frequently conserved in multi-exonic lncRNAs (Haerty and Ponting, 2015). Together these observations suggest that near equivalent levels of selective pressure act on the regulatory elements of mRNA and lncRNA genes. Therefore, not having to maintain codons for protein synthesis might allow lncRNAs to be more amenable to evolutionary changes in nucleotide sequence provided the structure and overall function of the transcript is preserved.

As progress is being made in assigning function to lncRNAs in organisms from a variety of taxa, it is also becoming evident that many functional lncRNAs retain little to no detectable primary sequence conservation between even the most closely related species (Pang *et al.*, 2006). A prominent example is the RNA component of the telomerase enzyme, an RNP complex with reverse transcriptase activity that contains a telomere repeat-containing lncRNA template to extend telomere length and protect chromosome ends from shortening (Lingner *et al.*, 1997). Despite this essential function in cells, telomerase RNA sequence and length is extremely variable between different eukaryotes with sizes ranging from as few as ~450 nt in vertebrates to >1,000 nt in many species of yeast (Theimer and Feignon, 2006). Therefore, an absence of detectable sequence conservation does not necessarily negate function for any given lncRNA.

To date, an increasing number of lncRNAs have been found to play diverse roles in cells. Most notably, many lncRNAs have been found to influence different steps in gene regulation (Geisler and Coller, 2013). Some of these functions include altering chromatin status to activate or silence transcription, recruiting or disrupting transcription factor and RNAPII binding, playing roles in co- and post-transcriptional processes as well as translation control, and even regulating RNA degradation. The mechanisms that underlie some of these functions are discussed below.

1.4.2 *lncRNAs as precursors for shorter functional RNAs*

In some cases, the product of lncRNA transcription is not functional in and of itself, but is instead processed into smaller regulatory RNAs. For example, multiple snoRNAs and snRNAs originate from lncRNA transcripts (Askarian-Amiri *et al.*, 2011; Fejes-Toth *et al.*, 2009). Additionally, studies in *S. pombe* have shown that lncRNAs transcribed from repeats flanking centromeres are processed into double-

stranded RNAs by RNA-dependent RNA polymerase (RDRP) activity and later into siRNAs by the RNAi machinery, which targets the H3K9 methyltransferase Clr4 complex (CLRC) to establish repressive pericentric heterochromatin (Bayne *et al.*, 2010; Motamedi *et al.*, 2004; Verdel *et al.*, 2004; Volpe *et al.*, 2002) (**Fig. 1.8**). In these and other cases, such lncRNAs do not represent functional transcripts *per se* but instead serve as precursors that are processed into other functional RNAs.

1.4.3 *Antisense lncRNA transcription regulates gene expression*

Eukaryotic gene expression is regulated at many different levels by the transcription of lncRNAs antisense to protein coding genes. Frequently these transcripts are targeted by RNA decay pathways and therefore exhibit low levels of expression. It is also important to note that antisense transcripts can also be derived by RDRP activity and that transcripts derived by this process can also play important roles in post-transcriptional gene regulation (Ahlquist, 2002; Lehmann *et al.*, 2007).

In many cases, the act of transcribing antisense lncRNAs represses genes on the sense strand (Bitton *et al.*, 2011; Xu *et al.*, 2011). This has been found to be the case in variety of organisms, including those that lack functional RNAi pathways, arguing against the idea that these transcripts form double-stranded sense-antisense pairs sensitive to RNAi activity. In other words, regulation in these cases cannot be explained by targeted degradation by the RNAi machinery. Although the mechanism(s) for regulation by antisense transcription is/are yet to be fully resolved, the effects might simply be the consequence of stronger transcription on one strand competing with the progression of RNA polymerase on the opposing strand in any given cell. This is a likely explanation since convergent RNA polymerases collide and are incapable of passing one another (Hobson *et al.*, 2012). Thus, controlling the balance between sense and antisense transcription might provide a simple yet

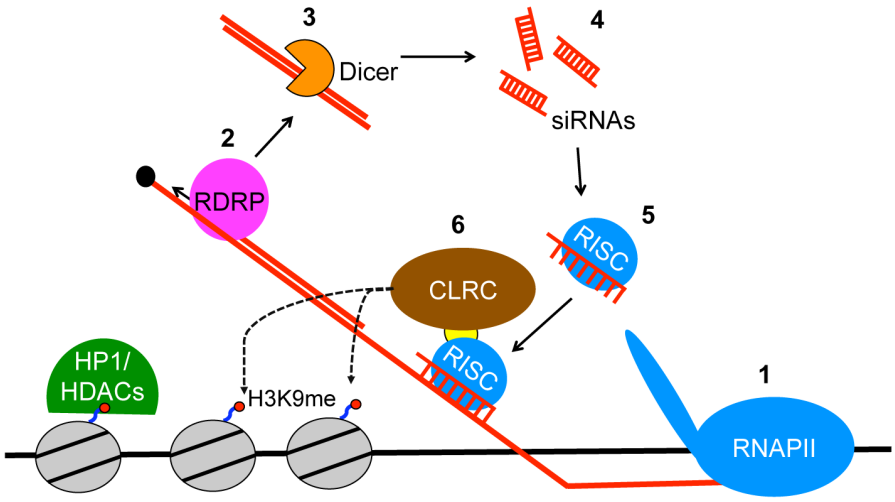

Figure 1.8. Model of pericentric heterochromatin formation in *S. pombe*. (1) RNAPII transcribes pericentric repeats into lncRNAs, which (2) are synthesized into double-stranded RNAs (dsRNA) by RNA-dependent RNA polymerase (RDRP) activity. (3) dsRNAs are targeted by Dicer for siRNA production. (4) siRNAs are loaded into the RNA induced silencing complex (RISC). (5) siRNA-RNA base-pairing allows RISC to associate with nascent transcripts at centromeres. (6) The Clr4 H3K9 methyltransferase complex (CLRC) is recruited to the nascent lncRNA via interactions with RISC to deposit the methyl-H3K9 constitutive heterochromatin mark at centromeres.

effective mechanism for regulating genes that need to be quickly activated or repressed. Beyond regulating transcription on the sense strand, upstream antisense transcription resulting from bidirectional promoters can also transmit regulatory activity to neighbouring genes as well (Wei *et al.*, 2011). This kind of transcriptional circuitry is emerging as an important aspect of many eukaryotic gene expression programs.

Additional roles for antisense transcripts involve complementary base pairing to regulate the corresponding sense RNA. In some cases, sense/antisense pairing has indeed been found to target the RNAi machinery to process double-stranded RNA into siRNAs that mediate further post-transcriptional silencing activities (Colmenares *et al.*, 2007). Alternatively, complementary sense/antisense pairing can influence other aspects of RNA biology, including splicing and protein translation (Beltran *et al.*, 2008; Carrieri *et al.*, 2012; Jabnoune *et al.*, 2013; Kawano *et al.*, 2007). In at least one instance, sense/antisense pairing has been reported to mask miRNA-binding sites in the BACE1 mRNA, which encodes an enzyme implicated in Alzheimer's disease (Faghihi *et al.*, 2010). In doing so, the BACE1-antisense transcript positively regulates BACE1 mRNA stability. Thus, the pairing of antisense transcripts with their mRNA counterparts can have a number of effects on gene expression that cannot simply be predicted based on the detection of an antisense transcription nor from simply analyzing nucleotide sequence.

Finally, some antisense lncRNAs have also been reported to regulate transcription by recruiting chromatin-modifying complexes and/or chromatin remodelers that alter local chromatin architecture in a manner that affects transcription from the sense strand (Camblong *et al.*, 2007; Houseley *et al.*, 2008; Swiezewski *et al.*, 2009; Yamanaka *et al.*, 2015). Indeed, this has emerged as a feature of many lncRNAs

transcribed from within introns and intergenic regions of eukaryotic genomes as well. For this reason, the diverse mechanisms by which antisense, intronic, and intergenic lncRNAs are thought to alter chromatin structure are described below.

1.4.4 *lncRNA-directed chromatin modifications*

An increasing number lncRNAs are thought to directly or indirectly associate with/recruit factors involved in altering chromatin status, and in doing so can either silence or activate target genes (**Fig. 1.9**). This phenomenon plays a significant role in *S. cerevisiae* where lncRNAs have been reported to aid the response of cells to specific changes in nutrient availability by recruiting chromatin-modifying complexes (e.g. HDACs) to dynamically regulate multiple stress-response genes (Camblong *et al.*, 2007; Houseley *et al.*, 2008; van Werven *et al.*, 2012). Related silencing mechanisms that utilize lncRNA-dependent recruitment of chromatin-modifying complexes have also been reported in multicellular eukaryotes. For example, the transcription of an intronic lncRNA in *Arabidopsis thaliana* termed *COLDAIR* recruits an HMT called the polycomb repressive complex 2 (PRC2) to silence the *Flowering Locus C* (*FLC*) gene by depositing the repressive H3K27me mark locally (Heo and Sung, 2011). Remarkably, the outcome of *FLC* regulation by this lncRNA is control over flowering time in this plant. lncRNAs in human and mouse have also been found to physically associate with and target PCR2 activity to bring about repressive H3K27me chromatin over target genes (Kotake *et al.*, 2011; Pandey *et al.*, 2008). In addition to these examples, the mammalian lncRNA *H19* has been shown to recruit H3K9 methyltransferases and the methyl-CpG-binding domain protein 1 (MBD1) to silence several imprinted genes (Monnier *et al.*, 2013). Other lncRNAs have been identified as playing even more direct roles in the regulation of DNA methylation. For example, the human lncRNA *ecCEBPA* promotes *CEBPA* gene activation by preventing the DNA methyltransferase DNMT1 from depositing cytosine methylation

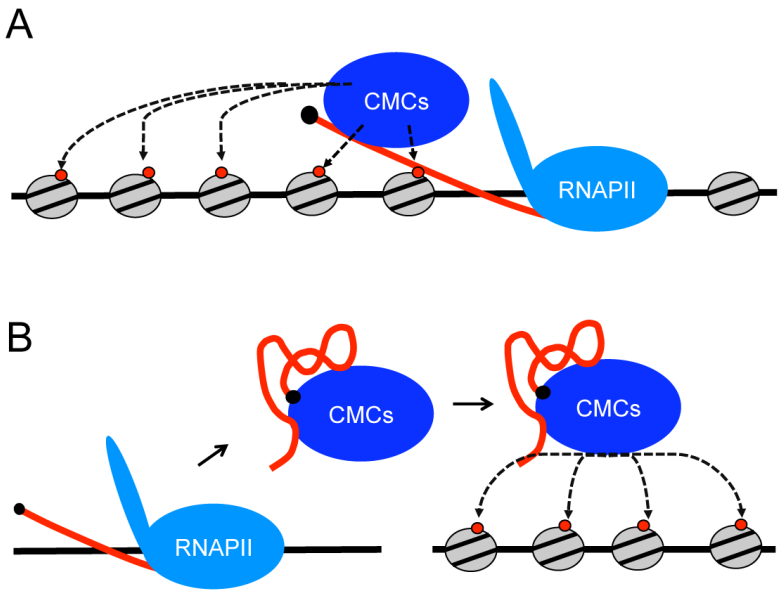

Figure 1.9. lncRNAs can direct chromatin modifications in *cis* and/or *trans*. (A) *cis*-acting lncRNAs, such as *HOTTIP*, interact with and recruit chromatin-modifying complexes (CMCs) to deposit histone and/or DNA modifications locally, regulating nearby gene expression. In contrast, *trans*-acting lncRNAs, such as *HOTAIR*, regulate direct chromatin-modifying activities to regulate genes at distal loci (B). Indeed, CMCs might conceivably be stabilized by interactions with nascent lncRNAs or mRNAs in addition to interactions with the RNAPII CTD.

over the promoter of this gene (Di Ruscio *et al.*, 2013). Related mechanisms have also been assigned to the lncRNA *Dali* in both human and mouse cells and to the lncRNA *Evf2* in mouse (Berghoff *et al.*, 2013; Bond *et al.*, 2009; Chalei *et al.*, 2014), both of which have been shown to control the expression of genes important in neural development and differentiation. Together these examples reveal that the transcription of lncRNAs can play important roles in gene silencing by directing repressive histone modifications or DNA methylation locally in *cis*.

An increasing number of lncRNAs have also been found to recruit parts of the transcription machinery and/or chromatin-modifying activities that deposit active histone marks to stimulate the expression of nearby genes. For example, the human lncRNA *HOTTIP* interacts with the WDR5 protein and targets a Set1-like H3K4 methyltransferase to stimulate transcription at developmental genes in the *HOXA* locus (Wang *et al.*, 2011). More generally, a number of enhancer-like lncRNAs in human cells have been reported to activate adjacent genes by mechanisms that involve altering local chromatin structure, facilitating chromatin looping, and/or by directly recruiting elements of the transcription machinery (Lai *et al.*, 2013; Li *et al.*, 2013; Mousavi *et al.*, 2013; Ørom *et al.*, 2010). While it is now accepted that most, if not all, enhancer elements in human cells are transcribed (Andersson *et al.*, 2014), it is not yet clear whether the bulk of these transcripts are the cause or consequence of enhancer action on nearby genes. Collectively, the above examples reveal that many lncRNAs regulate nearby genes in *cis* by directly or indirectly recruiting factors involved in controlling gene expression.

In special cases, *cis*-acting lncRNAs can alter the chromatin status of entire chromosomes. For example, two lncRNAs in *Drosophila*, *roX1* and *roX2*, are responsible for inducing an active chromatin state that facilitates hyper-transcription

from the single male X chromosome in order to achieve dosage compensation between the sexes (Ilik and Akhtar, 2009). An alternative strategy for dosage compensation is employed in mammals where the *cis*-acting *Xist* lncRNA indirectly recruits chromatin-modifying complexes that bring about repressive heterochromatin along one of the two X chromosomes in female mammals (McHugh *et al.*, 2015). Despite this highly conserved lncRNA-dependent mechanism of achieving dosage compensation in mammals, the *Xist* RNA itself is poorly conserved in nucleotide sequence among even closely related mammalian species (Pontier and Gribnau, 2011). This observation lends further weight to the argument that a lack of primarily sequence conservation does not necessarily rule out conservation of lncRNA function.

While the majority of lncRNAs found to influence chromatin status operate locally in *cis*, there is evidence that some lncRNAs direct chromatin-modifications at distant sites in *trans*. The first *trans*-acting lncRNA reported, the human lncRNA *HOTAIR*, is transcribed from the *HOXC* locus but targets PCR2 activity to silence developmental genes in the *HOXD* locus (Rinn *et al.*, 2007). Since the discovery of *HOTAIR*, additional lncRNAs have been reported to regulate nearby and/or distal genes in *trans*. Some of the most notable of these *trans*-acting lncRNA include the *S. cerevisiae* Ty1 retrotransposon regulatory RNA (Berretta *et al.*, 2008) and the *PHO84* antisense lncRNA (Camblong *et al.*, 2009), the mouse lncRNAs *Evf2* (Berghoff *et al.*, 2013), *Dali* (Chalei *et al.*, 2014), *NeST* (Gomez *et al.*, 2013), *Firre* (Hacisuleyman *et al.*, 2014), *Bvht* (Klattenhoff *et al.*, 2013), and *Paupar* (Vance *et al.*, 2014), and the human lncRNAs *Dali* (Chalei *et al.*, 2014) and *CTBP1-AS* (Takayama *et al.*, 2013). In many of these cases the lncRNA product is proposed to interact with chromatin-modifying complexes and direct histone and/or DNA modifications. However, this ability of lncRNAs to act in *trans* is frequently disputed

and has become increasingly controversial due to the difficulty of many current approaches and techniques to reliably distinguish between *cis* and *trans* effects (Bassett *et al.*, 2014). Improved experimental design is therefore required to conclusively establish *trans* functions for any given lncRNA.

1.4.5 *lncRNA transcription can influence nearby gene expression*

The mere act of intergenic lncRNA transcription itself, including accompanying chromatin modifications and resulting changes in nucleosome positioning and/or density (Li *et al.*, 2007), can have a profound impact on neighbouring gene expression (**Fig. 1.10**). In the simplest scenario, lncRNA expression can provide an environment that is either suitable or unsuitable for transcription factor binding. For example, in a process termed "transcriptional interference," serine mediated repression of the *SER3* gene in *S. cerevisiae* is brought about by lncRNA transcription into the gene promoter, which increases nucleosome density and prevents transcription factor access (Hainer *et al.*, 2011; Martens *et al.*, 2004; Thebault *et al.*, 2011). Mechanisms of transcriptional interference have since been observed in numerous other organisms. The *clr* gene in *Escherichia coli* (Zafar *et al.*, 2014), the *Ubx* gene in *Drosophila* (Petruk *et al.*, 2006), the human dihydofolate reductase gene (Martianov *et al.*, 2007), and the imprinted *Igf2r* gene in mammals (Latos *et al.*, 2012) are all repressed by lncRNA transcription into their respective promoters. Alternatively, promoter-associated lncRNA transcription has in some cases been observed to reposition nucleosomes in a manner that helps to activate gene expression. For example, lncRNA transcription immediately upstream of the *S. pombe* $fbp1^{+}$ gene is required to induce $fbp1^{+}$ (Hirota *et al.*, 2008), while lncRNA transcription antisense to the *PHO5* gene in *S. cerevisiae* displaces inhibitory nucleosomes in the promoter to facilitate *PHO5* induction (Uhler *et al.*, 2007). Taken together these examples illustrate the positive and negative influence that lncRNA

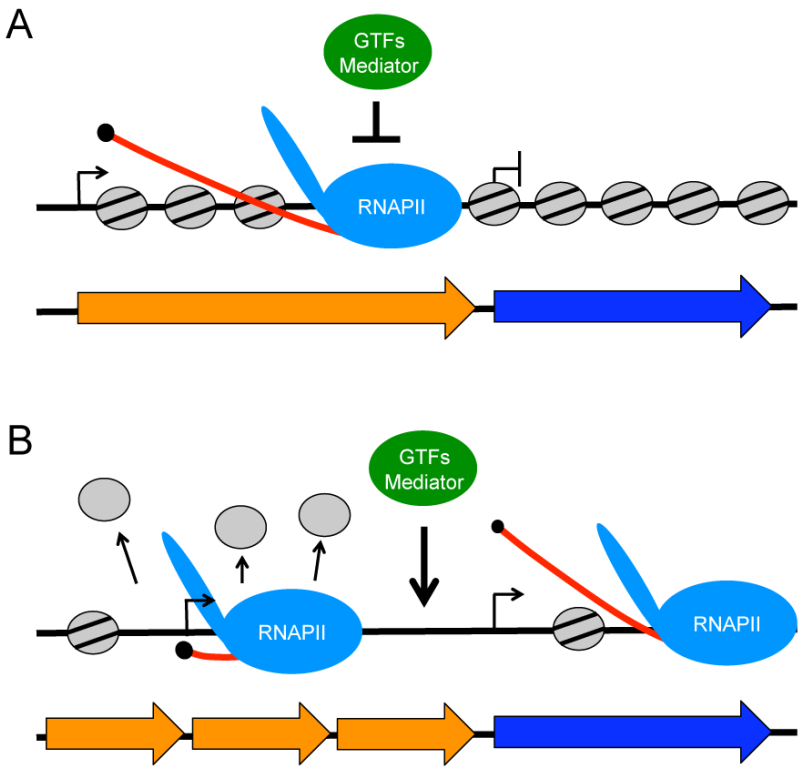

Figure 1.10. The act of lncRNA transcription can regulate nearby genes. (A) lncRNA transcription-associated changes in nucleosome density and histone modifications over a promoter can bring about a chromatin environment that prevents gene induction, a mechanism observed in many systems and termed "transcriptional interference". In other cases, upstream lncRNA can create an open chromatin structure to permit gene activation as is observed at the *S. pombe fbp1*$^{+}$ locus (B).

transcription can exert on the expression of adjacent genes but also the difficulty of assigning function to the mere detection of lncRNA transcription within a gene promoter since the outcomes are clearly locus-dependent. Importantly, these findings emphasize the requirement for experimental approaches to distinguish between outcomes that might simply be a consequence of lncRNA transcription and those that are mediated by functional lncRNA products.

1.5 *Schizosaccharomyces* as a model for studying lncRNA biology

The fission yeast genus *Schizosaccharomyces* is comprised of four known single-celled species: *S. pombe*, *S. octosporus*, *S. japonicus*, and *S. cryophilus* (Rhind *et al.*, 2011). Rather than dividing by asymmetric budding, as is the case for budding yeasts (*Saccharomycetes*), fission yeast cells grow length-wise and divide by medial fission. *S. pombe* is the best studied fission yeast species, having first been discovered in East African millet beer in 1893 (Nasim *et al.*, 1989); it is therefore fitting that the species name "*pombe*" means beer in Swahili. By the late 20th century *S. pombe* had become a powerful experimental model for studying eukaryotic biology. Significant advances in genetics and the understanding of eukaryotic cell cycle regulation stemmed from studies utilizing *S. pombe* (Wood *et al.*, 2002). These studies have been especially valuable since fission yeasts share many important biological processes with higher eukaryotes. These similarities include conserved cell cycle regulation, frequent intron splicing, chromosomes with large repetitive centromeres and telomeres, many shared heterochromatin proteins and histone modifications, and an active RNAi pathway (Rhind *et al.*, 2011). Moreover, the fission yeast genomes, although relatively small and condensed, encode an

abundance of uncharacterized lncRNA transcripts, making these organisms a useful system for studying lncRNA biology and evolution.

The *S. pombe* genome is predicted to contain greater than 1,500 stable lncRNAs (Wilhelm *et al.*, 2008; Rhind *et al.*, 2011). Since RNA decay pathways degrade many lncRNAs (Berretta and Morillon, 2009), there are likely to be more cryptic lncRNAs present in the *S. pombe* genome. Functionally, the balance between sense/antisense transcription is now known to control the expression of many genes involved in sexual differentiation and stress-response pathways in this organism (Bitton *et al.*, 2011; Leong *et al.*, 2014), but little is known about the functional significance of most intergenic lncRNAs in this organism. Although very few of the >500 putative intergenic lncRNAs in *S. pombe* show any detectible sequence homology with lncRNAs in related fission yeast species, many lncRNAs appear to reside in regions of conserved gene order (synteny) (Rhind *et al.*, 2011). As has been predicted in other organisms, it is possible that these types of positionally conserved lncRNAs are functionally conserved transcripts whose primary sequences might have diverged too much so as not to retain detectable homology.

1.6 Project aims

The primary objective of this project is to expand the repertoire of known functional lncRNAs transcribed by eukaryotic genomes by assigning function to some of the many uncharacterized intergenic lncRNAs in *S. pombe*. More specifically, the aims of this project are to (i) identify conserved intergenic lncRNAs in *S. pombe*, (ii) determine the consequence(s) of lncRNA loss on *S. pombe* growth and viability, and (iii) functionally characterize lncRNAs whose loss results in a clear phenotype.

CHAPTER 2

Materials and methods

2.1 Standard techniques and yeast protocols

2.1.1 *Bacterial growth conditions and media*

Single bacterial colonies were grown in Lysogeny broth (LB) medium at 37°C. For plasmid selection, bacterial colonies were grown in LB medium containing 50 μg/mL Carbenicillin (or 100 μg/mL Ampicillin).

LB: 1% w/v Bacto tryptone, 0.5% w/v Bacto yeast extract, 170 mM NaCl, and 15 g/L Bacto agar for LB/agar plates

2.1.2 *Yeast growth conditions and media*

Fission yeasts *S. pombe*, *S. octosporus*, and *S. japonicus* cultures and colonies were incubated at temperatures ranging from 18°C to 36°C and grown in either YES (Yeast extract plus supplements) medium or PMG (Pombe minimal glutamate) synthetic medium as indicated for each experiment. Budding yeast *S. cerevisae* cultures were grown in YPD (Yeast extract-peptone-dextrose) medium or SD Broth 2% Glucose medium (Formedium). For phosphate starvation experiments, *S. pombe*, *S. octosporus*, and *S. japonicus* cells were grown to mid-log phase in YES medium or PMG synthetic medium, washed twice in dH_2O to remove any residual phosphate, and then grown for indicated times in PMG lacking phosphate ($-PO_4$). In contrast, *S. cerevisiae* cells were grown in YPD medium or SD Brother 2% Glucose

(Formedium) to mid-log phase, washed twice in dH_2O, and then grown in SD Broth 2% Glucose without Phosphate (Formedium) for phosphate starvation experiments. For drug-sensitivity experiments, *S. pombe* cells were spotted onto YES agar or PMG agar with DMSO or 20 μg/mL thiabendazole (TBZ), 10 mM hydroxyurea (HU), 15 mM caffeine (CAF). For oxidative stress experiments, *S. pombe* cells were spotted onto YES agar in the presence of 1 mM hydrogen peroxide (H_2O_2). For UV-sensitivity experiments, *S. pombe* cells spotted on YES agar were UV-irradiated at 80J/m^2 with a Stratalinker® UV crosslinker (Stratagene) and grown in the dark at 25°C for 7+ days. Full repression of *nmt* promoters was achieved by growing *S. pombe* cells in the presence of 15 μM (~5 μg/mL) Thiamine. For growth curve analysis, *S. pombe* cells were dispensed in 96-well microplates that were read at OD_{595} every 15 mins for 24 hrs at 32°C with continuous shanking in a Sunrise™ plate reader (Tecan).

YES: 0.5% w/v yeast extract, 3% w/v glucose, 225 mg/L supplements (adenine, histidine, leucine, uracil, and lysine hydrochloride), and 20 g/L agar for YES/agar plates

YPD: 1% w/v yeast extract, 2% w/v Bacto peptone, 2% w/v glucose, and 20 g/L agar for YPD/agar plates

PMG: 14.7 mM potassium hydrogen phthalate, 15.5 mM Na_2HPO_4, 3.75 g/L L-glutamic acid (monosodium salt), 2% w/v glucose, 20 mL/L 50x salt stock, 1 mL/L 1,000x vitamin stock, 0.1 mL/L 10,000x mineral stock, plus supplements, and 20 g/L agar for PMG/agar plates

PMG minus phosphate (-PO_4): 14.6 mM NaOAc, 3.75 g/L L-glutamic acid (monosodium salt), 2% w/v glucose, 20 mL/L 50x salt stock, 1 mL/L 1,000x vitamin stock, 0.1 mL/L 10,000x mineral stock, plus supplements

ME agar plates: 30 g/L malt extract (OXOID), plus supplements, 20 g/L agar

50x Salt Stock: 260 mM $MgCl_2.6H_2O$, 4.99 mM $CaCl_2.2H_2O$, 670 mM KCl, 14.1 mM Na_2SO_4

1,000x Vitamin Stock: 4.20 mM pantothenic acid, 81.2 mM nicotinic acid, 55.5 mM inositol, 40.8 mM biotin

10,000x Mineral Stock: 80.9 mM boric acid, 23.7 mM $MnSO_4$,13.9 mM $ZnSO_4.7H_2O$, 7.40 mM $FeCl_2.6H_2O$, 2.47 mM molybdic acid, 6.02 mM KI, 1.60 mM $CuSO_4.5H_2O$, 47.6 mM citric acid

Supplement stocks: 5 g/L 50x Adenine, 10 g/L 100x Arginine, 10 g/L 100x Histidine, 10 g/L 100x Leucine, 2 g/L 20x Uracil

Table 2.2.1 Haploid *S. pombe* generation times

MEDIUM	TEMPERATURE	GENERATION TIME
YES	25°C	3 hrs
	32°C	2 hrs 10 mins
	36°C	2 hrs
PMG	25°C	4 hrs
	32°C	2 hrs 30 mins
	36°C	2 hrs 20 mins

2.1.3 *Spotting assay*

Spotting assays assessed the growth of *S. pombe* strains to different conditions. An equal number of cells were mixed into 200 μL of filter-sterilized dH_2O. Five serial (1:4) dilutions were made in sterile microtitre plates. Cells were spotted onto YES agar or PMG agar, allowed to dry, and grown at desired temperature.

2.1.4 *Lithium acetate transformation of S. pombe cells*

Genetic deletions and protein tagging were carried out by lithium acetate transformation of linear DNA fragments. 50 mL culture of *S. pombe* cells were grown to mid-log phase (~0.5-1x10^7 cells/mL) and harvested at 3,500 RPM for 2 mins. Cells were washed twice in 50 mL of 0.1 M LiAc and then resuspended to a density of ~1x10^9 cells/mL in 0.1 M LiAc. 100 μL aliquots per transformation were incubated at 32°C for 30 mins with shaking. 1-10 μg DNA (in no more than 15 μL) was added to samples, followed by 290 μL of pre-warmed PEG (50% w/v polyethylene glycol – 3350). Samples were mixed by vortexing and incubated for 30 to 45 mins at 32°C. Cells were then heat shocked at 42°C for 20 mins, centrifuged at 13,000 RPM for 1 min at 4°C, and resuspended in 1 mL non-selective media to grow for 1-2 hrs. For antibiotic selection, cells were grown overnight in non-selective media. 10, 50, and 200 μL of cells were pipetted onto selective plates. Selections were performed on PMG agar plates with according auxotrophy or on YES agar plates with appropriate antibiotic(s). Working concentrations of compounds used are as follows: 1 g/L 5-fluoro-orotic acid (FOA) (Sigma-Aldrich), 2.5 μg/mL phloxine B (Sigma-Aldrich), 100 μg/mL Nourseothricin (CloNAT) (Werner BioAgents), 100 μg/mL Geneticin (G418; not the same kanamycin used for bacterial selections, despite the use of the name kanR) (Gibco), 400 μg/mL hygromycin B (Life Technologies). Plates were allowed to dry for 30 to 60 mins and incubate inverted at 32°C for several days.

2.1.5 *Transformation of S. pombe cells by electroporation*

Plasmids were transformed into *S. pombe* cells by electroporation. 50 mL cultures of log phase cells (5x10^6 to 1x10^7 cells/mL) were harvested at 3,500 RPM for 2 mins. Cells were washed three times in ice-cold 1.2 M sorbitol and then resuspended to a

density of 10^9 cells/mL. 200 μL of cells were mixed with ~100 ng plasmid DNA in ice-cold cuvettes. Cells were pulsed using a Gene Pulser® II electroporation system (Bio-Rad) using the following *S. pombe*-specific settings: 2.25 kV, 200 Ω, and 25 μF. Immediately following pulse, 500 μL of ice-cold 1.2 M sorbitol was added and cells were pelleted. Cells were resuspended in 1 mL dH_2O and plated in different amounts on selective plates. For antibiotic markers cells were first grown overnight in non-selective media before plating. Plates were allowed to dry for 30 to 60 mins at room temperature and incubated inverted at 32°C for 3-5 days.

2.1.6 *Mating and crosses*

Crosses were performed on malt extract (ME) medium in order to starve cells of nitrogen and induce mating/sporulation. A similar amount of cells from two strains of opposite mating types (h^+/h^-) were mixed together and incubated for two days at 32°C (or 25°C for temperature sensitive strains). The presence of asci containing four spores was assessed by light microscopy. Asci were then resuspended in 300 μL of 1:10 diluted glusulase and incubated overnight at 36°C (or for two days at room temperature for temperature sensitive strains). Glusulase digests asci wall and vegetative cells so that only spores remain alive. Ethanol can also be added to kill any remaining vegetative cells. 10 mL dH_2O was then added and 2 μL, 20 μL, and 200 μL were pipetted onto YES agar plates and incubated for 2-4 days at 25-32°C (36°C inhibits germination). Single colonies were replica plated to selective media.

2.1.7 *Genetic screening*

The *S. pombe* Genome-wide Deletion Mutant Library (Bioneer) includes ~3,000 h^+ haploid strains bearing single non-essential gene deletions. This library was used to profile synthetic phenotypes (i.e. lethality, reduced cell growth, etc.). Manipulations

were carried out using a High Throughput Screening RoToR colony pinning robot (Singer Instruments). The library was arrayed in a 384-colony format, four colonies per deletion strain, on YES agar containing 100 μg/mL G418. The *SPNCRNA.808Δ::ura4*$^{+}$ tester strain was crossed with the PEM-2 strain and then also arrayed in 384-colony format on YES agar and containing 100 μg/mL CloNAT. All cells were grown at 30°C for 4 days. Cells from the Bioneer Genome-Wide Deletion Mutant Library collection and the tester strain were then combined to mate on PMG (full supplements) agar plates and incubated at 25°C for an additional 4 days. The resulting mix of cells and spores was then transferred directly to non-selective PMG (full supplements) agar plates containing 2.5 μg/mL phloxine B to detect proportion of dead cells (dark pink) and antibiotics CloNAT and G418 as a control for growth. The mix of cells and spores was also transferred to selective PMG (-uracil) agar plates containing phloxine B, antibiotics CloNAT and G418, and cyclohexamide for anti-diploid selection. Plates were incubated at 30°C for 5 days and then transferred to the fridge at 4°C for 2 days prior to visual analysis and imaging.

2.2 DNA protocols

2.2.1 *Bacterial transformation*

Plasmids were transformed into competent DH5α *E. coli* cells as follows. 50 μL of cells were thawed and incubated with DNA on ice for 30 min before heat shocking at 42°C for 42 sec. Cells were immediately incubated back on ice for 2 min. 500 μL of LB was added and cells were then grown at 37°C for 1 hr and pipetted onto selective plates. Plates were allowed to dry for 30 to 60 min and incubate inverted at 37°C overnight.

2.2.2 *Plasmid miniprep*

Plasmid DNA was isolated by miniprep as follows. Single bacterial colonies were grown in 5 mL LB plus appropriate antibiotic at 37°C overnight. Cells were harvested and miniprep was performed using the Qiaprep® Spin Miniprep Kit (Qiagen) according to manufacturer's instructions. Plasmid DNA was eluted from columns using TE buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA) and stored at -20°C.

2.2.3 *S. pombe genomic DNA isolation*

Genomic DNA was isolated from *S. pombe* cells as follows. Cells were grown to stationary phase in a 5-10 mL culture and harvested at 3,000 RPM for 2 mins. Pellets were resuspended in 250 μL SP1 buffer (1.2 M Sorbitol, 50 mM NaOAc, 50 mM Sodium Phosphate, 40 mM EDTA, pH 5.6) containing 0.4 mg/mL Zymolyase-100T (MP Biomedicals) and incubated 30 to 60 mins at 37°C. Cells were quickly pelleted at 13,000 RPM. Pellets were resuspended in 0.5 mL TE, 50 μL 10% SDS, and vortexed. 165 μL of 5M KOAc was added and samples were incubated on ice for 30 mins before centrifugation at 13,000 RPM at 4°C for 10 mins. Supernatants were added to 0.75 mL isopropanol, incubated on dry ice for 10 mins, and centrifuge. Pellets were resuspended in 0.3 mL TE containing 10 μg/mL RNase A (Roche) and incubated for 30 mins at 37°C. DNA was extracted with phenol/chloroform and precipitated by the addition of 1/10 volume of 3M NaOAc and 2-3 volumes of 100% ethanol. DNA pellets were resuspended in 30 μL TE and stored at -20°C.

2.2.4 *Rapid isolation of S. pombe genomic DNA by colony-PCR*

Genetic modifications to *S. pombe* were confirmed by colony PCR using Taq DNA polymerase (Roche) and oligonucleotide primer pairs (Sigma-Aldrich) over new

genome junctions. A very small amount of a single colony of *S. pombe* was suspended in 10 μL SPZ buffer (1.2 M sorbitol, 100 mM sodium phosphate, 2.5 mg/mL Zymolyase-100T) and incubated for 30 mins at 37°C. SPZ reactions were diluted 1/10 with 90 μL dH_2O. 5 μL of diluted crude genomic DNA was used as template for PCR reactions.

2.2.5 *Polymerase chain reaction (PCR)*

DNA was amplified by PCR as follows. Reactions were carried out in 0.2 mL thin walled PCR tubes (STARLAB) containing the following components: 10-100 ng template DNA, 10 mM primers, 2.5 mM dNTPs, 10x PCR buffer, 0.5 U Taq DNA Polymerase (Roche), and dH_2O. All oligonucleotide primers were purchases from Sigma-Aldrich or Integrated DNA technologies. For cloning purposes, a high fidelity DNA polymerases such as Platinum® Pfx DNA Polymerase (Life Technologies) or Phusion® High-Fidelity DNA Polymerase (New England Biolabs) was used according to manufacturer's instructions. PCR reactions were performed using a T3000 Thermocycler (Biometra).

2.2.6 *Agarose gel electrophoresis*

Mixtures of DNA were separated according to molecular size by agarose gel electrophoresis as follows. Agarose (1%-2% w/v) was dissolved in 1x TBE buffer (0.1 M Trizma® Base, 0.1 M boric acid, 2 mM EDTA pH 8.0) by heating in a microwave. Once cooled, 0.03 μg/mL ethidium bromide (Sigma-Aldrich) was added. 6x Orange G loading buffer (30% glycerol, 0.25% Orange G) was added to DNA samples and loading into wells within the agarose gel. An electric current (120-140 V) was applied to the gel for 30-60 mins. Being negatively charged, DNA moves towards the positively charged anode. Gels were visualized under a U:GENIUS UV

transilluminator (Synergene). However, in the case of preparative gels, 2 μg/mL crystal violet was added to 1% agarose in 1x TBE in order to visualize large amounts of stained DNA by eye and simplify excision of desired bands from gel.

2.2.7 *Quantitative real-time PCR (qPCR)*

Absolute amounts of DNA in a sample were measured by qPCR and performed using a LightCycler® 480 instrument (Roche). Product size was restricted to 80-120 bp for primer pairs used in qPCR experiments. Reactions were carried out in 10 μL volumes: 5 μL 2x SYBR® Green qPCR Master Mix (Roche), 0.5 μL/each 10 mM primers, 1 μL filter sterilized dH_2O, and 3 μL diluted DNA to be analyzed. qPCR program: 95°C for 2 mins, 45 cycles of 95°C for 20 secs, 55°C for 20 secs, 72°C for 20 secs, and final melting curve. Data was analyzed using the Second Derivative Maximum method available with LightCycler® 480 Software 1.5.0.39. This method identifies the maximum acceleration of the PCR reaction's fluorescence signal by calculating the maximum point of the second derivative of the amplification curve (i.e. the crossing point or Cp value). For all qPCR reactions, Cp values were obtained by calculating the mean Cp from three technical triplicates. To eliminate problems introduced by pipetting error, mean Cp values with standard deviations greater than 1.5 were excluded from analysis and repeated.

2.2.8 *Molecular cloning*

All sequence editing and primer designed was performed using SeqBuilder™ software in Lasergene Genomics Suite 11.0.0 (DNASTAR). The plasmids containing the *lacZ* gene under the control of the *nc-tgp1* and *nc-1343* bidirectional promoter were cloned as follows. This non-coding promoter was amplified from *S. pombe* genomic DNA in both orientations (using lacZ_1_F/lacZ_1_R and

lacZ_2_F/lacZ_2_R primer pairs). Restriction enzyme digestions of PstI and SalI restriction sites provided sticky ends for Quick T4 DNA Ligase (New England Biolabs) ligation of PCR products into the *pREP3x-LacZ* vector containing the *lacZ* gene. To test if *nc-tgp1* can repress *tgp1*$^{+}$ *in trans*, the *nc-tgp1* transcription unit was amplified from *S. pombe* genomic DNA (using nc-tgp1_SalI_F and nc-tgp1_XmaI_R primer pairs) and ligated into the *pREP3x* vector under the control of very strong, thiamine repressible *nmt1* promoter using SalI and XmaI restriction sites. All ligation reactions were transformed into DH5α competent *E. coli* cells. Plasmid DNA was isolated using the Qiaprep® Spin Miniprep Kit (Qiagen) according to manufacturer's instructions. To confirm positive clones, newly ligated vectors were test digested using strategically chosen restriction enzymes and were sequenced. DNA sequencing was performed by Edinburgh Genomics on BigDye® (Life Technologies) terminator sequencing reactions according to manufacturer's instructions.

2.3 RNA protocols

2.3.1 *RNA isolation*

RNA was isolated from yeast cells grown to mid-log phase using the RNeasy Mini- or Midi-Kits (Qiagen) and treated with DNase I from the RNase-free DNase Set (Qiagen) according to manufacturer's instructions. Depending on the application, RNA was quantified using a NanoDrop™ ND-2000c spectrophotometer (Thermo Scientific) or a Qubit™ fluorometer (Life Technologies) according to manufacturers' instructions.

2.3.2 *Northern analysis*

Northern analysis was performed to detect the size and abundance of RNA isolated from cells. All buffers for northern blotting were made fresh and autoclaved a day prior to use. Three volumes of denaturing RNA loading buffer (1X HEPES, 50% deionized formamide, BromoPhenol blue, ethidium bromide, 6% paraformaldehyde) was added to 10 μg of total RNA. Samples were denatured for 10 mins at 65°C. Denatured RNA was immediately transferred to cool on ice for 5-10 mins before loading on a Formaldehyde RNA gel (1% w/v agarose, 1X HEPES, 6% paraformaldehyde). RNA gels were run overnight for 16 hrs at 25 V followed by 1-2 hrs at 70 V the next morning. Gels were washed twice in dH_2O and imaged under UV-light. Gels were then soaked in 0.05 M NaOH for 20 mins, washed in dH_2O, and then soaked in 20x SSC (300 mM Na-Citrate pH 7.0, 3 M NaCl) for 40 mins. RNA was transferred overnight from gels onto nylon membrane (Hybond N, Amersham) by capillary action. The next day, membranes were quickly dried on Whatman filter paper and UV-crosslinked at 1200J with a Stratalinker® UV crosslinker (Stratagene). Crosslinked membranes were stored in the dark at room temperature for no more than two weeks before hybridization using UTP-[α^{32}P]-labelled RNA probes. To make RNA probes, DNA fragments specific to target transcripts were amplified from genomic DNA by PCR and gel-purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). The T7 promoter was equipped at the end of the DNA fragment using an oligonucleotide containing T7 promoter sequence at the 5'-end (TAATACGACTCACTATAGGGAGA). The T7 promoter containing PCR products were transcribed *in vitro* using the MaxiScript T7 Kit (Ambion) to produce UTP-[α^{32}P]-labelled RNA probes according to manufacturer's instructions. Unincorporated radionucleotides were removed using NucAway Spin columns (Life Technologies) according to manufacturer's instructions. The UTP-[α^{32}P]-labelled RNA probes were

hybridized to membranes overnight in church buffer (0.5 M Na_2HPO_4 pH 7.2, 1 mM EDTA, 7% SDS) at 68°C in a rotating oven. Hybridized membranes were washed twice in a pre-warmed buffer containing 2x SSC and 0.1% SDS for 30 mins at 68°C followed by two washes in a buffer containing 0.5x SSC and 0.1% SDS for 15 mins at 68°C. To detect transcripts, northern blots were analyzed after 1-2 days of exposure on a Phosphor Screen (Molecular Dynamics) using a Typhoon Phosphorimager (GE Healthcare Life Sciences).

2.3.3 *Quantitative reverse-transcriptase PCR (RT-qPCR)*

The relative abundance of specific RNA transcripts was quantified by RT-qPCR. First strand complimentary DNA (cDNA) synthesis was performed on 1 μg of Turbo™ DNase (Life Technologies) treated RNA using random hexamers and SuperScript® III reverse transcriptase (Invitrogen) according to manufacturer's instructions. Negative controls lacking the reverse transcriptase enzyme (-RT) were performed alongside all RT-qPCR experiments. cDNAs were diluted 1/20 with dH_2O. Quantitative analysis was performed by qPCR. For RT-qPCR experiments, all transcript levels were calculated by normalizing the product of interest to an internal reference gene mRNA (the highly transcribed housekeeping gene actin: *act1*$^+$) and expressed relative to levels detected in wild-type cells grown under normal conditions. The expression levels for each transcript of interest in different physiological conditions and/or mutant cells were expressed relative to the levels detected in wild-type cells.

2.3.4 *5'-RACE PCR*

Transcription start sites were mapped using the SMARTer® RACE cDNA Amplification Kit (Clontech) according to manufacturer's instructions. 5'-RACE-PCR

was performed on 1 μg of total DNase-treated RNA. Primers to the actin *act1*$^{+}$ gene were used as a positive control for these experiments. RACE PCR reactions were run on a 1% agarose gel containing ethidium bromide and imaged under UV-light. 5'-RACE fragments were excised and gel-purified using Wizard® SV Gel and PCR Clean-Up System (Promega) and cloned into a linearized pRACE vector using In-Fusion® HD (Clontech). Positive colonies were selected and plasmid DNA was isolated by plasmid miniprep. Plasmids containing 5'-RACE products were sequenced by Edinburgh Genomics with BigDye® (Life Technologies) terminator sequencing reactions according to manufacturer's instructions. Transcription start sites were detected as the first nucleotide following the known 5'-RACE adaptor sequence: 5'-AAGCAGTGGTATCAACGCAGAGTACATGGGG-3'.

2.3.5 *Strand-specific RNA sequencing library preparation*

Strand-specific RNA-seq libraries were made as follows. First, rRNA was depleted from total DNase-treated RNA using the Ribo-Zero-Magnetic Gold Kit (Yeast) (Epicentre-Illumina) according to the manufacturer's instructions. RT-qPCR control experiments confirmed that 95-99% of rRNA was removed using this methodology. 40 ng of rRNA-depleted RNA was fragmented by heating samples to 95°C in NEXTflex™ RNA Fragmentation Buffer (Bioo Scientific) for 10 mins. Samples were then immediately placed on ice. First strand reverse transcription and second strand synthesis reactions were performed using the NEXT-flex™ Rapid Directional mRNA-Seq Kit (Bioo Scientific), followed by end-repair, adenylation, and adapter ligation reactions following manufacturer's instructions. Directionality was achieved by the addition of deoxyuridine-trisphosphate (dUTP) during second strand synthesis step and subsequent cleavage of the uridine-containing strand by treatment of the sample with Uracil DNA Glycosylase for 30 mins at 37°C. Limited PCR amplification

(12-13 cycles) preceded PCR clean-up with Agencourt AMPure XP beads (Beckman Coulter). Whenever possible, reactions were perfomed in Eppendorf® RNA/DNA LoBind Microfuge Tubes (Sigma-Aldrich). RNA-seq libraries were quantified using a 2100 Bioanalyzer Instrument (Agilent Technologies), pooled to allow multiplexing (>5 ng in 25 μL), and shipped to either the Beijing Genomics Institute (Beijing, China) or to Edinburgh Genomics (Edinburgh, UK) for Illumina-based sequencing. Dr. Pin Tong in the Allshire lab performed all bioinformatic analyses.

2.4 Protein protocols

2.4.1 *S. pombe protein extraction*

Protein samples were extracted from *S. pombe* cells as follows. 50 mL cultures of *S. pombe* were grown to mid-log phase and harvested at 3,000 RPM at 4°C. Cell pellets were resuspended in 0.5 mL 2x NuPAGE® LDS Sample Buffer (Life Technologies) containing freshly added 2 mM PMSF, protease inhibitor cocktail (Sigma-Aldrich), Bond-Breaker® TCEP Solution (Thermo Scientific) and lysed by bead beating. Samples were boiled for 5-10 minutes at 95°C then spun at 13,000 RPM for 1 min to collect whole cell protein extract from pelleted beads and cell debris.

2.4.2 *Western analysis*

Proteins were separated from whole cell protein extract by polyacrylamide gel electrophoresis (PAGE) and analyzed by western blotting as follows. Protein samples were loading into pre-prepared NuPAGE® Bis-Tris Mini Gels (Life Technologies) in an assembled Novex Mini-Cell apparatus (Life Technologies). Protein gels were run at 200 V for 60 mins in 1x NuPAGE® MES Running Buffer or

1x NuPAGE® MOPS Running Buffer (Life Technologies) depending on desired resolution of protein sizes, with the former being better for the separation of smaller proteins while the latter is better for larger proteins. Following PAGE, proteins were transferred from the polyacrylamide protein gel to a nitrocellulose membrane (Schleicher and Schuell) using the XCell™ Blot Module (Invitrogen). Transfers were carried out in 1x NuPAGE® Transfer Buffer (Life Technologies) containing 10% methanol for 1-2 hrs at 30 V. Membranes were stained with Ponceau S (Sigma-Aldrich) to confirm protein transfer and imaged for documentation. Membranes were then blocked in blocking buffer (3% milk powder in PBS-T) for 1 hr at room temperature. The primary antibody was added to blocking buffer and incubated with membranes overnight at 4°C. Primary antibodies used here for western blotting include anti-GFP (Roche) and anti-α-tubulin (Sigma-Aldrich). Membranes were washed three times in PBS-T (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 0.1% Tween-20, pH 7.4), each wash lasting 15-20 mins. A secondary horseradish peroxide (HRP) conjugated antibody, either anti-mouse or anti-rabbit depending on the source of primary antibody used, was added to blocking buffer and incubated with membranes for 1 hr at room temperate. Membranes were again washed in PBS-T as before and then rinsed twice in dH_2O. Proteins were detected using Enhanced Chemi-Luminescence (Amersham) according to manufacturer's instructions and exposed on BioMax® light film (Kodak) in a dark room. Films were developed and fixed using an SRX-101A Tabletop Processor (Konica Minolta) according to manufacturer's instructions.

2.4.3 *Chromatin immunoprecipitation (ChIP)*

Protein/DNA interactions were measured by performing ChIP experiments as follows. $5x10^8$ cells were grown to mid-log phase at 32°C in YES per sample, unless

indicated otherwise. Cells were fixed with 1% paraformaldehyde (PFA) for 15 min at room temperature. Fixed cells were centrifuged at 3,500 RPM for 2 mins, washed twice with ice-cold PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). Cell pellets were flash frozen in dry ice and stored at -80°C. For phosphate starvation experiments, cells were grown to mid-log phase in PMG medium with full supplements, washed twice with dH_2O, and then grown in PMG lacking phosphates (-PO_4) for 4 hrs before fixation. Cells were lysed by bead beating (Biospec Products) in 350 μL of ChIP lysis buffer (50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1% Triton-X-100, 0.1% Na-Deoxycholate, 1 mM EDTA, 1 mM PMSF, and yeast protease inhibitor cocktail). Crude whole cell extract was collected by puncturing small holes in the tube using flame-headed needle and centrifugation into a new microfuge tube at 1,000 RPM for 1 min. Crude whole cell extract was briefly vortexed and then sonicated using a Bioruptor (Diagenode) sonicator at 5°C on high for a total of 20 min (30 sec ON/OFF cycles). Insoluble material was removed by centrifugation at 13,500 RPM for 20 mins. 30 μL samples of whole cell extracts were collected as total input controls ("Input") and frozen at -20°C. Soluble lysates were pre-cleared with IgG Dynabeads® (Life Technologies) for 1 hour at 4°C and then incubated with appropriate antibody and IgG beads overnight at 4°C. 5 μL of Rpb1 antibody (#2629; Cell Signaling), 2 μL GFP antibody (G10362; Life Technologies), 2 μL H3 antibody (ab1791; Abcam), and 1 μL of H3K9me2 antibody (m5.1.1; Nakagawachi *et al.*, 2003) were used for IPs. IPs were washed for in ChIP lysis buffer for 1 min, followed by high salt ChIP lysis buffer (50 mM Hepes-KOH pH 7.5, 0.5 M NaCl, 1% Triton-X-100, 0.1% Na-Deoxycholate, 1 mM EDTA) for 10 min, wash buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% Na-Deoxycholate, 1 mM EDTA) for 10 mins, and twice with TE for 5 mins. Beads following IP and 10 μL of Input samples were incubated with 100 μL of 1% Chelex®

100 Resin (Bio-Rad) in dH_2O, boiled to remove DNA-protein crosslinks for 12 minutes, and then treated with proteinase K (10 mg/mL) for 30 mins at 55°C. Samples were boiled for an additional 10 mins to denature proteinase K. 60 μL of supernatant was carefully pipetted using duckbilled pipettes into new microfuge tubes. Quantitative analysis was performed by qPCR on diluted samples. For ChIP analysis, Input DNA samples were diluted 1/60 in dH_2O while IP DNA samples were diluted 1/20. ChIP enrichments were calculated as the ratio of product of interest from IP sample normalized to the corresponding input sample and expressed as "%IP".

2.4.4 *ChIP-seq library preparation*

Genome-wide histone H3 lysine 9 methylation patterns were mapped by ChIP-seq. $1.25x10^9$ cells were fixed for 15 mins in 1% PFA and lysed in 1 mL ChIP Lysis Buffer by bead beating. Crude whole cell extract was collected by puncturing small holes in the tube using flame-headed needle and centrifugation into a new microfuge tube at 1,000 RPM for 1 min. Crude whole cell extract was briefly vortexed and then sonicated using a Bioruptor (Diagenode) sonicator at 5°C on high for a total of 20 min (30 sec ON/OFF cycles). Insoluble material was removed by centrifugation at 13,500 RPM for 20 mins. 100 μL samples of whole cell extracts were collected as total input control ("Input") and frozen at -20°C. 1 mL of soluble lysate was incubated overnight with 100 μL IgG Dynabeads® (Life Technologies) and 3 μL of H3K9me2 antibody (m5.1.1; Nakagawachi *et al.*, 2003) at 4°C. IPs were washed for in ChIP lysis buffer for 10 min, followed by high salt ChIP lysis buffer for 10 min, ChIP wash buffer for 10 mins, and twice with TE for 5 mins. Washed beads were resuspended in 200 μL ChIP Elution Buffer (10 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM EDTA, 1% SDS) and incubated overnight at 65°C to reverse crosslinks. For input controls,

200 μL of 1.5x ChIP elution buffer was added to the 100 μL input samples. Following reverse crosslinking, samples were cooled to 37°C and treated with 1 μL RNase A (Qiagen) for 1 hour before treatment with 30 μL of proteinase K (10 mg/mL) at 55°C for 2 hrs. Samples were collected and an additional 100 μL of pre-warmed ChIP elution buffer was added to beads for 15 mins. First and second elutions from beads were pooled and DNA was purified using the Qiagen PCR cleanup kit (Qiagen). Recovered DNA concentrations were measured using a Qubit™ fluorometer (Life Technologies) according to manufacturers' instructions. H3K9me2 enrichments were validated by qPCR. Illumina libraries were prepared using the TruSeq Nano DNA kit (Illumina) according to manufacturer's instructions. Briefly, 5-20 ng of DNA were blunt ended for 45 min at room temperature. DNA was purified by 1.6:1 AMPure XP beads (Beckman Coulter). DNA was A-tailed using klenow (exo-) for 30 min at 37°C. The enzyme was heat inactivated at 75°C for 5 mins before samples were placed on ice. NEXTflex (Bio Scientific) adapters with internal barcodes were ligated for 15 min at room temperature and purified by 1:1 AMPure XP bead (Beckman Coulter) selection. Limited PCR amplification (12-13 cycles) preceded PCR clean-up with AMPure XP beads (Beckman Coulter). Whenever possible, reactions were perfomed in Eppendorf® RNA/DNA LoBind Microfuge Tubes (Sigma-Aldrich). ChIP-seq libraries were quantified using a 2100 Bioanalyzer Instrument (Agilent Technologies), pooled to allow multiplexing (>5 ng in 25 μL), and shipped to Edinburgh Genomics (Edinburgh, UK) for Illumina-based sequencing. Dr. Pin Tong in the Allshire lab performed all bioinformatic analyses.

2.4.5 *RNA immunoprecipitation (RIP)*

Protein/RNA interactions were measured by RIP. All RIP experiments in this thesis were performed using a Hisx6-TEV-Protein A-tagged Mmi1 (Mmi1-HTP) strain

alongside untagged wild-type cells as a negative control. Cells were fixed, lysed, and sonicated as per ChIP experiments, with the following modifications. All RIP buffers were made fresh, autoclaved the day prior to performing RIP experiments, and were all supplemented with freshly added RNase inhibitor RNasin® Plus (Promega) immediately prior to use. Cells were lysed in RIP lysis buffer (50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton-X-100, 0.1% Na-Deoxycholate). Mmi1-HTP was captured from cell lysate with IgG Dynabeads® (Life Technologies) for 2 hours at 4°C. Samples were washed at 4°C for 10 mins in RIP lysis buffer, followed by 10 mins in RIP wash buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% Na-Deoxycholate, 1 mM EDTA), and then for a final two 10 mins washes in TE. Samples were eluted with 100 μL of preheated elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) for 15 mins at 65°C and centrifuged at 13,000 RPM. Supernatant was transferred to new tubes while the remaining pellet was resuspended in pre-warmed AE buffer (50 mM NaOAc pH 5.2, 10 mM EDTA, 0.67% SDS), vortexed, and added to the previous supernatant. Crosslinks were reversed by incubating elutions at 65°C for 6 hrs, adding fresh RNasin® Plus after the first 3 hrs. Samples were then treated with 100 μg proteinase K (5 μL of 10 mg/mL stock) for 30 mins at 55°C. Mmi1-bound RNA was isolated by acid phenol-chloroform extraction and ethanol precipitation. Isolated RNA was treated with Turbo™ DNase (Life Technologies) according to manufacturer's instructions in order to remove any residual DNA contamination. RNA clean up was performed using acid phenol-chloroform and followed by ethanol precipitation. Isolated RNA was reverse-transcribed into cDNA using SuperScript® III reverse transcriptase (Invitrogen) according to manufacturer's instructions. Controls lacking reverse transcriptase (-RT) were performed alongside all RIP experiments. Quantitative analysis was performed by qPCR. RIP enrichments were calculated as

per ChIP analysis with two additional steps. First, all levels were reported as fold enrichment over levels detected using primer pairs to *act1*$^{+}$, a negative control for Mmi1-binding. Next, this value was normalized to the corresponding values detected in cells lacking the HTP-tagged Mmi1 in order to determine the fold enrichment over absolute background (signal noise).

2.5 Enzymatic assay

2.5.1 *Liquid assay for β-galactosidase activity*

Assays for *β-galactosidase* activity were performed as follows. Yeast cells transformed with vectors expressing the *lacZ* gene for *β-galactosidase* under the control of various promoters were grown to log phase in selective PMG media. Vectors expressing *lacZ* under the control of thiamine-repressible promoters of known strength (*nmt1*: strong; *nmt41*: medium; *nmt81*: weak) were used as controls for *β-galactosidase* activity in these experiments. Cells were resuspended in 1 mL of ice-cold Z buffer (0.06 M Na_2HPO_4, 0.04M NaH_2PO_4, 0.01M KCl, 0.001M $MgSO_4$). Before use, add fresh 0.03 M β-mercaptoethanol and permeabolized by adding 1-2 drops of 0.01% SDS and 1-2 drops of chloroform. Cell extracts were equilibrated at 30°C for 5 min before the addition of ortho-Nitrophenyl-β-galactoside (ONPG; 4 mg/mL, filter-sterilized and stored in the dark at 4°C), the colourimetric substrate for detection of β-galactosidase activity. The reaction was stopped with 0.5 mL of 1M Na_2CO_3 once the solution turned yellow and elapsed time was recorded. Cell debris was spun and the OD_{420} was measured on an Ultrospec 2100 pro spectrophotometer (GE Healthcare Life Sciences). Units were calculated as follows: Units/OD = 1000 x (OD_{420}/Volume x Time x OD_{595}). Note: yeast cells growing in log phase should have an OD_{595} of ~0.5.

2.6 Oligonucleotides and strains used in this thesis

Table 2.6.1 5'-RACE oligonucleotides

PRIMER NAME	SEQUENCE
act1_GSP1	CGGCGTTTTCAAGACCCAAAGCTGAGGG
act1_NGSP1	CTTCAGGGGCACGGAAACGCTCG
act1_GSP2	CATGCGTCTTGATCTCGCCGGTCGTGAC
act1_NGSP2	TGACTGACTACCTTATGAAGATTCTC
1271.09_GSP1	CCAGTAAGGCACCAGGAAGGTAGAAGG
1271.09_NGSP1	GCGGTAGAAGCATCGGCGGGTA
1271.09_GSP2	CCCTGTGTACGGGTGCTTACGGCTAC
1271.09_NGSP2	AGGGCAGTATCAATGGCATGCTTTC
nctgp1_GSP1	GTCCTACACATGAGGCAACCATGCCG
nctgp1_NGSP1	GAGGTAATAGAATTGGTTGAAGTAG
nctgp1_GSP2	GCCGTCCGTTGTTTGTCACCCTCAAC
nctgp1_NGSP2	ATATCGACTCCGTGACTGTCATG
nc1343_GSP1	CGAGACGGCTTTGAGGCAACCGGGAATG
nc1343_NGSP1	GAAAACAACACGGCAAGTCCTTGG
nc1343_GSP2	GCAAGTCTCAGGACGCCGCTCAAGCCG
nc1343_NGSP2	TGACATTGATTGCGTATAGAAGAG

Table 2.6.2 Primer pairs for northern probes

PRIMER	SEQUENCE
Nb103_F	TTTGTGTTGTGGTTTGTTCG
Nb103_R	TAATACGACTCACTATAGGGAGAAAGGATAACAATGCAGCCAAA
Nb214_F	GTGCAAATTGTTGGCTGAA
Nb214_R	TAATACGACTCACTATAGGGAGACGCAAAGAATCCAAGTTCAA
Nb388_F	TCCCTCATCATCCAATATGTTTC
Nb388_R	TAATACGACTCACTATAGGGAGAAATGATTATGCGGGTGTTGT
Nb808_F	TCCATGGAGTCTTTGGATTT
Nb808_R	TAATACGACTCACTATAGGGAGAGATGCCGCATAAAGTATTATTCA
Nb879_F	TTGTGATGCGTTGCAATATG
Nb879_R	TAATACGACTCACTATAGGGAGATCCTGTAAAGAATGCAAGCAAA
Nb1343_F	CAAACCAAACAGCAAAGCAA
Nb1343_R	TAATACGACTCACTATAGGGAGACATTGCAATTTCGCAACACT
Nb1443_F	TATTTGTTTGGCTTGCATGG
Nb1443_R	TAATACGACTCACTATAGGGAGATCCACGTGTTCTTGCAATTT
Nb1556_F	TTAACCTAAGGAAGTTTCCGAGT
Nb1556_R	TAATACGACTCACTATAGGGAGATAAAGTATGCAGCTGGAATCACA
Nbtgp1_F	ATGGTTACTGCTCCAATTCAATCGG
Nbtgp1_R	TAATACGACTCACTATAGGGAGAAATCAATGGCACCGTCCGTAAC
Nbnctg_F	ATGCATTCCATCATTCCTCCTTG
Nbnctg_R	TAATACGACTCACTATAGGGAGAACAAGATTGGTATGCATAGTCAGT
Nbpho1_F	GGTGGAAATGCTGCTTTCGA
Nbpho1_R	TAATACGACTCACTATAGGGAGAGGGTAGTGAAATCATCCGCG

Table 2.6.3 PCR oligonucleotides

PRIMER NAME	SEQUENCE
qAct1_F	GGTTTCGCTGGAGATGATG
qAct1_R	ATACCACGCTTGCTTTGAG
qSme2_F	AAACAAGGGAGGTAAACAGACTTAG
qSme2_R	GCATGCATATTCCGTCTTACAATAG
q1271.10c_F	CGCTTCGTATCTTTCTCTTTCC
q1271.10c_R	CCAGTCCTCTTCTTCGGTTGTA
q1271.09a_F (PP: 1)	TCGGTTGGAATGTTCTAATCAATAC
q1271.09a_R (PP: 1)	AGACCGGTGATCAAACAATATTTAG
q1271.09b_F (PP: 2)	TGAAGTAGTTAGACAGGTTAGCGA
q1271.09b_R (PP: 2)	CTTGTCGTCCAACTTCTCTTCATC
qnctgp1c_F (PP: 3)	GGCAGTAAATCTATCTGTAGCGAGT
qnctgp1c_R (PP: 3)	TACACGGTAAATGTCAAGTCTGCTA
qnctgp1b_F (PP: 4)	CTGACAAACCAATTATCCCTACACG
qnctgp1b_R (PP: 4)	GTATTACGATTTGGCAACCTCATCC
qnctgp1a_F (PP: 5)	TTAAATGCTGCACTCACATACTGAC
qnctgp1a_R (PP: 5)	ACTCTCCCTTGGGTTCATTTGATTA
qnc1343_F (PP: 6)	ATACAGACGTGTGGATTGCAA
qnc1343_R (PP: 6)	CCTCTTCTATACGCAATCAATGTC
q1271.08c_F	TTCAAGGAGCATTTCAATTCTAAAC
q1271.08c_R	TATGTATCGTTAGTTATGCCTCGTG
qMug96_F	CATCCTATGTTTATTTGTCTGTTGC
qMug96_R	CTCATGATGGTCCTTAAACCTATTG
qPho1a_F	CTTTGGACCCTCTAATACATCCGAT
qPho1a_R	AAGAGTGTCAAAGTTCTGGATACCA
qPho1b_F	AAGATTCTAAGTACTATGTCCGCCA
qPho1b_R	ATCGGATGTATTAGAGGGTCCAAAG
qncPho1a_F	ATGATGTTTGAGATTTACGGGAAGT
qncPho1a_R	TTCTGTAAATGTGTCCCGAACCAAA
qncPho1b_F	ATGATGTTTGAGATTTACGGGAAGT
qncPho1b_F	TTCTGTAAATGTGTCCCGAACCAAA
qDg_F	AATTGTGGTGGTGTGGTAATAC
qDg_R	GGGTTCATCGTTTCCATTCAG
lacZ_1_F	TACTACGTCGACCGACTGACCTCAAACCAAACAGCA
lacZ_1_R	TACTACCTGCAGTCACTAATGTCATACTCGGCTTGAG
lacZ_2_F	TACTACCTGCAGCGACTGACCTCAAACCAAACAGCA
lacZ_2_R	TACTACGTCGACTCACTAATGTCATACTCGGCTTGAG
nc-tgp1_SalI_F	TACTACGTCGACCATATCCAAATATGGAAACT
nc-tgp1_XmaI_R	TACTACCCCGGGCTGCCGACTTACAAGTCTCG
qncRNA214_F	GGTGCAGTGTACGTGAGTCTTCTG
qncRNA214_R	ATTCGTTGTGATCTGACAAGCACTTA
qncRNA338_F	TATTTCTACAATGGCACAGCTCACA
qncRNA338_R	ATGATAGCGAAGGGTCATGGTTATT
qncRNA808_F	CCTAATCAAGTGCTCTAACTCGC
qncRNA808_R	AATCTCAGAACAACATTCGACC
qncRNA879_F	TGCTCTTTGCTGTTCTTGTCCTTAT
qncRNA879_R	CCACGGTAAAACGGGTATAAAGAAAG
qncRNA1443_F	ACTTGCATTCTACTTCCTTGCATTG
qncRNA1443_R	GTGTTGGCAATTTCCACTGTAAAAC
qncRNA1556_F	GAAGCATATCGCTGTCAAGGTAGAA
qncRNA1556_R	GGATGTGCTTCGTGTTACTAGTTGG
rga7_F	AAATACCACTTCCTCTGATGATTTC
rga7_R	ATTTAGGATTGCTAGACCAAGTTCC
SPBC23G7.10c_F	TTAGTGGATAAGTTTGTTGTTGCTG
SPBC23G7.10c_R	TTGACGATATAAGATAACCATGAGC

Table 2.6.3 PCR oligonucleotides (cont.)

PRIMER NAME	SEQUENCE
rpc2_F	ATGGTAGATCCTTTACAAGTCGATG
rpc2_R	AGATCCTCAAACAATAATGCCAAC
SPAC24C9.06c_F	ATATTGAACGTGGTGTCTCTTACTTG
SPAC24C9.06c _R	GACATAAATTGAAGGATAGCCATTTG
bgs2_F	TGGACTGACTTCTTTGTTGATTACC
bgs2_R	CCTATCCATGAGTTTATATGAGAACGTG
isp4_F	AGCTTCCTCCCAGAATCATGTTT
isp4_R	CAATCCAATACGCCGATCTGAAC
rpl26_F	ATGAAGTTCTCCAGGGATGTCAC
rpl26_R	CATTAACACACGGCGTACAGATG
mrp7_F	CAAATGGATGTGTCTCTTCGCAA
mrp7_R	GTTCATCTTTGGGCCTTGGTAAC
lid2_F	TAATTCTGCATCGCTTTCTCTTAAC
lid2_R	AATTGTTAGTCTTCCCTCTGAATCG
sjact1F	GACTCTGGTCATGGTGTTACTCA
sjact1R	TCAAGTAATCGGTCAAGTCACGA
sj3644F1	TGGCACTTGTTACTGGCTCTATT
sj3644R1	ACAGTCTCCAAACAATCCGAAGA
sj3644F2	AGTTGCCTTAGTCTCTGATGGAT
sj3644R2	TACCCACATTCTTCAGAGCACTA
sj3644F3	AAGCAAACGCATATACAACACAGA
sj3644R3	GGATGACGTCTAGAGTATGCTGA
sj3644F4	CATGTCTCCTCTAACGTCTCAGG
sj3644R4	ATCAGACGAATTTGAAACGGTCG
sj0232F	CTTCACAAGTTTCTCGTTGCACT
sj0232R	TGGTGATCACTGAACCGATTGTA
sj5325F	CAACCATTCAGAGCTACGCAAAT
sj5325R	AATACTAATAACCGCGCCAATGG
soact1F	GTTGACTGAAGCTCCTTTGAACC
soact1R	GACGGCTTGAATGGAAACGTAAA
sotgp1F1	ATTGCATTGGGATATGTTCTGGG
sotgp1R1	GACAGCTCCCAAACCTATCGATA
sotgp1F2	AGCTTTATGTGGAAGATTTG
sotgp1R2	TCCACGGCACTAATTCATTACG
sotgp1F3	GGAATCGCACTCTTTGTTGC
sotgp1R3	TTCCGTAACCAGCCTCAATAC
sotgp1F4	GTATGCCTCTTCCGTATTCAGG
sotgp1R4	TTTACAAGCGCCGTGGTCATAG
so4583F	TCTTTGCTCTGGTGCTTATGG
so4583R	CTCCTAAGCCGATACCAAGG
scact1F	GACTGAAGCTCCAATGAACCCTA
scact1R	TAGAAGGCTGGAACGTTGAAAGT
scgit1F1	TTGGTGTTGGTGCAGAATATCCTA
scgit1R1	GCAAATTTGTCACCATAACCAGG
scgit1F2	GGTTTATCTGCTGTGACTGG
scgit1R2	GTCCATCTTGCACCCAAATTATC
scgit1F3	TGCTATCGTGTCATCATTTCGTG
scgit1R3	CTCATCGTCATGCTCTAATGTG
scgit1F4	ACAGCTGCCTACTCAATTACGG
scgit1R4	TTTCCTCATTTGTGATTTCTGTCG
pho5F	GGTATTTCTCGTGATTTGCCTG
pho5R	CCAGACTGACAGTAGGGTATC

Table 2.6.4 Strains used in this thesis

STRAIN	ID #	GENOTYPE	SOURCE
wild-type	1645	*h+ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	Lab stock
wild-type	1646	*h- ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	Lab stock
rrp6Δ	7865	*h+ rrp6Δ::kan ade6-210 ura4-D18 leu1-32*	Lab stock
ago1Δ	8061	*h+ ago1Δ::ura4 otr1R(SphI):ade6+ ura4-D18 leu1-32 ade6-M210*	Lab stock
dcr1Δ	8146	*h? dcr1Δ::KAN (G418R) ade6-210*	Lab stock
swi6Δ	951	*h90 swi6Δ::ura4 ura4-D18*	Lab stock
clr4Δ	8435	*h- clr4Δ::ura4 his7-366 ade6-210/216 leu1-32 ura4-D18*	Lab stock
dis3-54	A1264	*h+ dis3-54 ade6-216 leu1-32 arg3-D4*	Lab stock
1343Δ::ura4$^+$	A9016	*h+ SPNCRNA.1343Δ::ura4*$^+$*ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
1343Δ	A9032	*h+ SPNCRNA.1343Δ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-18*	This thesis
tgp1Δ1343Δ	A9352	*h+ 1343Δ tgp1Δ::ura4*$^+$ *ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
red1Δ	A9392	*h90 red1Δ::kan leu1-32 ura4-D18 ade6-M210*	Sugiyama, T.
red5-2	A9396	*h90 red5-2 ura4-D18 ade6-M210*	Sugiyama, T.
mmi1Δ	A9393	*h- mmi1Δ::kan leu1-32 mei4-P572*	Sugiyama, T.
AΔ	A9520	*h+ 1343AΔ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
BΔ	A9522	*h+ 1343BΔ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
nc-tgp1:ura4$^+$	A9523	*h+ nc-tgp1:ura4+ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
Pho7-GFP	A9827	*h- pho7-GFP:NAT ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis

Table 2.6.4 Strains used in this thesis (*cont.*)

STRAIN	ID #	GENOTYPE	SOURCE
Pho7-GFP/ 1343Δ::ura4+	A9974	*h- pho7-GFP:NAT1343Δ::ura4+ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
103Δ::ura4+	A9011	*h+ SPNCRNA.103Δ::ura4+ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
214Δ::ura4+	A9012	*h+ SPNCRNA.214Δ::ura4+ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
388Δ::ura4+	A9013	*h+ SPNCRNA.388Δ::ura4+ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
808Δ::ura4+	A9014	*h+ SPNCRNA.808Δ::ura4+ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
879Δ::ura4+	A9015	*h+ SPNCRNA.879Δ::ura4+ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
1443Δ::ura4+	A9017	*h+ SPNCRNA.1443Δ::ura4+ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
1556Δ::ura4+	A9018	*h+ SPNCRNA.1556Δ::ura4+ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
Mmi1-HTP	B0398	*h+ mmi1-his6-TEV-ProA::KAN MX imr1R (NcoI)::ura4+ ura4D-18 ade6-M216 leu1-32*	Vasilieva, L.
103Δ	A9027	*h+ SPNCRNA.103Δ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-18*	This thesis
214Δ	A9028	*h+ SPNCRNA.214Δ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-18*	This thesis
388Δ	A9029	*h+ SPNCRNA.388Δ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-18*	This thesis
808Δ	A9030	*h+ SPNCRNA.808Δ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-18*	This thesis
879Δ	A9031	*h+ SPNCRNA.879Δ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-18*	This thesis
1443Δ	A9033	*h+ SPNCRNA.1443Δ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-18*	This thesis
1556Δ	A9034	*h+ SPNCRNA.1556Δ ade6-210 arg3-D4 his3-D1 leu1-32 ura4-18*	This thesis
nmt1-nc-tgp1	B0200	*h- nc-tgp1-promoter:nmt1-NAT ade6-210 arg3-D4 his3-D1 leu1-32 ura4-D18*	This thesis
cnp1-1	6960	*h- leu1-32 ura4-D18 cnp1Δ::ura4 lys1::cnp1-1*	Lab stock

Table 2.6.4 Strains used in this thesis (*cont.*)

STRAIN	ID #	GENOTYPE	SOURCE
cdc25-22 H3.2-HA/T7	A9823	*h- ura4::[4xTetO-ade6] cdc25-22 ars1:prad15-cre-EBD-LEU2 ade6-210 leu1-32 his3D1 arg4-D4 H3.2-low-HA-hygR-lox-T7*	Lab stock
S. cerevisiae	BY4741	*S. cerevisiae*	Tollervey, D.
RRP6Δ (*S. cerevisiae*)	yaeh236	*S. cerevisaie RRP6Δ::NATMX6*	Tollervey, D.
S. japonicus	A1855	*h? S. japonicus*	Lab stock
S. octosporus	A6970	*h90 S. octosporus*	Lab stock
S. cryophilus	A6972	*h90 S. cryophilus*	Lab stock

CHAPTER 3

Identification and characterization of positionally conserved lncRNAs in fission yeast

3.1 Introduction

In 2002, *S. pombe* became the sixth eukaryotic organism to have its genome sequenced (Wood *et al.*, 2002). Since then, the genomes of many natural *S. pombe* isolates collected throughout the world have also been sequenced (Avelar *et al.*, 2013; Brown *et al.*, 2011; Fawcett *et al.*, 2014; Jeffares *et al.*, 2015), along with the genomes of three other known fission yeast species (Rhind *et al.*, 2011). Together these resources provide a powerful tool for studying the relationship between genotype and phenotype across the *Schizosaccharomyces* clade.

Genome-wide studies have predicted that the *S. pombe* genome encodes >1,500 putative lncRNAs (Wilhelm *et al.*, 2008; Rhind *et al.*, 2011). Consistent with observations from higher eukaryotes, the majority of these transcripts are expressed at very low levels, frequently below the level of one copy per cell (Marguerat *et al.*, 2012). Relatively low expression levels do not negate functionality, however. For example, antisense transcripts are generally present at very low levels, yet the act of transcribing an antisense lncRNA can compete with transcription on the sense strand, which regulates many meiotic and stress-response genes in *S. pombe* (Bitton *et al.*, 2011; Leong *et al.*, 2014). Antisense transcription also appears to be a regulatory feature of many meiotic and stress-response genes in related fission

yeast species (Rhind *et al.*, 2011), suggesting this mechanism of gene regulation is well conserved within the *Schizosaccharomyces* clade. Although the functional significance of antisense transcription is relatively well established in these species, not much is known about the biological importance of most intergenic lncRNAs present in fission yeast genomes.

While little functional information is available for most intergenic regions in *S. pombe*, the transcription of telomeric and subtelomeric lncRNAs in *S. pombe* is known to be important for maintaining telomere integrity (Bah *et al.*, 2012). This is consistent with findings in other eukaryotes where telomere transcription has also been demonstrated to play a role in maintaining chromosome stability (Azzalin and Lingner, 2015). Chromosome stability is also maintained by lncRNAs originating from repetitive sequences flanking centromeres in *S. pombe* since these transcripts are processed into siRNA by the RNAi machinery and target the H3K9 methyltransferase Clr4 to establish pericentromeric heterochromatin (Bayne et al., 2010; Motamedi et al., 2004; Verdel et al., 2004; Volpe et al., 2002) (**See Fig. 1.8**). The most distantly related species in the *Schizosaccharomyces* genus, *S. japonicus*, also appears to have a related siRNA-dependent mechanism for directing heterochromatin to the transposon-rich repeats that flank centromeres (Rhind *et al.*, 2011). The importance of the RNAi pathway in heterochromatin formation has yet to be explored in the more closely related fission yeast species *S. octosporus* and *S. cryophilus*. Thus further analyses are needed in order to conclude whether processing pericentric lncRNAs into siRNAs is a conserved regulatory mechanism for silencing centromeres in all fission yeast species.

Functionally characterized intergenic lncRNAs in *S. pombe* include lncRNAs that prevent the spreading of centromeric heterochromatin into adjacent euchromatin

(Keller *et al.*, 2013), the RNA component of the telomerase complex *TER1* (Leonardi *et al.*, 2008; Webb and Zakian, 2008), and an lncRNA transcribed from the *sme2*$^{+}$ locus that controls entry into meiosis (Yamashita *et al.*, 1998). The *sme2*$^{+}$ lncRNA interacts with the meiotic regulator Mei2 and another RNA-binding protein, Mmi1. Mmi1 selectively binds RNAs containing specific DSR (determinant of selective removal) motifs and recruits the nuclear exosome to eliminate such transcripts (Harigaya *et al.*, 2006). DSR motifs in the *sme2*$^{+}$ lncRNA act as decoys to sequester Mmi1, allowing meiotic DSR-containing meiotic transcripts to accumulate and initiate sexual differentiation (Shichino *et al.*, 2014; Yamashita *et al.*, 2012). Remarkably, the lncRNA product of the *sme2*$^{+}$ gene is also proposed to help mediate sister-chromatid pairing during meiosis (Ding *et al.*, 2012). This latter finding suggests ncRNA-dependent mechanisms may control pairing at other chromosomal locations and that such a model could apply to sister-chromatid pairing in other organisms as well.

Very few of the >500 lncRNAs annotated as “intergenic” in the *S. pombe* genome are conserved at the sequence level in three divergent *Schizosaccharomyces* species (Rhind *et al.*, 2011). In fact, even within natural isolates of *S. pombe*, intergenic lncRNA genes experience a great deal of sequence variation (Jeffaries *et al.*, 2015). Despite exhibiting little conservation at the nucleotide level, ~138 intergenic lncRNAs in *S. pombe* retain conserved gene order with putative lncRNAs in at least one other fission yeast species (Rhind *et al.*, 2011). Here, eight discrete intergenic lncRNAs that are positionally conserved in at least three of the four known *Schizosaccharomyces* species were chosen for further study. Two such lncRNAs in *S. pombe* include the *TER1* telomerase RNA and the *sme2*$^{+}$ lncRNA. The fact that these two functionally characterized intergenic lncRNAs met the above

criteria provides encouraging evidence that such an approach could, at least in theory, be useful for the identification of other functional lncRNAs in fission yeast.

3.2 Results

3.2.1 *Identifying syntenic intergenic lncRNAs in fission yeast*

Discrete intergenic lncRNA candidates were selected for further analysis based on conserved gene order. The rationale behind this approach is that lncRNAs maintained in syntenic regions across the *Schizosaccharomyces* genus might be conserved in function but not necessarily conserved in primary nucleotide sequence. For example, the functionally characterized telomerase RNA in *S. pombe*, *TER1*, is an intergenic lncRNA that shares conserved gene order with putative telomerase RNAs of roughly equivalent length, but no detectible sequence homology, in all known fission yeast species (**Fig. 3.1A**). Thus other functional lncRNAs might be conserved in a similar manner.

Despite an absence of sequence conservation for most lncRNAs, ~138 intergenic lncRNAs predicted to be encoded by the *S. pombe* genome reside in regions of conserved gene order with lncRNAs in at least one other *Schizosaccharomyces* species (Rhind *et al.*, 2011). However, the principal criterion for defining intergenic lncRNAs in *S. pombe* is that they do not overlap protein-coding genes (Wilhem *et al.*, 2008; Rhind *et al.*, 2011). This is problematic for two main reasons. First, intergenic lncRNAs overlapping the untranslated regions (UTR) of nearby protein-coding genes might simply be alternative UTRs themselves. For example, the *SPNCRNA.1551* locus is predicted to encode an intergenic lncRNA that is conserved in synteny and sequence in all known fission yeast species yet it

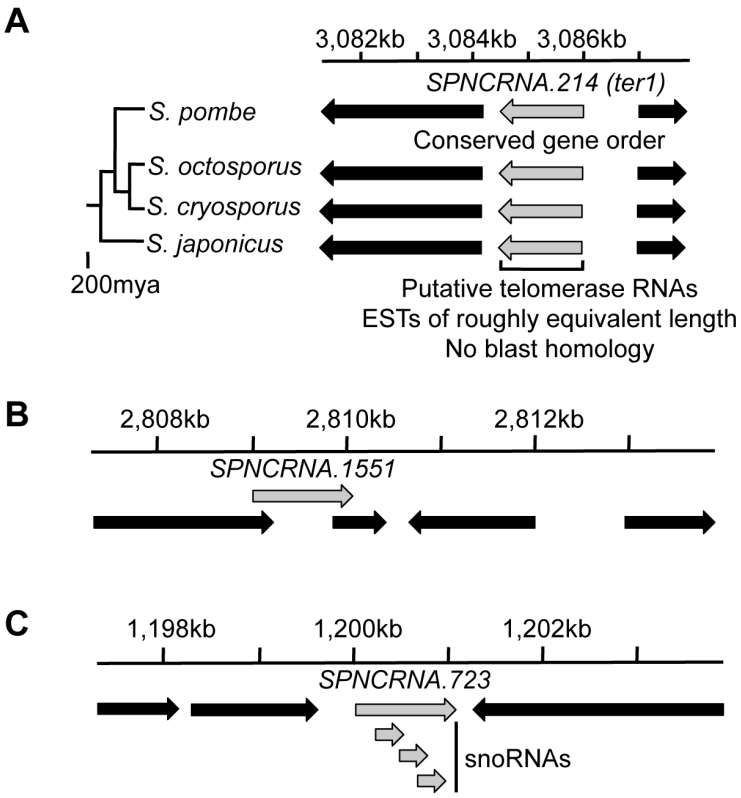

Figure 3.1 Conserved lncRNA positions. (A) Schematic representation of an *S. pombe* lncRNA (in this case: *SPNCRNA.214*, telomerase RNA or *ter1*$^{+}$) with conserved gene order (synteny) in related *Schizosaccharomycs* species. (B and C) Predicted intergenic lncRNA loci *SPNCRNA.1551* and *SPNCRNA.723* are depicted in order to illustrate the difficulty associated with making endogenous manipulations that do not alter other nearby transcripts and the possibility that some loci are misannotated. Note: black arrows represent protein-coding genes, while grey arrows represent non-coding genes.

overlaps the 5' UTR of the transcription factor TFIIH gene *tfb5*$^{+}$, which is an essential protein involved in RNAPII transcription (See **Fig. 3.1B**). Sequence conservation in this region is more likely due to the fact that this is also the site of the *tfb5*$^{+}$ promoter. The annotation of this transcript as an intergenic lncRNA is questionable. Notably, one third of all intergenic lncRNAs in *S. pombe* overlap adjacent protein-coding gene UTRs and therefore might not in fact encode intergenic products (Rhind *et al.*, 2011). Second, intergenic lncRNAs might have non-coding genes (e.g. snoRNAs, tRNAs, etc.) embedded within or antisense to the annotated locus. For example, the *SPNCRNA.723* locus is conserved in gene order and annotated as an intergenic lncRNA even though there are three snoRNAs within the gene: snR41, snR70, and snR51b (**Fig. 3.1C**). It is possible that an lncRNA transcribed from this locus is merely a precursor for snoRNA biogenesis and therefore not a *bone fide* functional lncRNA. Alternatively, sequencing reads over the snRNAs may have resulted in the misannotation of this locus. Another example of an annotated locus that overlaps a different ncRNA gene is *SPNCRNA.1366*, which exhibits synteny and sequence conservation with a putative lncRNA homolog in *S. cryophilus*. However, *SPNCRNA.1366* is antisense to the rRNA gene *SPRRNA.28*. In this case, it would be more appropriate for *SPNCRNA.1366* to be annotated as an antisense transcript. For these reasons, lncRNAs that overlap UTRs, other ncRNA genes, or simply reside too close to nearby genes to allow effective deletion were excluded from the list. These added criteria significantly reduced the number of available syntenic lncRNAs for further analysis (See **Table 3.2.1**).

Table 3.2.1 ***Candidate intergenic lncRNAs with conserved gene order/sequence***

INTERGENIC LNCRNA	GENE NAME	CH. & SIZE	CONSERVED GENE ORDER?	CONSERVED SEQUENCE?	COPIES / CELL**	RPKM wt, *dis3*
SPNCRNA.103	*sme2*$^+$	Ch. 2 667 nt	*S. cryophilus* *S. octosporus* -	No	0.11	1.1, 2.8
SPNCRNA.214	*ter1*$^+$	Ch. 1 1413 nt	*S. cryophilus* *S. octosporus* *S. japonicus*	No	1.6	19, 17
SPNCRNA.388	-	Ch. 2 1576 nt	*S. cryophilus* *S. octosporus* *S. japonicus*	Yes	1.2	17, 23
SPNCRNA.808	-	Ch. 1 290 nt	*S. cryophilus* *S. octosporus* -	Yes	60	100, 217
SPNCRNA.879	-	Ch. 2 1413 nt	*S. cryophilus* *S. octosporus* *S. japonicus*	Yes	0.087	0.61, 2.3
SPNCRNA.1343	-	Ch. 2 1543 nt	*S. cryophilus* *S. octosporus* *S. japonicus*	No	0.31	42, 30
SPNCRNA.1443	-	Ch. 2 2796 nt	*S. cryophilus* - *S. japonicus*	No	0.52	28, 24
SPNCRNA.1556	-	Ch. 2 458 nt	*S. cryophilus* *S. octosporus* -	No	0.11	30, 27

*CH.: Chromosome, **Marguerat *et al.*, 2012

Notably, the most promising eight lncRNAs candidates included two previously characterized lncRNAs discussed earlier: the telomerase RNA *TER1* and the *sme2*$^+$ lncRNA. The remaining six candidates have yet to be studied. In contrast to *TER1* and *sme2*$^+$, which are conserved only in gene order and not sequence, three of the genes (*SPNCRNA.388*, *SPNCRNA.808*, and *SPNCRNA.879*) are reported to have detectible levels of sequence conservation, in addition to conserved gene order, making them the most promising candidates for functional lncRNAs from the outset.

It is important to note that recent ribosome profiling analyses indicate that as many as a quarter of all transcripts annotated as non-coding in *S. pombe* interact with ribosomes (Duncan and Mata, 2014). While the interaction of an lncRNA with the ribosome is not direct evidence of active protein translation (Guttman *et al.*, 2013),

these analyses suggest it is possible that short ORFs within some lncRNAs might actually encode small protein products. Relevant to this study, ribosome profiling in *S. pombe* found that a 72 amino acid polypeptide might be translated from the *SPNCRNA.388* transcript, a 21 amino acid polypeptide might be translated from the *SPNCRNA.1343* transcript, and a 144 amino acid polypeptide might be translated from the *SPNCRNA.1443* (Duncan and Mata, 2014). Other annotated lncRNA loci studied here did not interact with ribosomes and are therefore likely to be truly non-coding.

3.2.2 *Initial characterization of candidate lncRNAs*

Previous genome-wide quantification of RNA levels in *S. pombe* showed that lncRNA abundance varies greatly from transcript to transcript (Marguerat *et al.*, 2012). Many of the syntenic lncRNA candidates chosen for further analysis are expressed at or below one copy per cell (**Table 3.2.1**). As mentioned above, low levels of expression do not rule out biological significance. Indeed, the *S. pombe* telomerase RNA *TER1* is present at levels only slightly above one copy per cell but is essential for telomerase function. In addition, the functionally characterized *sme2*$^{+}$ lncRNA (*SPNCRNA.103*) is present at levels far below one copy per cell, which is expected given that the *sme2*$^{+}$ lncRNA is rapidly targeted by Mmi1 for degradation by the nuclear exosome in vegetative cells (Yamashita *et al.*, 2012).

The highly conserved *SPNCRNA.808* gene produces an uncharacterized lncRNA that is unusually abundant at 60 copies per cell (Marguerat *et al.*, 2012). For comparison, the same study calculated that housekeeping genes actin (*act1*$^{+}$) and β-tubulin (*nda3*$^{+}$) are present at 180 and 25 copies per cell, respectively. To test whether the *SPNCRNA.808* transcript is regulated by the exosome, RNA levels

were quantified as reads per kilobase per million reads (RPKM) from published RNA-seq data in wild-type cells and cells with a cold-sensitive Dis3 mutation (*dis3-54*) (Choi *et al.*, 2011). This analysis revealed increased *SPNCRNA.808* transcript in *dis3-54* cells grown at the restrictive temperature (**See Table 3.2.1**). Northern analysis confirmed increased *SPNCRNA.808* transcript levels in *dis3-54* cells (**Fig. 3.2D**). Consistent with high levels of transcription, chromatin immunoprecipitation (ChIP) of Rpb1, the largest RNAPII subunit, detected roughly equivalent levels of RNAPII at the *SPNCRNA.808* gene and the highly expressed actin gene *act1*$^{+}$ (See **Fig. 3.4**). Together these findings suggest that the *SPNCRNA.808* gene is highly transcribed and that the levels of this lncRNA product are tightly controlled by the exosome.

Northern analysis of RNA isolated from asynchronous wild-type *S. pombe* cells detected a ~1.3 kb *TER1* transcript (*SPNCRNA.214*), consistent with previous reports (**Fig. 3.2B**; Leonardi *et al.*, 2013; Webb and Zakian, 2008). A ~1.2 kb *SPNCRNA.388* transcript (**Fig. 3.2C**), a ~0.9 kb *SPNCRNA.1343* transcript (**Fig. 3.2F**), and a ~1.2 kb *SPNCRNA.1443* transcript (**Fig. 3.2G**) were also detected. In contrast, northern analysis failed to detect transcripts corresponding to the *sme2*$^{+}$ (*SPNCRNA.103*), *SPNCRNA.879*, and *SPNCRNA.1556* genes (**Fig. 3.2A, 3.2E, and 3.2H**). However, a ~400 bp transcript corresponding to the *SPNCRNA.1556* gene was detected in cells with defective Dis3 activity, suggesting this transcript is actively degraded by the exosome in wild-type cells. In addition, losing Dis3 function clearly altered the size and abundance of the stable lncRNA transcribed from the *SPNCRNA.388* gene, suggesting the mature *SPNCRNA.388* transcript requires processing by the exosome. Although transcripts corresponding to the *sme2*$^{+}$ and *SPNCRNA.879* genes were not detected in Dis3 mutant cells by northern analysis,

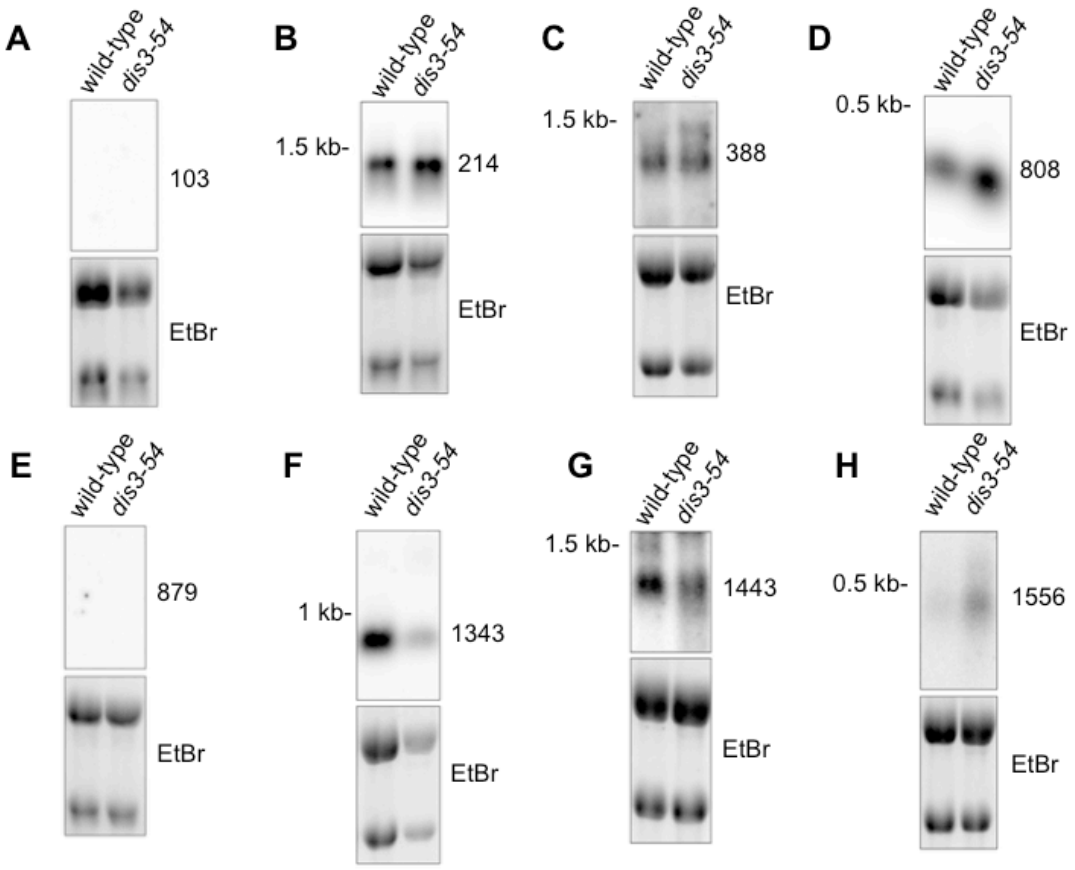

Figure 3.2. Analysis of *S. pombe* lncRNAs in wild-type and exosome-deficient cells. Northern blot analysis of transcripts encoded by lncRNA loci in wild-type and *dis3-54* cells.

RPKM quantification of RNA-seq experiments suggest transcripts at both loci increase modestly in the Dis3 mutant (**See Table 3.2.1**).

It is important to note that RPKM is not a robust quantification method (Wagner *et al.*, 2012). Instead, quantitative reverse-transcription polymerase chain reaction (RT-qPCR) experiments provide a much more accurate quantification of relative RNA levels and is much more sensitive to small changes in transcript abundance. For this reason, primer pairs were designed to lncRNA genes and RT-qPCR was performed to measure subtler changes in expression. RT-qPCR experiments revealed that the levels of the *sme2*$^{+}$ lncRNA increase slightly in Dis3 mutant cells but not nearly as much as in cells lacking Rrp6, the other catalytic subunit of the nuclear exosome complex (**Fig. 3.3A**). This result is consistent with a previous study reporting that Mmi1 preferentially targets the *sme2*$^{+}$ lncRNA for exosome degradation by Rrp6, not Dis3 (Chen *et al.*, 2011). In addition, RT-qPCR experiments revealed that the lncRNA encoded by *SPNCRNA.1343* accumulates exclusively in the absence of Rrp6, not in *dis3-54* cells (**Fig. 3.3F**). In contrast, transcripts encoded by *SPNCRNA.388*, *SPNCRNA.808*, and *SPNCRNA.1556* appear to be regulated by Dis3 and Rrp6 equally (i.e. both catalytic subunits of the nuclear exosome) (**Fig. 3.3C, 3.3D, and 3.3H**).

Although transcript levels from the *SPNCRNA.879* gene were below the level of detection by northern analysis, a small increase in transcript levels was detected in the Dis3 mutant by RT-qPCR (**Fig. 3.3E**). Unlike the *sme2*$^{+}$ lncRNA, which accumulates significantly in the absence of Rrp6, RNA levels from *SPNCRNA.879* increased relatively little in exosome-deficient cells. Rpb1 ChIP detected near

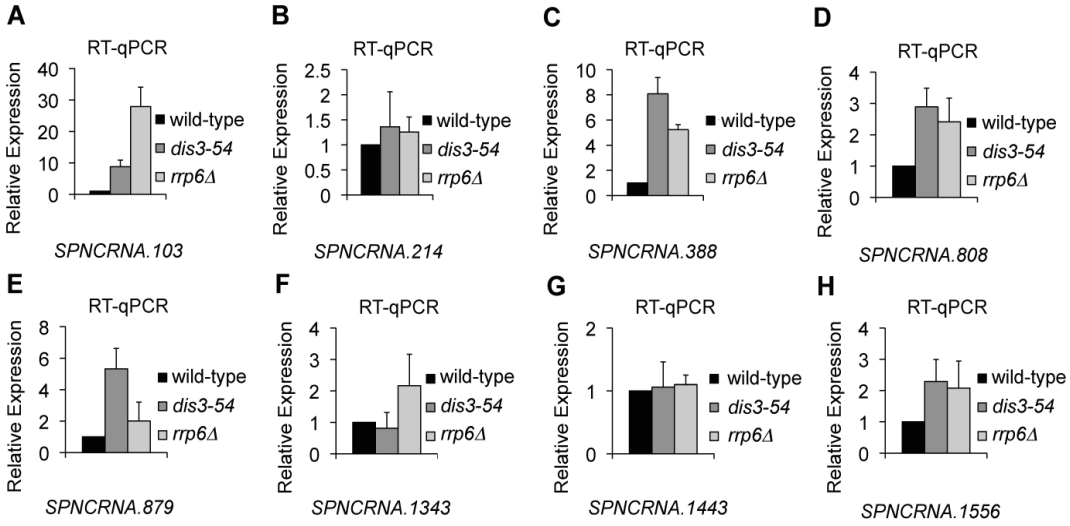

Figure 3.3. Quantitative analyses of lncRNA expression in exosome-deficient cells. (A-H) RT-qPCR experiments measuring lncRNA transcripts levels in wild-type and exosome-deficient cells (*dis3-54* and *rrp6Δ*). *dis3-54* cells were transferred to restrictive temperature (18°C) for 6 hours and RNA levels were normalized to those detected in wild-type cells grown in the same manner. Error bars represent SEM resulting from at least three independent replicates.

background levels of RNAPII at the *SPNCRNA.879* gene (**Fig. 3.4**), suggesting that the *SPNCRNA.879* gene is not actively transcribed in wild-type cells and thus not a significant target of the exosome. In contrast, RNAPII levels were detected above background at other lncRNA genes examined. Interestingly, the *SPNCRNA.879* gene is conserved at the sequence level with putative lncRNAs in all known species of the *Schizosaccharomyces* genus. Therefore, this transcript might only be produced in response to specific environmental or cellular conditions. Alternatively, this non-coding region might be the location of conserved DNA elements. Ultimately, it is unclear how the *SPNCRNA.879* gene was annotated as an intergenic lncRNA from RNA-seq datasets using asynchronous wild-type cultures since active transcription cannot be detected.

Mmi1 loss significantly induces many meiosis-specific genes in vegetative cells, including the *sme2*$^{+}$ lncRNA (Harigaya *et al.*, 2006). To test the possibility that other lncRNAs studied here are involved in meiosis, transcript levels in *mmi1Δ* cells were measured by RT-qPCR. *sme2*$^{+}$ lncRNA levels clearly accumulate in cells lacking Mmi1 (**Fig 3.5A**). Surprisingly, the relatively stable *SPNCRNA.388* lncRNA also accumulated roughly 3-fold in cells depleted of Mmi1 (**Fig. 3.5C**), suggesting it may at least partially be regulated by Mmi1-targeted degradation. A small increase in *SPNCRNA.879* transcript levels in cells lacking Mmi1 was also observed (**Fig. 3.5E**), yet the significance of this is unclear as this gene does not appear to be transcribed in wild-type cells. RNA immunoprecipitation (RIP) experiments using a strain containing an endogenously Hisx6-TEV-Protein A-tagged Mmi1 (Mmi1-HTP) revealed that Mmi1-HTP binds the *SPNCRNA.388* transcript, although this interaction was detected at very low levels compared to the interaction of Mmi1-HTP with the *sme2*$^{+}$ lncRNA (**Fig 3.5I and 3.5J**). Low levels of Mmi1-binding to the

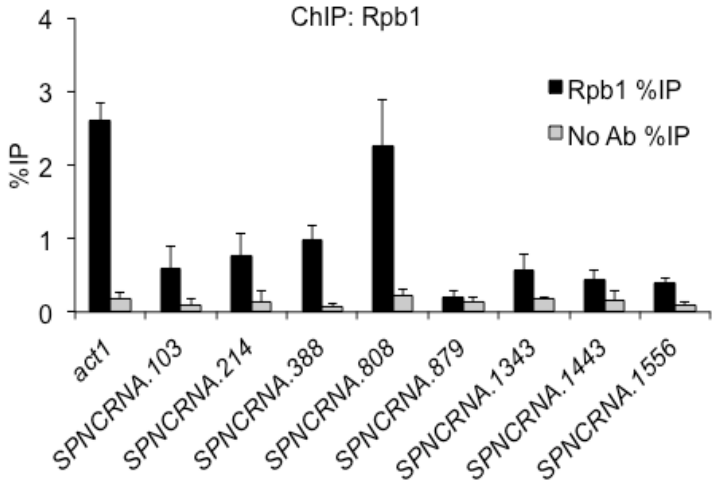

Figure 3.4. RNAPII occupancy at lncRNA genes. Rbp1 ChIP-qPCR analysis performed in wild-type cells. The housekeeping actin gene *act1*$^{+}$ is used as a positive control. No antibody represents negative control for these ChIP experiments. Error bars represent SEM resulting from at least three independent replicates.

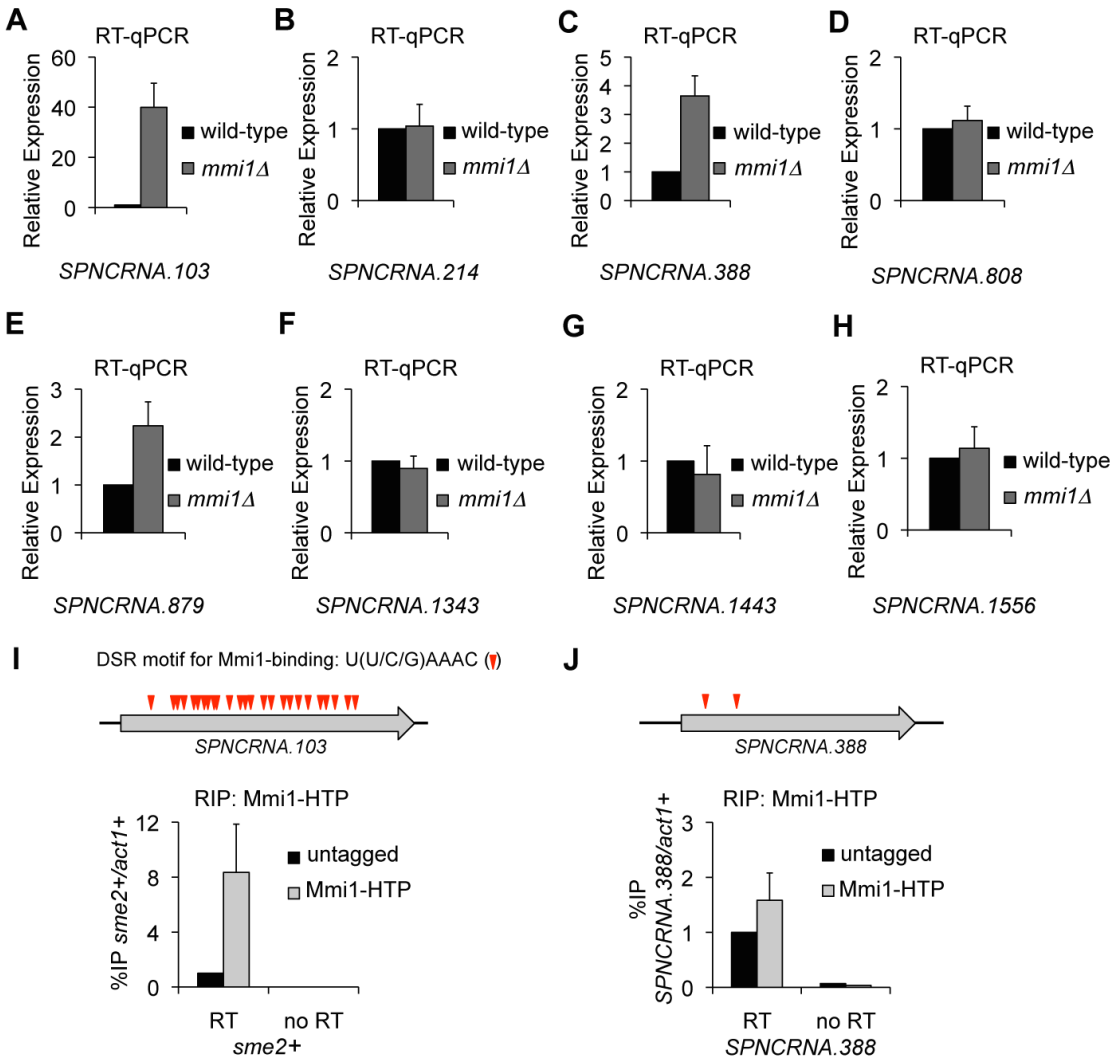

Figure 3.5. Analysis of lncRNA expression in cells lacking Mmi1. (A-H) RT-qPCR experiments measuring lncRNA transcripts levels in wild-type and *mmi1Δ* cells. Red triangles indicate predicted DSR motifs for Mmi1 binding in *SPNCRNA.103* and *SPNCRNA.388* loci. Error bars represent SEM resulting from at least three independent replicates. Mmi1-HTP RIP and quantification by RT-qPCR to detect binding of Mmi1 to lncRNAs encoded by (I) *SPNCRNA.103/sme2*$^{+}$ and (J) *SPNRNA.388* loci. Error bars represent standard deviation resulting from two independent experiments.

SPNCRNA.388 transcript are not surprising considering this lncRNA is relatively stable in wild-type cells. It is difficult to rule out the possibility that the *SPNCRNA.388* transcript is targeted for partial degradation simply because the locus contains two putative DSR motifs for Mmi1 binding. In contrast, the *sme2*$^{+}$ transcript is one of the primary targets of Mmi1 and contains over twenty DSR motifs (Shichino *et al.*, 2014). Since northern analysis clearly shows that the *SPNCRNA.388* transcript is processed by the exosome (**Fig. 3.2D**), Mmi1 might target processing activities over degradation, although such a role for Mmi1 has not yet been reported in the literature. Finally, other lncRNAs tested here did not show increased transcript levels in cells deleted for Mmi1, suggesting they are not targeted for degradation by this mechanism and thus unlikely to be involved in meiosis.

3.2.3 *Strategy for deleting lncRNA loci in S. pombe*

To assess cell viability following lncRNA loss, a *loxP* flanked *ura4*$^{+}$ cassette was integrated to replace candidate lncRNA genes (**Fig. 3.6A**). Positive integrations were confirmed by PCR amplification over new DNA junctions and by northern analysis to confirm transcript loss (**Fig. 3.6B and 3.6D**). *loxP* sites were recombined by exogenous over-expression of the Cre-Recombinase enzyme, which removed the *ura4*$^{+}$ marker leaving a short *loxP* footprint (**Fig. 3.6C**). Again, PCR amplification over new DNA junctions was performed, in addition to growing cells on synthetic medium lacking uracil, to confirm *ura4*$^{+}$ loss following *loxP* recombination (**Fig. 3.6D and 3.6E**). The benefit of this strategy is that *lncRNAΔ::ura4*$^{+}$ strains maintain active transcription at non-coding loci, as the act of transcription alone might serve a biological function, while *lncRNAΔ* strains represent full deletions of the annotated locus.

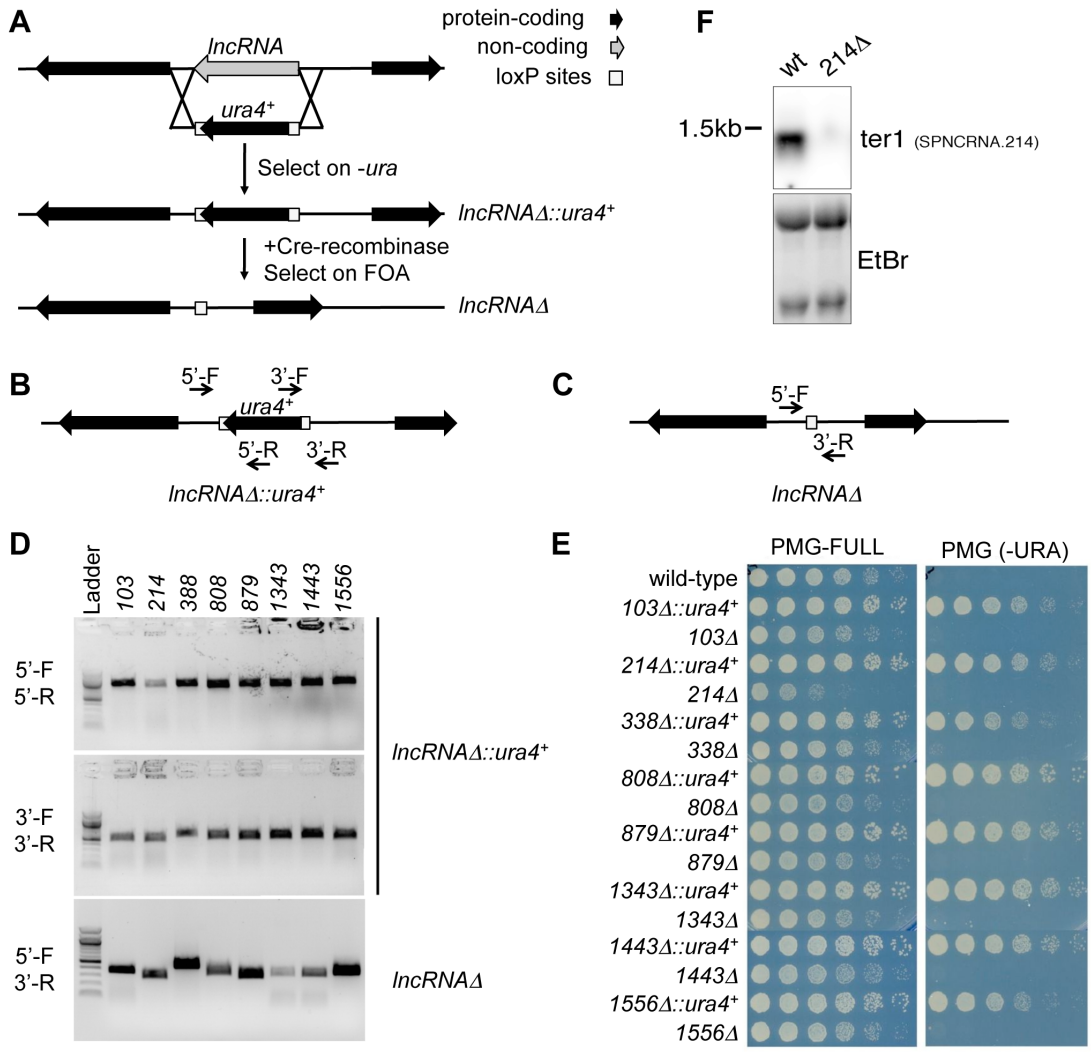

Figure 3.6. Strategy for deleting positionally conserved lncRNAs in *S. pombe*. (A) Schematic diagram of the strategy employed to delete lncRNAs in *S. pombe*. (B and C) The location of primer pairs to check new DNA junctions following the manipulation of lncRNA loci. (D) Colony PCRs run on 1.5% agarose gel electrophoresis to confirm correct genetic manipulations. (E) Serial dilutions of wild-type cells and lncRNA deletions were spotted on PMG medium with or without uracil present. (F) Northern analysis was performed to confirm lncRNA deletions (in this case the loss of telomerase RNA in *SPNCRNA.214Δ* cells).

3.2.4 *Assessing cell viability and growth following lncRNA deletions*

The cold temperature-sensitive *dis3-54* strain was grown alongside wild-type cells and lncRNA deleted cells as a control to assess possible growth abnormalities resulting from lncRNA loss. With the exception of cells lacking telomerase RNA *TER1* (*214Δ*), all lncRNA deletions were viable and grew similar to wild-type cells (**Fig 3.7A**). Together these findings suggest that even some of the most conserved lncRNAs predicted in the fission yeast clade are non-essential for normal cell growth and viability in *S. pombe*.

An increasing number of lncRNAs are thought to regulate gene expression in response to environmental changes and stress (Bitton *et al.*, 2011; Leong *et al.*, 2014). Given that lncRNAs might play more subtle roles in cells, lncRNA deleted cells were grown in the presence of the following stresses: temperature extremities, the microtubule destabilizing drug thiabendazole (TBZ), the DNA synthesis-inhibitor hydroxyurea (HU), UV-induced DNA damage, H_2O_2-induced oxidative stress, and caffeine, a potent inhibitor of cAMP phosphodiesterase. Only cells lacking the *SPNCRNA.1343* gene displayed a clear phenotype in these conditions: hypersensitivity to TBZ, HU, and caffeine, but not to temperature changes, UV-irradiation, or oxidative stress (**Fig. 3.7B and 3.7C**). Further characterization of the *SPNCRNA.1343* gene, and this drug sensitivity phenotype, make up the central focus of Chapter 4.

3.2.5 *Effects of lncRNA deletion on neighbouring gene expression*

Many lncRNAs have been demonstrated to regulate the expression of nearby genes in *cis* (Guil and Esteller, 2012). For this reason, the expression levels of protein-coding genes flanking lncRNA genes were measured before and after lncRNA

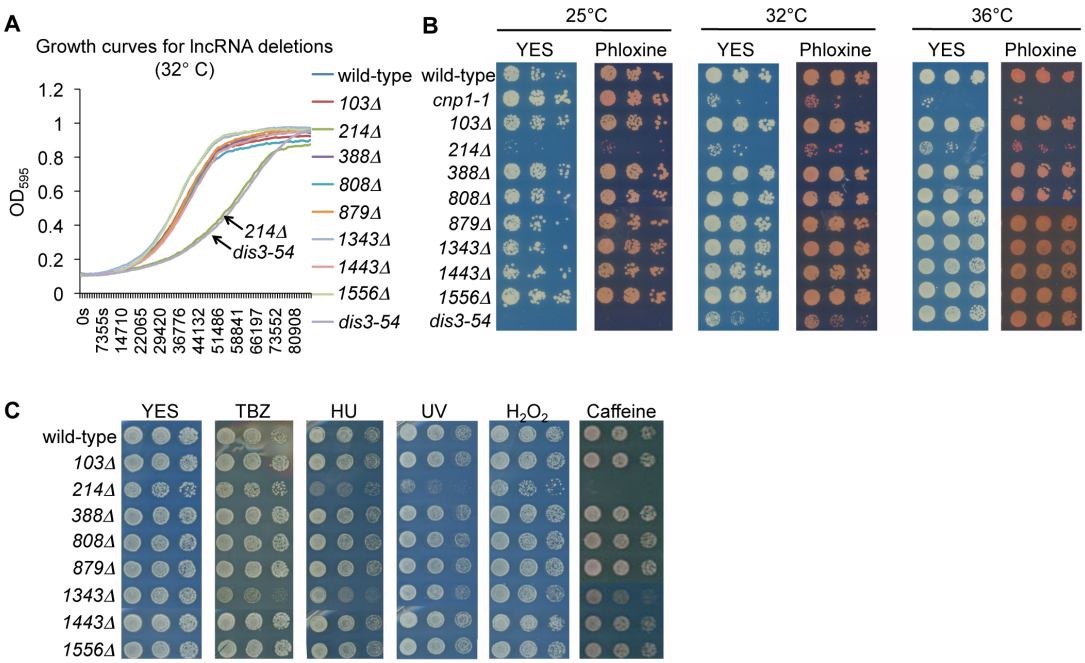

Figure 3.7. Deletion of *SPNCRNA.1343* results in sensitivity to multiple drugs. (A) Growth curves of cells deleted for positionally conserved lncRNAs grown in liquid media at 32°C. *dis3-54* is a cold-sensitive strain and control for perturbed growth. (B) Serial dilutions of lncRNA deletions were spotted on non-selective YES medium or on plates containing phloxine B, which indicates the proportion of dead cells (dark pink) in a colony. *cnp1-1* and *dis3-54* are hot and cold-sensitive control strains, respectively. (C) Serial dilutions of lncRNA deletions were spotted on non-selective YES medium or in the presence of various stresses, including exposure to the microtubule destabilizing drug thiabendazole (TBZ; 20 μg/mL), DNA synthesis inhibitor hydroxyurea (HU; 10 mM), UV-irradiation (80 J/m^2), oxidative stress (H_2O_2; 1 mM), or caffeine (15 mM).

deletion by RT-qPCR. Deleting *SPNCRNA.1343* caused the expression levels of an adjacent non-essential glycerophosphodiester permease gene *SPBC1271.09* to increase >50-fold (**Fig. 3.8D**). Remarkably, interrupting all other candidate lncRNA genes resulted in little to no change in the expression of flanking genes (**Fig. 3.8**). These results suggest that, with the exception of the *SPNCRNA.1343* gene, other lncRNA genes studied here do not regulate neighbouring gene expression.

3.2.6 *SPNCRNA.808 encodes a conserved and highly expressed lncRNA of unknown function*

The highly abundant ~290 bp RNA transcribed from the *SPNCRNA.808* locus shares conserved gene order and a great deal of sequence similarity with putative lncRNA homologs in related fission yeast species *S. octosporus* and *S. cryophilus*, but not the more distantly related *S. japonicus* species (Rhind *et al.*, 2011) (**Fig. 3.9A and 3.9B**). The fact that deleting *SPNCRNA.808* had no significant effect on the expression levels of neighbouring genes suggests that the RNA product of this gene does not act in *cis* (**Fig. 3.9B**). It is therefore possible that the *SPNCRNA.808* transcript could act to regulate genes in *trans*. However, before considering that possibility, one must rule out whether or not *SPNCRNA.808* actually encodes a short peptide from a predicted 49 amino acid ORF present in the gene sequence. Despite this possibility, previous genome-wide analyses found that the *SPNCRNA.808* transcript lacks a poly-A tail (Marguerat *et al.*, 2012), unusual for an mRNA. Furthermore, recent ribosome profiling analyses in *S. pombe* did not detect translation of the *SPNCRNA.808* transcript (Duncan and Mata, 2014). Taken together, these findings suggest that the *SPNCRNA.808* transcript is likely to be an lncRNA. As such, the *SPNCRNA.808* transcript is one of the most abundant and well-conserved lncRNAs in *S. pombe*, making it a great candidate for further analyses.

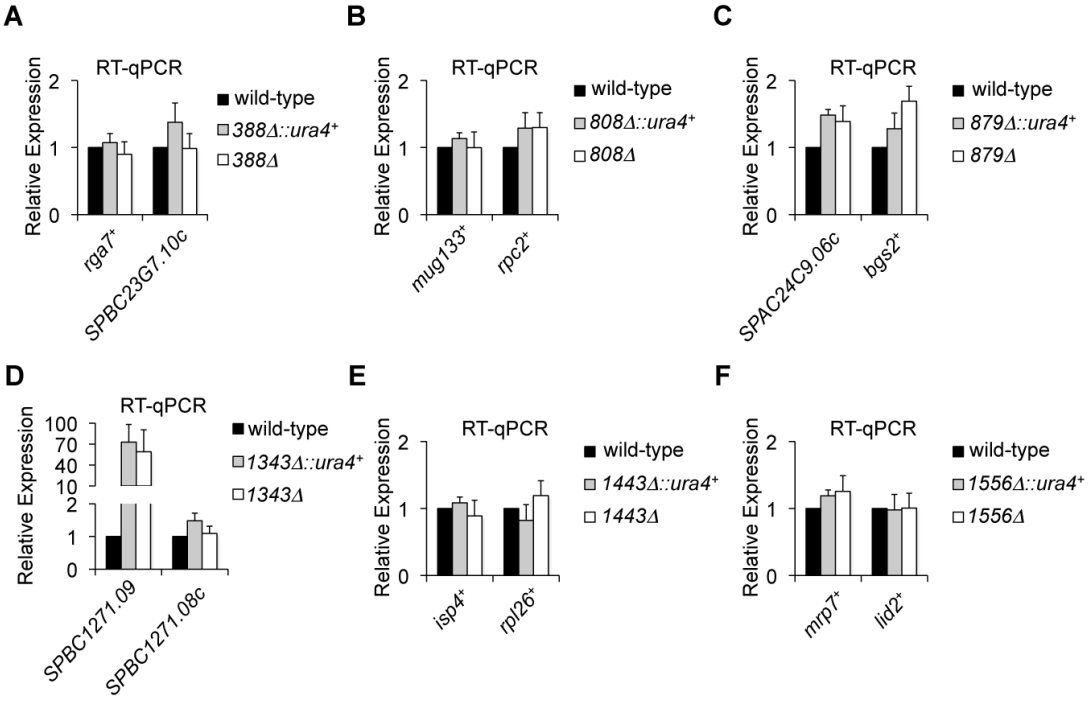

Figure 3.8. Deleting the *SPNCRNA.1343* gene induces the expression of a neighbouring permease-encoding gene. (A-F) The expression levels of adjacent genes were measured by RT-qPCR before and after lncRNA deletion. Error bars represent SEM resulting from at least three independent replicates.

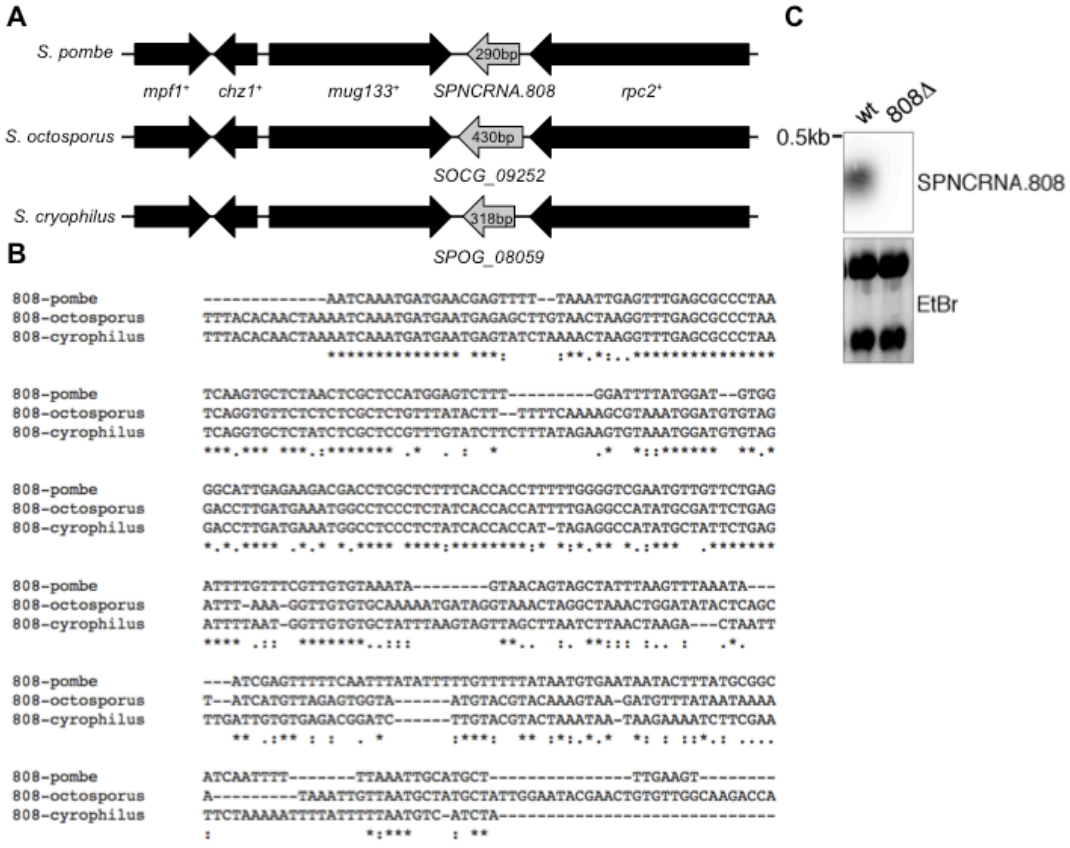

Figure 3.9. The *SPNCRNA.808* gene is highly conserved. (A) The *SPNCRNA.808* gene is conserved in position between three of the four known fission yeast species. (B) Primary sequence conservation detected between the *S. pombe SPNCRNA.808* gene and orthologs in *S. octosporus* and *S. cryophilus* using Clustal Omega software (Sievers *et al.*, 2011). (C) Northern analysis was performed to confirm *SPNCRNA.808* deletion.

As observed above, deleting *SPNCRNA.808* did not significantly disrupt cell growth or viability, and failed to reveal any detectable phenotype in the conditions tested (**Fig. 3.7**). However, it is important to note that many protein-coding gene deletions in *S. pombe* are also viable and show no overt phenotype. For this reason, cells lacking the *SPNCRNA.808* gene were screened against the Bioneer *S. pombe* Genome-Wide Deletion Mutant Library, which includes ~3,000 strains bearing single non-essential gene deletions. It is estimated that ~17.5% of the ~4,900 predicted protein-coding genes in the *S. pombe* genome are essential, leaving ~4,000 non-essential protein-coding genes. Thus, this version of the Bioneer Deletion Library covers roughly 75% of the non-essential genes in *S. pombe*. The purpose of this approach is to uncover possible genetic interactions that might provide functional evidence for the *SPNCRNA.808* lncRNA. However, the genetic screen did not show synthetic sickness or synthetic lethality when cells lacking the *SPNCRNA.808* gene were crossed into any strain in the library. In sum, the functional significance of this unusually well conserved and highly expressed lncRNA remains elusive.

3.3 Discussion

It is now clear that many eukaryotic genomes, from yeast to human, produce an abundance of lncRNAs. Growing interest has therefore been placed on understanding the functional significance of these transcripts. Here, eight discrete lncRNAs in *S. pombe* that show a conserved gene order in at least two of the other three known *Schizossacharomyces* species were selected for further analysis and characterization. While some of the transcripts were stably expressed in vegetative wild-type cells, the majority showed some degree of processing/degradation by the exosome complex. In general, these experiments validated lncRNA abundance

estimated from the quantification of RNA-seq experiments, while northern analysis revealed that annotated transcript size predicted from RNA-seq data was frequently incorrect. This latter finding highlights the shortcomings of genome-wide transcriptome profiling to provide detailed, locus-specific information and emphasizes the need for comprehensive analyses of lncRNAs in order to characterize individual transcripts.

Excluding the telomerase RNA control (*SPNCRNA.214*), lncRNA deletions performed here revealed that even some of the most conserved intergenic lncRNAs in *S. pombe* are not required for normal cell growth and viability. However, this does not rule out function. Indeed, the lncRNA product of the *sme2*$^{+}$ gene, which helps to mediate sister-chromatid pairing during meiosis, has negligible defects in chromosome pairing when deleted (Ding *et al.*, 2012). These results imply that redundant mechanisms likely overcome lncRNA loss in this case. Thus, lncRNAs might play subtler roles in cells. This appears to be the case for *SPNCRNA.1343*, which exhibits a definitive phenotype when cells lacking this gene are grown in the presence of various compounds. Therefore, other conditions need to be tested in order to identify phenotypes that might emerge following lncRNA loss.

Neighbouring gene expression levels were largely unaltered following lncRNA deletions, with the notable exception to this being *SPNCRNA.1343* loss. Deleting *SPNCRNA.1343* resulted in the strong induction of the nearby permease-encoding gene *SPBC1271.09*. It is unclear whether the increase in mRNA levels of this gene are a direct result of deleting the *SPNCRNA.1343* lncRNA itself or simply the consequence of manipulating this locus. An addition possibility is that a short 21 amino acid ORF in the *SPNCRNA.1343* transcript might be translated and account for this phenotype. However, ribosome profiling analyses suggest the probability

score for such a peptide is very low (Duncan and Mata, 2014). Moreover, the *SPNCRNA.1343* transcript is present at very low levels, well below one copy per cell (Marguerat *et al.*, 2012). Instead, it is far more plausible that the increased expression of the nearby permease-encoding gene is responsible for causing drug sensitivity in cells lacking the *SPNCRNA.1343* gene, rather than lncRNA loss itself. Such an explanation is wholly consistent with the observation that *1343Δ* cells are no more perturbed than wild-type cells following exposure to UV-irradiation, oxidative stress, or changes in temperature. Indeed, the phenotype was drug-specific. This finding suggests that expression of the *SPBC1271.09* permease gene in *1343Δ* cells might lead to greater drug uptake and account for cell death. The roles of *SPNCRNA.1343* and *SPBC1271.09* in regulating *S. pombe* drug tolerance are explored in Chapter 4.

Surprisingly, deleting the highly conserved *SPNCRNA.808* gene, which encodes one of the most abundantly expressed lncRNAs in *S. pombe*, failed to show any discernible phenotype in the conditions tested. Furthermore, deleting *SPNCRNA.808* had no detectible effect on nearby gene expression, nor did it reveal any synthetic phenotypes when crossed with a non-essential gene deletion library. While these results rule out numerous possible functions for the lncRNA produced from this gene (e.g. it does not appear to regulate nearby genes in *cis*), they provide no indication what its function might be. To explore the possibility that the *SPNCRNA.808* transcript might regulate gene expression in *trans*, genome-wide RNA levels should be measured by RNA-seq to compare expression levels in cells before and after *SPNCRNA.808* deletion. In order to determine what role(s) this transcript might play in cells, future studies should try to identify what cellular compartment this lncRNA localizes to and what proteins it binds. High expression

levels and sequence conservation suggest that this gene has undergone a great deal of selective pressure and is therefore likely to encode a functional transcript. Thus, further work is required to determine the function of this lncRNA and others produced from fission yeast genomes.

CHAPTER 4

lncRNA transcription over a permease gene promoter confers drug tolerance in fission yeast

4.1 Introduction

An increasing number of lncRNAs have been found to play central roles in the regulation of gene expression and diverse regulatory mechanisms have been attributed to these functions. Numerous lncRNAs have been proposed to directly/indirectly interact with and/or recruit chromatin-modifiers that alter chromatin status, while other lncRNAs are proposed to recruit transcriptional activators, repressors, or components of the transcription machinery itself (Geisler and Coller, 2013). Although there is evidence that some lncRNAs regulate distant loci in *trans*, lncRNAs more frequently influence nearby gene expression in *cis* (Guil and Esteller, 2012). In fact, the simple act of transcribing an lncRNA can have a significant impact on the expression of neighbouring genes by altering local chromatin accessibility to create environments that are either suitable or unsuitable for transcription initiation (Kornienko *et al.*, 2013). It is therefore paramount that rigorous *in vivo* manipulations of lncRNA loci are performed to determine whether the lncRNA product itself or merely the process of lncRNA transcription mediates any observed changes in gene regulation. In addition, experiments must be designed so as to distinguish *trans* from *cis* effects.

4.2 Results

4.2.1 *Drug sensitivity is a direct result of increased tgp1$^+$ levels in 1343Δ cells*

Deleting the putative lncRNA locus *ncRNA.1343* caused *S. pombe* cells to acquire hypersensitivity to growth in the presence of various compounds. Moreover, RT-qPCR experiments revealed replacing the *ncRNA.1343* gene with a *ura4*$^+$ marker gene (*1343Δ*::*ura4*$^+$) or outright deletion (*1343Δ*) induced expression of *tgp1*$^+$, a phosphate regulated permease gene ~2 kb upstream, while other nearby genes were unaffected by these manipulations (**Fig. 4.1A and 4.1B**). Northern analysis confirmed the *tgp1*$^+$ mRNA was indeed induced in *1343Δ* cells but not wild-type cells, both grown in the presence of phosphate (repressed condition) (**Fig. 4.1C**).

To determine whether the drug sensitivity phenotype observed in *1343Δ* cells directly resulted from increased *tgp1*$^+$ expression, the *tgp1*$^+$ gene was deleted in cells already lacking *ncRNA.1343* (*tgp1Δ1343Δ*). This manipulation restored TBZ, HU, and caffeine tolerance to levels comparable with wild-type cells (**Fig. 4.1D**). In conclusion, this finding reveals that increased *tgp1*$^+$ expression is indeed directly responsible for the drug sensitivity phenotype observed in cells lacking *ncRNA.1343*.

4.2.2 *Bidirectional lncRNA promoter upstream of tgp1$^+$*

Previous RNA-seq analyses identified a putative lncRNA transcribed in the sense orientation upstream of *tgp1*$^+$ in cells lacking Rrp6 and the Mmi1-associated factor Red1 (Lee *et al.*, 2013). RNA-seq analyses performed here also detect increased transcript levels upstream of *tgp1*$^+$ in *rrp6Δ* and *mmi1Δ* cells (**Fig. 4.2A**), suggesting this promoter region is actively transcribed in wild-type cells but the RNA product is sensitive to Mmi1-directed degradation by the nuclear exosome complex. Again

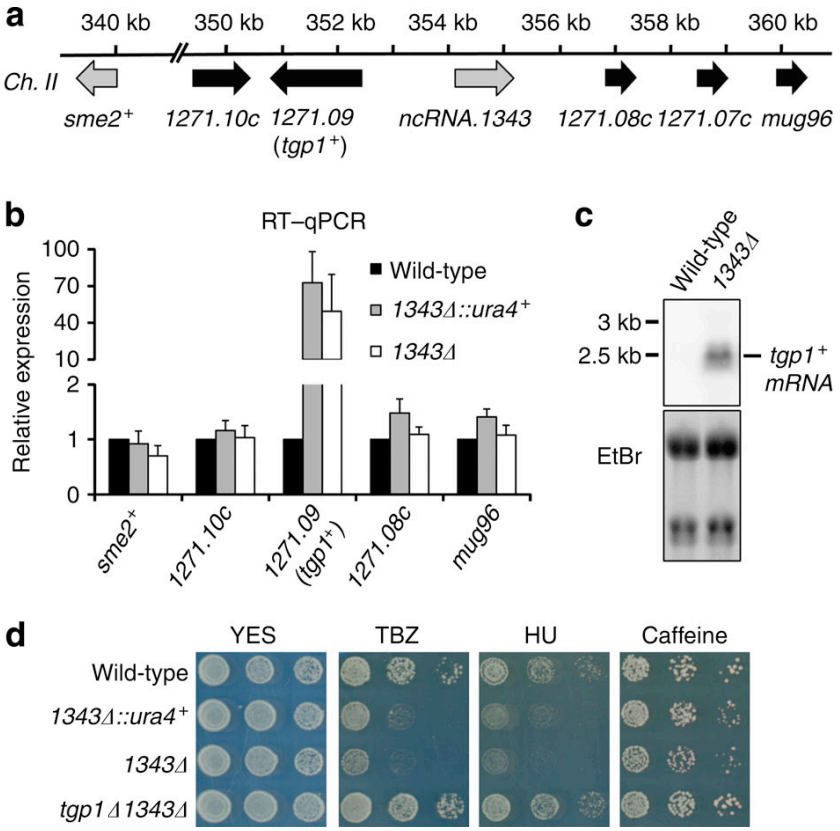

Figure 4.1. Drug sensitivity following *ncRNA.1343* deletion is due to increased *tgp1*$^{+}$ expression. (A) Schematic representation of genes flanking *ncRNA.1343*. (B) RT-qPCR experiments measured transcript levels for nearby gene in wild-type cells and following replacement of *ncRNA.1343* with *ura4*$^{+}$ (*1343Δ::ura4*$^{+}$) or deletion (*1343Δ*). Error bars represent SEM resulting from at least three independent replicates. (C) Northern analysis of *tgp1*$^{+}$ transcript levels in wild-type and *1343Δ* cells grown in the presence of phosphate. rRNA bands visualized by ethidium bromide (EtBr) represent controls for equal loading. (D) Serial dilutions of wild-type, *1343Δ::ura4*$^{+}$, *1343Δ*, and *tgp1Δ1343Δ* double mutant spotted on non-selective YES medium or in the presence of TBZ (20 μg/mL), HU (10 mM), or caffeine (15 mM).

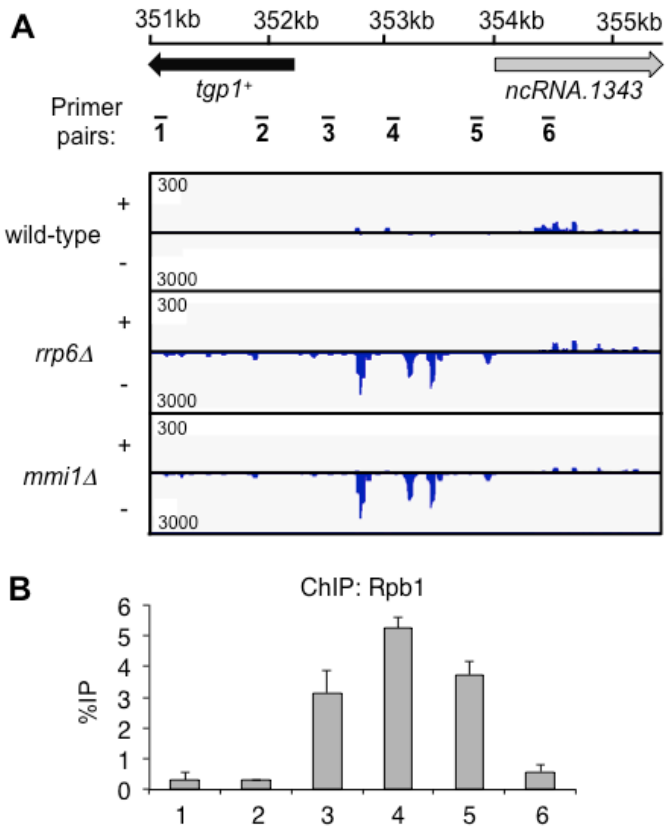

Figure 4.2. lncRNA transcription upstream of *tgp1*$^{+}$. (A) Strand-specific RNA-seq at the SPBC1271.09/*tgp1*$^{+}$ locus in wild-type, *rrp6Δ*, and *mmi1Δ* cells. Bioinformatic analyses performed by Pin Tong. Location of qPCR primer pairs are shown below. (B) Rbp1 ChIP-qPCR experiments performed in wild-type cells. Error bars represent SEM resulting from at least three independent replicates.

consistent with active RNAPII transcription in this region, Rpb1 ChIP analysis using primers spaced over the *tgp1*$^{+}$ gene and up to ~3 kb upstream revealed RNAPII enrichment between the *tgp1*$^{+}$ gene and *ncRNA.1343* in wild-type cells (**Fig. 4.2B**). 5'-Rapid amplification of cDNA ends (5'-RACE) experiments identified two divergent transcriptional start sites (TSS) arising within the *ncRNA.1343* locus: one RNA transcribed towards the *tgp1*$^{+}$ gene (*nc-tgp1*) and the other in the opposite orientation (*nc-1343*) (**Fig. 4.3A and 4.3B**). Unlike the region immediately upstream of the *nc-1343* TSS, a putative TATA box element is present ~25 bp upstream of the *nc-tgp1* TSS (**Fig. 4.3C and 4.3D**). In order to measure the strength of the bidirectional promoter positioned ~2 kb upstream of *tgp1*$^{+}$, the *ncRNA.1343* promoter was cloned (in both orientations) into a plasmid to control the expression of a *lacZ* reporter gene (**Fig. 4.3E**). *lacZ* reporter assays demonstrate that the bidirectional promoter drives stronger transcription in the *nc-tgp1* direction (**Fig. 4.3D**), consistent with Rpb1 ChIP experiments showing elevated RNAPII levels over the *nc-tgp1* transcription unit and much lower RNAPII levels present over the *nc-1343* transcription unit (**Fig. 4.2B**). In addition, a greater number of RNA-seq reads map to *nc-tgp1* in cells with defective Mmi1-targeted exosome degradation (**Fig. 4.2A**). Together these results support the conclusion that the *ncRNA.1343* bidirectional promoter primarily drives expression of the unstable *nc-tgp1* RNA.

Despite the detection of ample RNAPII occupancy over the *nc-tgp1* transcription unit in wild-type cells, previous RNA-seq analyses failed to annotate a transcript at this locus. The transcript corresponding to *nc-tgp1* can be detected in *rrp6Δ*, *mmi1Δ*, and *red1Δ* cells, but not in wild-type cells (**Fig. 4.2A**; Lee *et al.*, 2013), suggesting the *nc-tgp1* RNA is an unstable substrate of the nuclear exosome and that the Mmi1/Red1 pathway is involved in targeting it for degradation. Indeed, a consensus DSR motif

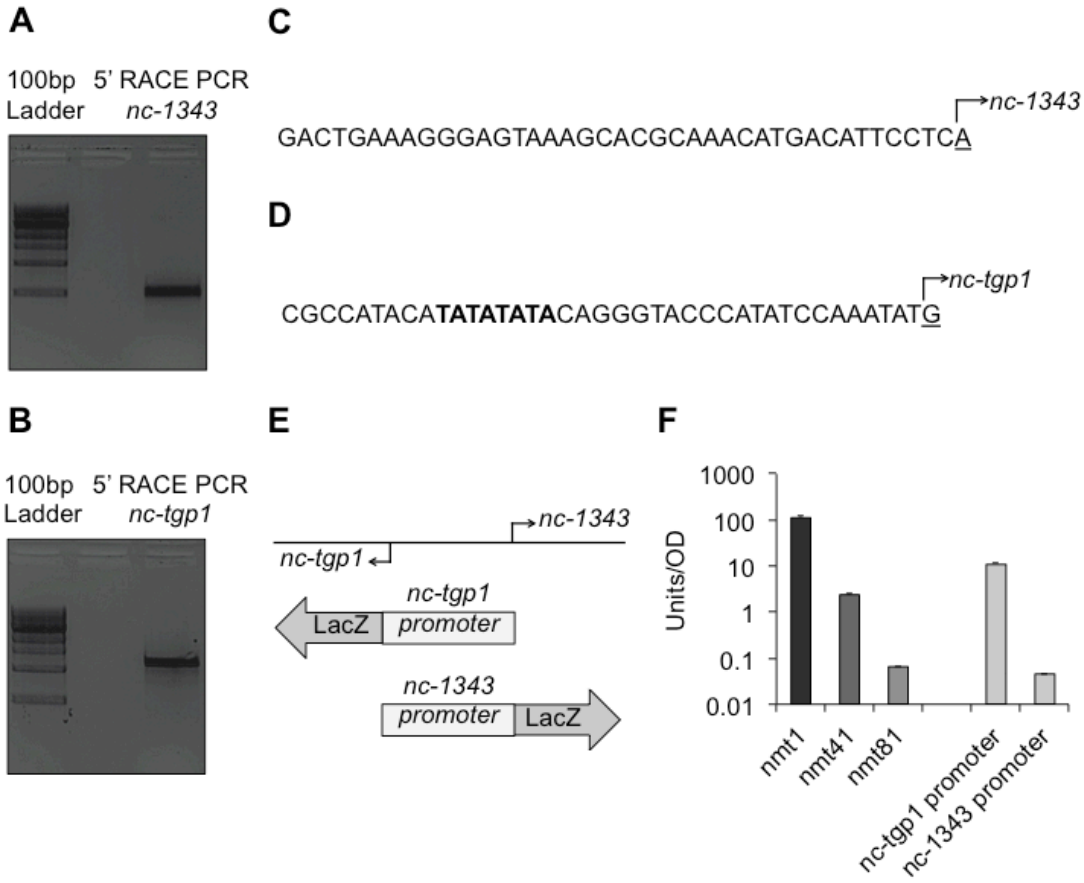

Figure 4.3. Two distinct lncRNAs are transcribed from a bidirectional promoter upstream of *tgp1*$^{+}$. (A and B) 5'-RACE PCR products for *nc-tgp1* and *nc-1343* RNAs. (C and D) TSSs (underlined) identified by sequencing 5'-RACE products for *nc-tgp1* and *nc-1343* RNAs. Bold letters indicate the position of a putative TATA box element 22-30 bp upstream of the *nc-tgp1* TSS. (E) Schematic representation of divergent transcription start sites in the *ncRNA.1343* locus and diagrams of the *LacZ* reporter gene under the control of this bidirectional promoter (in both orientations). (F) *β-galactosidase* assays from wild-type cells transformed with *LacZ* vectors. *nmt81*, *nmt41*, and *nmt1* are control promoters of increasing strength that drive *LacZ* expression. Error bars indicate standard deviation from three independent experiments.

for Mmi1 binding was detected at position +820 nt within the *nc-tgp1* transcript (**Fig. 4.4A**). RIP experiments confirmed a direct interaction between endogenously Hisx6-TEV-Protein A-tagged Mmi1 (Mmi1-HTP) and the *nc-tgp1* RNA (**Fig. 4.4B**). Consistent with these findings and RNA-seq data, northern analysis detected a ~1.9 kb *nc-tgp1* RNA in *rrp6Δ* and *mmi1Δ*, but not wild-type cells (**Fig. 4.5B**). This observation was confirmed by RT-qPCR, where increased *nc-tgp1* lncRNA levels were detected in cells lacking Rrp6 and Mmi1, and to a lesser extent in cells lacking Mmi1-associated factors Red1 and Red5 (**Fig. 4.5B and 4.5C**). Additionally, loss of Dis3 function failed to induce a significant increase in *nc-tgp1* levels, consistent with the observation that the majority of Mmi1 targets are preferentially degraded by the Rrp6 subunit of the nuclear exosome, not Dis3 (Chen *et al.*, 2011; Hiriart *et al.*, 2012). More recent genome-wide profiling of Mmi1 binding also detected direct binding between Mmi1 and DSR motifs in the *nc-tgp1* transcript (Kilchert *et al.*, 2015). In contrast to Mmi1-directed degradation of the *nc-tgp1* RNA, a stable ~0.9 kb *nc-1343* transcript was readily detected in wild-type cells (**Fig 4.5D**). The size and levels of the *nc-1343* transcript increased in nuclear exosome defective *rrp6Δ* cells, but not cells lacking Mmi1, Red1, Red5 or Dis3 (**Fig. 4.5D and Fig. 4.5E**). In sum, both *nc-1343* and *nc-tgp1* transcripts are processed by the exosome, but only *nc-tgp1* lncRNA is regulated by Mmi1-mediated recruitment of the nuclear exosome.

A moderate increase in *tgp1*$^{+}$ transcript levels has previously been reported in cells lacking Mmi1 (Hiriart *et al.*, 2012). In agreement with this, a similar increase (~4-fold) in *tgp1*$^{+}$ mRNA levels was detected in *mmi1Δ* and exosome (*rrp6Δ* or *dis3-54*) mutant cells by RT-qPCR (**Fig. 4.5G**). This increase, however, is significantly less than the >50-fold upregulation of *tgp1*$^{+}$ observed in *1343Δ* cells (**Fig. 4.5F and 4.5G**). Moreover, northern analysis failed to detect the *tgp1*$^{+}$ transcript in *rrp6Δ* cells

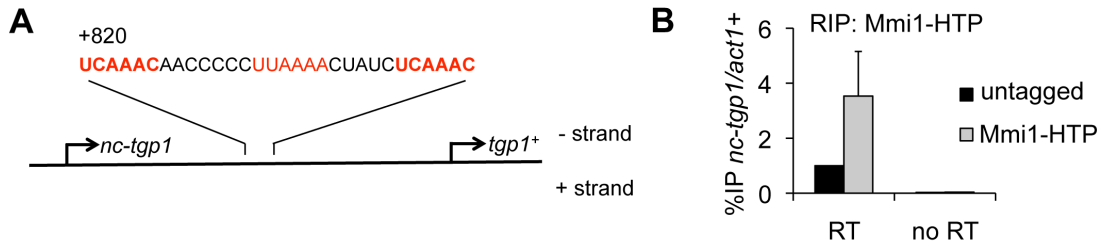

Figure 4.4. *nc-tgp1* lncRNA contains putative DSR sites for Mmi1-binding. (A) Schematic showing three putative DSR elements (two canonical: bold red text; one suboptimal: red text) embedded within the *nc-tgp1* transcription unit. (B) Mmi1-HTP RIP and quantification by RT-qPCR for *nc-tgp1* binding. Error bars indicate standard error from two independent experiments.

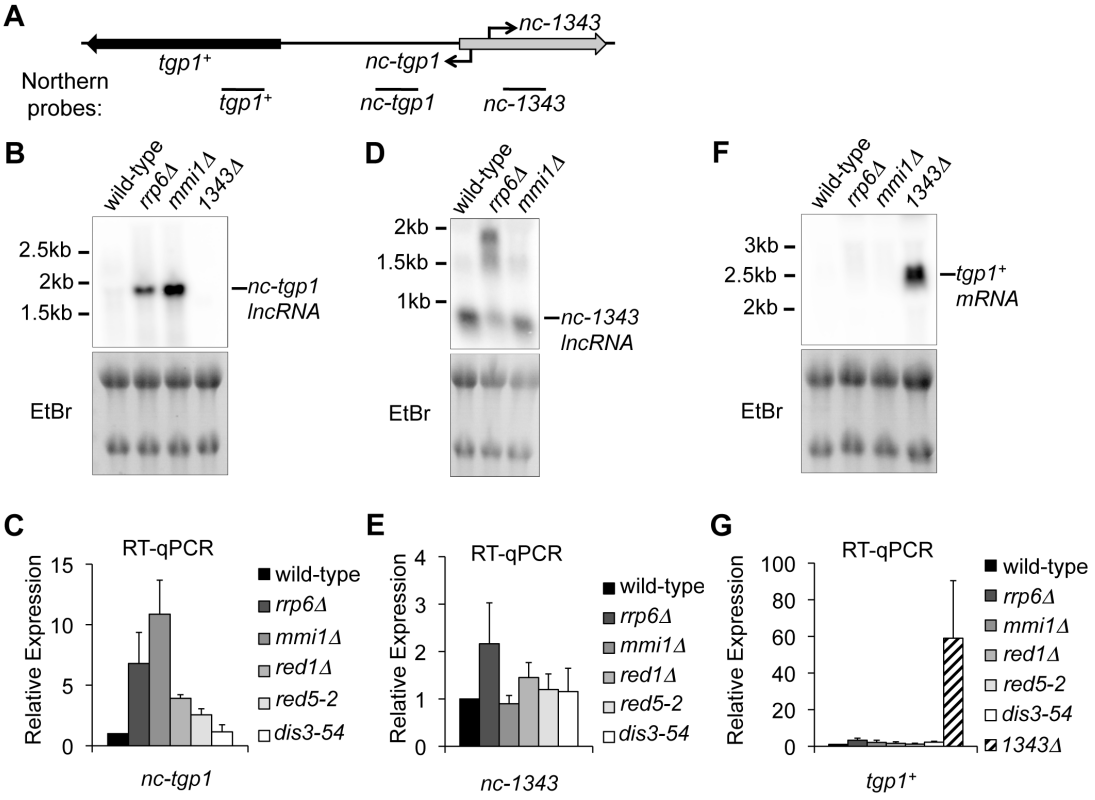

Figure 4.5. *nc-tgp1* is targeted for exosome-mediated degradation by Mmi1. (A) Schematic representation of the *tgp1*$^+$ locus, including the sites of northern probes. Northern analysis of (B) *nc-tgp1*, (D) *nc-1343* and (F) *tgp1*$^+$ transcript levels in wild-type, *rrp6Δ*, *mmi1Δ*, and *1343Δ*. rRNA bands visualized by ethidium bromide (EtBr) represent controls for equal loading. RT-qPCR experiments measured (C) *nc-tgp1*, (E) *nc-1343* and (G) *tgp1*$^+$ transcript levels in wild-type, *rrp6Δ*, *mmi1Δ*, *red1Δ*, *red5-2*, *dis3-54* and *1343Δ* cells using primer pairs 1 (*tgp1*$^+$), 5 (*nc-tgp1*), 6 (*nc-1343*) (See Figure 4.2). Error bars represent SEM resulting from at least three independent replicates.

or *mmi1Δ* cells, indicating the *tgp1*$^{+}$ gene is not induced in the absence of these factors. Thus, Mmi1-mediated Rrp6 degradation is not the predominant mechanism involved in directly silencing the *tgp1*$^{+}$ gene.

4.2.3 *tgp1*$^{+}$ *is repressed by nc-tgp1, not nc-1343*

The presence of the unstable *nc-tgp1* RNA upstream of *tgp1*$^{+}$ suggests that either *nc-tgp1*, *nc-1343*, or both, regulate *tgp1*$^{+}$ expression. To test the involvement of these lncRNAs in *tgp1*$^{+}$ regulation, a series of strategic genetic manipulations were performed (**Fig. 4.6A**). Truncations of *nc-1343* (i.e. *AΔ* and *BΔ*) that retain its 5' end did not result in the drug sensitivity phenotype observed in *1343Δ* cells (**Fig. 4.6B**) and, similarly, did not induce *tgp1*$^{+}$ expression (**Fig. 4.6C**). This indicates that full-length *nc-1343* is not required for *tgp1*$^{+}$ repression. We next tested if *nc-tgp1* is involved in repressing *tgp1*$^{+}$. 5'-RACE analysis shows that transcription of the *nc-tgp1* lncRNA starts within the *ncRNA.1343* transcription unit (**Fig. 4.3**), meaning that deletion of the entire locus (*1343Δ*) removes the *nc-tgp1* promoter, and the 5' end of its transcript, resulting in the observed loss of *nc-tgp1* expression (**Fig. 4.6C**). The *AΔ* and *BΔ* truncations of *nc-1343*, which retain the *nc-tgp1* promoter and TSS, do not affect *nc-tgp1* transcription or relieve *tgp1*$^{+}$ repression. In contrast, interrupting the *nc-tgp1* transcription unit by the insertion of the *ura4*$^{+}$ marker gene (*nc-tgp1:ura4*$^{+}$) after the TSS prevented *nc-tgp1* transcription elongation over the *tgp1*$^{+}$ promoter and induced *tgp1*$^{+}$ expression to levels equivalent to those observed in *1343Δ* levels, thereby increasing the sensitivity of these cells to TBZ, HU, and caffeine exposure (**Fig. 4.6B and 4.6C**). These analyses demonstrate that it is the rapidly degraded *nc-tgp1* lncRNA, not the stable *nc-1343* lncRNA, which is critical for repressing the *tgp1*$^{+}$ gene.

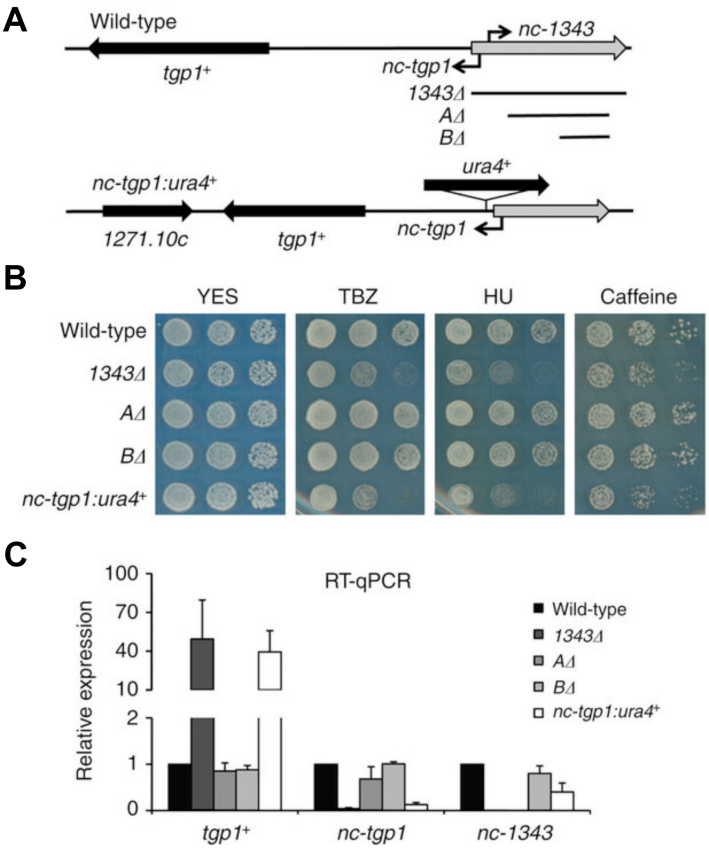

Figure 4.6. *nc-tgp1*, not *nc-1343*, represses *tgp1*$^+$ to confer drug tolerance. (A) Schematic diagram showing strategic manipulations of lncRNAs upstream of *tgp1*$^+$, including *1343Δ*, shorter deletions of *ncRNA.1343* (*AΔ* and *BΔ*), and *ura4+* integration within the *nc-tgp1* lncRNA locus (*nc-tgp1:ura4*$^+$) in wild-type background. (B) Serial dilutions of wild-type, *1343Δ, AΔ*, *BΔ*, and *nc-tgp1:ura4*$^+$ were spotted on non-selective YES medium or in the presence of TBZ (20 μg/mL), HU (10mM), or caffeine (15mM). (C) RT-qPCR experiments measured *tgp1*$^+$, *nc-tgp1*, and *nc-1343* transcript levels in wild-type, *1343Δ, AΔ*, *BΔ*, and *nc-tgp1:ura4*$^+$ cells using primer pairs 1 (*tgp1*$^+$), 5 (*nc-tgp1*), 6 (*nc-1343*) (See Figure 4.2). Error bars represent SEM resulting from three independent replicates.

4.2.4 *nc-tgp1 represses the tgp1$^+$ gene in cis*

The full-length *nc-tgp1* was cloned into a *pREP* vector under the control of the strong *nmt1* promoter in order to exogenously overexpress this lncRNA and examine the possibility that it might repress *tgp1*$^+$ in *trans*. This plasmid was transformed into wild-type cells and *1343Δ* cells. Exogenous expression of the *nc-tgp1* RNA from a plasmid failed to repress the increased *tgp1*$^+$ levels found in *1343Δ* cells (**Fig. 4.7A**). The *tgp1*$^+$ gene is normally induced when cells are exposed to low external phosphate concentrations (Carter-O'Connell *et al.*, 2012). Notably, increased *tgp1*$^+$ mRNA levels following phosphate limitation correlated with a small but detectable reduction in *nc-tgp1* levels (**Fig. 4.7B**). This observation is consistent with the idea that alleviating repressive *nc-tgp1* transcription permits *tgp1*$^+$ expression. Importantly, introducing high levels of the *nc-tgp1* lncRNA from a plasmid in phosphate-starved wild-type cells failed to repress the induction of *tgp1*$^+$ expression (**Fig. 4.7B**). Taken together these results rule out the possibility that the *nc-tgp1* RNA operates in *trans* to silence *tgp1*$^+$.

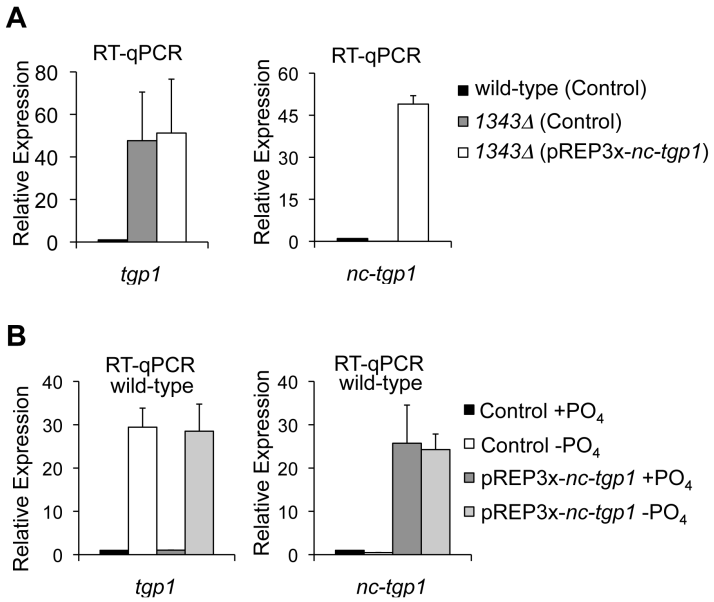

Figure 4.7. *nc-tgp1* does not repress *tgp1*$^+$ *in trans*. RT-qPCR experiments to measure *tgp1*$^+$ mRNA and *nc-tgp1* lncRNA levels in (a) *1343Δ* cells and (b) wild-type responding to phosphate availability, each transformed with an empty pREP3x vector (Control) or pREP3x vector containing *nc-tgp1* under the control of the strong *nmt1* promoter (pREP3x-*nc-tgp1*). Cells were grown in the absence of thiamine. Error bars indicate the standard deviation resulting from two independent experiments.

4.3 Discussion

Genome-wide RNA sequencing has allowed for the detection of a large number of previously unknown lncRNA species in a variety of organisms. However, it remains unclear what proportion of these lncRNAs are functional transcripts that act to influence gene expression and/or chromatin landscapes. Examples such as the *Xist* lncRNA in mammals and *roX* lncRNAs in *Drosophila* represent functional transcripts that are critical for establishing dosage compensation by altering chromatin status and expression levels from sex chromosomes (Lee and Bartolomei, 2013). However, enthusiasm for lncRNA function has been somewhat dampened by recent reports showing that deleting some of the best-characterized lncRNAs in animal models (for example: *HOTAIR*, *MALAT1*, *Kcnq1ot1*, and *NEAT1*) exhibited far less dramatic or undetectable phenotypes (Eißmann *et al.*, 2012; Korostowski *et al.*, 2012; Nakagawa *et al.*, 2011; Nakagawa *et al.*, 2012; Schorderet and Duboule, 2011; Zhang *et al.*, 2012). Such findings suggest that other factors might compensate for lncRNA loss and/or act redundantly in the context of the whole organism. Alternatively, it is possible that the functional significance of some lncRNAs characterized by RNAi knockdown and/or over-expression studies in cells might be overstated. Deleting lncRNA loci in their entirety is not without its own drawbacks since it can make it difficult to attribute any observed phenotypes resulting from such a manipulation to the actual RNA product itself. It is equally possible that such deletions might result in the loss of important DNA elements embedded in the lncRNA gene. It is therefore unsurprising that there have recently been calls for the strategic manipulation of endogenous lncRNA loci that distinguish between the roles played by lncRNA products, the effects that might result simply

from lncRNA transcription alone, and the influence of overlapping DNA elements (Bassett *et al.*, 2014).

The function of an intergenic lncRNA transcribed from the *S. pombe sme2*$^{+}$ locus is well established. Numerous independent groups have used various strategies and approaches to reveal that the *sme2*$^{+}$ lncRNA hosts dozens of DSR-motifs for Mmi1 binding that allow it to be a major target for Mmi1/Red1-directed exosome degradation (Harigaya *et al.*, 2006; Hiriart *et al.*, 2012; Shichino *et al.*, 2014; Yamashita *et al.*, 2012; Yamashita *et al.*, 2013). The consequence of this regulation is that the *sme2*$^{+}$ lncRNA behaves as a decoy to sequester Mmi1 and allow meiotic Mmi1-target genes to accumulate and initiate sexual differentiation. Another purported functional lncRNA gene in *S. pombe*, *SPNCRNA.1164*, is much less characterized. Although deleting the non-conserved *SPNCRNA.1164* gene has been shown to cause *S. pombe* cells to acquire a mild resistance to osmotic stress (Leong *et al.*, 2014), the mechanism of action was not explored further. This is problematic for many reasons, but the most important reason is that this region is predicted to encode three distinct lncRNAs, one mapping to the annotated locus (*SPNCRNA.1164*) and two on the opposite strand (*prl6* and *SPNCRNA.1165*). This ambiguity makes it unclear whether one or more of the putative transcripts originating from this locus is/are actually involved in controlling the cellular response to osmotic stress in *S. pombe*. It is equally possible that there are one or more DNA elements present in this locus or that these transcripts might encode short peptides important for the response of *S. pombe* cells to environmental changes in osmolarity. To eliminate this kind of ambiguity, detailed analyses of the *ncRNA.1343* locus, including mapping transcription start sites, determining transcript length, identifying factors responsible for transcript processing/turnover, and informed genetic manipulations were all performed here to identify whether lncRNA

transcription indeed accounts for the drug sensitivity phenotype observed in cells lacking *ncRNA.1343*.

The induction of a nearby phosphate-regulated permease gene (*tgp1*$^{+}$) in *S. pombe* cells that lack the *ncRNA.1343* gene was found to be directly responsible for the decreased tolerance of these cells to growth in the presence of different compounds. Closer inspection of the *ncRNA.1343* locus revealed that the *ncRNA.1343* promoter is bidirectional and that transcription from this bidirectional promoter preferentially favours the production of a previously unannotated and unstable lncRNA (*nc-tgp1*) transcribed towards the *tgp1*$^{+}$ gene under repressive conditions. Additional experiments were required to show that deletion of *ncRNA.1343* actually affected the expression of this divergent transcript. Only after further strategic manipulations and analyses could it be concluded that the transcription of *nc-tgp1* over the *tgp1*$^{+}$ promoter interferes with the expression of *tgp1*$^{+}$ downstream and that the function of this lncRNA is limited to *cis* regulation.

The fact that the unstable *nc-tgp1* transcript is the functional partner of the apparently non-functional stable *nc-1343* RNA, which is transcribed from the same bidirectional promoter, demonstrates the importance of comprehensive analyses of lncRNAs and the unpredictable consequences of their deletion. Based on the analyses performed here, low-level expression of the *nc-1343* RNA, which is predicted to be present at much less than one copy per cell (Marguerat *et al.*, 2012), could merely represent transcriptional noise resulting as a byproduct of ample *nc-tgp1* transcription.

Genome-wide approaches are extremely powerful and can rapidly catalogue the presence and response of various lncRNAs to different conditions. Despite these

strengths, much more detailed locus-specific analyses are required to rule out the possibility that any lncRNA might simply represent transcriptional noise. Additional experiments are also required to pinpoint the function of individual functional lncRNAs with respect to *cis* regulation of nearby genes or *trans* regulation of genes at other loci.

CHAPTER 5

Two phosphate-regulated genes in fission yeast are repressed by transcriptional interference

5.1 Introduction

Cells depend on their external environment for supplying nutrients essential for growth and survival. Accordingly, cells have evolved complex strategies to sense external nutrient levels and to integrate this information into a transcriptional response that controls the expression of specific genes that help maintain nutrient homeostasis. Genome-wide fluctuations in gene expression accompanying nutrient limitation have been observed in many systems and generally include the induction of general stress-response genes as well as genes specific to overcoming different nutrient deficiencies (Brauer *et al.*, 2008).

While stress-specific transcription factors are important to initiate gene activation in response to nutrient starvation, accumulating evidence indicates that lncRNA transcription also helps to maintain nutrient homeostasis by coordinating changes in gene expression. For example, the balance of sense/antisense lncRNA transcription at stress-response genes is critical for many yeast species to appropriately respond to the reduced availability of various nutrients (Yassour *et al.*, 2010). Importantly, nutrient limitation can drive other cellular behaviour. In particular, nitrogen starvation stimulates sexual differentiation in *S. pombe*, in part by alleviating repressive antisense lncRNA transcription at a number of meiotic genes (Bitton *et al.*, 2011).

lncRNA-dependent mechanisms are also responsible for controlling sexual differentiation following nitrogen limitation in *S. cerevisiae*, where the central inducer of meiosis, the *IME1* gene, is repressed in the presence of nitrogen and fermentable sugars by upstream lncRNA transcription (van Werven *et al.*, 2012). Intergenic lncRNA transcription has also been reported to regulate nearby stress-response genes in other organisms as well, including *S. pombe* where cascading lncRNA transcription upstream of the *fbp1*$^{+}$ gene, which encodes the metabolic enzyme fructose-1,6-bisphosphatase, is required to create an open chromatin environment and induce *fpb1*$^{+}$ expression following glucose starvation (Hirota *et al.*, 2008). In contrast, the *S. cerevisiae SER3* gene is repressed by transcriptional interference in the presence of serine (Martens *et al.*, 2004). In this case, intergenic lncRNA transcription into the *SER3* promoter increases nucleosome density, prohibiting transcription factor access (Hainer *et al.*, 2011; Thebault *et al.*, 2011). Other lncRNA-dependent regulatory mechanisms have been reported in higher eukaryotes as well. Notably, the human lncRNA *Gas5* acts as a decoy for the glucocorticoid receptor (GR) by competing with DNA for binding to prevent target gene activation following nutrient starvation (Kino *et al.*, 2010).

lncRNA products themselves have been reported to recruit chromatin-modifying activities to regulate nearby genes. This form of regulation also appears to play a pivotal role in maintaining nutrient homeostasis in many organisms. For example, antisense transcription through the *S. cerevisiae GAL* cluster produces an lncRNA product that is thought to recruit HDAC activity to silence *GAL* genes when external glucose concentrations are sufficiently high (Houseley *et al.*, 2008). Low glucose levels stimulate *GAL* gene expression, in part by reducing transcription of this repressive lncRNA. Recent studies in *S. pombe* suggest that the phosphate-regulated *pho1*$^{+}$ gene is silenced by transient heterochromatin brought about by an

overlapping lncRNA (Lee *et al.*, 2013; Shah *et al.*, 2014). In this case, Mmi1-binding to the *pho1*$^{+}$-regulatory lncRNA was proposed to also recruit components of the RNAi machinery and the H3K9 methyltransferase Clr4 to distribute the methyl-H3K9 mark over the locus, silencing the *pho1*$^{+}$ gene when phosphate is readily available to cells (Shah *et al.*, 2014). Together these examples illustrate the importance of lncRNA transcription in diverse stress-response pathways that control nutrient homeostasis.

5.2 Results

5.2.1 *Phosphate starvation induces tgp1*$^{+}$ *by repressing nc-tgp1*

S. pombe cells grown in a phosphate-limited environment induce the expression of several specialized genes that help cells harvest inorganic phosphate from the external environment, including *tgp1*$^{+}$*, pho1*$^{+}$, and *pho84*$^{+}$ (Carter-O'Connell *et al.*, 2012). To determine how the transcription of the regulatory lncRNA *nc-tgp1* is altered in response to phosphate, and how it might influence *tgp1*$^{+}$ expression in this natural physiological stress, expression levels were assessed by northern blotting and RT-qPCR in phosphate rich (+PO_4) and phosphate deprived (-PO_4) conditions. As expected, the mRNA levels of *tgp1*$^{+}$ increased upon phosphate starvation (**Fig. 5.1A and 5.1B**). Notably, prolonged phosphate-starvation induced greater levels of the *tgp1*$^{+}$ mRNA than those observed in *1343Δ* cells (**Fig. 4.1A**). In contrast, the levels of both *nc-tgp1* and *nc-1343* lncRNAs decreased substantially in the absence of extracellular phosphate (**Fig. 5.1A, 5.1C, and 5.1D**). Since *nc-1343* is transcribed from the same promoter that generates *nc-tgp1*, reduced *nc-1343* RNA levels are likely a consequence of decreased *nc-tgp1* transcription. Importantly, the observed reduction in *nc-tgp1* RNA levels is wholly consistent with and further supports the

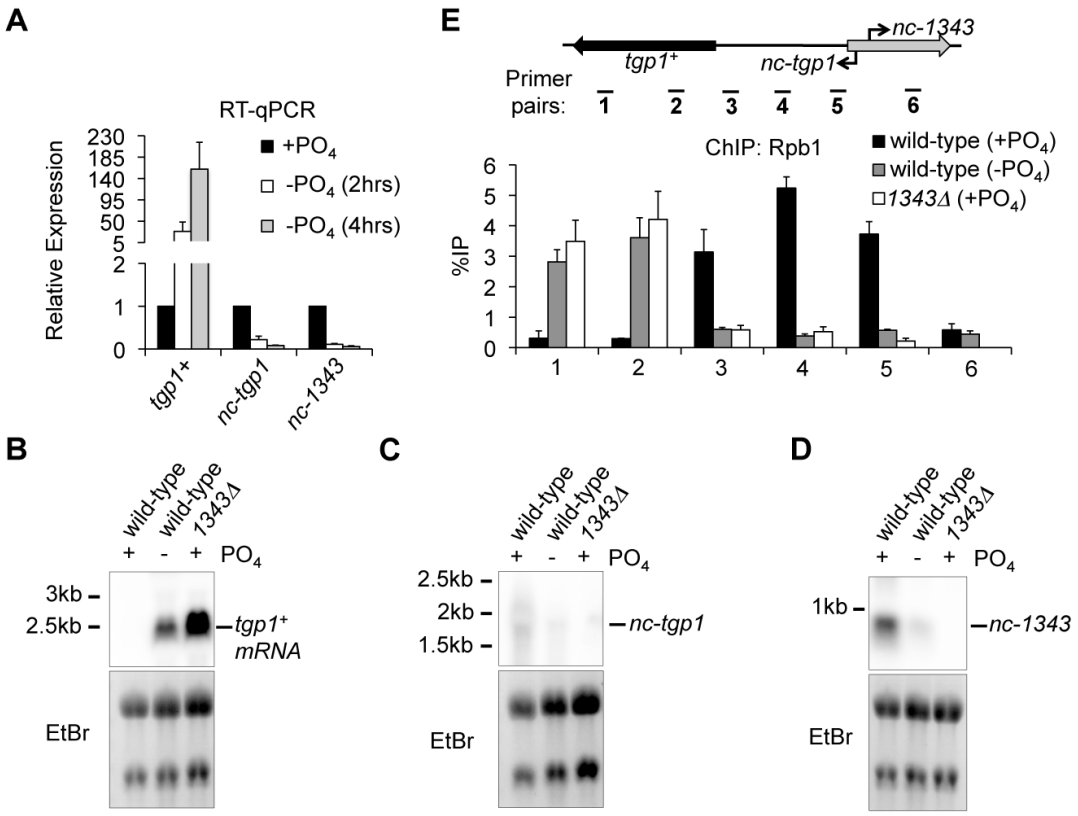

Figure 5.1. Phosphate starvation induces *tgp1*$^{+}$ and reduces lncRNA transcription. (A) RT-qPCR experiments measured *tgp1*$^{+}$, *nc-tgp1*, and *nc-1343* transcript levels in wild-type cells grown in phosphate-rich medium (+PO_4) or in the absence of phosphate (-PO_4) for the indicated times. (B-D) Northern analyses of the *tgp1*$^{+}$ mRNA and lncRNAs *nc-tgp1* (cryptic) and *nc-1343* in wild-type cells grown in the presence of phosphate or following two hours of phosphate starvation, as well as in *1343Δ* cells grown in normal phosphate-rich conditions. rRNA bands visualized by ethidium bromide (EtBr) represent controls for equal loading. (E) Rbp1 ChIP-qPCR experiments performed in wild-type cells grown in the presence or absence of phosphate, and *1343Δ* cells grown in the presence of phosphate. Error bars represent SEM resulting from three independent experiments.

hypothesis that loss or reduction of *nc-tgp1* transcription permits *tgp1*$^+$ induction. In agreement with this, significantly less RNAPII associates with the *nc-tgp1* transcription unit in both phosphate-starved wild-type cells and phosphate-replete *1343Δ* cells, which do not transcribe *nc-tgp1* (**Fig. 5.1E**). Thus, preventing *nc-tgp1* transcription in phosphate-rich medium (repressive conditions) appears to recapitulate the changes in RNAPII occupancy that normally accompany *tgp1*$^+$ induction upon phosphate deprivation.

5.2.2 *RNAi-directed heterochromatin does not regulate tgp1*$^+$

Cells with defective exosome function (e.g. *rrp6Δ*) accumulate non-coding RNAs, some of which have been reported to attract Mmi1-dependent RNA elimination factors, along with RNAi components and the Clr4 H3K9 methyltransferase, leading to the formation of transiently regulated HOODs (heterochromatin domains) (Yamanaka *et al.*, 2013). The *tgp1*$^+$ gene was reported to be located within HOOD-17 and forms a region of Mmi1-directed transient heterochromatin in *rrp6Δ* cells (Lee *et al.*, 2013; Yamanaka *et al.*, 2013). The *nc-tgp1* transcript is clearly regulated by Mmi1-directed exosome degradation (**Fig. 4.4B**), however quantitative ChIP analyses detected very low levels of methyl-H3K9 (H3K9me2) over the *tgp1*$^+$, *nc-tgp1,* or *nc-1343* genes within HOOD-17 in wild-type cells (**Fig. 5.2A**). These low levels of H3K9me2 did not drop appreciably upon *tgp1*$^+$ induction in phosphate-starved cells (-PO_4). Moreover, equivalent low signals were detected in cells lacking Clr4, the sole *S. pombe* H3K9 methyltransferase, suggesting that the signal detected represents experimental noise/background. Equivalent background levels of H3K9me2 were detectable on another Mmi1-targeted lncRNA gene (*sme2*$^+$) and on the highly expressed euchromatic actin gene (*act1*$^+$). In contrast, H3K9me2 was enriched approximately 100-fold over background at centromeric outer repeats (*dg*),

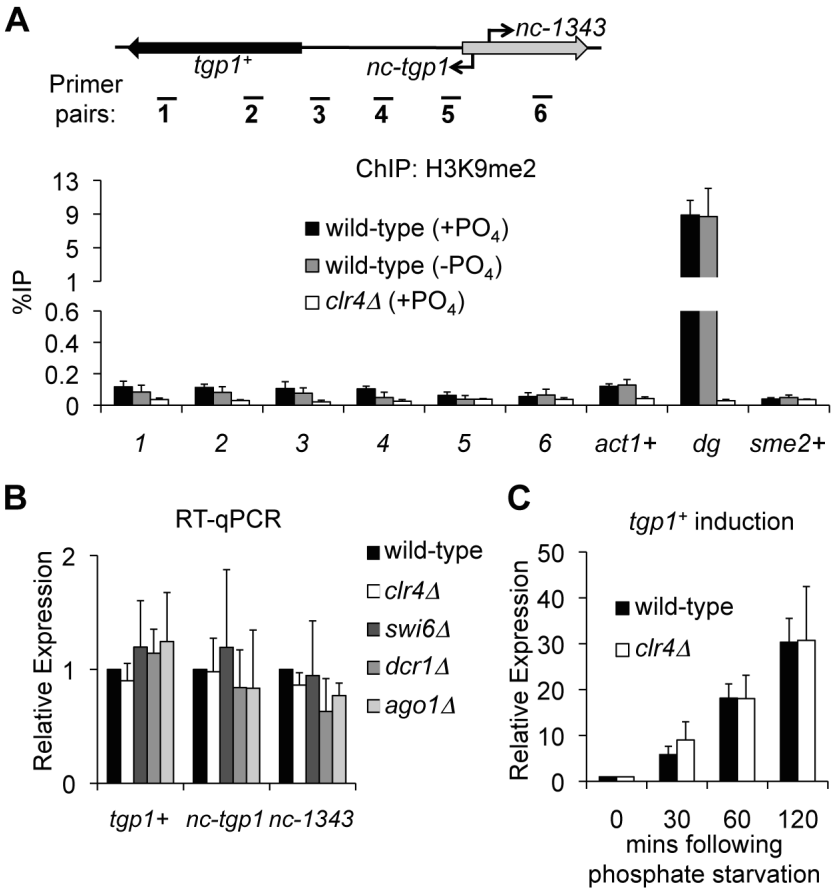

Figure 5.2. *tgp1*$^{+}$ is not regulated by RNAi/heterochromatin. (A) H3K9me2 ChIP-qPCR experiments performed in the presence or absence of phosphate. *clr4Δ* cells were used as a negative control. The euchromatic actin gene (*act1*$^{+}$) and pericentric repeats (*dg*) are negative and positive controls for methyl-H3K9 chromatin, respectively. The *sme2*$^{+}$ gene encodes a lncRNA target of Mmi1 that is not reported to accumulate H3K9 methylation and therefore an additional negative control for methyl-H3K9 chromatin. (B) RT-qPCR experiments measured *tgp1*$^{+}$, *nc-tgp1*, and *nc-1343* transcript levels in wild-type cells and cells lacking factors involved in heterochromatin formation and stability, including the H3K9 methyltransferase Clr4, the HP1 homolog Swi6, as well as the Dicer and Argonaut homologs Dcr1 and Ago1, respectively. (C) RT-qPCR analysis of *tgp1*$^{+}$ mRNA induction kinetics following phosphate starvation in wild-type and *clr4Δ* cells. Error bars represent SEM resulting from at least three independent replicates.

while H3K9me2 levels at *dg* repeats reduced to background levels in *clr4Δ* cells, indicating that H3K9-methylated chromatin had been efficiently immunoprecipitated. Collectively, these findings are in agreement with published genome-wide analyses where high levels of H3K9 methylation were present at regions of constitutive heterochromatin (e.g. centromeres) but only background levels were present at the *tgp1*$^{+}$ gene (Wang *et al.*, 2015; Yamanaka *et al.*, 2013). Consistent with a lack of H3K9me2, the transcript levels of *tgp1*$^{+}$, *nc-tgp1*, and *nc-1343* were unaffected by the loss of RNAi (e.g. *ago1Δ* or *dcr1Δ*) or heterochromatin components (e.g. *clr4Δ* or *swi6Δ*) (**Fig. 5.2B**). In addition, the kinetics of *tgp1*$^{+}$ mRNA induction following phosphate-starvation were not noticeably altered in cells lacking heterochromatin (**Fig. 5.2C**). Together these results agree with previous expression profiling analyses that found unaltered *tgp1*$^{+}$ mRNA levels in cells lacking RNAi/heterochromatin (Hansen *et al.*, 2005). In contrast, *nc-tgp1* and *sme2*$^{+}$ RNA levels were clearly elevated in cells lacking Mmi1-mediated exosome degradation (**Fig. 3.3, 3.4, and 4.4**). Although H3K9 methylation is reported to accumulate at particular euchromatic regions in *rrp6Δ* cells (e.g. HOOD-17: *tgp1*$^{+}$), these findings demonstrate that RNAi and heterochromatin play no appreciable role in regulating *tgp1*$^{+}$ under normal physiologically repressive conditions or during induction.

Consistent with the above findings, profiling H3K9me2 levels genome-wide by ChIP-seq analyses showed high enrichment of H3K9 methylation at centromeres in wild-type cells (**Fig. 5.3A**), but significant levels of this mark could not be detected above background (*clr4Δ*) at the *tgp1*$^{+}$ gene (**Fig. 5.3B**). This mapping also revealed no significant enrichment of H3K9 methylation at the *pho1*$^{+}$ gene (**Fig. 5.3C**). In addition, only modest levels of this mark were detected at the meiotic *mei4*$^{+}$ gene (**Fig. 5.3D**). This result is surprising since *mei4*$^{+}$ has been proposed to form an

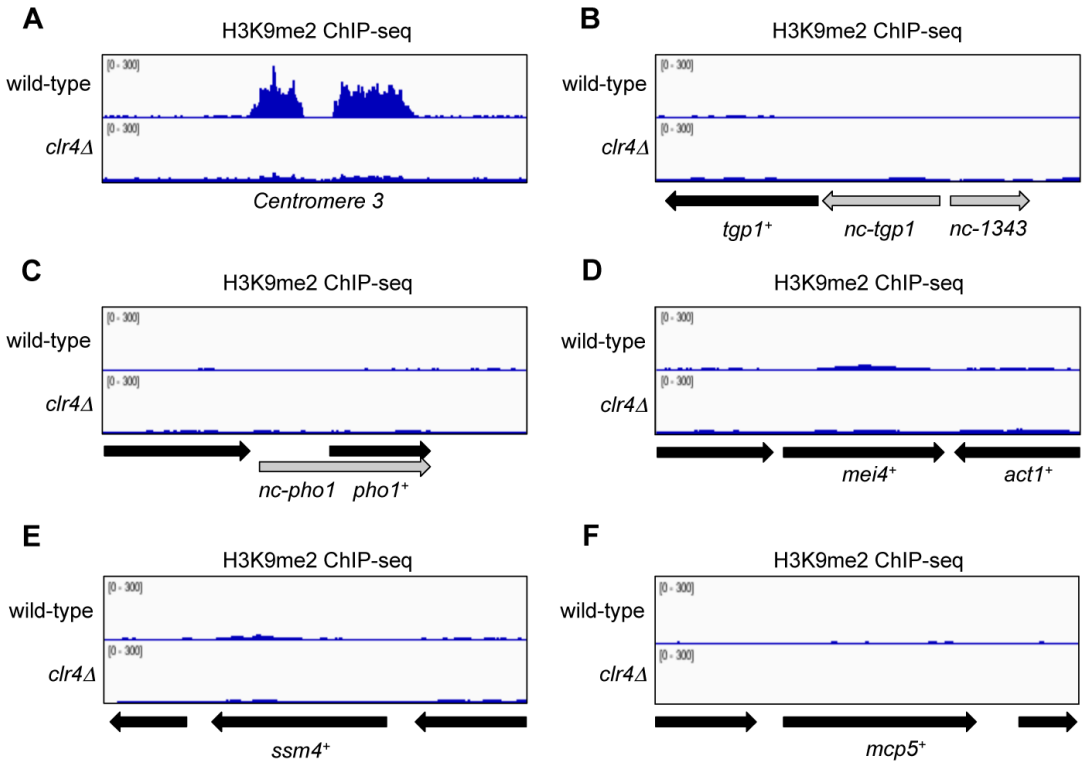

Figure 5.3. Low levels of H3K9 methylation at a representative heterochromatin islands and two HOODs. H3K9me2 ChIP-seq experiments performed in wild-type and *clr4Δ* cells. (A) High enrichment of the H3K9me2 mark at pericentric heterochromatin, but not at (B) *tgp1*$^{+}$ (HOOD-17), (C) *pho1*$^{+}$ (HOOD-23), or heterochromatin islands (D) *mei4*$^{+}$, (E) *ssm4*$^{+}$, or (F) *mcp5*$^{+}$. Bioinformatic analyses performed by Pin Tong.

RNAi-independent heterochromatin island in vegetative cells (Hiriart *et al.*, 2012; Zofall *et al.*, 2012). Interestingly, other proposed "facultative heterochromatin islands" in *S. pombe* showed equally low levels by H3K9me2 ChIP-seq analyses (**Fig. 5.3E and 5.3F**). Collectively, these findings lead one to question the real biological significance of low levels of H3K9 methylation reported at euchromatic loci.

5.2.3 *nc-tgp1 transcription increases nucleosome density and prevents Pho7 transcription factor binding*

The above analyses indicate that *nc-tgp1* is transcribed into the *tgp1*$^+$ promoter and that production of this upstream lncRNA represses expression of the *tgp1*$^+$ gene. However, it is unclear how the *nc-tgp1* RNA interferes with the induction mechanism of *tgp1*$^+$ in response to phosphate availability. The Pho7 transcription factor has previously been shown to engage phosphate-response gene promoters in phosphate-deficient cells (Carter-O'Connell *et al.*, 2012; Henry *et al.*, 2011). The Pho7 protein was C-terminally tagged with green fluorescent protein (GFP) in wild-type cells and *1343Δ* cells (**Fig. 5.4A**). anti-GFP ChIP analyses confirmed that Pho7-GFP accumulates over the region upstream of *tgp1*$^+$ when activated in cells starved of phosphate (**Fig. 5.4B**). However, in cells unable to transcribe *nc-tgp1* (*1343Δ*), higher levels of Pho7-GFP associate with the region upstream of *tgp1*$^+$ even in repressive conditions (i.e. +PO_4). These findings suggest that loss of *nc-tgp1* expression, either due to phosphate starvation or by artificially preventing production of this lncRNA in repressive phosphate-replete conditions (as seen in *1343Δ*), allows Pho7 binding and subsequently *tgp1*$^+$ expression. These results imply that Pho7 is already primed to bind the *tgp1*$^+$ promoter in repressed conditions but lncRNA transcription actively destabilizes this interaction.

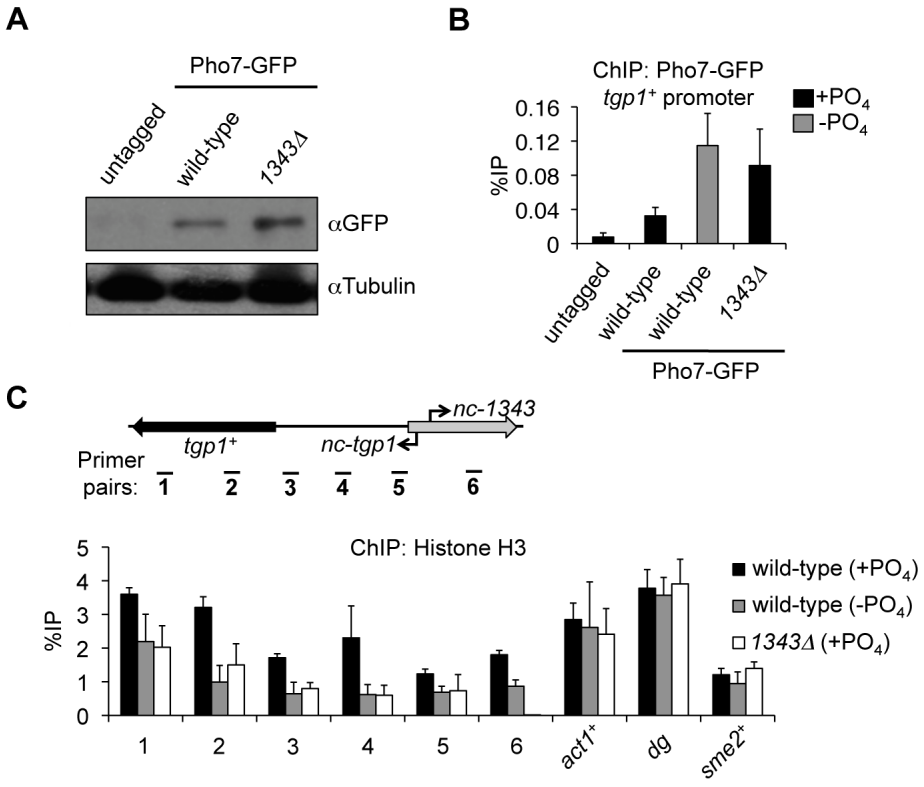

Figure 5.4. *nc-tgp1* transcription prevents stable Pho7 binding and increases nucleosome density upstream of *tgp1*$^+$. (A) Western blot analysis of C-terminal GFP-tagged Pho7 in wild-type and *1343Δ* backgrounds. Tubulin was used as a loading control. (B) Pho7-GFP ChIP-qPCR experiments were performed in the presence or absence of phosphate. An untagged strain was used as a negative control. Primer pair #3 was used to detect Pho7 binding at the *tgp1*$^+$ promoter. (C) Nucleosome density was measured by histone H3 ChIP-qPCR experiments in wild-type cells grown in the presence or absence of phosphate, and in *1343Δ* cells grown in phosphate-rich conditions. Error bars represent SEM resulting from three independent replicates.

Active RNAPII promoters display reduced nucleosome density (Yuan *et al.*, 2005). In some cases, lncRNA transcription over promoters has been found to increase nucleosome density, obstructing transcription factor binding and thus preventing gene induction (Hainer *et al.*, 2011; Thebault *et al.*, 2011; van Werven *et al.*, 2012). Histone H3 ChIP revealed greater nucleosome density over the *tgp1*$^+$ locus in repressive conditions (+PO_4) compared to when *tgp1*$^+$ is expressed (**Fig. 5.4C**). Thus, upstream lncRNA transcription increases nucleosome density over the *tgp1*$^+$ promoter, which is consistent with a transcriptional interference mechanism that alters the chromatin landscape to prevent access to the key phosphate-response transcription factor Pho7.

To directly test if transcriptional interference of *tgp1*$^+$ by *nc-tgp1* is responsible for *tgp1*$^+$ repression, the *nc-tgp1* promoter was replaced with the strong, thiamine-regulated *nmt1* promoter (*nmt1-nc-tgp1*) (**Fig. 5.5A**). Transcription of *nc-tgp1* from the *nmt1* promoter is rendered unresponsive to phosphate (**Fig. 5.5B**). Instead, *nc-tgp1* is repressed or derepressed in the presence or absence of thiamine, respectively. When *nc-tgp1* was transcribed from the *nmt1* promoter, *tgp1*$^+$ remained repressed regardless of phosphate availability. A weaker *nmt81* promoter driving lower levels of *nc-tgp1* transcription failed to repress *tgp1*$^+$ (**Fig. 5.5C**), indicating that high levels of lncRNA transcription are required to repress downstream gene expression. Importantly, repression of *nmt1*-driven *nc-tgp1* by the addition of thiamine to minimal growth medium resulted in the induction of *tgp1*$^+$ expression in phosphate-replete conditions and consequently caused such cells to acquire drug sensitivity (**Fig. 5.5B and 5.5D**). Additionally, histone H3 levels over the region upstream of *tgp1*$^+$ were high when *nc-tgp1* was transcribed but were reduced when *nc-tgp1* transcription was repressed by thiamine (**Fig. 5.5E**), consistent with increased nucleosome density at the *tgp1*$^+$ promoter when the *nc-tgp1* lncRNA is

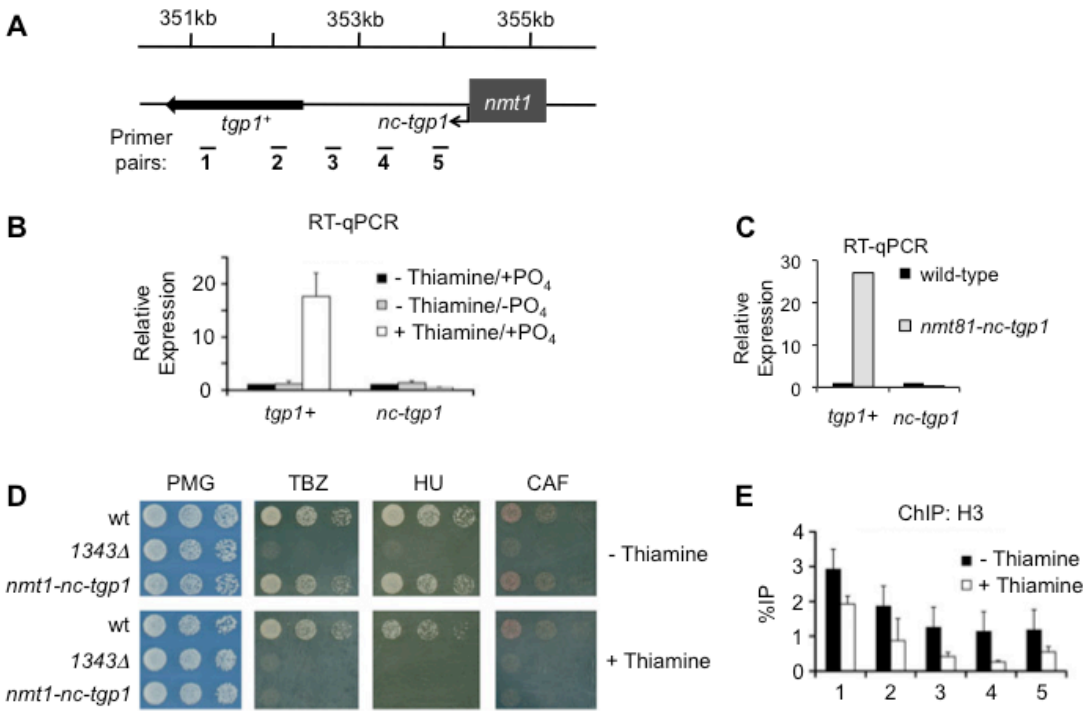

Figure 5.5. *nmt1* controlled *nc-tgp1* alters drug tolerance in response to thiamine. (A) Schematic diagram of *nc-tgp1* under the control of the strong, thiamine-repressible *nmt1* promoter. (B) RT-qPCR experiments measured *tgp1*$^{+}$ and *nc-tgp1* levels in response to thiamine and phosphate availability using *nmt1-nc-tgp1* cells. (C) RT-qPCR experiments measured *tgp1*$^{+}$ and *nc-tgp1* levels in wild-type cells and cells with *nc-tgp1* under the control of the weak *nmt81* promoter (*nmt81-nc-tgp1*). (D) Serial dilutions of wild-type, *1343Δ*, and *nmt1-nc-tgp1* cells were spotted on non-selective PMG medium or in the presence of TBZ, HU, or caffeine, with or without thiamine as indicated. (E) H3 ChIP-qPCR experiments in *nmt1-nc-tgp1* cells grown in the presence or absence of thiamine. Error bars represent SEM resulting from three independent replicates.

transcribed in wild-type cells grown in a phosphate-rich environment (**Fig. 5.4D**). Collectively, these findings confirm that it is the transcription of *nc-tgp1* over the *tgp1*$^{+}$ promoter that alters nucleosome density to regulate *tgp1*$^{+}$ induction. An inadvertent consequence of this regulation is drug tolerance control.

5.2.4 *Repressive lncRNA transcription over the pho1*$^{+}$ *gene promoter*

The *S. pombe pho1*$^{+}$ gene encodes a secreted acid phosphatase important for cells to adapt to low extracellular phosphate concentrations. Similar to the *S. cerevisiae* homolog *PHO5*, which is activated upon phosphate-starvation (Bergman *et al.*, 1986), the *pho1*$^{+}$ gene is tightly regulated in response to phosphate availability (Schweingruber *et al.*, 1992). Rpb1 ChIP experiments were performed to measure RNAPII occupancy over the *pho1*$^{+}$ locus in response to changes in phosphate availability. While RNAPII levels were enriched over the *pho1*$^{+}$ gene and upstream region in repressed conditions, phosphate-starvation reduced upstream RNAPII levels (**Fig 5.6A**). Phosphate depletion resulted in accumulating RNAPII levels over the *pho1*$^{+}$ gene, which corresponded with increased *pho1*$^{+}$ mRNA levels, as detected by RT-qPCR and northern analysis (**Fig. 5.6A, 5.6B, and 5.6C**).

Two independent groups recently found that *S. pombe pho1*$^{+}$ repression in response to phosphate availability is mediated by an unstable lncRNA transcription originating upstream of the *pho1*$^{+}$ gene (Lee *et al.*, 2013; Shah *et al.*, 2014). Rrp6 or Mmi1 loss results in the accumulation of this overlapping lncRNA, termed here *nc-pho1* (**Fig. 5.7A and 5.7B**), reminiscent of *tgp1*$^{+}$ regulation by the upstream *nc-tgp1* RNA (**Fig. 4.5**). In addition, the *nc-pho1* lncRNA contains three DSR-motifs for Mmi1 binding and RIP experiments confirmed direct binding between Mmi1-HTP and *nc-pho1* (**Fig. 5.7C and 5.7D**). These results are consistent with published studies which concluded that Mmi1 targets the repressive lncRNA transcribed upstream

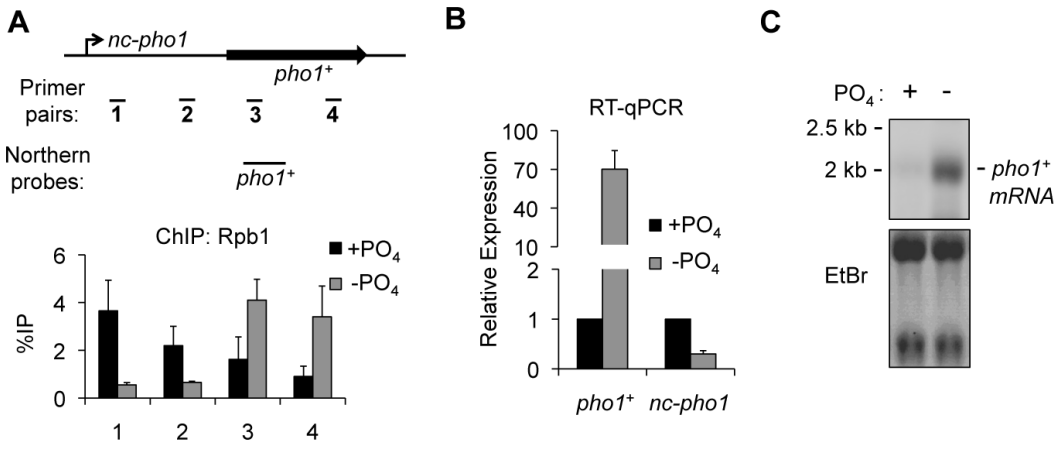

Figure 5.6. lncRNA transcription upstream of *pho1*$^+$ responds to phosphate availability. (A) Schematic representation of the *pho1*$^+$ locus, including the sites of northern probe and qPCR primer pairs, Rbp1 ChIP-qPCR experiments performed in wild-type cells grown in the presence or absence of phosphate. (B) RT-pPCR experiments measured *pho1*$^+$ mRNA (primer pair #3) and upstream lncRNA *nc-pho1* levels (primer pair #1) in wild-type cells grown in the presence or absence of phosphate. (C) Northern analysis of the *pho1*$^+$ mRNA in phosphate-rich and phosphate-depleted wild-type cells. rRNA bands visualized by ethidium bromide (EtBr) represent controls for equal loading. Error bars indicate standard error from two independent experiments.

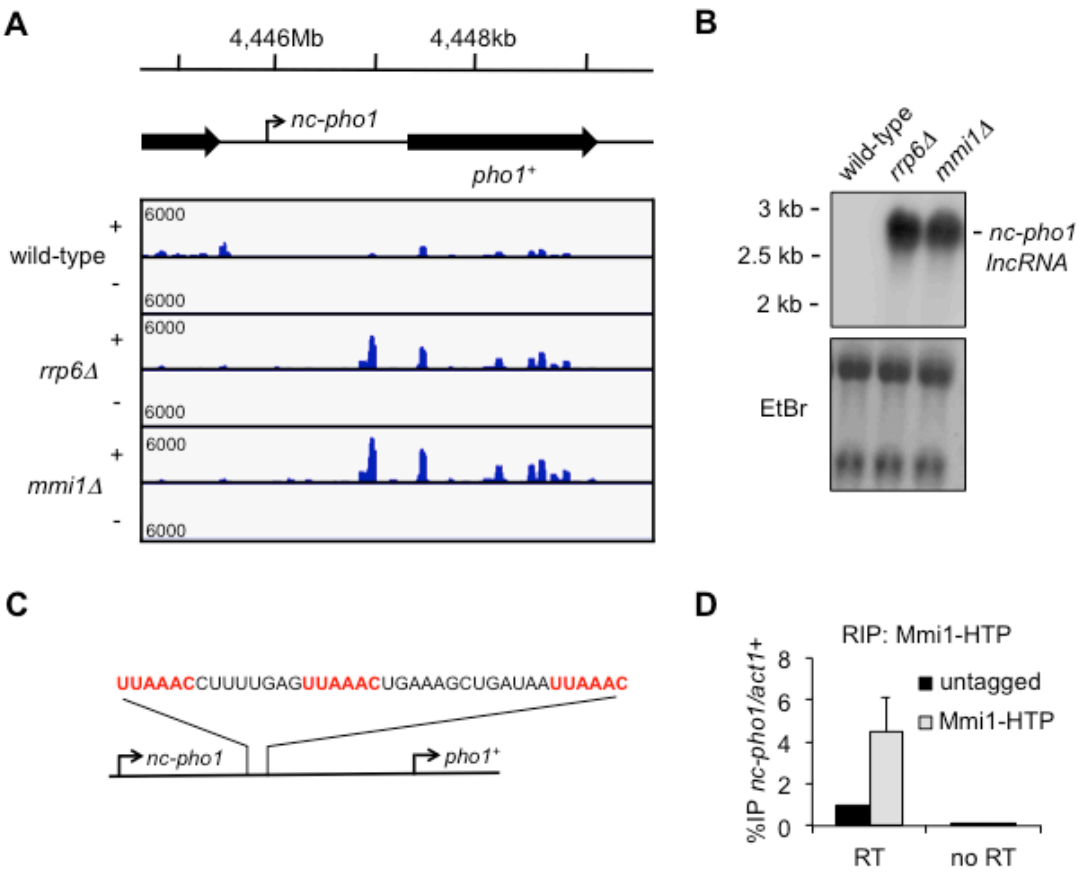

Figure 5.7. lncRNA overlapping the *pho1*$^{+}$ gene is targeted for exosome-mediated degradation by Mmi1. (A) Strand-specific RNA-seq at the *pho1*$^{+}$ locus in wild-type, *rrp6Δ*, and *mmi1Δ* cells. Location of qPCR primer pairs and probes for northern analysis are shown below. Bioinformatic analyses performed by Pin Tong. (B) Northern analysis of the *nc-pho1* lncRNA with the same probe used in Figure 5.6 in wild-type, *rrp6Δ*, and *mmi1Δ* cells. rRNA bands visualized by ethidium bromide (EtBr) represent controls for equal loading. (C) Schematic representation of the *pho1*$^{+}$ locus with putative DSR motifs (UUAAAC) in the *nc-pho1* lncRNA. (D) Mmi1-HTP RIP and quantification by RT-qPCR for *nc-pho1* binding. Error bars indicate standard error from two independent experiments.

from the *pho1*$^{+}$ gene for degradation by the nuclear exosome (Lee *et al.*, 2013; Shah *et al.*, 2014). Importantly, these findings suggest that *pho1*$^{+}$ and *tgp1*$^{+}$ are both regulated in a similar lncRNA-dependent manner.

5.2.5 *pho1*$^{+}$ *is repressed by transcriptional interference*

The repression of both *tgp1*$^{+}$ and *pho1*$^{+}$ by upstream lncRNAs degraded by Mmi1-recruited exosome activity implies a similar regulatory mechanism might control expression of both phosphate-response genes. However, in contrast to *nc-tgp1*-dependent transcriptional interference at the *tgp1+* locus, it has recently been proposed that the lncRNA upstream of the *pho1*$^{+}$ gene recruits components of the RNAi machinery and Clr4 via direct interactions with Mmi1 to deposit transient heterochromatin over the *pho1*$^{+}$ locus in response to phosphate availability (Lee *et al.*, 2013; Shah *et al.*, 2014). However, H3K9me2 mapping by ChIP-seq failed to detect this mark at the *pho1*$^{+}$ gene in repressed wild-type cells (**Fig. 5.3C**). Phosphate starvation only slightly reduced the marginal H3K9me2 levels at the *pho1*$^{+}$ promoter, but quantitative ChIP analyses indicate that wild-type levels of this mark were not significantly enriched at the *pho1*$^{+}$ locus when compared to *clr4Δ* control cells (**Fig. 5.8A and 5.3C**). Consistent with previous expression profiling analyses showing unaltered *pho1*$^{+}$ levels in the absence of RNAi/heterochromatin (Hansen *et al.*, 2005), cells lacking RNAi/heterochromatin failed to induce expression or alter the induction kinetics of the *pho1*$^{+}$ gene (**Fig. 5.8B and 5.8C**). Importantly, published genome-wide analyses using ChIP-chip show background levels of H3K9me2 over *pho1*$^{+}$ in wild-type cells grown in normal, repressive conditions (Wang *et al.*, 2015; Yamanaka *et al.*, 2013).

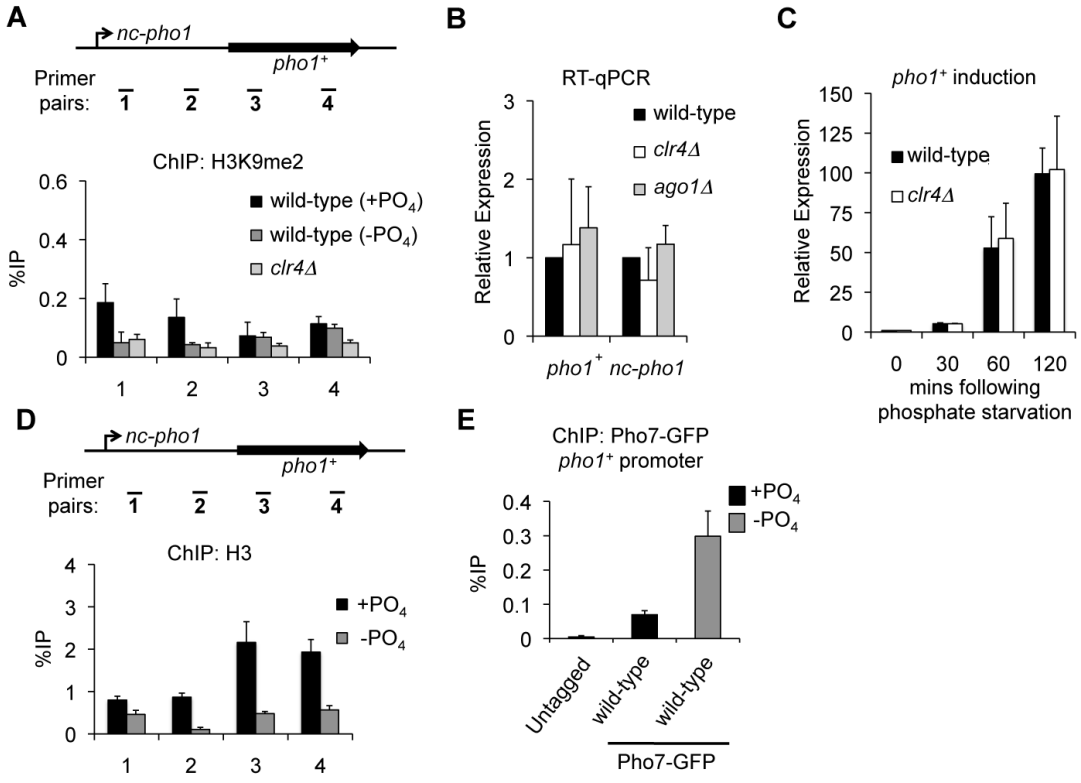

Figure 5.8. *pho1*$^+$ is repressed by transcriptional interference, not transient heterochromatin. (A) H3K9me2 ChIP-qPCR experiments performed in the presence and absence of phosphate. *clr4Δ* was used as a negative control. (B) RT-qPCR analysis of *pho1*$^+$ and *nc-pho1* transcript levels in wild-type cells and cells lacking factors involved in heterochromatin formation and stability. (C) RT-qPCR experiments measured *tgp1*$^+$ mRNA induction kinetics following phosphate depleition in wild-type and *clr4Δ* cells. (D) Nucleosome density was measured by histone H3 ChIP-qPCR experiments in wild-type cells grown in the presence or absence of phosphate. (E) Pho7-GFP ChIP-qPCR experiments were performed in the presence or absence of phosphate in cells. An untagged strain was used as a negative control. Primer pair #2 was used to detect Pho7 binding at the *pho1*$^+$ promoter. Error bars represent SEM resulting from at least three independent replicates.

Histone H3 ChIP was performed to test if the *pho1*$^{+}$ gene might be regulated by transcriptional interference. These analyses show that nucleosome density decreases over the *pho1*$^{+}$ locus in response to reduced lncRNA transcription when cells are starved of phosphate (**Fig. 5.8D**). As observed at the *tgp1*$^{+}$ locus, decreased nucleosome density over the *pho1*$^{+}$ promoter also correlated with increased Pho7-GFP binding (**Fig. 5.8E**). Together these results argue against a role for heterochromatin in the repression of *pho1*$^{+}$ in wild-type cells. Rather, these analyses suggest that *pho1*$^{+}$ is repressed in response to phosphate availability by a mechanism of transcriptional interference that is analogous to *tgp1*$^{+}$ regulation. Thus, two central regulators of the phosphate-response in *S. pombe* appear to be controlled by related regulatory mechanisms involving cryptic upstream lncRNA transcription that limits expression in phosphate-replete environments.

5.2.6 *H3K9 methylation increases at tgp1*$^{+}$ *and pho1*$^{+}$ *genes in rrp6Δ cells*

Previously published genome-wide mapping of H3K9 methylation showed the presence of RNAi-dependent heterochromatin at *tgp1*$^{+}$ and *pho1*$^{+}$ in cells lacking Rrp6 (**Figure 5.9A and 5.9B**; Yamanaka *et al.*, 2013). In agreement with these findings, H3K9me2 ChIP detected increased levels of H3K9 methylation at *tgp1*$^{+}$ and *pho1*$^{+}$ in *rrp6Δ* cells (**Fig. 5.9C and Fig 5.9D**). However, the levels of H3K9me2 detected were still very low when compared to that observed at *bone fide* heterochromatin. The fact that *mmi1Δ* cells also showed increased H3K9 methylation levels at some sites within the *tgp1*$^{+}$ and *pho1*$^{+}$ loci is not compatible with the proposed role for Mmi1 in recruiting the RNAi machinery in exosome-deficient cells. Additionally, similar marginal increases in H3K9 methylation were detected at the euchromatic actin gene (*act1*$^{+}$) in cells lacking Rrp6 and Mmi1 (**Fig. 5.9E**). Moreover, H3K9 methylation levels decreased substantially at *dg* repeats in

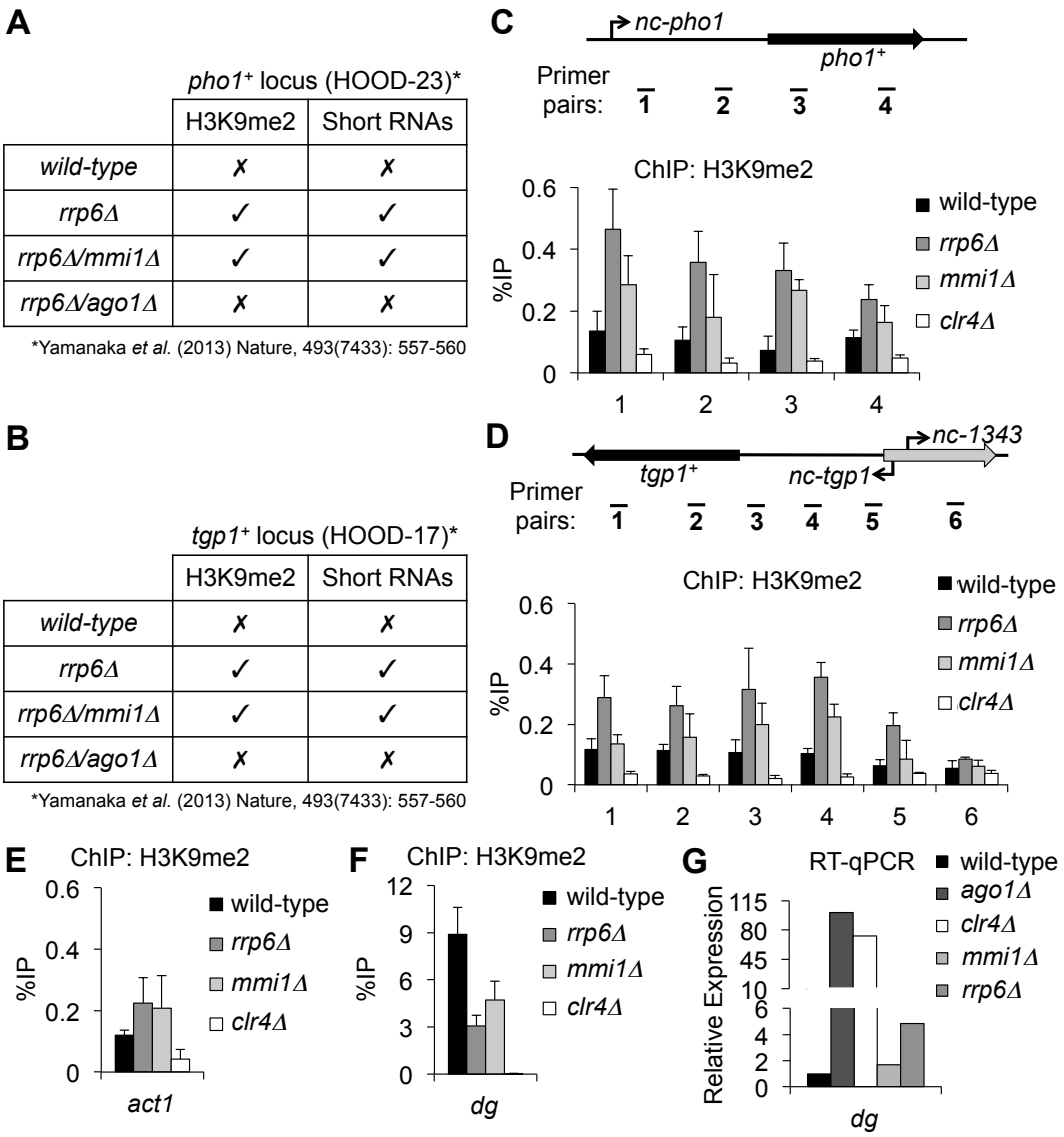

Figure 5.9. Rrp6 loss causes H3K9 methylation to increase slightly at *pho1*$^+$ and *tgp1*$^+$ genes. Tables show the detected presence or absence of H3K9me2 and/or siRNAs at the *pho1*$^+$ gene (A) and the *tgp1*$^+$ gene (B) in a previous study (Yamanaka *et al.*, 2013). (C - F) H3K9me2 ChIP-qPCR experiments performed in wild-type, *rrp6Δ*, and *mmi1Δ*. *clr4Δ* cells were used as a negative control. (G) RT-qPCR experiments measured pericentromeric (*dg*) transcript levels in wild-type, *ago1Δ*, *clr4Δ*, mmi1Δ, and *rrp6Δ*. Error bars represent SEM resulting from at least three independent replicates.

these cells (**Fig. 5.9F**), which is consistent with previous reports showing reduced centromeric heterochromatin in cells following Rrp6 loss (Reyes-Turcu *et al.*, 2011). Compromised heterochromatin at centromeres in *rrp6Δ* cells corresponded with significantly increased transcript levels emanating from *dg* repeats, as detect by RT-qPCR (**Fig. 5.9G**). The presence of increased H3K9 methylation levels at the *tgp1*$^{+}$ and *pho1*$^{+}$ genes in *rrp6Δ* and *mmi1Δ* cells correlates with reduced RNAPII levels (**Fig. 5.10A and Fig 5.10B**), as detected by Rpb1 ChIP. Despite this, *nc-tgp1* and *nc-pho1* are stabilized and accumulate when Mmi1-dependent degradation is missing (**Fig. 4.4 and 5.7**). Thus transcription of these lncRNAs is not effectively silenced as a result of slight increases in H3K9me2 levels in *rrp6Δ* and *mmi1Δ* cells.

Slightly increased RNAPII levels were detected over the 3'-ends of *tgp1*$^{+}$ and *pho1*$^{+}$ genes bodies in cells lacking Mmi1 or Rrp6 (**Fig. 5.10A and 5.10B**). These results suggest that the absence of exosome- mediated degradation of regulatory lncRNAs might lead to greater transcription read-through. This is a plausible explanation since transcription read-through occurs widely in *S. pombe* cells with compromised exosome activity (Lemay *et al.*, 2014), and has been shown at the *pho1*$^{+}$ gene in *rrp6Δ* cells (Shah *et al.*, 2014). Therefore, reduced RNAPII levels over the *tgp1*$^{+}$ and *pho1*$^{+}$ promoters might not necessarily indicate less transcription as a consequence of increased H3K9me2 levels in these mutants, but instead represent decreased RNAPII stalling in cells lacking co-transcriptional exosome degradation. Thus, it is unclear if the low H3K9 methylation levels detected in exosome-compromised cells would be sufficient to reduce RNAPII transcription. Interestingly, *pho1*$^{+}$ and *tgp1*$^{+}$ induction was significantly delayed in *rrp6Δ* cells transferred to phosphate-free medium (**Fig. 5.10C and 5.10D**). These findings are in agreement with a previous study that showed much slower *pho1*$^{+}$ induction kinetics in cells lacking Rrp6 (Shah

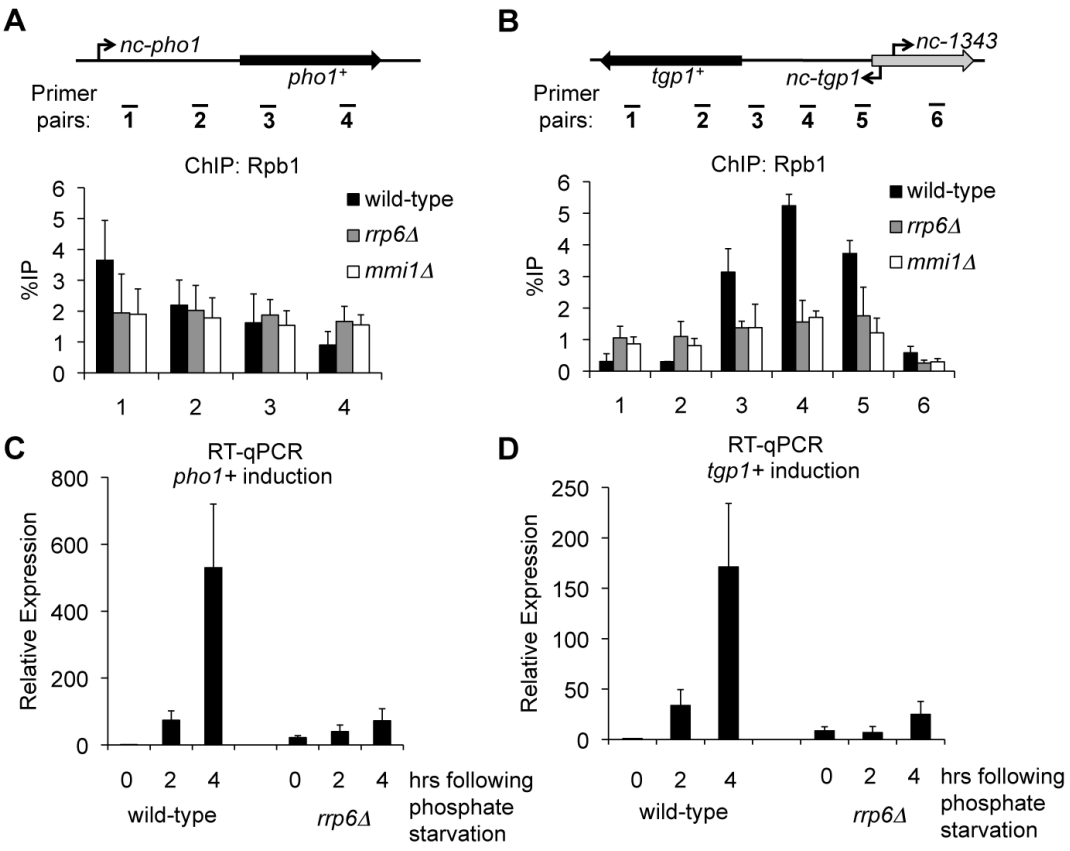

Figure 5.10. Rrp6 loss significantly attenuates induction of *pho1*$^+$ and *tgp1*$^+$. (A and B) Rbp1 ChIP-qPCR experiments performed in wild-type, *rrp6Δ*, and *mmi1Δ* cells. RT-qPCR experiments measured (C and D) *pho1*$^+$ mRNA and *tgp1*$^+$ mRNA induction in wild-type and *rrp6Δ* cells. Error bars represent SEM resulting from at least three independent replicates.

et al., 2014). Taken together, these findings suggest that exosome-mediated degradation of upstream lncRNAs might play role in *tgp1*$^{+}$ and *pho1*$^{+}$ induction following phosphate-starvation, but silencing by H3K9 methylation is unlikely to account for this delay.

5.3 Discussion

An increasing number of lncRNAs have been found to influence eukaryotic gene expression control in response to intra- and extra-cellular changes that require rapid, integrated responses at the level of transcription. While it is now well established that antisense transcription controls genes involved in various stress-response pathways in *S. pombe* (Bitton *et al.*, 2011; Leong *et al.*, 2014), the role of intergenic lncRNAs in the regulation of these or other pathways is understudied.

Recent studies in *S. pombe* have implicated certain nascent mRNAs and lncRNAs in gene repression by mechanisms involving transient RNAi–dependent and –independent heterochromatin formation (Hiriart *et al.*, 2012; Lee *et al.*, 2013; Zofall *et al.*, 2012). For example, the DSR-containing lncRNA transcribed upstream and overlapping the *pho1*$^{+}$ gene has been proposed to recruit Mmi1 and the RNAi machinery to locally deposit H3K9 methylation and thereby repress *pho1*$^{+}$ in response to phosphate availability (Shah *et al.*, 2014). However, these findings differ from genome-wide mapping which shows background levels of H3K9 methylation at *pho1*$^{+}$ and *tgp1*$^{+}$ (**Fig. 5.3**; Wang et al., 2015; Yamanaka *et al.*, 2013). In fact, these genes only accumulate RNAi-directed H3K9 methylation in mutants with defective RNA processing/degradation, not in wild-type cells grown under normal repressive conditions (**Fig. 5.9**; Yamanaka *et al.*, 2013). The significance of *rrp6Δ*-dependent

heterochromatin at *pho1*$^+$ and *tgp1*$^+$ genes is therefore unclear. Cells lacking Rrp6 accumulate aberrant RNAs and exhibit disrupted heterochromatin globally. In *rrp6Δ* cells, H3K9 methylation levels are significantly reduced at centromeres and increase elsewhere in euchromatin regions of the genome (Reyes-Turcu *et al.*, 2011; Yamanaka *et al.*, 2013), including at the housekeeping actin gene *act1*$^+$ (**Fig. 5.9**). For this reason, caution must be exercised when interpreting the analyses of mutants with such severe defects in RNA processing/degradation. Even low H3K9me2 levels at a subset of meiotic Mmi1-target genes (i.e. levels equivalent to or greater than those found at *pho1*$^+$ and *tgp1*$^+$ genes in *rrp6Δ* cells) have recently been shown to be insufficient to repress RNAPII transcription (Egan *et al.*, 2014). Instead, accumulating evidence seems to indicate that Mmi1, in concert with Red1, the exosome complex, and other accessory factors, primarily silence target DSR-containing genes at the post-transcriptional level, not by the formation of transient heterochromatin islands.

The absence of H3K9me2 enrichment on the *pho1*$^+$ and *tgp1*$^+$ promoters/genes in wild-type cells grown under repressive (phosphate-rich) conditions (**Fig. 5.2, 5.3, and 5.8**; Wang *et al.*, 2015; Yamanaka *et al.*, 2013), and the fact that *pho1*$^+$ and *tgp1*$^+$ expression is unaffected by loss of RNAi/heterochromatin (**Fig. 5.2 and 5.8**; Hansen *et al.*, 2005), are together wholly inconsistent with these genes being repressed by transient heterochromatin. Rather, the results presented in this chapter suggest that both *nc-tgp1* and *nc-pho1* mediate repression of downstream genes (*tgp1*$^+$ and *pho1*$^+$, respectively) by transcriptional interference (**Fig. 5.11**). This conclusion is based on the following findings: (i) *tgp1*$^+$ and *pho1*$^+$ expression is unaffected by loss of RNAi/heterochromatin; (ii) H3K9me2 is not detected at *tgp1*$^+$ or *pho1*$^+$ loci in wild-type cells; (iii) *nc-tgp1* and *nc-pho1* transcription declines upon

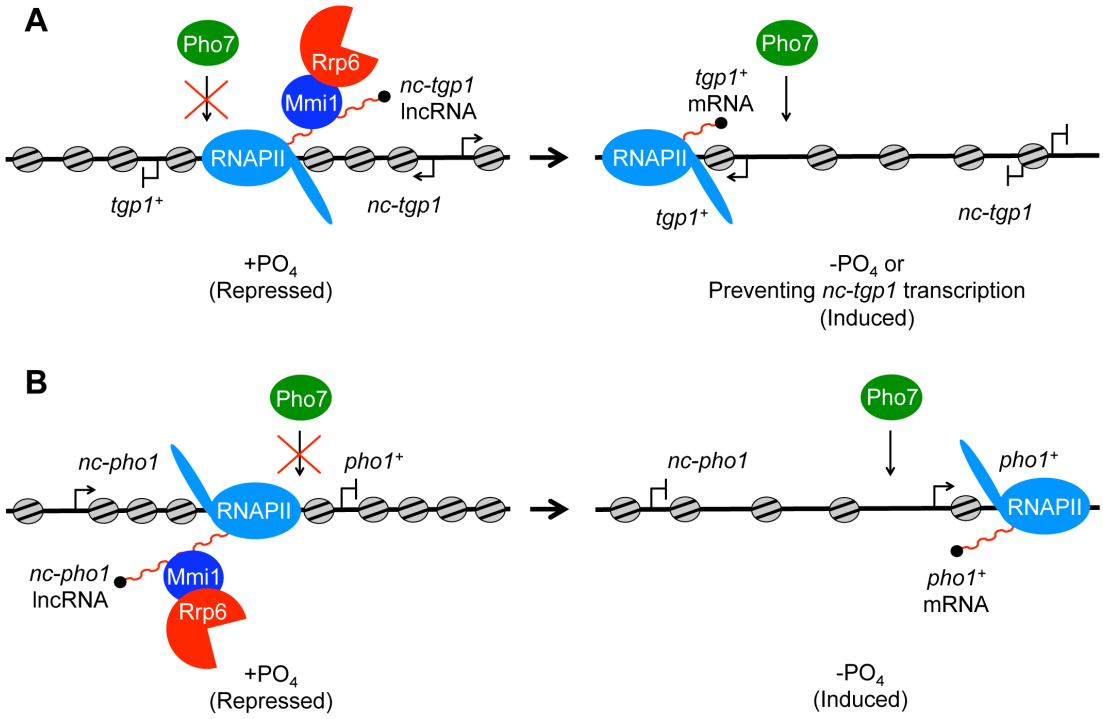

Figure 5.11. Model of transcriptional interference at *tgp1*$^+$ and *pho1*$^+$. The presence of phosphate induces the transcription unstable lncRNAs targeted by Mmi1/exosome degradation upstream of phosphate-regulated genes (A) *tgp1*$^+$ and (B) *pho1*$^+$. lncRNA transcription increases nucleosome density and occludes Pho7 transcription factor binding and thus represses downstream genes. lncRNA expression is reduced following phosphate starvation, decreasing nucleosome density, and allowing Pho7 to stably engage the promoter and induce expression.

tgp1$^+$ and *pho1*$^+$ induction (-PO_4); (iv) loss of lncRNA transcription upstream induces *tgp1*$^+$ and *pho1*$^+$ in repressive medium (+PO_4); (v) transcription of *nc-tgp1* by a thiamine-repressible promoter results in *tgp1*$^+$ being controlled by thiamine, rather than phosphate; (vi) RNAPII and nucleosome density is increased over *tgp1*$^+$ and *pho1*$^+$ promoters when the repressive upstream lncRNAs are transcribed; and (vii) the Pho7 activator binds the *tgp1*$^+$ and *pho1*$^+$ promoter regions when upstream lncRNA transcription is lost.

Transcriptional interference is well established in many systems. In the bacterium *Escherichia coli*, the gene encoding the *clr* transcription activator is repressed in response to nitrogen starvation by the act of lncRNA transcription from an alternate upstream promoter (Zafar *et al.*, 2014). In the single-celled yeast *S. cerevisiae*, non-coding transcription over the promoters of *SER3*, *IME1*, *GAL7*, and *FLO11* has been found to repress gene induction (Bumgarner *et al.*, 2009; Greger *et al.*, 2000; Martens *et al.*, 2004; van Werven *et al.*, 2012). Analogous mechanisms have been also reported in multicellular eukaryotes. The *Drosophila Ubx* gene, the human dihydrofolate reductase gene, and the imprinted *Igf2r* gene in mammals are all repressed independent of RNAi and heterochromatin by lncRNA transcription into their respective promoters (Latos *et al.*, 2012; Martianov *et al.*, 2007; Petruk *et al.*, 2006). These examples illustrate that transcriptional interference is a simple, conserved mechanism for modulating specific genes.

An outstanding question in the regulation of these two phosphate-regulated genes is the requirement of exosome-mediated degradation of upstream lncRNAs. It is difficult to entirely rule out a role for the RNA product in this mechanism since exosome recruitment by lncRNA-Mmi1 interactions appears to have an impact on *pho1*$^+$ and *tgp1*$^+$ activation following phosphate-starvation. One possible explanation

is that exosome-mediated degradation is simply required to clear high levels of these lncRNAs from chromatin since the accumulation of these transcripts increases the possibility of lncRNA-DNA duplex formation. Such duplexes between RNA and DNA, termed R loops, can have profound consequences on gene expression and are therefore tightly controlled (Skourti-Stathaki and Proudfoot, 2014). Future studies should investigate if exosome-mediated degradation directly influences the induction of *pho1*$^{+}$ and *tgp1*$^{+}$ or whether attenuated activation in *rrp6Δ* cells results from indirect effects. Deleting DSR-motifs from these lncRNAs should alleviate the concern of indirect effects owing to the loss of exosome activity and should therefore help to elucidate the significance of this regulation on *pho1*$^{+}$ and *tgp1*$^{+}$ activation. Importantly, even though the lncRNAs transcribed upstream of *pho1*$^{+}$ and *tgp1*$^{+}$ are rapidly degraded, the mere act of transcription is critical for regulation. Notably, these two genes represent the first documented examples of transcriptional interference in *S. pombe*.

Alleviating the repression of *pho1*$^{+}$ and *tgp1*$^{+}$ by transcriptional interference requires phosphate-starved cells to reduce repressive upstream lncRNA transcription. It is currently unknown how phosphate-starved cells accomplish this. The same genetic screen that identified Pho7 as a positive *pho1*$^{+}$ gene activator in *S. pombe* also identified the cyclin-dependent kinase activating kinase Csk1 as a negative regulator of *pho1*$^{+}$ activation in phosphate-replete conditions (Henry *et al.*, 2011). Cells lacking Csk1 have also been reported to exhibit reduced growth in the presence of drugs such as hydroxyurea and rapamycin (Hayles *et al.*, 2013; Doi *et al.*, 2015), and increased *tgp1*$^{+}$ levels in *csk1Δ* cells might at least partially account for this drug sensitivity phenotype. Therefore, Csk1 might be responsible for

silencing *pho1*$^{+}$ and *tgp1*$^{+}$ in phosphate-rich environments by stimulating upstream lncRNA transcription.

Csk1 prevents the full activation of the transcription factor Pho7 but does not directly regulate Pho7 promoter enrichment (Carter-O'Connell *et al.*, 2012). Basal Pho7 levels at the *pho1*$^{+}$ promoter have been shown to be sufficient to induce expression in *csk1Δ* cells (Carter-O'Connell *et al.*, 2012), analogous to the finding here that stable Pho7 levels accumulate at the *tgp1*$^{+}$ promoter in the absence of *nc-tgp1* transcription (**Fig. 5.4**). Importantly, prolonged phosphate limitation leads to further increases in Pho7 promoter binding and stimulates *pho1*$^{+}$ and *tgp1*$^{+}$ induction beyond the levels detected in phosphate-replete cells lacking transcriptional interference (**Fig. 5.1**; Carter-O'Connell *et al.*, 2012; Shah *et al.*, 2014). It is therefore conceivable that Csk1 signaling through an unknown pathway stimulates repressive lncRNA transcription upstream of *pho1*$^{+}$ and *tgp1*$^{+}$, and this activity is somehow lost when cells are starved of phosphate. Decreased lncRNA transcription over the *pho1*$^{+}$ and *tgp1*$^{+}$ gene promoters stabilizes Pho7 binding.

Finally, the regulation of phosphate-response genes by lncRNAs is not limited to *S. pombe*. Transcription factor binding to the promoter of *PHO5*, the *S. cerevisiae* homolog of *pho1*$^{+}$, is obstructed by increased nucleosome density in phosphate-rich conditions (Venter *et al.*, 1994). However, unlike transcriptional repression of *pho1*$^{+}$ by an interfering lncRNA in *S. pombe*, antisense lncRNA transcription is thought to be needed to reposition nucleosomes within the *PHO5* promoter in order to favour *PHO5* expression in *S. cerevisiae* cells starved of phosphate (Uhler *et al.*, 2007). In addition, repression of a different phosphate-response gene in *S. cerevisiae*, termed *PHO84*, results from the recruitment of HDAC activity by antisense lncRNA

transcription in phosphate-replete conditions (Camblong *et al.*, 2007). There is also evidence in multicellular organisms for lncRNA-dependent repression of phosphate-regulated genes. In *Arabidopsis*, the *PHO2* gene is suppressed in phosphate-rich environments by the microRNA miR399 (Bari *et al.*, 2006). Phosphate starvation in this plant induces the expression of *IPS1*, an lncRNA that acts as a target decoy for miR399 and allows *PHO2* mRNA levels to accumulate (Franco-Zorrilla *et al.*, 2007). Additionally, phosphate starvation in the rice plant *Oryza sativa* leads to the expression of an antisense lncRNA at the *PHO1;2* gene that promotes translation of the *PHO1;2* mRNA, a central component of the phosphate response in this organism (Jabnoune *et al.*, 2013). Collectively these studies show that different unicellular eukaryotes and sessile multicellular organisms utilize diverse lncRNA-dependent regulatory mechanisms to maintain phosphate homeostasis.

CHAPTER 6

tgp1^{+} homologs in related fission yeast species are not regulated by transcriptional interference

6.1 Introduction

Inorganic phosphate is an essential nutrient required by all living organisms. Maintaining stable cellular phosphate levels is often a challenge for microorganisms and multicellular organisms alike since inorganic phosphate availability can fluctuate unpredictably. To combat this challenge, organisms have evolved complex strategies to sense extracellular phosphate levels and communicate this information into a transcriptional response (Bergwitz and Jüppner, 2011). The transcriptional response required to maintain phosphate homeostasis in eukaryotic cells is best understood in budding yeast *S. cerevisiae*, and to a lesser degree in fission yeast *S. pombe*. Despite being separated by hundreds of millions of years of evolution (Hedges, 2002), these two unicellular fungi have evolved parallel signal transduction pathways that respond to phosphate limitation by inducing a conserved core regulon (Carter-O'Connell *et al.*, 2012).

In *S. cerevisiae*, the transcriptional response following the exposure of cells to low phosphate availability is mediated by the transcription factor Pho4. When extracellular phosphate is plentiful, the Pho85-Pho80 complex phosphorylates Pho4, which is thought to inactivate Pho4 and retain it in the cytoplasm (O'Neill *et al.*, 1996). When phosphate levels are depleted, however, the Pho85-Pho80

complex is inhibited, which allows the unphosphorylated form of Pho4 to accumulate in the nucleus and induce phosphate-response genes that help scavenge inorganic phosphate from the environment (Schneider *et al.*, 1994; Kaffman *et al.*, 1998). Core components of the phosphate regulon in *S. cerevisiae* include the secreted acid phosphatase gene *PHO5*, the inorganic phosphate transporter gene *PHO84*, and the glycerophosphodiester membrane permease gene *GIT1* (Almaguer *et al.*, 2003; Thomas and O'Shea, 2005). Likewise, *S. pombe* cells adjust to phosphate starvation by inducing a core phosphate regulon comprising *pho1*$^{+}$, *pho84*$^{+}$, and *tgp1*$^{+}$, homologs of *S. cerevisiae PHO5*, *PHO84,* and *GIT1*, respectively (Carter-O'Connell *et al.*, 2012). The transcriptional response to phosphate limitation in *S. pombe*, however, is achieved by a non-homologous signal transduction pathway and is activated by the transcription factor Pho7 (Carter-O'Connell *et al.*, 2012; Henry *et al.*, 2011), which lacks an ortholog in *S. cerevisiae*. Unlike the Pho4 transcription factor in *S. cerevisiae*, which is retained in the cytoplasm of cells that are grown in the presence of phosphate (O'Neill *et al.*, 1996), findings presented in Chapter 5 suggest that the *S. pombe* transcription factor Pho7 is able to activate the transcription of target genes *tgp1*$^{+}$ and *pho1*$^{+}$ in repressive, phosphate-rich conditions, provided upstream repressive lncRNA transcription is lost. Instead, lncRNA transcription over *tgp1*$^{+}$ and *pho1*$^{+}$ promoters is required to prevent stable Pho7 binding and subsequent gene activation. It is therefore plausible that this transcriptional interference mechanism might be preserved to regulate phosphate-response genes in related fission yeast species.

6.2 Results

6.2.1 *tgp1+ orthologs in different Schizosaccharomyces species*

The phosphate response pathways of *S. octosporus*, *S. cryophilus*, and *S. japonicus* have yet to be characterized. Since the *S. pombe* phosphate regulon is conserved in budding yeast *S. cerevisiae*, one might predict that other *Schizosaccharomyces* species have a similar core regulon. Curiously, however, *pho1*$^{+}$ homologs could not be identified in the genomes of *S. octosporus*, *S. cryophilus*, or *S. japonicus*. In addition, another phosphate-regulated gene in *S. pombe*, *pho84*$^{+}$, appears to have been lost in the *S. octosporus* and *S. cryophilus* lineage. Another striking difference between *S. pombe* and related species is that *S. octosporus*, *S. cryophilus*, and *S. japonicus* each appear to have more than one copy of the *tgp1*$^{+}$ gene (**Fig. 6.1A and Fig. 6.1B**). It is currently unclear whether these represent true orthologs that originated from gene duplication following speciation. Importantly, some of these putative *tgp1*$^{+}$ orthologs and paralogs are reported to have stable, divergent lncRNA transcription upstream (**Fig. 6.1C**). Indeed, this conservation of lncRNAs upstream of *tgp1*$^{+}$ genes was the principle criterion for selecting the *S. pombe ncRNA.1343* gene for deletion in Chapter 3. It is therefore plausible that syntenic transcripts represent stable byproducts of bidirectional promoters that primarily drive unstable lncRNA transcription over *tgp1*$^{+}$ promoters in these organisms, homologous to *nc-tgp1* repression of *tgp1*$^{+}$ in *S. pombe*.

The analysis of previously published strand-specific RNA-seq datasets revealed that *tgp1*$^{+}$ copies *SOCG_04583* in *S. octosporus* and *SPOG_03676* in *S. cryophilus* are constitutively expressed in cells grown in normal, phosphate-containing medium (**Fig. 6.2A and 6.2B**; Rhind *et al.*, 2011). According to phylogenetic analysis of the

A

	tgp1⁺ ortholog	Identity
S. octosporus	*SOCG_04583*	68%
	SOCG_01135	67%
S. cryophilus	*SPOG_03676*	67%
	SPOG_01604	67%
S. japonicus	*SJAG_05325*	63%
	SJAG_03644	59%
	SJAG_00232	59%

B

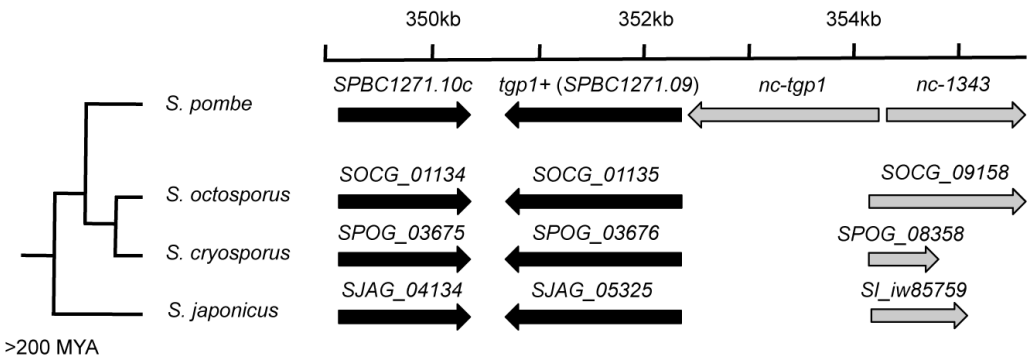

Figure 6.1. lncRNA transcription upstream of *tgp1*⁺ homologs in related fission yeast species. (A) Table displaying *tgp1*⁺ homologs in other *Schizosaccharomyces* species, including percentage amino acid similarity with *S. pombe tgp1*⁺ (Identity). Highlighted in grey are the copies of *tgp1*⁺ with putative lncRNA transcription detected upstream (Rhind *et al.*, 2011). (B) Phylogenetic tree and schematic representation of *tgp1*⁺ genes with upstream lncRNA transcription in different fission yeast species (Rhind *et al.*, 2011).

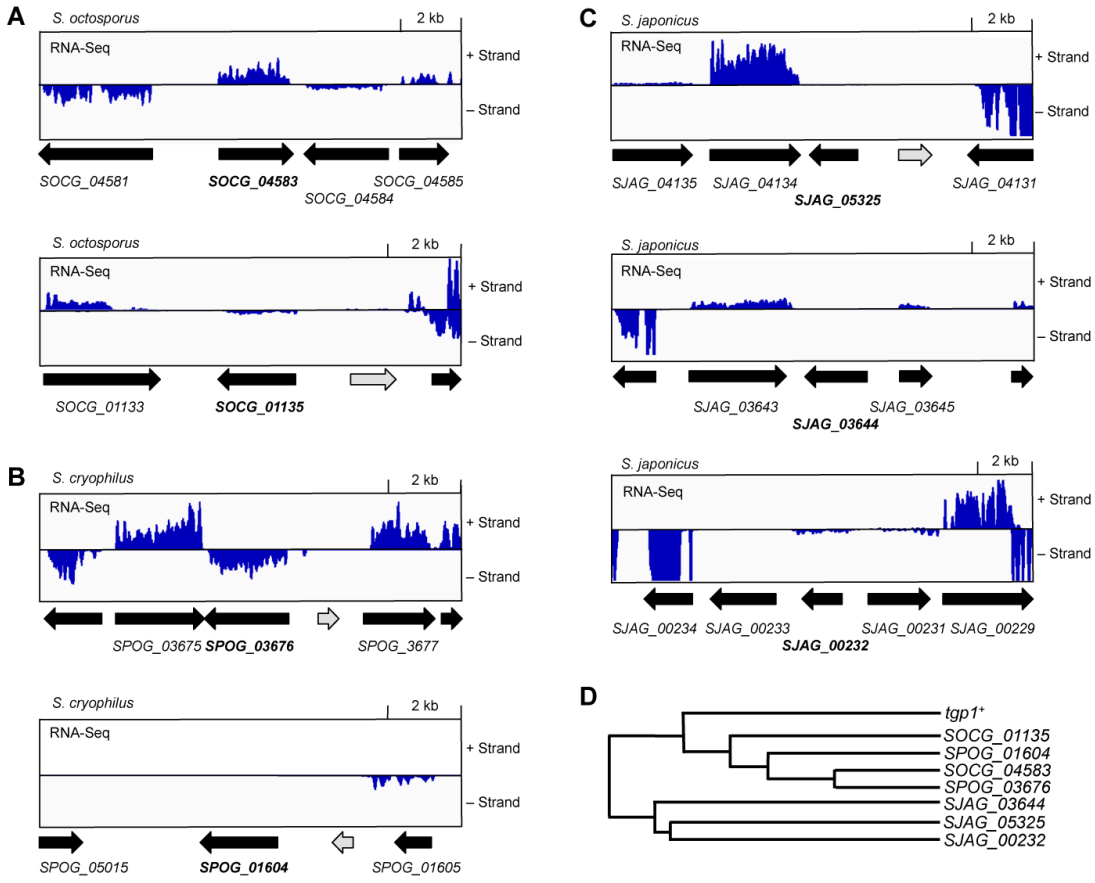

Figure 6.2. Transcription profiles for *tgp1*$^{+}$ orthologs. (A-C) Previously published strand-specific RNA-seq analyses in *S. cryophilus*, *S. octosporus*, and *S. japonicus* showing some *tgp1*$^{+}$ orthologs/paralogs are expressed while others are repressed (Rhind *et al.*, 2011). Black arrows indicate protein-coding genes, while grey arrows represent predicted lncRNA genes. Bioinformatic analyses performed by Dr. Pin Tong. (D) *tgp1*$^{+}$ gene family analysis in the genus *Schizosaccharomyces* (Rhind *et al.*, 2011).

Schizosaccharomyces tgp1^{+} gene family (Rhind *et al.*, 2011), *SOCG_04583* and *SPOG_03676* originated from a *tgp1*$^{+}$ gene duplication event unique to the *S. octosporus* and *S. cryophilus* lineage (**Fig. 6.2D**). On the other hand, repressed genes *SOCG_01135* in *S. octosporus* and *SPOG_01604* in *S. cryophilus* are more closely related to the ancestral *Schizosaccharomyces tgp1*$^{+}$ gene. Uniquely, two *tgp1*$^{+}$ duplications appear to have occurred in *S. japonicus*, with the most ancestral copy of *tgp1*$^{+}$ predicted by this analysis to be the repressed *S. japonicus* gene *SJAG_03644* (**Fig. 6.2D**).

The *S. octosporus SOCG_01135* gene resides in a region of conserved synteny, including a predicted lncRNA conserved in position upstream (**Fig. 6.1B**). While *S. cryophilus SPOG_01604* is also downstream of a predicted lncRNA locus, this gene does not share gene order with *tgp1*$^{+}$ in *S. pombe*. In *S. japonicus*, two copies of *tgp1*$^{+}$ (*SJAG_00232* and the more ancestral *SJAG_03644*) are not present in a region of conserved gene order (**Fig. 6.2C**). Instead, the synteny conserved *SJAG_05325* might in fact be more closely related to *S. pombe tgp1*$^{+}$ than *SJAG_03644*. Accordingly, blastp analyses identified greater amino acid sequence homology between *S. pombe tgp1*$^{+}$ and *S. japonicus SJAG_05325* (**Fig. 6.1A**). Unlike *tgp1*$^{+}$ genes in *S. octosporus* and *S. cryophilus*, all three copies of *tgp1*$^{+}$ in *S. japonicus* are repressed in rich growth medium (**Fig. 6.2C**). Finally, published H3K9me2 ChIP analyses indicate that this heterochromatin mark is absent from all *tgp1*$^{+}$ genes in *S. octosporus*, *S. cryophilus*, and *S. japonicus* (**Fig. 6.3**), consistent with H3K9 methylation not having a role in *S. pombe tgp1*$^{+}$ regulation. Notably, this analysis also revealed that *mei4*$^{+}$ genes in *S. octosporus*, *S. cryophilus*, and *S. japonicus* do not accumulate H3K9 methylation heterochromatin islands, as is proposed for *S. pombe mei4*$^{+}$ (Hiriart *et al.*, 2012; Zofall *et al.*, 2012).

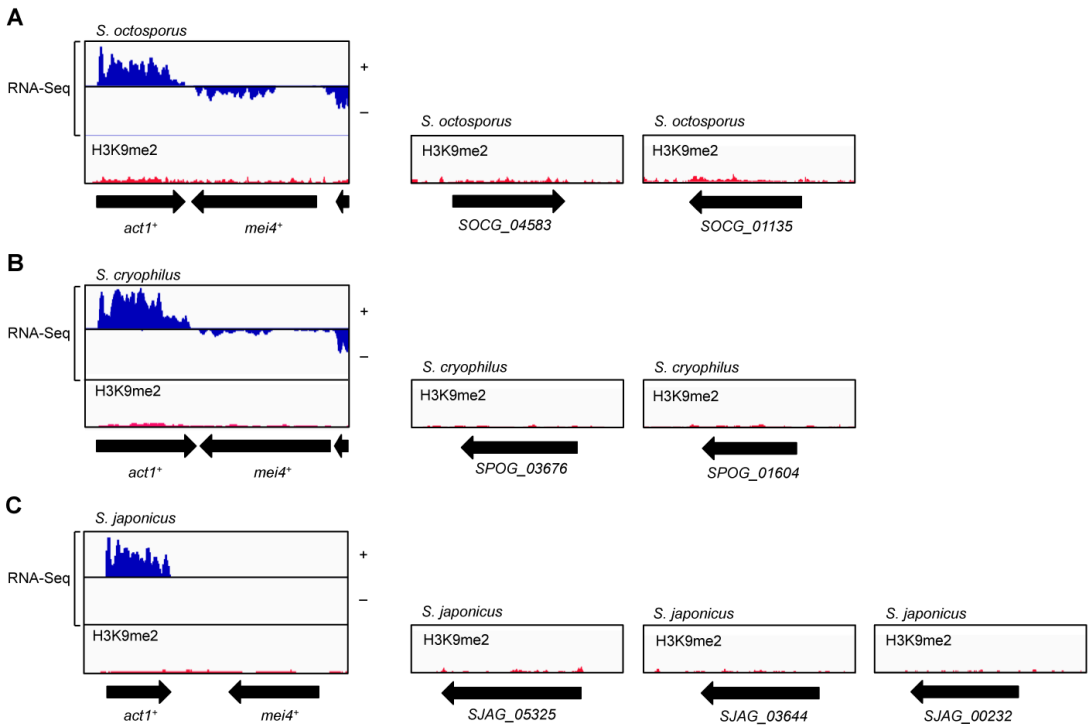

Figure 6.3. H3K9 methylation is not detected at *tgp1*$^{+}$ orthologs. (A-C) Previously published RNA-seq and genome-wide H3K9me2 mapping in *S. cryophilus*, *S. octosporus*, and *S. japonicus* (Rhind *et al.*, 2011) showing no significant levels of H3K9 methylation at *tgp1*$^{+}$ orthologs/paralogs nor at *mei4*$^{+}$ orthologs in these organisms. Bioinformatic analyses performed by Dr. Pin Tong.

This observation suggests that heterochromatin islands might not be conserved between different fission yeast species.

6.2.2 *No evidence of transcription upstream of tgp1$^+$ in S. octosporus*

S. pombe tgp1$^+$ and the *S. cerevisae* homolog *GIT1* are repressed by the presence of extracellular phosphate and induced when external phosphate levels are depleted. *S. octosporus* cells were grown in phosphate rich (+PO_4) and phosphate deprived (-PO_4) conditions to determine whether *SOCG_01135*, the repressed copy of *tgp1$^+$* in *S. octosporus*, responds to changes in phosphate availability. RT-qPCR analysis showed that *SOCG_01135* transcript levels do indeed accumulate in phosphate-starved cells (**Fig. 6.4A**). In contrast, the constitutively expressed copy of *tgp1$^+$* in *S. octosporus*, *SOCG_04583*, failed to respond to phosphate starvation (**Fig. 6.4B**). These results suggest that *SOCG_04583* is likely to have evolved a new function after duplication, which might also explain its lower amino acid sequence conservation. Consistent with *SOCG_01135* induction following phosphate starvation, Rpb1 ChIP detected increased levels of RNAPII over the *SOCG_01135* gene in phosphate-depleted conditions (**Fig. 6.4D**). RNAPII levels at a control gene remained unaffected by phosphate starvation (**Fig. 6.4E**). Importantly, unlike the profile of RNAPII occupancy observed at the *tgp1$^+$* promoter in *S. pombe* (**Fig. 4.2B**), RNAPII levels over the region upstream of *SOCG_01135* were relatively low and did not significantly change after starving cells of phosphate (**Fig. 6.4D**). Given that transcriptional interference mechanisms require high levels of RNAPII transcription to effectively silence a downstream gene (Palmer *et al.*, 2011), low levels of RNAPII transcription over the *SOCG_01135* promoter suggests that this *tgp1$^+$* gene is not regulated by transcriptional interference. Attempts at performing endogenous genetic manipulations in *S. octosporus* were unsuccessful,

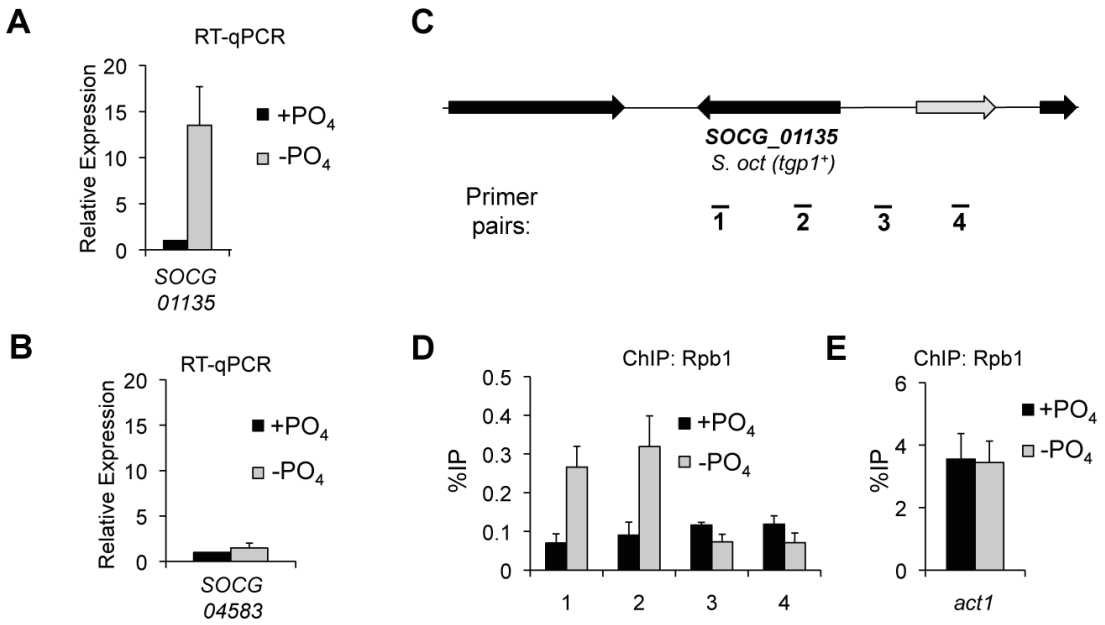

Figure 6.4. No evidence of repressive transcription over the *tgp1*$^+$ promoter in *S. octosporus*. (A) RT-qPCR experiments measured *tgp1*$^+$ homolog *SOCG_01135* mRNA levels in *S. octosporus* cells grown in phosphate-rich medium (+PO_4) or in the absence of phosphate (-PO_4). (B) RT-qPCR experiments measured the mRNA levels of putative paralog *SOCG_04583* in *S. octosporus* cells grown in response to changing phosphate availability. (C) Schematic representation of the *SOCG_01135* and depictions of primer pair locations. (D) Rbp1 ChIP-qPCR experiments performed in *S. octosporus* cells grown in the presence or absence of phosphate. (E) Rbp1 ChIP-qPCR controls experiments at the *S. octosporus act1*$^+$ locus. Error bars represent standard deviation resulting from two biological replicates, each done in technical triplicate.

making it difficult to identify the mechanism by which this phosphate-regulated gene is repressed in phosphate-rich conditions.

6.2.3 *S. japonicus tgp1$^+$ is not regulated by transcriptional interference*

All copies of *tgp1*$^+$ found in *S. japonicus* are repressed in normal growth conditions. To examine whether one or more of these genes is induced by phosphate starvation, *S. japonicus* cells were grown in phosphate rich (+PO_4) and phosphate deprived (-PO_4) conditions. RT-qPCR analysis revealed that the *SJAG_05325* gene was not significantly induced in response to phosphate starvation (**Fig. 6.5A**), despite sharing synteny and greater sequence homology with *S. pombe tgp1*$^+$ than other *tgp1*$^+$ copies. This suggests that *SJAG_05325* is likely to have evolved a new function independent of the phosphate response. The *SJAG_00232* gene also failed to respond to changes in phosphate availability (**Fig. 6.5A**). Only the *SJAG_03644* gene showed increased expression levels in phosphate-depleted conditions (**Fig. 6.5A**). RNAPII occupancy, as measured by Rpb1 ChIP, showed no significant level of transcription over the *SJAG_03644* promoter (**Fig. 6.5B**). Take together, these results rule out transcriptional interference as a mechanism for repressing *SJAG_03644* in phosphate-replete conditions. Future manipulations of this locus are required in order to identify how this gene is regulated at the transcriptional level.

6.2.4 *S. cerevisiae GIT1 is not regulated by transcriptional interference*

The budding yeast homologs of *pho1*$^+$ (*PHO5*) and *tgp1*$^+$ (*GIT1*) have previously been shown to respond to phosphate availability (Almaguer *et al.*, 2003). Antisense transcription at the *PHO5* locus reorganizes nucleosomes in the *PHO5* promoter to permit gene activation in phosphate-starved cells (Uhler *et al.*, 2007). It is not yet known whether non-coding transcription also influences *GIT1* induction. RT-qPCR experiments confirmed that *PHO5* and *GIT1* are significantly induced following

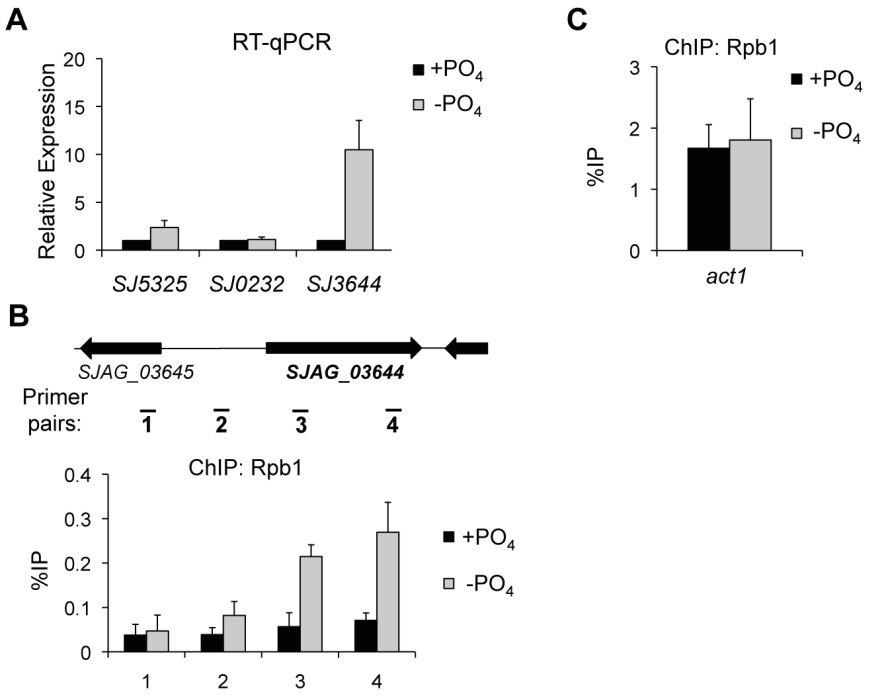

Figure 6.5. *tgp1*$^{+}$ homolog in *S. japonicus* is not repressed by upstream transcription. (A) RT-qPCR experiments measured mRNA levels of *tgp1*$^{+}$ copies *SJAG_05325*, *SJAG_00232*, and *SJAG_03644* mRNA levels in *S. japonicus* cells grown in phosphate-rich medium (+PO_4) or in the absence of phosphate (-PO_4). (B) Schematic representation of the *SJAG_03644* locus and depictions of primer pair locations and Rbp1 ChIP-qPCR experiments performed in *S. japonicus* cells grown in the presence or absence of phosphate. (C) Rbp1 ChIP-qPCR controls experiments at the *S. japonicus act1*$^{+}$ locus. Error bars represent standard deviation resulting from two biological replicates, each done in technical triplicate.

phosphate starvation (**Fig. 6.6A**). The nearest protein-coding ORF is located over 2 kb upstream of *GIT1*. Since the *S. cerevisiae* genome is highly condensed, this large intergenic region is unusual and might contribute to the regulation of *GIT1*. Indeed, transcriptional interference of *S. pombe tgp1*$^{+}$ occurs over 2 kb region upstream. However, unlike the pattern of RNAPII observed upstream of *S. pombe tgp1*$^{+}$, RNAPII levels over the *GIT1* promoter actually increased in phosphate-starved cells (**Fig. 6.6B**). It is therefore clear that upstream transcription does not repress *GIT1*. Instead, upstream transcription might favour induction. Since these experiments do not detect strand specificity it is unclear whether this is tandem upstream transcription or divergent transcription originating from the activated *GIT1* promoter. Alternatively, it is possible that a mechanism related to *PHO5* regulation requiring antisense transcription might be involved. Due to time limitations, these possibilities were not investigated in greater detail. However, the preliminary data obtained suggest that the budding yeast homolog of *tgp1*$^{+}$ is not regulated by transcriptional interference. Future work is therefore required to compare and contrast the regulatory mechanisms responsible for regulating genes involved in the phosphate response in *S. cerevisiae*.

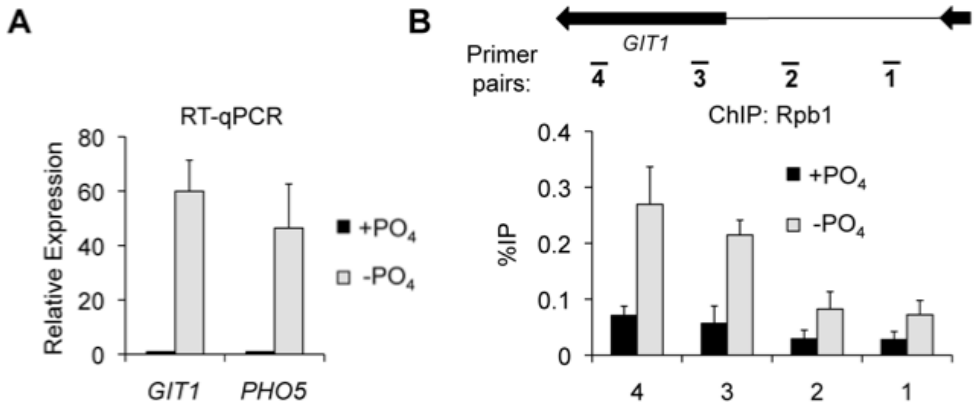

Figure 6.6. *S. cerevisiae GIT1* is not regulated by transcriptional interference. (A) RT-qPCR experiments measured mRNA levels of phosphate-regulated *S. cerevisiae* genes *GIT1* and *PHO5* in cells grown in fully supplemented SD medium or SD medium lacking phosphate. (B) Schematic representation of the *GIT1* gene, including 2 kb of intergenic space upstream. Below, primer pair locations and Rbp1 ChIP-qPCR experiments performed in *S. cerevisiae* cells grown in the presence or absence of phosphate. Error bars represent standard deviation resulting from two biological replicates, each done in technical triplicate.

6.3 Discussion

In natural environments, organisms are frequently exposed to suboptimal nutrient levels. To survive, cells sense fluctuations in the availability of essential nutrients and implement rapid responses by eliciting rapid and highly integrated changes in gene expression. Studies utilizing the budding yeast *S. cerevisiae* have revealed that different nutrient signals elicit common transcriptional responses, often triggering a transient quiescent state, while specific genes are also induced to overcome specific nutrient deficiencies (Conway *et al.*, 2012). In the case of phosphate starvation, it is remarkable that organisms as distantly related as *S. cerevisiae* and *S. pombe* have maintained an evolutionarily conserved core regulon to overcome reduced phosphate availability (Carter-O'Connell *et al.*, 2012). Despite the conservation of genes induced by phosphate starvation, these two organisms have evolved markedly different signal transduction pathways to sense external phosphate levels and integrate that information into a transcriptional response. This observation supports the notion that signaling pathways responsible for regulating phosphate homeostasis in these organisms are far more malleable to change over the course of evolution than the genetic response itself.

Mechanisms of transcriptional regulation also appear to be malleable. The *pho1*$^{+}$ and *tgp1*$^{+}$ genes in *S. pombe* are silenced in phosphate-rich conditions by transcriptional interference to prevent Pho7 transcription factor binding (Chapter 5). However, their homologs in *S. cerevisiae*, *PHO5* and *GIT1,* might not require such a mechanism as the Pho4 transcription factor crucial to the phosphate response in this organism is generally thought to be sequestered in the cytoplasm when extracellular phosphate levels are high (O'Neill *et al.*, 1996; Kaffman *et al.*, 1998).

However, it is known that antisense transcription at the *PHO5* locus corresponds with gene induction in phosphate-starved *S. cerevisiae* cells by reorganizing nucleosomes in the *PHO5* promoter (Uhler *et al.*, 2007), while silencing of a different phosphate-response gene, *PHO84*, requires antisense transcription in phosphate-rich conditions (Camblong *et al.*, 2007; Castelnuovo *et al.*, 2013). These findings indicate that the presence of phosphate alone is not sufficient to repress these two genes in *S. cerevisiae*. Indeed, there is some evidence that Pho4 can localize to the nucleus in phosphate-rich conditions and regulate multiple genes by inducing antisense and intergenic lncRNA transcription (Nishizawa *et al.*, 2008). Moreover, Pho4 has also been reported to play a role in mediating the transcriptional response to glucose, phosphate, and nitrogen limitation in *S. cerevisiae* (Conway *et al.*, 2012). These findings support the idea that different post-translational modifications might modulate Pho4 activity and selectivity in response to different nutrient deficiencies (Springer *et al.*, 2003). It is therefore evident that further experimental analyses are still required to fully characterize the signaling events and mechanisms of transcriptional regulation in *S. cerevisiae* that are responsible for countering nutrient starvation.

The phosphate-regulated *tgp1*$^{+}$ gene in *S. octoporus* appears not to be regulated by transcriptional interference. Upstream RNAPII levels were low over this region and did not respond to changes in external phosphate levels. Experiments performed in *S. japonicus* cells show that the phosphate-regulated *tgp1*$^{+}$ gene in this organism lacks any detectible upstream lncRNA transcription, ruling out transcriptional interference as a regulatory mechanism. An inability to effectively manipulate genetic loci in *S. octosporus*, *S. cryophilus, S. japonicus* hindered further analyses of *tgp1*$^{+}$ regulation in these species. However, the preliminary findings presented

here are sufficient to conclude that transcriptional interference is unlikely to be involved in regulating *tgp1*$^{+}$ orthologs in these species.

Future genetic manipulations of endogenous loci encoding *tgp1*$^{+}$ genes in other fission yeasts will be required to dissect the differences in the regulatory mechanisms responsible for controlling the phosphate response in these organisms. It is also worth exploring whether *tgp1*$^{+}$ regulation by transcriptional interference is preserved in different natural isolates of *S. pombe* (Jeffares *et al.*, 2015). Beyond *tgp1*$^{+}$ regulation, it is surprising that *pho1*$^{+}$ homologs are absent in *S. octosporus*, *S. cryophilus*, and *S. japonicus*, and that *pho84*$^{+}$ is missing in the *S. octosporus* and *S. cryophilus* lineage. It is possible that sequencing or genome assembly errors might have caused the omission of these genes in the database. However, going on currently available data, the absence of *pho1*$^{+}$ and/or *pho84*$^{+}$ orthologs implies these species have evolved alternative strategies to harvest inorganic phosphate from low phosphate environments. The most effective way of studying the phosphate response in these organisms would be to grow each fission yeast species in both phosphate-rich and phosphate-starved conditions and measure the genome-wide transcriptional response to phosphate limitation in these related organisms. Since the transcription factor Pho7 is conserved in all fission yeast genomes, but absent in budding yeast *S. cerevisiae* genome, it is also worth investigating whether the signaling pathways that stimulate the phosphate response in *S. pombe* are functionally conserved across the *Schizosaccharomyces* clade.

CHAPTER 7

Discussion

7.1 *Assigning function to lncRNAs*

Eukaryotic genomes produce an abundance of lncRNAs transcribed antisense to protein-coding genes, from within introns, as well as from regions of the genome previously thought to be transcriptionally silent (Ponting *et al.*, 2009). Although it is still unclear what proportion of the lncRNAs detected in various organisms serve genuine biological functions, substantial progress has been made to assign function to many individual lncRNAs. However, this has not been a trivial task. New studies regularly overturn the interpretations of previous ones (Cech and Steitz, 2014). Even the mechanism by which the *Xist* lncRNA initiates X-inactivation in mammals is a matter of ongoing debate (Cerase *et al.*, 2015), despite having first been discovered in the early 1990s (Kay *et al.*, 1993). One model posits that *Xist* interacts with and recruits PRC2 (Zhao *et al.*, 2008), which deposits H3K27me on the inactive X chromosome (Plath *et al.*, 2003). Although PRC2 has been proposed to interact with *Xist* and many other lncRNAs (Khalil *et al.*, 2009; Zhao *et al.*, 2008), more thorough analyses found that PCR2 binds RNA non-specifically in many common assays (Davidovich *et al.*, 2013). These findings have introduced some doubt as to the significance of previously reported interactions between PRC2 and different lncRNAs, including *Xist*. In fact, it has recently been demonstrated by super-resolution microscopy that PRC2 and *Xist* are spatially separated in cells (Cerase *et al.*, 2014), arguing against the direct recruitment model. Even more recently, two independent groups could not detect a direct interaction between *Xist* and PRC2

(Chu *et al.*, 2015; McHugh *et al.*, 2015). Instead, *Xist* appears to initiate X-inactivation by interacting with a protein called SHARP that directs HDACs to the X chromosome targeted for inactivation (McHugh *et al.*, 2015). Importantly, HDAC recruitment by *Xist*/SHARP precedes PRC2 recruitment. While PRC2 reinforces silencing of the inactive X chromosome in female mammals, it is still unclear how it is recruited. After more than two decades of research into *Xist* function there are still many unanswered questions.

Predictably, many of the lncRNAs identified in recent years have also suffered similar disputes regarding their functional significance and mechanisms of action. Notably, the loss of the *HOTAIR* lncRNA in mouse was first reported to have no significant effect on *HOXD* regulation or development (Schorderet and Duboule, 2011), suggesting the *trans* function that had been reported for human *HOTAIR* is not conserved. However, later studies found evidence to the contrary and proposed that the *trans* function is indeed conserved in mouse (Lai *et al.*, 2015; Li *et al.*, 2013). In these conflicting studies the mouse *HOTAIR* gene had been disrupted using different strategies, which the authors argue might account for the contradictory conclusions reached. This explanation, while not particularly satisfying, is telling since it highlights the complexity and consequences of examining lncRNA function *in vivo*. Similar controversies have also emerged after deletion of the transcription units encoding other well-characterized lncRNAs, such as *MALAT1*, *Kcnq1ot1*, and *NEAT1*, resulted in less dramatic or even undetectable phenotypes in animal models (Eißmann *et al.*, 2012; Korostowski *et al.*, 2012; Nakagawa *et al.*, 2011; Nakagawa *et al.*, 2012; Zhang *et al.*, 2012). Although one cannot rule out the possibility that additional factors may act redundantly and compensate for the loss of these lncRNAs in the context of whole organisms, it is still unclear what proportion of the lncRNAs detected in high-throughput genome-wide studies have

real biological roles in organisms. These controversies also raise concerns about assigning function to lncRNAs by methods relying principally on over-expression and/or RNAi knockdown in cells. Future attempts to characterize lncRNAs must therefore utilize complementary approaches to rule out/in specific functions.

In this thesis, the preliminary characterization of synteny conserved intergenic lncRNAs in *S. pombe* revealed that deleting some of the most conserved lncRNAs in this organism had little effect on normal cell growth or viability (Chapter 3). Although loss of the *ncRNA.1343* gene rendered cells hypersensitive to various compounds, no obvious phenotypes emerged from other lncRNA deletions performed here. However, this work is not exhaustive and numerous other conditions/stresses need to be tested in order to identify other possible phenotypes emerging from loss of these and other lncRNAs. Interestingly, a recent study found that relatively small genetic differences in natural isolates of *S. pombe*, such as single nucleotide polymorphisms (SNPs) and insertions/deletions, account for clear phenotypic differences when exposing these strains to a wide spectrum of stresses (Jeffares *et al.*, 2015). Combining this type of large-scale phenotypic screening approach with an intergenic lncRNA deletion library will no doubt accelerate the discovery of functional lncRNAs in *S. pombe*. Such an unbiased approach would also be useful since non-conserved lncRNAs unique to *S. pombe* might have recently emerged as functional transcripts. Any phenotypes associated with the loss of a specific lncRNA gene will require further experimental validation to reduce the ambiguity and confusion that can result from the failure to perform detailed locus-specific analyses.

Some of the strategies required to properly characterize lncRNAs include the reliable identification of transcription start and stop sites, along with possible introns,

and accurate measurement of transcript abundance and regulation. Identifying subcellular localization patterns and protein partners for stable lncRNAs is also required. For those lncRNAs implicated in gene expression regulation, it is important that experiments are designed to distinguish between effects that might arise as a consequence of lncRNA transcription from those played by the lncRNA product. Importantly, endogenous manipulations of lncRNA genes should be made that prevent lncRNA transcription while limiting the disruption of any overlapping DNA elements. Such manipulations might include deleting/altering/swapping promoters and/or truncating transcripts by inserting transcriptional stop sequences or ribozyme sites. The development of CRISPR-Cas9 systems for rapid genome editing has made such targeted genetic manipulations much easier to perform in diverse organisms, including higher eukaryotes (Sander and Young, 2014). There is no doubt that this powerful new technology provides the tools needed to better understand lncRNA function *in vivo*. The possibility that any given lncRNA might act in *trans* should be tested by exogenously expressing the lncRNA from a plasmid or a distant locus, while genome-wide transcript levels must be profiled in loss and gain of function approaches. In addition, *trans*-acting lncRNA localization should also be confirmed by microscopy and/or methods that provide insight into the three-dimensional structure of chromosomes, such as chromosome conformation capture (3C) and variants thereof (Ay and Noble, 2015), to identify whether a gene encoding a *trans*-acting lncRNA might actually be positioned in close proximity to target genes located elsewhere on the same or different chromosome. Together, such strategies should help to pinpoint lncRNA functions and control for indirect effects that might result from any individual method.

Many of the concerns described above were taken into consideration while following up the observation that *ncRNA.1343* loss reduced *S. pombe* growth in the presence

of various compounds (Chapter 4). Detailed analyses were required to determine that deleting the *ncRNA.1343* gene removed a bidirectional promoter that, in addition to generating the stable *nc-1343* lncRNA, initiates the transcription of a previously unannoted, exosome-sensitive lncRNA transcribed in the opposite orientation (*nc-tgp1*). Additional analyses, including strategic genetic manipulations, were needed to characterize the transcripts produced from this bidirectional promoter and to explore their influence on *tgp1*$^{+}$ regulation. Ultimately, these experiments revealed that the drug sensitivity phenotype first observed in cells lacking the *ncRNA.1343* gene was directly due to accumulating levels of the *tgp1*$^{+}$ permease resulting from the loss of repressive *nc-tgp1* transcription. Notably, *nc-1343* was entirely disposable for *tgp1*$^{+}$ regulation. Accordingly, deleting the *nc-1343* gene in a manner that did not interrupt *nc-tgp1* transcription had no effect on *tgp1*$^{+}$ levels or drug tolerance. The findings presented in Chapter 4, in particular, illustrate some of the unexpected consequences of making poorly informed manipulations of an lncRNA-encoding gene: if the annotation of *ncRNA.1343* had more accurately predicted the true 5'-end of the *nc-1343* lncRNA, deleting this gene would not have disrupted *nc-tgp1* transcription and the drug sensitivity phenotype would not have been identified. Regarding the lack of a defined function for the stable *nc-1343* lncRNA, one cannot conclusively rule out the possibility that it provides some function. Indeed, the *nc-1343* transcript is conserved in position, despite no sequence similarities in related fission yeasts (Rhind et al., 2011). The preliminary analyses presented in Chapter 6 suggest that these putative lncRNA orthologs may not be the stable byproducts of a promoter that initiates transcriptional interference in the opposite orientation. Thus, future work is required to determine whether the *nc-1343* transcript and these putative orthologs have some other genuine biological function that has been conserved or whether they might simply represent transcriptional noise.

7.2 *Gene regulation by lncRNA transcription*

Part of the reason that many lncRNAs continue to escape detection is that RNAPII transcription frequently fails to produce stable RNA products (Berretta *et al.*, 2009). Remarkably, the quality and depth of RNA-seq permits the detection of short transcripts produced during stalled transcription initiation events (Nechaev *et al.*, 2010). To a lesser degree, RNA-seq can even identify the presence of some longer unstable transcripts. For years, these cryptic transcripts had only been observed in cells lacking factors involved in RNA decay pathways (Houseley *et al.*, 2006). Less obstructive methods are now available to detect active RNAPII transcription genome-wide. For example, nascent elongating transcript sequencing (NET-seq) captures native RNAPII-DNA-RNA complexes from cells and sequences from the 3' most nucleotide of nascent transcripts in order to visualize active transcription with strand-specificity and single-nucleotide resolution (Churchman and Weissman, 2011). Although these and other genome-wide approaches have corroborated the conclusion that eukaryotic genomes are pervasively transcribed by RNAPII, the biological significance of much of this transcription – especially cryptic unstable transcription – is not well understood.

One of the major findings of this thesis is that two phosphate-regulated genes in *S. pombe* (*tgp1*$^+$ and *pho1*$^+$) are regulated by cryptic lncRNA transcription into their respective promoters (Chapter 5). The mechanism of *tgp1*$^+$ and *pho1*$^+$ regulation resembles that of the *S. cerevisiae SER3* gene, whereby stable lncRNA transcription into the *SER3* promoter, or heterologous promoters, represses gene induction (Martens *et al.*, 2004). Mechanistically, *SER3* repression by lncRNA transcription requires histone chaperones, such as Spt6 and FACT, to bring about increased nucleosome density over the *SER3* promoter and prevent transcription

factor binding (Hainer *et al.*, 2011; Thebault *et al.*, 2011). Increased nucleosome density at repressed *tgp1*$^{+}$ and *pho1*$^{+}$ promoters suggests a possible role for lncRNA transcription-coupled chromatin remodelers in the regulation of these *S. pombe* genes as well. In *S. pombe*, the Spt6 histone chaperone is thought to reposition nucleosomes and facilitate Set2-dependent H3K36 methylation, both of which help to reduce intragenic transcription from cryptic promoters in gene bodies (DeGennaro *et al.*, 2013). Given that gene promoters occluded by interfering lncRNAs can also be thought of as "cryptic promoters" within the lncRNA transcription unit, it is reasonable to suggest that the mechanisms that prevent intragenic transcription initiation may also contribute to the effectiveness of transcriptional interference. Thus, repression by interfering lncRNA transcription might be reinforced by H3K36 methylation, which is deposited by the elongating RNAPII-associated HMT Set2 and recruits HDACs (Carrozza *et al.*, 2005; Keogh *et al.*, 2005; Venkatesh *et al.*, 2012). Indeed, the repression of the *S. cerevisiae IME1* gene by an interfering lncRNA requires Set2 activity (van Werven *et al.*, 2012). However, this mechanism does not appear to be universal since Set2 is not required for *SER3* repression (Hainer *et al.*, 2011). This difference is likely explained by the fact that Set2 predominantly represses the initiation of intragenic transcription within long genes (Li *et al.*, 2007), and the relatively short *SER3*-regulatory lncRNA (~500 nt) might not be long enough to utilize this mechanism. It is also plausible that different transcription factors might also be more or less sensitive to specific chromatin features present in any given promoter. Since some transcription factors can interact with RNA (Cassiday and Maher, 2002; Sigova *et al.*, 2015), relatively stable nascent lncRNAs might attract or repel such factors from promoters. Additional experiments are needed to determine whether histone chaperones, H3K36 methylation, and/or other histone modifications or factors such as the lncRNA transcripts themselves participate in *tgp1*$^{+}$ and *pho1*$^{+}$ regulation in *S. pombe*. Replacing either gene with a marker gene (e.g. GFP) and

crossing such a strain against the Bioneer *S. pombe* non-essential gene deletion library to look for suppressors of transcriptional interference (i.e. GFP expression) should facilitate this aim. Importantly, the finding that RNAi/heterochromatin plays no appreciable role in repressing *tgp1*$^{+}$ and *pho1*$^{+}$ in *S. pombe* lends support to the idea that the mechanism of repression by transcriptional interference is at least partially conserved between *S. pombe*, which retains active RNAi, and *S. cerevisiae*, where the RNAi pathway is absent. Further mechanistic insight may be gained by comparing how transcriptional interference is achieved in these two model organisms.

Transcriptional interference has been observed in diverse systems, including *E. coli* (Zafar *et al.*, 2014), *S. cerevisiae* (Bird *et al.*, 2006; Bumgarner *et al.*, 2009; Greger *et al.*, 2000; Martens *et al.*, 2004; van Werven *et al.*, 2012), *S. pombe* (Chapter 5), plants (Hedtke and Grimm, 2009), *Drosophila* (Petruk *et al.*, 2006), and in mammals (Abarrateui and Krangel, 2007; Latos *et al.*, 2012; Martianov *et al.*, 2007). In addition to these many examples, transcriptional interference contributes to the genetic disease alpha thalassemia, which is caused by an intergenic SNP that creates a new promoter and initiates novel transcription that interferes with the expression of the downstream alpha globin gene (De Gobbi *et al.*, 2006). Transcriptional interference has also been demonstrated to maintain human immunodeficiency virus HIV-1 latency (Han *et al.*, 2008; Lenasi *et al.*, 2008). Collectively, these findings demonstrate that transcriptional interference is a simple, conserved mechanism for modulating specific genes. While pervasive transcription in eukaryotes suggests that this mechanism might be a general feature of eukaryotic gene regulation and contribute to human health and disease, it is still not clear how widespread concerted gene regulation by transcriptional interference actually is.

Greater mechanistic insight is therefore required in order to determine the prevalence of transcriptional interference.

Research from diverse organisms suggests that transcription elongation is itself too rapid to mediate strong repression of downstream genes (Palmer *et al.*, 2011). In bacteria, interference appears to be achieved by either dislodging transcription factors and/or by transcription pausing that occludes underlying promoter sequences. As described above, eukaryotic interference mechanisms frequently involve transcription-coupled changes in chromatin status. If a few of these basic mechanistic features are found to be universally required for eukaryotic transcriptional interference, the presence of such features could be used to indicate how widespread this regulation mechanism is. To achieve this level of understanding, genome-wide approaches will be required to better predict additional examples of gene regulation by interfering lncRNA transcription. For example, NET-seq provides an unparalleled view of nascent transcription in cells and is therefore among the best available tools to identify additional lncRNA-transcribed promoters that might repress downstream genes by transcriptional interference. In addition, a powerful new transcript profiling method called transcript isoform sequencing (TIF-seq) sequences transcription start/end sites simultaneously and provide a detailed global picture of transcript diversity (Pelechano *et al.*, 2013). Specifically, TIF-seq can distinguish altered transcript isoforms from upstream lncRNAs that overlap promoters and/or downstream genes (overlapping transcripts in particular are under represented in conventional RNA-seq/NET-seq datasets). Thus, TIF-seq might be a useful technique to better identify the prevalence of upstream interfering lncRNAs in any given genome. Genome-wide mapping of nucleosome positions and/or specific transcription-coupled histone modifications (e.g. H3K36me3) by ChIP-seq might also prove to be a valuable tool for discovering new longer interfering lncRNAs.

Importantly, locus-specific experiments and genetic manipulations will be required to validate such genome-wide approaches and provide added mechanistic insight. Further attention needs to be placed on distinguishing the importance of histone chaperones, specific histone modifications, and other regulatory factors from the mere presence of elongating RNAPII over promoters in the regulation of eukaryotic genes by transcriptional interference. Further studies should reveal why some acts of upstream transcription are inhibitory while others, such as lncRNA transcription at enhancers or upstream of the *S. pombe fpb1*$^{+}$ gene (Hirota *et al.*, 2008; Ørom *et al.*, 2010), appear to favour downstream gene activation.

7.3 *Final thoughts*

Advances in RNA sequencing and improved methods for mapping the position of proteins, and post-translational histone modifications, and RNA on a genome-wide scale have uncovered many complex levels of eukaryotic gene regulation (Chu *et al.*, 2012; Johnson *et al.*, 2007; Wang *et al.*, 2009). Notably, eukaryotic genomes pervasively transcribe lncRNAs and while some of these transcripts are highly expressed, the majority of lncRNAs are present at very low levels and are frequently targeted for degradation by various RNA decay pathways (Ponting *et al.*, 2009). In fact, most eukaryotic genomes studied to date show evidence of widespread cryptic lncRNA transcription (Berretta *et al.*, 2009). While the low steady-state levels and poor primary sequence conservation of most lncRNAs was initially suggested to be evidence for their lack of function (Struhl, 2007), numerous studies have since found that both high and low abundance lncRNAs can play important roles in cells (Geisler and Coller, 2013).

A more recent challenge to lncRNA research has been the question of whether transcripts annotated as lncRNAs are truly non-coding. Following the development

of ribosome profiling to map active translation along mRNAs (Ingolia *et al.*, 2009), numerous studies have since found that ribosomes regularly associate with lncRNAs as well (Bazzini *et al.*, 2014; Brar *et al.*, 2011; Chew *et al.*, 2013; Duncan and Mata, 2014; Ingolia *et al.*, 2009; Ingolia *et al.*, 2011; Juntawong *et al.*, 2014). While it has been proposed that some lncRNAs might act as decoys for the ribosome and not actually be translated (Guttman *et al.*, 2013), recent proteomics studies in a variety of organisms have detected short peptides translated from regions of the genome previously annotated as non-coding (Ruiz-Orera *et al.*, 2014; Slavoff *et al.*, 2013; Smith *et al.*, 2014; Vanderperre *et al.*, 2013). Although the function of most short peptides is unknown and might simply represent the equivalent of "translational noise", emerging evidence indicates that an accumulating number of short peptides are functional and conserved (Anderson *et al.*, 2015; Andrews and Rothnagel, 2014; Crappé *et al.*, 2014). It is therefore apparent that some transcripts annotated as non-coding encode small functional peptides. Thus, studies investigating different lncRNAs must consider this possibility. It is also worth revisiting functionally characterized lncRNAs to determine whether any of these transcripts are translated and if their function might be mediated by their protein product, rather than the transcript itself as had been originally proposed.

To further complicate matters, there is no reason to assume that coding and non-coding functions for any given transcript are mutually exclusive. Although difficult to distinguish, it is reasonable to expect that some mRNAs possess lncRNA-like functions since nascent coding mRNAs should be equally capable of recruiting factors that might influence local chromatin structure, as has been proposed for numerous meiotic genes in *S. pombe* (Zofall *et al.*, 2012). In fact, such flexibility could be the driving force behind the evolution of some functional lncRNAs. Indeed, *Xist* is thought to have evolved from an ancestral protein-coding gene (Duret *et al.*,

2006). It is therefore possible that other functional lncRNAs have evolved from protein-coding genes. This scenario is likely since loss-of-function protein-coding genes would retain promoters and other regulatory elements that continue to drive transcription. Over time, the now stable lncRNA product might be free to acquire new roles in cells. Improved computational strategies are required to test this hypothesis directly and accelerate the identification of additional lncRNAs that might have originated in this manner.

Conversely, it is a possibility that novel proteins could emerge from lncRNAs that associate with ribosomes and have evolved short or long ORFs. Indeed, new proteins appear to arise *de novo* from non-coding DNA at a much greater frequency than originally thought (Cai *et al.*, 2008; Carvunis *et al.*, 2012; Knowles and McLysaght, 2009; Levine *et al.*, 2006; Murphy and McLysaght, 2012; Reinhardt *et al.*, 2013; Ruiz-Orera *et al.*, 2014; Toll-Riera *et al.*, 2009; Wu *et al.*, 2011; Xie *et al.*, 2012). It has since been proposed that a subset of low abundance ncRNAs might provide the raw material needed to generate new protein-coding genes with entirely novel functions (Wilson and Masel, 2011). If true, even transcriptional noise resulting from low RNAPII fidelity might actually provide an adaptive advantage to organisms. This might, at least in part, explain the reason that most eukaryotes contain an abundance of pervasively transcribed non-coding DNA. It does not, however, answer how it is that a complex multicellular organism such as *U. gibba* benefits from having discarded most of its non-coding DNA (Ibarra-Laclette *et al.*, 2013). Future work should help to answer such questions, but will no doubt raise many more.

In short, assigning biological functions to lncRNAs has been much more challenging and contentious than it has been for other classes of ncRNAs, such as short

regulatory RNAs (e.g. miRNAs, siRNAs, etc.) or rRNAs. Indeed, many new studies investigating lncRNA biology frequently contradict the interpretations of prior analyses (Bassett *et al.*, 2014; Cech and Steitz, 2014). It is critical that future studies differentiate the influence played by the act of transcription and/or genomic locus itself and distinguish these from any roles that are attributed to the lncRNA product. In addition, *cis* and *trans* mechanisms for any given lncRNA should also be addressed by designing experiments that adequately distinguish between these possibilities. Finally, given that low RNAPII fidelity might produce spurious lncRNAs with no function, or that other transcripts annotated as non-coding might actually encode short ORFs that are translated into functional micropeptides, it is essential that detailed analyses of individual lncRNAs be performed in order to rule out these possibilities before concluding any given lncRNA itself serves a genuine biological role. Despite these many challenges, an accumulating body of evidence has revealed that a great number of lncRNAs are important for gene regulation in our cells, either as functional products themselves or simply as a result of being transcribed.

REFERENCES

Abarrategui I and Krangel MS (2007) Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination. EMBO J, 26(20): 4380-4390.

Ahlquist P (2002) RNA-dependent RNA polymerases, viruses, and RNA silencing. Science, 296(5571): 1270-1273.

Alexander RD, Innocente SA, Barrass JD, and Beggs JD (2010) Splicing-dependent RNA polymerase pausing in yeast. Mol. Cell, 40(4): 582-593.

Alföldi J and Linblad-Toh K (2013) Comparative genomes as a tool to understand evolution and disease. Genome Res., 23(7): 1063-1068.

Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D *et al.* (1999) The yeast exosome and human PM-Sci are related complexes of 3' -> 5' exonucleases. Genes and Dev., 13(16): 2148-2158.

Allshire RC, Nimmo ER, Ekwall K, Javerzat JP, and Cranston G (1995) Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. Genes Dev., 9(2): 218-233.

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J *et al.* (2014) An atlas of active enhancers across human cell types and tissues. Nature, 507(7493): 455-461.

Anish R, Hossain MB, Jacobson RH, and Takada S (2009) Characterization of transcription from TATA-less promoters: idenfitication of a new core promoter element XCPE2 and analysis of factor requirements. PLoS One, 4(4): e5103.

Askarian-Amiri ME, Crawford J, French JD, Smart CE, Smith MA *et al.* (2011) SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. RNA, 17(5): 878-891.

Astier-Manifacier S and Cornuet P (1971) RNA-dependent RNA polymerase in Chinese cabbage. Biochim Biophys Acta., 232(3): 484-493.

Audergon PN, Catania S, Kagansky A, Tong P, Shukla M *et al.* (2015) Restricted epigenetic inheritance of H3K9 methylation. Science, 348(6230): 132-135.

Ay F and Noble WS (2015) Analysis methods for studying the 3D architecture of the genome. Genome Biol., 16:183.

Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature, 226(5252): 1209-1211.

Bari R, Datt Pant B, Stitt M, and Scheible WR (2006) PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. Plant Physiol., 141(3): 988-999.

Bartel DP (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell, 116(2): 281-297.

Bassett AR, Akhtar A, Barlow DP, Bird AP, Brockdroff N *et al.* (2014) Considerations when investigating lncRNA function in vivo. eLife, 3: e03058.

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L *et al.* (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature, 520: 243-247.

Bayne EH, White SA, Kagansky A, Bijos DA, Sanchez-Pulido L *et al.* (2010) Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. Cell, 140(5): 666-677.

Bell SP and Dutta A (2002) DNA replication in eukaryotic cells. Annu. Rev. Biochem., 71: 333-374.

Beltran M, Puig I, Pena C, Garcia M, Alvarez AB *et al.* (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev., 22(6): 756-769.

Berghoff EG, Clark MF, Chen S, Cajigas I, Leib DE *et al.* (2013) Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. Development, 140(21): 4407-4416.

Bergman LW, Stranathan MC, and Preis LH (1986) Structure of the transcriptionally repressed phosphate-repressible acid phosphatase gene PHO5 of Saccharomyces cerevisiae. Mol. Cell Biol., 6(1): 38-46.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell, 125(2): 315-326.

Bernstein E and Allis CD (2005) RNA meets chromatin. Genes Dev., 19(14): 1635-1655.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature, 409(6818): 363-366.

Berretta J, Pinskaya M, and Morillon A (2008) A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev., 22(5): 615-626.

Berretta J and Morillon A (2009) Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO Rep., 10(9): 973-982.

Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev., 16(1): 6-21.

Bird AJ, Gordon M, Eide DJ and Winge DR (2006) Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. EMBO J., 25: 5726-5734.

Bitton DA, Grallert S, Scutt PJ, Yates T, Li Y *et al.* (2011) Programmed fluctuations in sense/antisense transcript ratios drive sexual differentiation in S. pombe. Mol. Syst. Biol., 7: 559.

Bond AM, Vangompel MJ, Sametsky EA, Clark MF, Savage JC *et al.* (2009) Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. Nat. Neurosci., 12(8): 1020-1027.

Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC *et al.* (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. Mol. Biol. Cell, 19(1): 352-367.

Bres V, Yoh SM, and Jones KA (2008) The multi-tasking P-TEFb complex. Curr. Opin. Cell Biol., 20(3): 334-340.

Brickner DG, Cajigas I, Fondufe-Mittendorf Y, Ahmed S, Lee PC *et al.* (2007) H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcription state. PLoS Biol., 5(4): e81.

Brodsky AS, Meyer CA, Swinburne IA, Hall G, Keenan BJ *et al.* (2005) Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. Genome Biol., 6: R64.

Browning DF and Busby SJW (2004) The regulation of bacterial transcription initiation. Nat. Rev. Microbiol., 2: 57-65.

Buckley BA, Burkhart KB, Gu SG, Spracklin G, Kershner A *et al.* (2012) A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. Nature, 489(7416): 447-451.

Bühler M, Haas W, Gygi SP, and Moazed D (2007) RNAi-dependent and –independent RNA turnover mechanisms contribute to heterochromatic gene silencing. Cell, 129(4): 707-721.

Bühler M, Spies N, Bartel DP, and Moazed D (2008) TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway. Nat. Struct. Mol. Biol., 15(10): 1015-1023.

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD *et al.* (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science, 273(5278): 1058-1073.

Bumgarner SL, Dowell RD, Grisafi P, Gifford DK, and Fink GR (2009) Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. Proc. Natl. Acad. Sci., 106(43): 18321-18326.

Bushnell DA, Westover KD, David RE, and Kornberg RD (2004) Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 angstroms. Science, 303(5660): 983-988.

Camblong J, Iglesias N, Fickentscher C, Dieppois G, and Stutz F (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell, 131(4): 706-717.

Camblong J, Beyrouthy N, Guffanti E, Schlaepfer G, Steinmetz LM *et al.* (2008) Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. Genes Dev., 23: 1534-1545.

Capuano F, Mulleder M, Kok R, Blom HJ, and Ralser M (2014) Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. Anal. Chem., 86(8): 3697-3702.

Caputi M and Zahler AM (2002) SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. EMBO J., 21(4): 845-855.

Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S *et al.* (2012) Long non-coding antisense RNA controls Uchl1 translation through an embeeded SINEB2 repeat. Nature, 491(7424): 454-457.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK *et al.* (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell, 123(4): 581-592.

Carter-O'Connell I, Peel MT, Wykoff DD, and O'Shea EK (2012) Genome-wide characterization of the phosphate starvation response in Schizosaccharomyces pombe. BMC Genomics, 13: 697.

Cassiday LA and Maher LJ (2002) Having it both ways: transcription factors that bind DNA and RNA. Nucl. Acids Res., 30: 4118-4126.

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

Catania S and Allshire RC (2014) Anarchic centromeres: deciphering order from apparent chaos. Curr. Opin. Cell Biol., 26(100): 41-50.

Cech TR and Steitz JA (2014) The noncoding RNA revolution-trashing old rules to forge new ones. Cell, 157, 77-94

Cerase A, Pintacuda G, Tattermusch A and Avner P (2015) Xist localization and function: new insights from multiple levels. Genome Biol., 16: 166.

Cerase A, Smeets D, Tang YA, Gdula M, Kraus F *et al.* (2014) Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. Proc. Natl. Acad. Sci., 111:2235-2240.

Chalei V, Sansom SN, Kong L, Lee S, Montiel JF *et al.* (2014) The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. eLife, 3: e04530.

Chen FX, Woodfin AR, Gardini A, Rickels RA, Marshall SA *et al.* (2015) PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. Cell, 162(5): 1003-1015.

Chen HM, Futcher B, and Leatherwood J (2011) The fission yeast RNA binding protein Mmi1 regulates meiotic genes by controlling intron specific splicing and polyadenylation coupled RNA turnover. PLoS One, 6(10): e26804.

Cheng H, Dufu K, Lee CS, Hsu JL, Dias A *et al.* (2006) Human mRNA export machinery recruited to the 5' end of mRNA. Cell, 127(7): 1389-1400.

Chiu YL, Ho CK, Saha N, Schwer B, Shuman S *et al.* (2002) Tat stimulates cotranscription capping of HIV mRNA. Mol. Cell, 10: 585-597.

Choy MK, Movassagh M, Goh HG, Bennett MR, Down TA *et al.* (2010) Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. BMC Genomics, 11: 519.

Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M *et al.* (2015) Systemic discovery of Xist RNA binding proteins. Cell, 161(2): 404-416.

Churchman LS and Weissman JS (2011) Nascent transcript sequencing visualizes transcription at single nucleotide resolution. Nature, 469: 368-373.

Coller J and Parker R (2004) Eukaryotic mRNA decapping. Annu. Rev. Biochem., 73: 861-890.

Colmenares SU, Bunker SM, Buhler M, Dlakic M, and Mazed D (2007) Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. Mol. Cell, 27(3): 449-461.

Cramer P, Pesce CG, Baralle FE, and Kornblihtt AR (1997) Functional association between promoter structure and transcript alternative splicing. Proc. Natl. Acad. Sci., 94(21): 11456-11460.

Crick F (1970) Central dogma of molecular biology. Nature, 227: 561-563.

Crick F, Barnett L, Brenner S, and Watts-Tobin RJ (1961) General nature of the genetic code for proteins. Nature, 192(4809): 1227-1232.

Csink AK and Henikoff S (1996) Genetic modification of heterochromatic association and nuclear organization in Drosophila. Nature, 381: 529-531.

Davidovich C, Zheng L, Goodrich KJ, Cech TR (2013) Promiscuous RNA binding by Polycomb repressive complex 2. Nat. Struct. Mol. Biol., 20(11): 1250-1257.

de la Mata M and Kornbilhtt AR (2006) RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. Nat. Struct. Mol. Biol., 13: 973-980.

De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ *et al.* (2006) A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science, 312(5777): 1215-1217.

DeGennaro CM, Alver BH, Marguerat S, Stepanova E, Davis CP *et al.* (2013) Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. Mol. Cell Biol., 33(24): 4779-4792.

Dernburg AF, Sedat JW, and Hawley RS (1996) Direct evidence of a role for heterochromatin in meiotic chromosome segregation. Cell, 86(1): 135-146.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S *et al.* (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res., 22: 1775-1789.

Di Giammartino DC, Nishida K, and Manley JL (2011) Mechanisms and consequences of alternative polyadenylation. Mol. Cell, 43(6): 853-866.

Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA *et al.* (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. Nature, 503(7476): 371-376.

Dieci G, Preti M, and Montanini B (2009) Eukaryotic snoRNAs: a paradigm for gene expression flexibility. Genomics, 94: 83-88.

Dill KA and MacCallum JL (2012) The protein-folding problem, 50 years on. Science, 338(6110): 1042-1046.

Doi A, Fujimoto A, Sato S, Uno T, Kanda Y *et al.* (2015) Chemical genomics approach to identify genes associated with sensitivity to rapamycin in the fission yeast Schizosaccharomyces pombe. Genes Cells, 20: 292-309.

Duda CT, Zaitlin M, and Siegel A (1973) In vitro synthesis of double-stranded RNA by an enzyme system isolated from tobacco leaves. Biochim Biophys Acta., 319(1): 62-71.

Duret L, Chureau C, Samain S, Weissenbach J, and Avner P (2006) The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science, 312: 1653-1655.

Egan ED, Braun CR, Gygi SP, and Moazed D (2014) Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. RNA, 20(6): 867-881.

Eick D and Geyer M (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. Chem. Rev., 113(11): 8456-8490.

Elgar G and Vavouri T (2008) Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. Trends Genet., 24(7): 344-352.

Emami KH, Burke TW, and Smale ST (1998) Sp1 activation of a TATA-less promoter requires a species-specific interaction involving transcription factor IID. Nucleic Acids Res., 26(3): 839-846.

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature, 489: 57-74.

Eißmann M, Gutschner T, Hämmerle M, Günther S, Caudron-Herger M *et al.* (2012) Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. RNA Biol., 9(8): 1076-1087.

Esteve PO, Chin HG, Smallwood A, Feehery GF, Gangisetty O *et al.* (2006) Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev., 20: 3089-3103.

Fabian MR, Sonenberg N, and Filipowicz W (2010) Regulation of mRNA translation and stability by microRNAs. Annu. Rev. Biochem., 79: 351-379.

Fejes-Toth K, Sotirova V, Sachidanandam R, Assaf G, Hannon GJ *et al.* (2009) Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature, 457: 1028-1032.

Feng S, Jacobsen SE, and Reik W (2010) Epigenetic reprogramming in plant and animal development. Science, 330: 622-627.

Fiers W, Contreras R, Haegemann G, Rogiers R, Van de Voorde A *et al.* (1978) Complete nucleotide sequence of SV40 DNA. Nature, 273(5658): 113-120.

Flanagan JF, Mi LZ, Chruszcz M, Cymborowski M, Clines KL *et al.* (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature, 438(7071): 1181-1185.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF *et al.* (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science, 269(5223): 496-512.

Fong N, Bird G, Vigneron M, Bentley DL (2003) A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing, and 3' end processing. EMBO J., 22: 4274-4282.

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI *et al.* (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. Nat. Genet., 39(8): 1033-1037.

Gao F, Liu X, Wu XP, Wang XL, Gong D *et al.* (2012) Differential DNA methylation in discrete developmental stages of the parasitic nematode Trichinella spiralis. Genome Biol., 13: R100.

Gardini A, Baillat D, Cesaroni M, Hu D, Marinis JM *et al.* (2014) Integrator regulates transcription initiation and pause release following activation. Mol. Cell, 56(1): 128-139.

Gaydos LJ, Wang W, and Strome S (2014) H3K27me and PRC2 transmit a memory of repression across generations and during development. Science, 345: 1515-1518.

Geisler S and Coller J (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat. Rev. Mol. Cell Biol., 14(11): 699-712.

Gilber W (1986) Origin of life: the RNA world. Nature, 319(6055): 618.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B *et al.* (1996) Life with 6000 genes. Science, 274(5287): 546, 563-567.

Gomez JA, Wapinski OL, Yang YW, Bureau JF, Gopinath S *et al.* (2013) The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. Cell, 152(4): 743-754.

Görnemann J, Kotovic KM, Hujer K, and Neugebauer KM (2005) Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol. Cell, 19(1): 53-63.

Greer EL, Beese-Sims SE, Brookes E, Spadafora R, Zhu Y *et al.* (2014) A histone methylation network regulates transgenerational epigenetic memory in *C. elegans*. Cell Rep., 7(1): 113-126.

Greger IH, Aranda A, and Proudfoot N (2000) Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci., 97(15): 8415-8420.

Guil S and Esteller M (2012) Cis-acting noncoding RNAs: friends and foes. Nat. Struct. Mol. Biol., 19: 1068-1075.

Gusarov I and Nudler E (1999) The mechanism of intrinsic transcription termination. Mol. Cell, 3: 495-504.

Guttman M, Amit I, Garber M, French C, Lin MF *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature, 458(7235): 223-227.

Haag JR and Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat. Rev. Mol. Cell Biol., 12: 483-492.

Haberland M, Montgomery RL, and Olson EN (2009) The many roles of histone deacetylases in development and physiology: implications for disease and therapy. Nat. Rev. Genet., 10(1): 32-42.

Hacisuleyman E, Goff LA, Trapnell C, William A, Henao-Mejia J *et al.* (2014) Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat. Struct. Mol. Biol., 21(2): 198-206.

Haerty W and Ponting CP (2015) Unexpected selection to retain high GC content and splicing enhancers within exons of multiexonic lncRNA loci. RNA, 21(3): 333-346.

Hainer SJ, Pruneski JA, Mitchell RD, Monteverde RM, and Martens JA (2011) Intergenic transcription causes repression by directing nucleosome assembly. Genes Dev., 25(1): 29-40.

Han Y, Lin YB, An W, Xu J, Yang HC *et al.* (2008) Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. Cell Host Microbe., 4: 134-146.

Hannon GJ (2002) RNA interference. Nature, 418(6894): 244-251.

Hansen KH, Bracken AP, Pasini D, Dietrich N, Gehani SS *et al.* (2008) A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol., 10(11): 1291-1300.

Hansen KR, Burns G, Mata J, Volpe TA, Martienssen RA *et al.* (2005) Global effects on gene expression in fission yeast by silencing and RNA interference machineries. Mol. Cell Biol., 25(2): 590-601.

Harigaya Y, Tanaka H, Yamanaka S, Watanabe Y, Tsutsumi C *et al.* (2006) Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature, 442(7098): 45-50.

Harmston N and Lenhard B (2013) Chromatin and epigenetic features of long-range gene regulation. Nucleic Acids Res., 41(15): 7185-7199.

Hayles J, Wood V, Jeffery L, Hoe KL, Kim DU *et al.* (2013) A genome-wide resource of cell cycle and cell shape of fission yeast. *Open Biol.*, 3: 130053.

Heard E and Disteche CM (2006) Dosage compensation in mammals: fine-tuning the expression of the X chromosome. Genes Dev., 20(14): 1848-1867.

Hedtke B and Grimm B (2009) Silencing of a plant gene by transcriptional interference. Nucleic Acids Res., 37(11): 3739-3746.

Heo JB and Sung S (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science, 331(6013): 76-79.

Henry TC, Power JE, Kerwin CL, Mohammed A, Weissman JS *et al.* (2011) Systematic screen of Schizosaccharomyces pombe deletion collection uncovers parallel evolution of phosphate signal transduction pathway in yeasts. Eukaryot. Cell, 10(2): 198-206.

Hiriart E, Vavsseur A, Touat-Todeschini L, Yamashita A, Gilquin B *et al.* (2012) Mmi1 RNA surveillance machinery directs RNAi complex RITS to specific meiotic genes in fission yeast. EMBO J, 31(10): 2296-2308.

Hirose Y and Manley JL (1998) RNA polymerase II is an essential mRNA polyadenylation factor. Nature, 395: 93-96.

Hirose Y, Tacke R, and Manley JL (1999) Phosphorylated RNA polymerase stimulates pre-mRNA splicing. Genes Dev., 13(10): 1234-1239.

Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T *et al.* (2008) Stepwise chromatin remodeling by a cascade of transcription initiation of non-coding RNAs. Nature, 456(7218): 130-134.

Ho C and Shuman S (1999) Distinct effector roles for Ser2 and Ser5 phosphorylation of the RNA polymerase CTD in the recruitment and allosteric activation of mammalian capping enzyme. Mol. Cell, 3: 405-411.

Hobson DJ, Wei W, Steinmetz LM, and Svejstrup JQ (2012) RNA polymerase II collision interrupts convergent transcription. Mol. Cell, 48(3): 365-374.

Holoch D and Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. Nat. Rev. Genet., 16: 71-84.

Hong L, Schroth GP, Matthews HR, Yau P, and Bradbury EM (1993) Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 “tail” to DNA. J. Biol. Chem., 268(1): 305-314.

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(5): 685-695.

Houseley J and Tollervey D (2009) The many pathways of RNA degradation. Cell, 136(4): 763-776.

Hsin JP and Manley JL (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev., 26(19): 2119-2137.

Huang Y and Carmichael GC (1996) Role of polyadenylation in nucleocytoplasmic transport of mRNA. Mol. Cell Biol., 16(4): 1534-1542.

Ibarra-Laclette E, Lyons E, Hernandez-Guzman G, Perez-Torres CA, Carretero-Paulet L *et al.* (2013) Architecture and evolution of a minute plant genome. Nature, 498: 94-98.

Ilik I and Akhtar A (2009) roX RNAs: non-coding regulators of the male X chromosome in flies. RNA Biol., 6(2): 113-121.

Jabnoune M, Secco D, Lecampion C, Robaglia C, Shu Q *et al.* (2013) A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. Plant Cell, 25(10): 4166-4182.

Jackson V, Granner DK, and Chalkley R (1975) Deposition of histones onto replicating chromosomes. Proc. Natl. Acad. Sci., 72(11): 4440-4444.

Jacquier A (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat. Rev. Genet., 10(12): 833-844.

Januszyk K and Lima CD (2014) The eukaryotic RNA exosome. Curr. Opin. Struct. Biol., 24: 132-140.

Ji X and Fu XD (2012) The mediator couples transcription and splicing. Mol. Cell, 45(4): 433-434.

Jonkers I and Lis JT (2015) Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol., 16: 167-177.

Karpen GH and Allshire RC (1997) The case for epigenetic effects on centromere identify and function. Trends Genet., 13(12): 489-496.

Kawano M, Aravind L, and Storz G (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. Mol. Microbiol., 64(3): 738-754.

Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V *et al.* (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell, 123(4): 593-605.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A *et al.* (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci., 106(28): 11667-11672.

Kilchert C, Wittmann S, Passoni M, Shah S, Granneman S *et al.* (2015) Regulation of mRNA levels by decay-promoting introns that recruit the exosome specificity factor Mmi1. Cell Reports, 13: 1-12.

Kim TK, Ebright RH, and Reinberg D (2000) Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science, 288(5470): 1418-1422.

Kino T, Hurt DE, Ichijo T, Nader N, and Chrousos GP (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci. Signal., 3(107): ra8.

Kiss DL and Andrulis ED (2010) Genome-wide analysis reveals distinct substrate specificities of Rrp6, Dis3, and core exosome subunits. RNA, 16(4): 781-791.

Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA *et al.* (2013) Braveheart, a long non-coding RNA required for cardiovascular lineage commitment. Cell, 152(3): 570-583.

Köhler A and Hurt E (2007) Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol., 8: 761-773.

Komarnitsky P, Cho EJ, and Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev., 14(19) 2452-2460.

Koonin EV and Novozhilov AS (2009) Origin and evolution of the genetic code: the universal enigma. IUBMB Life, 61(2): 99-111.

Kornienko AE, Guenzl PM, Barlow DP and Pauler FM (2013) Gene regulation by the act of long non-coding RNA transcription. BMC Biology, 11: 59.

Korostowski L, Sedlak N, and Engel N (2012) THE Kcnq1ot1 long non-coding RNA affects chromatin conformation and expression of Kcnq1, but does not regulate its imprinting in the developing heart. PLoS Genet., 8(9): e1002956.

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N *et al.* (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15INK4B tumor suppressor gene. Oncogene, 30(16): 1956-1962.

Kowalik KM, Shimada Y, Flury V, Stadler MB, Batki J *et al.* (2015) The Paf1 complex represses small-RNA-mediated epigenetic gene silencing. Nature, 520(7646): 248-252.

LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E *et al.* (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. 121(5): 713-724.

Lai F, Ørom UA, Cesaroni M, Beringer M, Taatjes DJ *et al.* (2013) Activating RNAs associated with Mediator to enhance chromatin architecture and transcription. Nature, 494(7438): 497-501.

Lai K-MV, Gong G, Atanasio A, Rojas J, Quispe J *et al.* (2015) Diverse phenotypes and specific transcription patterns in twenty mouse lines with ablated lincRNAs. Plos One ,10(4): e0125522.

Laird-Offringa IA, de Wit CL, Elfferich P, and van der Eb AJ (1990) Poly(A) tail shortening is the translation-dependent step in c-myc mRNA degradation. Mol. Cell Biol., 10(12): 6132-6140.

Lamb JC and Birchler JA (2003) The role of DNA sequence in centromere formation. Genome Biol., 4: 214.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC *et al.* (2001) Initial sequencing and analysis of the human genome. Nature, 409: 860-921.

Lasda E and Parker R (2014) Circular RNAs: diversity of form and function. RNA, 20(12): 1829-1842.

Latos PA, Pauler FM, Koerner MV, Senergin HB, Hudson QJ *et al.* (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science, 338(6113): 1469-1472.

Lebreton A, Tomecki R, Dziembowski A, and Séraphin B (2008) Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature, 456: 993-996.

Lee JH and Skalnik DG (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J. Biol. Chem., 280(50): 41725-41731.

Lee JT and Bartolomei MS (2013) X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell, 152(6): 1308-1323.

Lee KK and Workman JL (2007) Histone acetyltransferase complexes: one size doesn't fit all. Nat. Rev. Mol. Cell Biol., 8(4): 284-95.

Lee NN, Chalamcharla VR, Reyes-Turcu F, Mehta S, Zofall M *et al.* (2013) Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. Cell, 155(5): 1061-1074.

Lee Y, Kim M, Han J, Yeom KH, Lee S *et al.* (2004) MicroRNA genes are transcribed by RNA polymerase II. EMBO J., 23(20): 4051-4060.

Lehmann E, Brueckner F, and Cramer (2007) Molecular basis of RNA-dependent RNA polymerase activity. Nature, 450: 445-449.

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L *et al.* (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr. Biol., 13(14): 1192-1200.

Lemay JF, Larochelle M, Marguerat S, Atkinson, S, Bähler J *et al.* (2014) The RNA exosome promotes transcription termination of backtracked RNA polymerase II. Nat. Struct. Mol. Biol., 21: 919-926.

Lenasi T, Contreras X and Peterlin BM (2008) Transcriptional interference antagonizes proviral gene expression to promote HIV latency. Cell Host Microbe., 4: 123-133.

Leong HS, Dawson K, Wirth C, Li Y, Connolly Y *et al.* (2014) A global non-coding RNA system modulates fission yeast protein levels in response to stress. Nat. Commun., 5: 3947.

Levine M (2011) Paused RNA polymerase II as a developmental checkpoint. 145(4): 502-511.

Lewis BA and Reinberg D (2003) The mediator coactivator complex: functional and physical roles in transcription regulation. J. Cell Sci., 116: 3667-3675.

Liebers R, Rassoulzadegan M, and Lyko F (2014) Epigenetic regulation by heritable RNA. PLoS Genet., 10(4): e1004296.

Li B, Howe L, Anderson S, Yates JR, and Workman JL (2003) The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. J. Biol. Chem., 278(11): 8897-8903.

Li B, Carey M, and Workman JL (2007) The role of chromatin during transcription. Cell, 128(4): 707-719.

Li B, Gogol M, Cary M, Pattenden SG, Seidel C *et al.* (2007) Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. Genes Dev., 21: 1422-1430.

Li F, Martienssen R, and Cande WZ (2011) Coordination of DNA replication and histone modification by the Rik1-Dos2 complex. Nature, 475: 244-248.

Li G and Reinberg D (2011) Chromatin higher-order structures and gene regulation. Curr. Opin. Genet. Dev., 21(2): 175-186.

Li H, Ilin S, Wang W, Duncan EM, Wysocka J *et al.* (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature, 442: 91-95.

Li L, Lio B, Wapinski OL, Tsai MC, Qu K *et al.* (2013) Targeted disruption of Hotair leads to homeotic transformation and gene de-repression. Cell Rep., 5(1): 3-12.

Li W, Notani D, Ma Q, Tanasa B, Nunez E *et al.* (2013) Functional roles of enhance RNAs for oestrogen-dependent transcriptional activation. Nature, 489: 516-520.

Lilley DM and Pardon JF (1979) Structure and function of chromatin. Annu. Rev. Genet., 13: 197-233.

Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V *et al.* (1997) Reverse transcriptase motifs in the catalytic subunit of telomerase. Science, 276(5312): 561-567.

Liu Q, Greimann JC, and Lima CD (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell, 127(6): 1223-1237.

Luco RF and Misteli T (2011) More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. Curr. Opin. Genet. Dev., 21: 366-372.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature, 389(6648): 251-260.

Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze SH *et al.* (2002) Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. Science, 298: 1747-1752.

Makino DL, Baumgärtner M, and Conti E (2013) Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. Nature, 495(7439): 70-75.

Maison C and Almouzni G (2004) HP1 and the dynamics of heterochromatin maintenance. Nat. Rev. Mol. Cell Biol., 5(4): 296-304.

Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R *et al.* (2012) Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell, 151(3): 671-683.

Martens JA, Laprade L, and Winston F (2004) Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature, 429(6991): 571-574.

Martianov I, Ramadass A, Serra Barros A, Chow N, and Akoulitchev A (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature, 445(7128): 666-670.

Matlin AJ, Clark F, Smith CW (2005) Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol., 6(5): 386-398.

Matoulkova E, Michalova E, Vojtesek B, and Hrstka R (2012) The role of the 3' untranslated region in post-translational regulation of protein expression in mammalian cells. RNA Biol., 9(5): 563-576.

McHugh CA, Chen CK, Chow A, Surka CF, Tran C *et al.* (2015) The Xist lncRNA directly interacts with SHARP to silence transcription through HDAC3. Nature, 521: 232-236.

McManus CJ and Graveley BR (2011) RNA structure and the mechanisms of alternative splicing. Curr. Opin. Genet. Dev., 21(4): 373-379.

Mello CC and Conte D (2004) Revealing the world of RNA interference. Nature, 431: 338-342.

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

Miller G and Hahn S (2006) A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. Nat. Struct. Mol. Biol., 13(7): 603-610.

Milutinovic S, Zhuang Q, and Szyf M (2002) Proliferating cell nuclear antigen associates with histone deacetylase activity, integrating DNA replication and chromatin modification. J. Biol. Chem., 277(23): 2097-2098.

Mitchell SF, Walker SE, Algire MA, Park EH, Hinnebusch AG *et al.* (2010) The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. Mol. Cell, 39(6): 950-962.

Miyakoshi M, Chao Y, and Vogel J (2015) Regulatory small RNAs from the 3' regions of bacterial mRNAs. Curr. Opin. Microbiol., 24: 132-139.

Mohr S, Bakal C, and Perrimon N (2010) Genomic screening with RNAi: results and challenges. Annu. Rev. Biochem., 79: 37-64.

Monnier P, Martinet C, Pontis J, Stancheva I, Ait-Si-Ali S et al. (2013) H19 lncRNA controls gene expression of the imprinted gene network by recruiting MBD1. Proc. Natl. Acad. Sci., 110(51): 20693-20698.

Motamedi MR, Verdel A, Colmenares SU, Gerber SA, Gygi SP *et al.* (2004) Two RNAi complexes, RITS and RDRC, physically interact and localized to noncoding centromeric RNAs. Cell, 119(6): 789-802.

Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G *et al.* (2013) eRNAs promoter transcription by establishing chromatin accessibility at defined genomic loci. Mol. Cell, 51(5): 606-617.

Murakami H, Goto DB, Toda T, Chen ES, Grewal SI *et al.* (2007) Ribonuclease activity of Dis3 is required for mitotic progression and provides a possible link between heterochromatin and kinetochore function. PLoS One, 2(3): e317.

Nagaike T and Manley JL (2011) Transcriptional activators enhance polyadenylation of mRNA precursors. RNA Biol., 8(6): 964-967.

Nakagawa S, Ip JY, Shioi G, Tripathi V, Zong X *et al.* (2012) Malat1 is not an essential component of nuclear speckles in mice. RNA, 18(8): 1487-1499.

Nakagawa S, Naganuma T, Shioi G, and Hirose (2011) Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. J. Cell Biol., 193: 31-39.

Nasim A, Young P, and Johnson BF (1989) Molecular biology of the fission yeast. Academic Press, San Diego.

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T *et al.* (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature, 505(7485): 635-640.

Ng HH, Robert F, Young RA, and Struhl K (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell, 11(3): 709-719.

Ng RK and Gurdon JB (2008) Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. Nat. Cell Biol., 10(1): 102-109.

Ni Z, Saunders A, Fuda NJ, Yao J, Suarez JR et al. (2008) P-TEFb is critical for the maturation of RNA polymerase into productive elongation in vivo. Mol. Cell Biol., 28(3): 1161-1170.

Nishioka K, Chuikov S, Sarma K, Erdjument-Bromage H, Allis CD *et al.* (2002) Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev., 16(4): 479-489.

Nissen P, Hansen J, Ban N, Moore PB, and Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science, 289(5481): 920-930.

Obbard DJ, Gordon KH, Buck AH, and Jiggins FM (2009) The evolution of RNAi as a defence against viruses and transposable elements. Philos. Trans. R. Soc. Lond. B. Biol. Sci., 364(1513): 99-115.

Olson S, Blanchette M, Park J, Savva Y, Yeo GW *et al.* (2007) A regulator of Dscam mutually exclusive splicing fidelity. Nat. Struct. Mol. Biol., 14(12): 1134-1140.

Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A *et al.* (2010) Long noncoding RNAs with enhancer-like function in human cells. Cell, 143(1): 46-58.

O'Sullivan RJ and Karlseder J (2010) Telomeres: protecting chromosomes against genome instability. Nat. Rev. Mol. Cell Biol., 11(3): 171-181.

Palmer AC, Egan JB, and Shearwin KE (2011) Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. Transcription, 2(1): 9-14.

Pan T and Sosnick T (2006) RNA folding during transcription. Annu. Rev. Biophys. Biomol. Struct., 35: 161-175.

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L *et al.* (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage specific transcriptional silencing through chromatin-level regulation. Mol. Cell, 32: 232-246.

Pang KC, Frith MC, and Mattick JS (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet., 22(1): 1-5.

Pelechano V, Wei W, Steinmetz LM (2013) Extentsive transcriptional heterogeneity revealed by isoform profiling. Nature, 497: 127-131.

Petruk S, Sedkov Y, Johnston DM, Hodgson JW, Black KL *et al.* (2012) TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. Cell, 150(5): 922-933.

Petruk S, Sedkov Y, Riley KM, Hodgson J, Schweisguth F *et al.* (2006) Transcription of bxd noncoding RNAs promoter by trithorax represses Ubx in cis by transcriptional interference. Cell, 127(6): 1209-1221.

Phatnani HP and Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev., 20(21): 2922-2936.

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA *et al.* (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science, 300: 131-135.

Ponting CP, Oliver PL, and Reik W (2009) Evolution and functions of long noncoding RNAs. Cell, 136(4): 629-641.

Pontier DB and Gribnau J (2011) Xist regulation and function explored. Hum. Genet., 130(2): 223-236.

Preiss T and Hentze MW (1998) Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. Nature, 392(6675): 516-520.

Preiss T, Muckenthaler M, and Hentze MW (1998) Poly(A)-tail-promoted translation in yeast: implications for translation control. RNA, 4(11): 1321-1331.

Ragunathan K, Jih G, and Moazed D (2015) Epigenetic inheritance uncoupled from sequence-specific recruitment. Science, 348(6230): 1258699.

Ramakrishnan V (2002) Ribosome structure and the mechanism of translation. Cell, 108(4): 557-572.

Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C *et al.* (2009) A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell, 138(1): 114-128.

Rands CM, Meader S, Ponting CP, and Lunter G (2014) 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. PLoS Genet., 10(7): e1004525.

Rasmussen EB and Lis JT (1993) In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc. Natl. Acad. Sci., 90: 7923-7927.

Reyes-Turcu FE, Zhang K, Zofall M, Chen E, Grewal SI (2011) Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. Nat. Struct. Mol. Biol., 18(10): 1132-1138.

Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ *et al.* (2011) Comparative functional genomics of the fission yeasts. Science, 332(6032): 930-936.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell, 129(7): 1311-1323.

Robinson A and van Oijen AM (2013) Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. Nat. Rev. Microbiol., 11: 303-315.

Rosonina E, Bakowski MA, McCracken S, and Blewcowe BJ (2003) Transcriptional activators control splicing and 3'-end cleavage levels. J. Biol. Chem., 278: 43034-43040.

Sander JD and Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol., 32(4): 347-355.

Sanger F, Nicklen S, and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci., 74(12): 5463-5467.

Sarraf SA and Stancheva I (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol. Cell, 15(4): 595-605.

Saxonov S, Berg P, and Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc. Natl. Acad. Sci., 103(5): 1412-1417.

Schmeing TM and Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. Nature, 461(7268): 1234-1242.

Schneider C, Anderson JT, and Tollervey D (2007) The exosome subunit Rrp44 plays a direct role in RNA substrate recognition. Mol. Cell, 27(2): 324-331.

Schorderet P and Duboule D (2011) Structural and functional differences in the long non-coding RNA hotair in mouse and human. PLoS Genet., 7(5): e1002071.

Schweingruber ME, Edenharter E, Zurlinder A, and Stockmaier KM (1992) Regulation of pho1-encoded acid phosphatase of Schizosaccharomyces pombe by adenine and phosphate. Curr. Genet., 22(4): 289-292.

Seizi M, Hartmann H, Hoeg F, Kurth F, Martin DE *et al.* (2011) A conserved GA element in TATA-less RNA polymerase II promoters. PLoS One, 6(11): e27595.

Shah S,Wittmann S, Kilchert C, and Vasiljeva L (2014) lncRNA recruits RNAi and the exosome to dynamically regulate pho1 expression in response to phosphate levels in fission yeast. Genes Dev., 28(3): 231-244.

Shichino Y, Yamashita A, and Yamamoto M (2014) Meiotic long non-coding meiRNA accumulates as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmi1. Open Biol., 4(6): 140022.

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Cluastal Omega. Mol. Syst. Biol., 7:536.

Sigova AA, Abraham BJ, Molinie B, Hannett NM, Guo YE *et al.* (2015) Transcription factor trapping by RNA in gene regulatory elements. Science, 350: 978-981.

Skourti-Stathaki and Proudfoot (2014) A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes and Dev., 28: 1384-1396.

Slotkin RK and Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet., 8: 272-285.

Somboonthum P, Ohta H, Yamada S, Onishi M, Ike A *et al.* (2005) cAMP-responsive element in TATA-less core promoter is essential for haploid-specific gene expression in testis. Nucleic Acids Res., 33(10): 3401-3411.

Soppe WJ, Jasencakova Z, Houben A, Kakutani T, Meister A *et al.* (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J, 21(23): 6549-6559.

Spitz F and Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet., 13: 613-626.

Stadelmayer B, Micas G, Gamot A, Martin P, Malirat N *et al.* (2014) Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. Nat. Commun., 5: 5531.

Struhl K (2007) Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol., 14: 103-105.

Svejstrup JQ (2007) Contending with transcriptional arrest during RNAPII transcript elongation. Trends Biochem. Sci., 32(4): 165-171.

Svejstrup JQ, Li Y, Fellows J, Gnatt A, Bjorkland S *et al.* (1997) Evidence for a mediator cycle at the initiation of transcription. Proc. Natl. Acad. Sci., 94: 6075-6078.

Swiezewski S, Liu F, Magusin A, and Dean C (2009) Cold-induced silencing by long antisense transcripts of an Arabidopsis polycomb target. Nature, 462(7274): 799-802.

Takayama KI, Horie-Inoue K, Katayama S, Suzuki T, Tsutsumi S *et al.* (2013) Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. EMBO J, 32(12): 1665-1680.

Temin HM and Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. 226(5252): 1211-1213.

Thebault P, Boutin G, Bhat W, Rufiange A, Martens J *et al.* (2011) Transcription regulation by the noncoding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. Mol. Cell Biol., 31(6): 1288-1300.

Theimer CA and Feigon J (2006) Structure and function of telomerase RNA. Curr. Opin. Struct. Biol., 16(3): 307-318.

Thomas MC and Chiang CM (2006) The general transcription machinery and general cofactors. Crit. Rev. Biochem. Mol. Biol., 41(3): 105-178.

Thomas MJ, Platas AA, and Hawley DK (1998) Transcriptional fidelity and proofreading by RNA polymerase II. Cell, 93(4): 627-637.

Tran H, Schilling M, Wirbelauer C, Hess D, and Nagamine Y (2004) Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. Mol. Cell, 13(1): 101-111.

Tremethick DJ (2007) Higher-order structures of chromatin: the elusive 30 nm fiber. Cell, 128(4): 651-654.

Tuck AC and Tollervey D (2013) A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell, 154(5): 996-1009.

Uhler JP, Hertel C, and Svejstrup JQ (2007) A role for noncoding transcription in activation of the yeast PHO5 gene. Proc. Natl. Acad. Sci., 104(19): 8011-8016.

Ulitsky I, Shkumatava A, Jan CH, Sive H, and Bartel DP (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell, 147(7): 1537-1550.

Usheva A, Maldonado E, Goldrin A, Lu H, Houbavi C *et al.* (1992) Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. Cell, 69(5): 871-881.

van Werven FJ, Neuert G, Hendrick N, Lardenois A, Buratowski S *et al.* (2012) Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. Cell, 150(6): 1170-1181.

Vance KW, Sansom SN, Lee S, Chalei V, Kong L *et al.* (2014) The long non-coding RNA Paupar regulates the expression of both local and distal genes. EMBO J, 33(4): 296-311.

Vasiljeva L, Kim M, Terzi N, Soares LM, and Buratowski S (2008) Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. Mol. Cell, 29(3): 313-323.

Venkatesh S, Smolle M, Li H, Gogol MM, Saint M *et al.* (2012) Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature, 489(7516): 452-455.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ *et al.* (2001) The sequence of the human genome. Science, 291(5507): 1304-1351.

Venter U, Svaren J, Schmitz J, Schmid A, and Horz W (1994) A nucleosome precludes binding of the transcription factor Pho4 in vivo to a critical target site in the PHO5 promoter. EMBO J., 13(2): 4848-4855.

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S *et al.* (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science, 303(5658): 672-676.

Verzijlbergen KF, Menendez-Benito V, van Welsem T, van Deventer SJ, Lindstrom DL *et al.* (2010) Recombination-induced tag exchange to track old and new proteins. Proc. Natl. Acad. Sci., 107(1): 64-68.

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI *et al.* (2002) Regulation of heterochromatin silencing and histone H3 lysine-9 methylation by RNAi. Science, 297(5588): 1833-1837.

Volpe T and Martienssen RA (2011) RNA interference and heterochromatin assembly. Cold Spring Harb. Perspect. Biol., 3(9): a003731.

Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T *et al.* (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev., 12(3): 343-356.

Wagschal A, Rousset E, Basavarajaiah P, Contreras X, Harwig A *et al.* (2012) Microprocessor, SetX, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. Cell, 150(6): 1147-1157.

Wang J, Reddy BD, and Jia S (2015) Rapid epigenetic adaptation to uncontrolled heterochromatin spreading. eLife, 06179.

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R *et al.* (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature, 472(7341): 120-124.

Wang L, Lewis MS, and Johnson AW (2005) Domain interactions within the Ski2/3/8 complex and between the Ski complex and Ski7p. RNA, 11(8): 1291-1302.

Wang Y, Jensen RC, and Stumph WE (1996) Role of TATA box sequence and orientation in determining RNA polymerase II/III transcription specificity. Nucleic Acids Res., 24(15): 3100-3106.

Wang Z, Gerstein M, and Snyder M (2009) RNA-seq: a revolutionary tool for trascriptomics. Nat. Rev. Genet., 10(1): 57-63.

Watson JD and Crick F (1953) A structure for deoxyribose nucleic acid. Nature, 171: 737-738.

Wei W, Pelechano V, Jarvelin AI, and Steinmetz LM (2011) Functional consequences of bidirectional promoters. Trends Genet., 27(7): 267-276.

Werner F and Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. Nat. Rev. Microbiol., 9(2): 85-98.

West S, Gromak N, and Proudfood NJ (2004) Human 5' -> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature, 432: 522-525.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V *et al.* (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature, 453(7199): 1239-1243.

Will CL and Luhrmann R (2011) Spliceosome structure and function. Cold Spring Harb. Perspect. Biol., 3(7). pii: a003707.

Wilusz JE, Sunwoo H, and Spector DL (2009) Long noncoding RNAs: functional surprises from the RNA world. Genes Dev., 23(13): 1494-1504.

Winter J, Jung S, Keller S, Gregory RI, and Diederichs S (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat. Cell Biol., 11(3): 228-234.

Xiao T, Hall H, Kizer DO, Shibata Y, Hall MC *et al.* (2003) Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev., 17(5): 654-663.

Xu Z, Wei W, Gagneur J, Clauder-Münster S, Smolik M *et al.* (2011) Antisense expression increases gene expression variability and locus interdependency. Mol. Syst. Biol., 7: 468.

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S *et al.* (2009) Bidirectional promoters generate pervasive transcription in yeast. Nature, 457(7232): 1033-1037.

Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A *et al.* (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell, 97(1): 41-51.

Yamanaka S, Mehta S, Reyes-Turcu FE, Zhuang F, Fuchs RT *et al.* (2013) RNAi triggered by specialized machinery silences developmental genes and retrotransposons. Nature, 493: 557-560.

Yamanaka Y, Faghihi MA, Magistri M, Alvarez-Garcia O, Lotz M *et al.* (2015) Antisense RNA controls LRP1 sense transcript expression through interaction with a chromatin-associated protein, HMGB2. Cell Rep., 11(6): 967-976.

Yamashita A, Shichino Y, Tanaka H, Hiriart E, Touat-Todeschini L *et al.* (2012) Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. Open Biol., 2(3): 120014.

Yamashita A, Takayama T, Iwata R, and Yamamoto M (2013) A novel factor Iss10 regulates Mmi1-mediated selective elimination of meiotic transcripts. Nucleic Acids Res., 41(21): 9680-9687.

Yang C, Bolotin E, Jiang T, Sladek FM, and Martinez E (2007) Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene, 389(1): 52-65.

Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnirke A *et al.* (2010) Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. Genome Biol., 11(8): R87.

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF *et al.* (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science, 309(5734): 626-630.

Zafar MA, Carabetta VJ, Mandel MJ, and Silhavy TJ (2014) Transcriptional occlusion caused by overlapping promoters. Proc. Natl. Acad. Sci., 111(4): 1557-1561.

Zee BM, Levin RS, DiMaggio PA, and Garcia BA (2010) Global turnover of histone post-translational modifications and variants in human cells. Epigenetics Chromatin, 3: 22.

Zhang B, Arun G, Mao YS, Lazar Z, Hung G *et al.* (2012) The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. Cell Rep., 2(1): 111-123.

Zhang B and Cech TR (1997) Peptide bond formation by in vitro selected ribozymes. Nature, 390(6655): 96-100.

Zhang K, Fischer T, Porter RL, Dhakshnamoorthy J, Zofall M *et al.* (2011) Clr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. Science, 331(6024): 1624-1627.

Zhao J, Kessley M, Helmling S, O'Connor JP, and Moore C (1999) Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor. Mol. Cell Biol., 19(11): 7733-7740.

Zhao J, Sun BK, Erwin JA, Song JJ, and Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science, 332(5902): 750-756.

Zofall M, Yamanaka S, Reyes-Turcu FE, Zhang K, Rubin C *et al.* (2012) RNA elimination machinery targeting meiotic mRNAs promoters facultative heterochromatin formation. Science, 335(6064): 96-100.