The Fission Yeast Non-Coding Transcriptome

Sophie Radha Atkinson

A thesis submitted for the degree of

Doctor of Philosophy

University College London

September 2014

Declaration

I, Sophie Radha Atkinson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Long non-coding RNAs (lncRNAs) are emerging as important regulators of gene expression, although it remains unclear to what extent they contribute overall to the information flow from genotype to phenotype. Using strand-specific RNA-sequencing, I identify thousands of novel unstable, or cryptic, lncRNAs in *Schizosaccharomyces pombe*. The nuclear exosome, the RNAi pathway and the cytoplasmic exonuclease Exo2 represent three key pathways regulating lncRNAs in *S. pombe*, defining the overlapping classes of CUTs, RUTs and XUTs, respectively. The nuclear exosome and the RNAi pathway act cooperatively to control nuclear lncRNA expression, while the cytoplasmic Exo2 pathway is more distinct. Impairing both the nuclear exosome and the cytoplasmic exonuclease Exo2 is lethal in *S. pombe*.

Importantly, I show that CUTs, RUTs and XUTs are stabilised under physiologically relevant growth conditions, with three key groups emerging: late meiotic RUTs/XUTs, early meiotic CUTs and quiescent CUTs. Late meiotic RUTs/XUTs tend to be antisense to protein-coding genes, and anti-correlate in expression with their sense loci. In contrast, early meiotic and quiescent CUTs tend to be transcribed divergently from protein-coding genes and positively correlate in expression with their mRNA partners.

The current study provides an in-depth survey of the lncRNA repertoire of *S. pombe*, and the pathways that regulate their expression. It seems likely that any regulatory functions mediated by most of these lncRNAs are in *cis*, nuclear and co-transcriptional. The current study provides a rich and comprehensive resource for future studies of lncRNA function.

Acknowledgements

I would like to thank Jürg Bähler for being an excellent supervisor and for giving me the opportunity to work in his lab. I have greatly benefited from his vision and insight, and constant willingness to offer his time and support throughout my PhD. I would also like to thank Samuel Marguerat for his invaluable input and guidance, especially during the initial stages of my PhD, and for teaching me everything I need to know about R.

Thank you to all past and present members of the Bähler group for all their help and advice. I'm grateful to Xavi for his patience in teaching me how to do tetrad dissections, and to Babis, Maria and Michal for always being so willing to help me troubleshoot wet-lab experiments. Thank you to Danny for providing countless discussions on the intricacies of analysing RNA-seq data, and to Sandra for managing the smooth running of the lab. In particular, I'd like to thank Maria, Michal, Danny and Dave for their friendship and support – I hope the climbing and pub trips will continue.

Thank you to all my friends for keeping me sane throughout this whole process, providing valuable distractions when necessary, and helping me keep my eye on the bigger picture. Thank you to my brother and sister-in-law, Stephen and Karen, and my beautiful niece, India, for always being there. Most importantly, thank you to my Mum and Dad for their constant love and support.

Publications arising from work done in this thesis

Atkinson SR, Marguerat S, Bähler J. 2012. Exploring long non-coding RNAs through sequencing. Semin Cell Dev Biol 23: 200-205.

Atkinson SR, Marguerat S, Lemay JF, Bachand F, Bähler J (manuscript in preparation). The fission yeast non-coding transcriptome.

Bitton DA, **Atkinson SR**, Rallis C, Smith GC, Ellis DA, Jeffares DC et al. (manuscript in preparation). Skip and bin: pervasive alternative splicing is degraded by nuclear RNA surveillance in fission yeast.

Lemay JF, Larochelle M, Marguerat S, **Atkinson SR**, Bähler J, Bachand F. 2014 (in press). The RNA exosome promotes transcription termination of backtracked RNA polymerase II. Nat Struct Mol Biol.

Abbreviations

AS	Antisense
bp	Base-pair
CAGE	Cap analysis of gene expression
CC	Correlation coefficient
cds	Coding sequence
cDTA	Comparative dynamic transcriptome analysis
ceRNAs	Competitive endogenous RNAs
CESR	Core environmental stress response
CFU	Colony forming units
ChIP-seq	Chromatin immunoprecipitation with sequencing
ChIRP-seq	Chromatin isolation by RNA purification
circRNAs	Circular RNA
CUT	Cryptic unstable transcript
dChIRP	Domain specific chromatin isolation by RNA purification
DE	Differentially expressed
DSR	Determinant of selective removal

eCUT	Elongated CUT
EMM	Edinburgh minimal media
eRNA	Enhancer RNA
GO	Gene ontology
GRO-seq	Global run-on sequencing
GTF	General transcription factor
GWAS	Genome-wide association studies
lincRNA	Long intergenic non-coding RNA
lncRNA	Long non-coding RNA
MEA	Malt extract agar
miRNA	MicroRNA
mRNA	Messenger RNA
mRNP	Messenger ribonucleoprotein particle
ncRNA	Non-coding RNA
NFR	Nucleosome free region
NGD	No-go decay
NGS	Next-generation sequencing

NH_4Cl	Ammonium chloride
NMD	Nonsense mediated decay
NNS	Nrd1-Nab3-Sen1
NPD	Non parental ditype
NSD	Non-stop decay
nt	Nucleotide
NUT	Nrd1-unterminated transcript
OD	Optical density
ORF	Open reading frame
PD	Parental ditype
PIC	Pre-initiation complex
piRNA	Piwi interacting RNA
Pol II	RNA polymerase II
rDNA	Ribosomal DNA
RIP	RNA immunoprecipitation
RNA-seq	RNA sequencing
RNAi	Interference RNA
RPKM	Reads per kilobase per million

rpm	Revolutions per minute
siRNA	Small-interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
SUT	Stable unannotated transcript
T	Tetratype
ts	Temperature sensitive
TSS	Transcription start site
TTS	Transcription termination site
UTR	Untranslated region
WT	Wild-type
XUT	Xrn1-dependent unstable transcript
YES	Rich media

Table of Contents

Table of Figures

Table of Tables

Chapter 1 Introduction

1.1 Overview

Long non-coding RNAs (lncRNAs) are emerging as important regulators of gene expression, although it remains unclear to what extent they contribute to the information flow from genotype to phenotype.

Since lncRNAs are pervasively transcribed, in this Chapter I will firstly consider unifying features of lncRNAs at a genome-wide level, focusing on their distribution throughout the genome and their post-transcriptional regulation. I will then discuss known functional mechanisms of lncRNAs, with a specific emphasis on our current understanding of lncRNAs in yeast model systems of the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe*. Based on outstanding questions in the field, I will then outline the aims of this thesis.

1.2 Genomes are pervasively transcribed

1.2.1 Transcriptomics reveals pervasive lncRNA transcripts

A profusion of microarray and RNA-sequencing (RNA-seq) studies, on species ranging from microbes to humans, has revealed that the transcribed portions of genomes are more pervasive and complex than anticipated (Carninci et al. 2005, Mortazavi et al. 2008, Okazaki et al. 2002, Wilhelm et al. 2008). For example, less than 2% of the human genome encodes proteins, yet as much as 80% of all DNA is transcribed (Birney et al. 2007, Mattick 2011).

Such pervasive transcription leads to a multitude of previously unknown non-protein coding RNAs (ncRNAs), which have an arbitrary minimal length cut-off of 200 nucleotides (due to RNA-seq library preparation protocols that exclude small RNAs), and are subsequently referred to as long ncRNAs (lncRNAs). These transcripts are distinct from small regulatory ncRNAs such as miRNAs and piRNAs, which are typically conserved, shorter than 30 nucleotides, and are known to be involved in transcriptional and post-transcriptional gene silencing through specific base-pairing with their mRNA targets (Ghildiyal and Zamore 2009).

1.2.2 General features of lncRNAs

Like mRNAs, lncRNAs are generally considered to be RNA Polymerase II (Pol II)-transcribed (Rhee and Pugh 2012), capped (Neil et al. 2009) and polyadenylated (David et al. 2006).

lncRNAs tend to be much more lowly expressed and less conserved in sequence than protein-coding genes. Such features may argue against a functional role for lncRNAs, and have led to the speculation that the majority of lncRNAs likely represent some form of basic transcriptional noise, generated at low levels throughout the genome due to the inability of the transcription machinery to identify true promoters (Struhl 2007). Since many lncRNAs are associated with mRNA genes (see Section 1.3), it is possible that their transcription may simply reflect the accessibility of the chromatin structure in such regions of the genome, but the transcripts themselves possess no obvious function.

Despite this, however, there is some evidence that lncRNAs have been subject to purifying selection, hinting at their possible functionality (Guttman et al. 2009, Ponjavic et al. 2009). Furthermore, when using conservation as an indicator of functionality, functional sequences have often been equated with sequence conservation across a diverse range of species. By analyzing sequence constraint between pairs of eutherian genomes over a range of divergences, Rands et al. (2014) have recently shown that noncoding sequences become increasingly mutually constrained as species pairs become more closely related. Thus, some noncoding sequences are conserved, but over shorter evolutionary timescales than protein-coding sequences. Together with the observation that protein-coding genes are largely conserved in number and function between organisms of highly different complexities, while the extent of lncRNAs increases much more with organismal complexity (Mattick 2011), lncRNAs have been speculated to contribute to species specific biology. Thus, lncRNAs may contribute significantly to phenotypic differences between species, possibly through differences in genome regulation that they may mediate (Kutter et al. 2012, Mattick 2011, Rands et al. 2014).

While a high expression level and a high degree of sequence conservation are generally considered to indicate how important a protein-coding gene's function is, the same may not be true for lncRNAs if one considers that their mode of action may be fundamentally different. For example, if lncRNA functions are exerted through the act of being transcribed, rather than the transcript itself, then high steady state expression levels and a high degree of sequence conservation may not be necessary for, or indicative of, function. Emerging evidence that at least some lncRNAs may function in this way is discussed in Section 1.5.4.

Finally, although, by definition, lncRNAs are not considered to be translated, several recent studies are now challenging this view. Ribosomal profiling – the deep sequencing of ribosome-protected RNAs – has been used to determine whether any lncRNAs are found in association with ribosomes (Duncan and Mata 2014, Ingolia et al. 2011, Ingolia et al. 2014, Wilson and Masel 2011). Remarkably, these studies have revealed that a significant proportion of annotated lncRNAs in *S. cerevisiae* (Wilson and Masel 2011) and mouse embryonic stem cells (Ingolia et al. 2011, Ingolia et al. 2014) are exported to the cytoplasm, where they are engaged by the protein translation machinery. A recent study in *S. pombe* has employed ribosome profiling to demonstrate that 24% of annotated lncRNAs are engaged with ribosomes (Duncan and Mata 2014).

However, the extent to which this engagement with ribosomes results in the production of small peptides, and subsequently the extent to which lncRNAs might act via peptides they encode rather than via RNA itself, remains an open question. Although it is unclear whether such translation is functionally important, such studies are leading to the concept that in addition to the genome being pervasively transcribed, the transcriptome is also pervasively translated.

In summary, lncRNAs are pervasively transcribed throughout genomes, generally not translated, lowly expressed and poorly conserved. Additional features, which may inform function, and that will be considered in the following sections, include: a non-random genome distribution (Section 1.3); regulation by RNA degradation pathways (Section 1.4). I will then consider known functional mechanisms of lncRNAs (Section 1.5), before outlining the aims of this thesis (Section 1.6).

1.3 Genomic origins of lncRNA transcription

lncRNAs are non-randomly distributed throughout the genome, and can be transcribed antisense to protein-coding genes, from promoters of coding transcripts, from enhancers or repetitive elements, or separate from coding genes (Figure 1.1). Each of these origins of lncRNAs will be considered in the following sections.

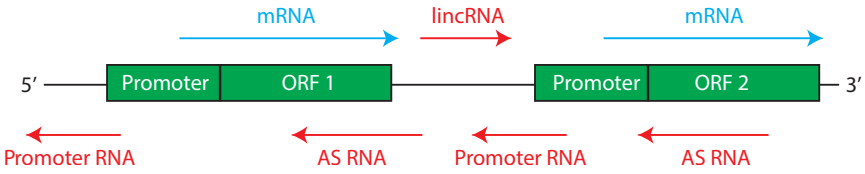

Figure 1.1: Different types of lncRNAs by location.

Schematic showing different genomic locations of lncRNAs. Green: DNA, with promoters and open reading frames (ORFs) depicted as boxes; blue arrows: messenger RNAs (mRNAs); red arrows: lncRNAs. Examples for different lncRNA locations are antisense (AS) RNAs, promoter RNAs transcribed in the opposite direction to mRNAs, and a long intergenic RNA (lincRNA). Some lncRNAs are associated with enhancer elements or with repeated regions. See main text for details. Adapted from Atkinson et al. (2012).

1.3.1 Antisense lncRNAs

Standard RNA-seq libraries do not preserve information about transcriptional direction. However, several methods now exist for strand-specific RNA-seq (Levin et al. 2010), revealing complex overlapping transcription in several genomes, with many lncRNAs being transcribed from the complementary strand of protein-coding genes. Such lncRNAs are referred to as antisense (AS) transcripts.

Initial studies using whole-genome tiling arrays in both *S. pombe* and *S. cerevisiae* have revealed extensive AS transcription in rapidly proliferating cells (David et al. 2006, Dutrow et al. 2008, Wilhelm et al. 2008, Xu et al. 2011). Application of strand-specific RNA-seq to these eukaryotic microbes has now provided maps of AS transcription across different growth conditions at higher resolution (Ni et al. 2010, Rhind et al. 2011, Yassour et al. 2010). Such studies provide accurate detection of

AS transcript boundaries, uncovering overlaps with variable portions of coding genes. In addition, AS transcript co- or anti-regulation (or lack thereof) with neighboring genes suggest that they can arise from novel transcription start sites, from bidirectional transcription at promoters, or via transcriptional read-through in the case of convergent genes. The same studies have additionally reported that detectable AS transcripts are more likely to overlap genes involved in sexual differentiation or stress response (Ni et al. 2010, Rhind et al. 2011, Yassour et al. 2010), as well as genes with higher variability in transcript levels (Xu et al. 2011). Notably, certain sense/AS transcript pairs and their co-regulation are conserved across several yeast species (Rhind et al. 2011, Yassour et al. 2010). Taken together, these data suggest that AS transcription is a phenomenon of pervasive gene expression with diverse features and impacts on gene regulation.

There are many described functions of individual antisense lncRNAs and only a few will be briefly described here. Most described antisense functions have an inhibitory effect on expression of their sense loci. For example, a recent study has elucidated a role for an antisense lncRNA in circadian clock function in *Neurospora crassa* (Xue et al. 2014). A lncRNA antisense to the circadian gene frequency (frq), was found to oscillate in expression antiphase to the frq sense RNA. In fact, mutual inhibition of sense and antisense transcripts was found to form a double negative feedback loop that is required for robust and sustained circadian rhythmicity. Furthermore, the authors suggest that antisense transcription inhibits sense expression by mediating chromatin modifications and premature transcription termination.

Similar examples of an inhibitory effect of antisense lncRNAs on their sense loci, mediated via chromatin modifications, have been described in budding yeast. Stabilisation of *S. cerevisiae PHO84* or *GAL1-10* antisense transcripts, during chronological ageing or repressive nutrient conditions respectively, leads to repression of their respective sense loci via the recruitment of histone deacetylases (Camblong et al. 2007, Houseley et al. 2008).

Although expression of antisense lncRNAs and their sense loci are generally considered to anti-correlate, mechanisms by which antisense lncRNAs can mediate positive effects on expression of their sense loci have also been described. For

example, in *S. cerevisiae* the stress-activated protein kinase Hog1 induces transcription of a lncRNA antisense to the cyclin-dependent kinase *CDC28* (Nadal-Ribelles et al. 2014). Transcription of the antisense lncRNA mediates the establishment of gene looping and the relocalisation of Hog1 from the 3'UTR of *CDC28* to the +1 nucleosome, where it induces *CDC28* expression. This enables cells to reenter the cell cycle more efficiently after stress, and has been proposed to represent a general mechanism to prime expression of genes needed after stresses are alleviated.

Microarray- and next generation sequencing (NGS)-based approaches have also revealed AS transcription in mammalian genomes (Birney et al. 2007, Carninci et al. 2005, Guttman et al. 2010). However, to what degree and by what mechanisms AS transcripts control sense transcripts on a global level remain to be fully elucidated.

1.3.2 Bidirectional lncRNAs

NGS-based approaches have helped to reveal an unanticipated property of Pol II: transcription initiation can be bidirectional. For example, cryptic unstable transcripts (CUTs) are Pol II-dependent transcripts defined in *S. cerevisiae*, which largely result from bidirectional transcription. CUTs are degraded by the nuclear exosome shortly after synthesis (Wyers et al. 2005). Two recent studies have generated genome-wide maps of CUTs using an NGS-based approach (Neil et al. 2009) or high-density tiling arrays (Xu et al. 2009). Both studies have revealed that CUTs are well-defined transcriptional units that tend to be transcribed from gene promoters in an opposite direction to the coding RNAs. Moreover, evidence suggests that such divergent lncRNAs emanate from distinct pre-initiation complexes (PICs) within the same nucleosome free region (NFR). Thus, in *S. cerevisiae*, high-resolution genome-wide studies of general transcription factor (GTF) occupancies have demonstrated that, as for divergent mRNA-mRNA pairs, divergent mRNA-lncRNA pairs employ independent PICs (Rhee and Pugh 2012).

Divergent transcription from promoters is not limited to unstable transcripts that are rapidly targeted by the nuclear exosome: a class of lncRNA in *S. cerevisiae* termed

stable unannotated transcripts (SUTs) can also arise from divergent transcription from known promoters (Xu et al. 2009).

Importantly, similar transcriptional patterns have been observed in other eukaryotes. Exosome depletion in human fibroblasts followed by tiling array analysis has revealed lncRNAs which map upstream of known protein-coding genes. Such promoter-upstream transcripts have been termed PROMPTs (Preker et al. 2008). In addition, stable transcripts mapping to both strands of promoters are detected in metazoan cells. RNA-seq of short RNAs (~20-200 nucleotides) from mouse embryonic stem cells has shown that many of these transcripts originate from promoters (transcription start site-associated RNAs) and are transcribed in a non-random, divergent direction (Seila et al. 2008). Furthermore, global run-on sequencing (GRO-seq) has been used to identify nascent RNAs associated with actively transcribing Pol II in human fibroblasts, independently of nascent transcript length or stability (Core et al. 2008). This analysis has revealed that 77% of transcriptionally active protein-coding genes display significant divergent transcription from promoters.

Divergent mRNA-lncRNA pairs tend to positively correlate in expression (Neil et al. 2009, Xu et al. 2009), which raises the question as to whether such lncRNAs are simply by-products of transcription, or whether they represent functional transcriptional units. There have been suggestions that bidirectional transcription may facilitate protein-coding gene expression by promoting an open-chromatin structure at promoters, or by recruiting positive or negative transcriptional regulators. Adachi and Lieber (2002) found bidirectional loci in the human genome to be enriched for essential housekeeping genes, whose promoters are associated with CpG islands, and depleted for TATA-elements. Wang G. Z. et al. (2011) extend this observation, postulating that transcription of divergent lncRNAs is a noise-reduction, expression-priming mechanism. As such, divergent lncRNAs are more associated with essential genes and less associated with stress-response genes.

Wu and Sharp (2013) propose that divergent transcription may shape the evolution of the genome by enabling new gene origination. By examining the distribution of U1 snRNP binding sites and poly(A)-signal motifs, which are known to promote and

inhibit transcription, respectively, the authors propose that the act of transcription at bidirectional promoters renders the upstream antisense region of an active gene vulnerable to accumulation of G and T content. This means that A-rich poly(A)-signal motifs are likely to be lost, while GT-rich U1 snRNP binding sites are likely to be gained. Consequently upstream antisense transcriptional activity is enhanced, which may eventually drive origination of a new-antisense-oriented gene. In support for such a scenario, evolutionary age-dependent progressive gain of U1 sites and loss of poly(A)-signal motifs within promoter-proximal regions of mouse genes has been reported (Almada et al. 2013).

In summary, bidirectional transcription from promoters seems to be a widespread phenomenon conserved across evolution. Although various functions of bidirectional lncRNAs have been proposed, on a global level it is unclear whether bidirectional lncRNAs are non-functional by-products of coding transcription, or whether they play regulatory roles.

1.3.3 Enhancer associated lncRNAs

Enhancers spatially and temporally regulate protein-coding transcription from a distance, and in an orientation-independent manner, relative to the regulated gene. Initial studies of the β-globin locus (Ling et al. 2005) have revealed that the hypersensitive site 2 (HS2) enhancer is transcribed into a lncRNA. Subsequently, several NGS-based studies have revealed widespread transcription from enhancers, resulting in a class of lncRNAs termed enhancer RNAs (eRNAs).

Using chromatin immunoprecipation with sequencing (ChIP-seq) to map genomic binding sites of the enhancer protein CBP, Kim T. K. et al. (2010) identified over 12000 stimulus-dependent enhancers in mouse neurons. Furthermore, ChIP-seq showed Pol II to be present at over 25% of those enhancers. RNA-seq of total RNA from neurons has revealed lncRNAs that are produced in both directions from such Pol II-bound enhancers, and the expression levels of these eRNAs are correlated with mRNA synthesis from nearby genes (Kim T. K. et al. 2010).

In a related study, De Santa et al. (2010) have applied ChIP-seq of Pol II in macrophages has been used to reveal actively elongating Pol II bound to putative enhancer sites upstream of lipopolysaccharide-inducible genes. Many of these Pol II-bound enhancers are actively transcribed, and eRNA expression correlates with the expression of neighbouring genes.

Notably, eRNA synthesis frequently precedes the induction of the adjacent coding gene (De Santa et al. 2010), raising the possibility that enhancer function may in fact be mediated through transcribed eRNAs. While it has been speculated that eRNAs may recruit enhancer-associated proteins, or perhaps facilitate chromatin looping to provide contact between the enhancer region and the promoter of the regulated gene, further studies are required to determine any biological functions and mechanisms of action of eRNAs.

1.3.4 Long intergenic ncRNAs (lincRNAs)

In mammals, large projects such as those of the FANTOM and ENCODE consortiums (Birney et al. 2007, Carninci et al. 2005) have uncovered widespread intergenic transcription which does not overlap with protein-coding genes. Guttman et al. (2009) have developed a method to systematically identify such long intergenic ncRNAs (lincRNAs) using genome-wide chromatin-state maps, generated by ChIP-seq, to identify discrete transcriptional units occurring between protein-coding genes. Based on the observation that Pol II-transcribed genes display H3K4Me3 marks at their promoters and H3K36Me3 marks along their bodies, 'K4-K36' domains lying outside of known protein-coding genes were uncovered. This approach identified 1600 lincRNAs in four mouse cell types.

More recently, a comprehensive annotation of human lincRNAs has been achieved using transcriptome assembly of RNA-seq data from 24 human tissues and cell types (Cabili et al. 2011). Importantly, the identified lincRNAs are expressed in a highly tissue-specific manner, much more so than protein-coding genes. In addition, protein-coding genes proximal to lincRNAs are disproportionately associated with development and transcriptional regulation (Guttman et al. 2010, Guttman et al.

2009, Ponjavic et al. 2009). Both of these observations hint at possible functional roles of lincRNAs.

One of the most well-characterised lincRNAs is HOTAIR, which is transcribed within the HOXC cluster and represses genes in the HOXD cluster by binding and recruiting the chromatin-modifying complex polycomb repressive complex 2 (PRC2) (Rinn et al. 2007). Khalil et al. (2009) show that many lincRNAs are bound by PRC2 and other chromatin-modifying complexes. Moreover, RNAi-based depletion of various PRC2-associated lincRNAs results in activation of genes known to be repressed by PRC2. Such studies have led to the suggestion that lincRNAs guide chromatin-modifying complexes to specific genomic loci (see Section 1.5.1).

It is worth noting that eRNAs and lincRNAs are generally lacking in yeast model systems, due to the lack of annotated enhancer elements, and smaller more compact genomes resulting in fewer clearly intergenic regions.

1.3.5 Repetitive element-associated lncRNAs

Repetitive elements such as retrotransposons comprise 30-50% of mammalian genomes, and the advent of NGS technologies has uncovered transcriptional activity associated with such elements.

Using a 'deep-CAGE' (cap analysis of gene expression) method to globally map transcription start sites (cleavage of ~20-nucleotide tags from extreme 5' and 3' ends of cDNAs, followed by sequencing), the FANTOM4 project has revealed extensive transcription of retrotransposons in human and mouse genomes (Faulkner et al. 2009). Retrotransposons are expressed in a tissue-specific manner and proximal to protein-coding genes, leading to the suggestion of roles in regulation of protein-coding genes (Faulkner and Carninci 2009).

Pseudogenes form another class of repetitive elements that can be transcribed into lncRNAs, as has recently been shown for the PTEN and KRAS loci, which can regulate expression of their corresponding protein-coding genes by competing for regulatory miRNA binding (see Section 1.5.3) (Poliseno et al. 2010, Tay et al. 2011).

In summary, the genome-wide distribution of lncRNAs is non-random. Although lncRNAs can be described by their locations in the genome, it is not yet clear whether, at a global level, such distinct locations reflect distinct biological functions. Importantly, RNA degradation pathways post-transcriptionally regulate lncRNAs. There is emerging evidence that lncRNAs targeted by certain degradation pathways could represent regulatory classes of lncRNAs. This will be considered in the next section.

1.4 Regulation of pervasive transcription by RNA degradation

A universal feature of pervasive non-coding transcription is that it is regulated. Uncontrolled accumulation of noncoding RNA can adversely affect genome stability (Li and Manley 2006). Such RNA may be toxic due to its interaction with complementary DNA sequences, possibly interfering with the process of transcription itself. Alternatively, RNA accumulation may be toxic if inappropriately translated, or if RNA-binding factors are sequestered (Jensen et al. 2013). Eukaryotes have therefore evolved multiple strategies to regulate lncRNA expression. These pathways often comprise RNA degradation and processing.

1.4.1 Overview of RNA degradation pathways

RNA degradation is important for regulating RNA levels and preventing deleterious accumulation of aberrantly transcribed RNAs. I will firstly focus on known pathways for mRNA surveillance and degradation (Figure 1.2), and then discuss the relevance of these pathways to lncRNA expression (Section 1.4.3).

The following is not an exhaustive list of all factors involved in RNA degradation, but aims to provide an overview of the key steps in eukaryotic RNA degradation, with a specific focus on the yeast model systems of *S. cerevisiae* and *S. pombe*. Key emerging themes of RNA degradation will be highlighted in Section 1.4.2.

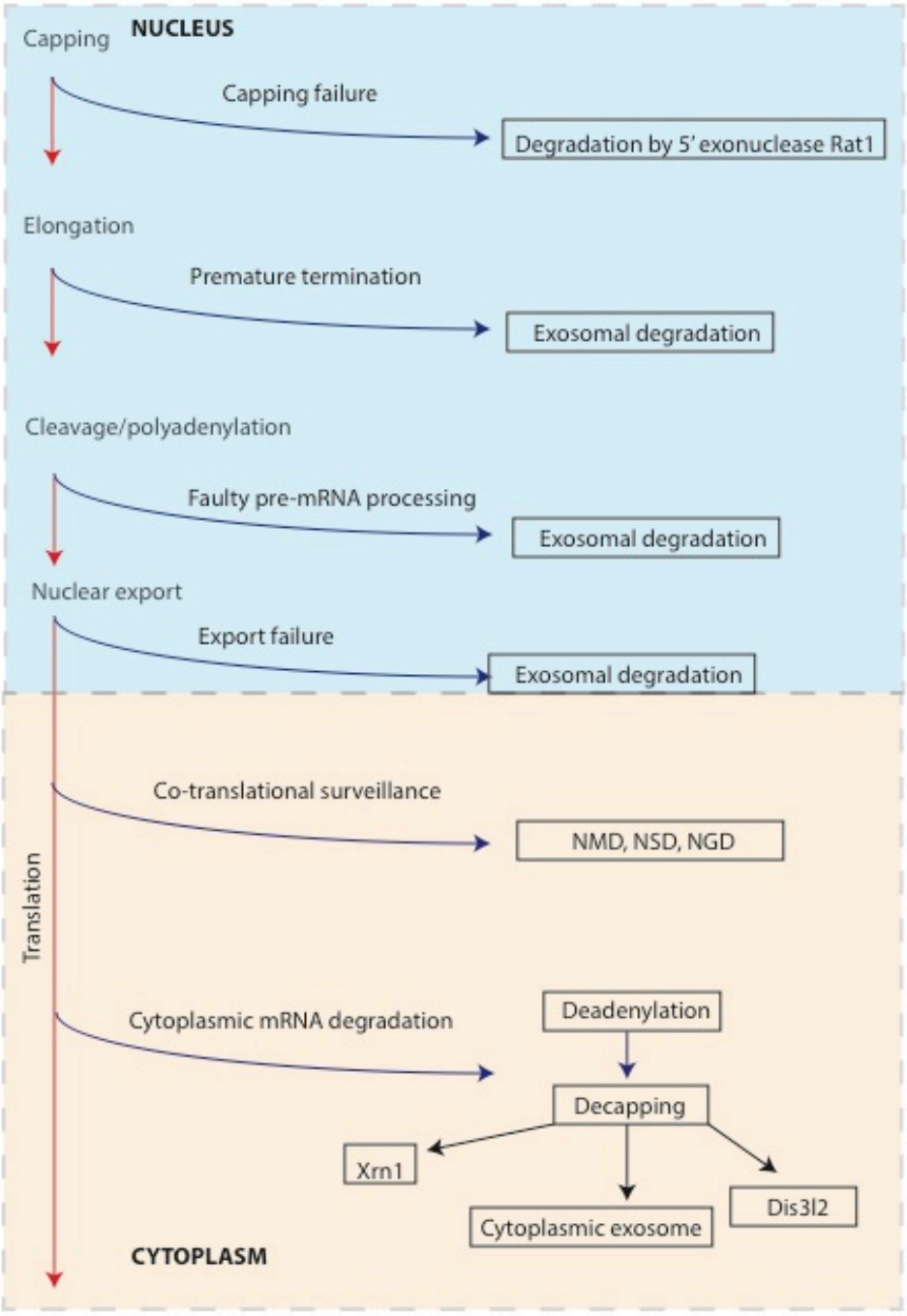

Figure 1.2: Overview of some key steps of mRNA surveillance and degradation.

Adapted from (Houseley and Tollervey 2009).

The exosome multi-protein complex is central to RNA degradation. The exosome degrades RNAs in a 3'-5' direction and is highly conserved across eukaryotes (Houseley and Tollervey 2009). The exosome acts on almost all RNAs at some

point during their lifetime, contributing to the processing, quality control and turnover of RNAs in both the nucleus and the cytoplasm. It consists of nine inactive subunits plus an active ribonuclease, Dis3, which has both 3'-5' exonucleolytic as well as endonucleolytic activity. The exonuclease and endonuclease activities of Dis3 likely cooperate on most substrates, with the endonucleolytic activity thought to play a role in releasing stalled exosome substrates (Schneider et al. 2012)

Rrp6 is a 3'-5' exonuclease which associates with the exosome to confer nuclear specificity. Rrp6 has overlapping and complementary roles to Dis3 in RNA degradation (Gudipati et al. 2012, Schneider et al. 2012). In contrast, the Ski multi-protein complex confers cytoplasmic specificity to the exosome. Ski2, Ski3 and Ski8 assemble in a tetramer with 1:1:2 stoichiometry, with Ski2 constituting the helicase core of the complex (Halbach et al. 2013). The Ski complex is thought to effectively channel single-stranded cytoplasmic RNA substrates into the exosome, where they are subsequently degraded by Dis3. Ski7 is a cytoplasmic exosome cofactor which connects the exosome to the Ski complex (Wang L. et al. 2005).

In addition to the exosome, other degradation factors act to target substrates to the exosome, to activate the exosome, or in a complementary manner to exosomal activities. I will firstly focus on nuclear factors, before moving onto cytoplasmic factors.

The TRAMP (Trf4/Air2/Mtr4) polyadenylation complex acts as a major cofactor for the nuclear exosome in eukaryotic cells (LaCava et al. 2005). The TRAMP complex consists of a non-canonical poly(A)-polymerase (Trf4 or Trf5 in *S. cerevisiae*; Cid14 in *S. pombe*), an RNA-binding protein (Air1 or Air2), and an RNA helicase (Mtr4). In the nucleus, aberrant RNAs are recognized by the TRAMP complex, which add a short oligo(A) tail to enable exosome-dependent degradation. All exonucleases have problems initiating degradation close to stable stem structures, and so the oligo(A) tail likely acts as a "landing pad", enabling the exonucleotytic activity of the nuclear exosome (Houseley and Tollervey 2009).

In addition to the nuclear exosome, aberrant nascent RNAs with an incomplete 5' cap structure are degraded by the nuclear 5'-3' exonuclease Dhp1 in *S. pombe*

(Shobuike et al. 2001), which is a homologue of Rat1 in *S. cerevisiae* (Amberg et al. 1992).

As mRNAs are packaged into messenger ribonucleoprotein particles (mRNPs) for export to the cytoplasm, several surveillance systems monitor nuclear mRNP biogenesis and export. Imperfect mRNPs can be targeted and degraded by the nuclear exosome together with the TRAMP complex. Other factors, such as the endonuclease Swt1, which transiently associates with nuclear pore complexes, are also important for targeting aberrant mRNPs (Skruzny et al. 2009).

In the cytoplasm, mRNA degradation is closely linked to translation. There are three predominant forms of co-translational mRNA surveillance: nonsense-mediated decay (NMD), no-go decay (NGD) and nonstop decay (NSD) (Shoemaker and Green 2012). An mRNA that causes the ribosome to stall is subjected to NGD; mRNAs that contain a nonsense codon are subjected to NMD; and mRNAs that lack a termination codon are subjected to NSD.

In the cytoplasm, the 5'-3' and 3'-5' mRNA decay pathways are generally preceded by deadenylation of the poly(A) tail (by Ccr4-Not and Pan2-Pan3 complexes), followed by removal of the 5'-cap (Yamashita et al. 2005). The 5'-cap is removed by the decapping enzyme Dcp2, which is promoted by its co-activator Dcp1 (Chang et al. 2014), and other associated factors including the Lsm1-7 complex (He and Parker 2000).

Following deadenylation and decapping, mRNAs are accessible to the Ski complex and exosome for 3' degradation, or 5' degradation by the cytoplasmic 5'-3' exonuclease Xrn1 (homologue of Exo2 in *S. pombe*) (Muhlrad et al. 1994). In addition to cytoplasmic degradation by the exosome and Xrn1, the 3'-5' exonuclease Dis3l2 has been shown to degrade RNAs in the cytoplasm in an exosome-independent manner. Importantly, Dis3l2 is conserved from *S. pombe* (Malecki et al. 2013) to humans (Lubas et al. 2013).

1.4.2 Themes of RNA degradation pathways

The following aims to highlight some of the complexities and emerging themes of RNA degradation.

It is important to note that many RNA degradation factors additionally participate in RNA processing as opposed to degradation. Thus a single enzyme can precisely process some RNA species to generate defined ends while retaining the capacity to degrade other RNAs entirely. For example, the Nrd1-Nab3-Sen1 (NNS) complex triggers transcription termination on *S. cerevisiae* snRNAs and snoRNAs, releasing a non-polyadenylated transcript that undergoes 3' exonucleolytic processing by the nuclear exosome to produce mature 3' ends of snoRNAs and snRNAs (Steinmetz et al. 2001, Vasiljeva et al. 2008). How the exosome distinguishes between substrates targeted for processing and substrates destined for complete degradation remains an open question (Houseley and Tollervey 2009). The exosome possesses a catalytically inert core with the RNases Dis3 and Rrp6 bound at either end. The helicase components of either the nuclear TRAMP complex or the cytoplasmic Ski complex are thought to thread RNA substrates through the internal channel of this core to the RNases. However, Dis3 can also associate directly with substrates via channel-independent routes, and it has been recently suggested that processed substrates are more likely to be targeted directly to catalytic subunits, while substrates that are directed towards discard pathways are more likely to be threaded through the exosome core (Mitchell 2014).

Another common theme in RNA degradation pathways is that they form a complex interwoven network (Sun M. et al. 2013a, Tuck and Tollervey 2013), characterized by extensive redundancy but also substrate preferences. For example, Sun et al. (2013) revealed differences in mRNA substrate preference between the two deadenylase complexes, Ccr4-Not and Pan2-Pan3, as well as between NMD factors Upf2 and Upf3, while Dis3 and Rrp6 have been shown to have overlapping and specific substrates (Gudipati et al. 2012, Schneider et al. 2012).

RNA degradation pathways are also highly interconnected with other aspects of RNA metabolism including transcription termination, translation, and even

transcription itself. In fact, although many RNA processing factors have traditionally been thought of as being primarily involved in degradation of transcripts, recent studies have made it clear that RNA levels are regulated by a network of these factors via a dynamic interplay of regulated synthesis as well as degradation.

A recent global analysis of mRNA synthesis and degradation upon deletion of 46 degradation factors in *S. cerevisiae* revealed that mRNA levels are buffered upon perturbation of degradation rates. Thus a decreased degradation rate is accompanied by a compensatory decreased synthesis rate, and vice versa (Sun M. et al. 2013a, Sun M. et al. 2012). Crucially, all mutant strains analyzed displayed such mRNA buffering, except for the strain lacking Xrn1, indicating that buffering requires Xrn1. The authors propose a model whereby global synthesis and degradation rates are controlled by Xrn1-dependent induction of global transcription repressors such as Nrg1.

Importantly, other studies also point to key role for Xrn1 in mRNA synthesis as well as degradation. In *S. cerevisiae*, Xrn1 was found to shuttle between the cytoplasm and nucleus where it also functions as a genome-wide transcription factor, binding to the 5' region of a number of genes (Haimovich et al. 2013). In support of this, Xrn1 has also been recently described as a 'synthedegradase' referring to its regulation not only of mRNA degradation, but also mRNA synthesis, possibly by preventing Pol II backtracking during elongation (Medina et al. 2014).

1.4.3 RNA degradation pathways regulating lncRNAs

Understanding how lncRNA expression is regulated ultimately informs the likelihood, and potential mechanism, of molecular function. Recent studies, particularly in *S. cerevisiae*, have revealed classes of lncRNAs, which are stabilised upon impairment of various degradation factors. There is evidence that such lncRNAs, normally targeted for rapid degradation, are stabilised and functional under relevant growth conditions, thus raising the possibility that environmental perturbation could modulate RNA degradation factors, which in turn modulate regulatory lncRNA expression.

The transcriptomes of *S. cerevisiae rrp6* mutants - defective for the nuclear exosome - have revealed a class of transcripts referred to as CUTs (cryptic unstable transcripts), which are rapidly degraded after synthesis, and arise largely from bidirectional transcription from mRNA promoters (Neil et al. 2009, Xu et al. 2009). There is evidence that physiological conditions may affect the unstable nature of CUTs and render them stable. For example, loss of the Rrp6 protein in *S. cerevisiae* leads to stabilisation of a lncRNA antisense to the *PHO84* gene, and subsequent repression of *PHO84* transcription. Intriguingly, the same phenotype is observed during chronological ageing, as the Rrp6 protein shows weaker association with the *PHO84* locus under these conditions (Camblong et al. 2007). More recently, the study has been extended to a genome-wide analysis of histone modification mutants in a *rrp6*-deletion background, revealing that at least 28 *S. cerevisiae* genes appear to be repressed by antisense CUTs in a similar manner to PHO84 (Castelnuovo et al. 2014). In addition, a class of budding yeast meiotic lncRNAs is actively degraded during the mitotic cell-cycle by the nuclear exosome and becomes stabilised as the cell proceeds into meiotic differentiation (Lardenois et al. 2011). These findings indicate that modulation of Rrp6 function by external, physiologically relevant cues could contribute to gene regulation via regulation of lncRNA levels. Furthermore, unstable transcripts could represent regulatory classes of lncRNAs.

In *S. cerevisiae*, the NNS complex is composed of the RNA-binding proteins Nrd1 (Seb1 in *S. pombe*) and Nab3, and the helicase Sen1. In budding yeast, the NNS complex has been shown to play important roles in the transcription termination of sn/sno-RNAs and their subsequent targeting to the exosome for 3'-end processing (Steinmetz et al. 2001, Vasiljeva et al. 2008). More recently it has also been shown to play a role in the transcription termination of CUTs and their subsequent targeting to the exosome for degradation (Arigo et al. 2006, Thiebaut et al. 2006). In fact, a recent genome-wide study identified ~1500 NUTs (Nrd1-unterminated transcripts) detectable upon rapid nuclear depletion of Nrd1 (Schulz et al. 2013). NUTs extensively overlap CUTs (68%), but poorly overlap mRNAs. Moreover Nrd1 and Nab3 binding sites are depleted in mRNAs but enriched in NUTs, indicating that NNS is essential for selectively terminating this class of lncRNAs. NUTs largely arise from bidirectional transcription and are proposed to inhibit mRNA expression via transcriptional interference and antisense repression.

A recent structural study has found that the same domain of Nrd1 interacts with Pol II and Trf4 (TRAMP complex component) in a mutually exclusive manner (Tudek et al. 2014). Moreover, Nrd1-Trf4 interaction was shown to be required for optimal exosome activity. Thus there are two alternative forms of the NNS complex – one associated with Pol II and functioning in termination, and the other associated with TRAMP and promoting exosomal degradation. This provides a structural mechanism for the functional coupling of transcription termination and degradation of lncRNAs.

A further class of unstable lncRNAs in budding yeast has been identified using strand-specific RNA-seq of an *xrn1* exonuclease mutant (van Dijk et al. 2011). This study identified 1658 Xrn1p-sensitive unstable transcripts (XUTs) degraded by the 5'-3' cytoplasmic exonuclease, Xrn1. More than 50% of the identified XUTs are AS lncRNAs and are proposed to have repressive effects on their sense loci. Moreover, XUTs accumulate in lithium-containing media indicating a possible role in adaptive responses to changing growth conditions. XUTs overlap substantially with CUTs and NUTs (Schulz et al. 2013), and although targeting by Xrn1 indicates XUTs are exported to the cytoplasm, it is unclear whether any function they may have is exerted in the nucleus or cytoplasm (Jensen et al. 2013).

The above studies illustrate that, although certain degradation factors play roles in degrading both mRNAs and lncRNAs, their mode of action on these substrates likely differ. Both mRNAs and lncRNAs are Pol II-transcribed, and acquire a 5' cap and a poly(A) tail. Despite these similarities, their fate and function are clearly different - while mRNAs are exported to the cytoplasm for translation, a large fraction of lncRNAs are proposed to have diverse, largely nuclear, functions in regulating mRNA expression (see Section 1.5), and factors such as the exosome are proposed to regulate their expression. Recent analyses of mRNA and ncRNA maturation in budding yeast suggest that transcript fate depends on the use of distinct mechanisms of 3' end formation (Tuck and Tollervey 2013). While mRNAs have well-defined 3' ends, formed by cleavage and polyadenlylation at the poly(A)-site by canonical 3'-end processing machinery (cleavage and polyadenylation factor, CPF, and cleavage factors CFI/II), lncRNAs often have heterogeneous 3' ends. This is due to a distinct early transcription termination pathway by NNS, and subsequent targeting to the

exosome. Tuck and Tollervey (2013) additionally propose that the cleavage and polyadenylation factors Hrp1 and Nab2 have dual roles in differentiating between mRNAs and lncRNAs. Sequence nonspecific binding of these factors may be a default activity that is associated with recruitment of the nuclear surveillance system to lncRNAs. In contrast, Hrp1 binding to poly(A)-consensus motifs, and Nab2 bound to poly(A) at the 3' end of mRNAs in the context of the canonical 3' end processing machinery, may assist in the generation of stable mRNAs which are exported to the cytoplasm.

In summary, lncRNAs are post-transcriptionally regulated by some of the same factors that target mRNAs, but in a distinct manner leading to distinct outcomes. There is emerging evidence that unstable lncRNAs, which are rapidly targeted for degradation, may be stabilised and functional under certain physiologically relevant growth conditions. Thus, systematic RNA-seq analysis of mutants defective for RNA processing pathways may represent a powerful approach to reveal novel, possibly regulatory, classes of lncRNAs (see Section 1.6).

1.5 Emerging themes of lncRNA functional mechanisms

Determining the nature and possible biological functions of lncRNAs has been a rapidly developing field over the past decade (Figure 1.3), pioneered by mechanistic studies of single genes. Such studies have unraveled a variety of functions of lncRNAs in gene regulation, some of which are discussed below. The following will not aim to exhaustively cover all functional mechanisms elucidated for lncRNAs, but instead aims to highlight some of the key emerging themes.

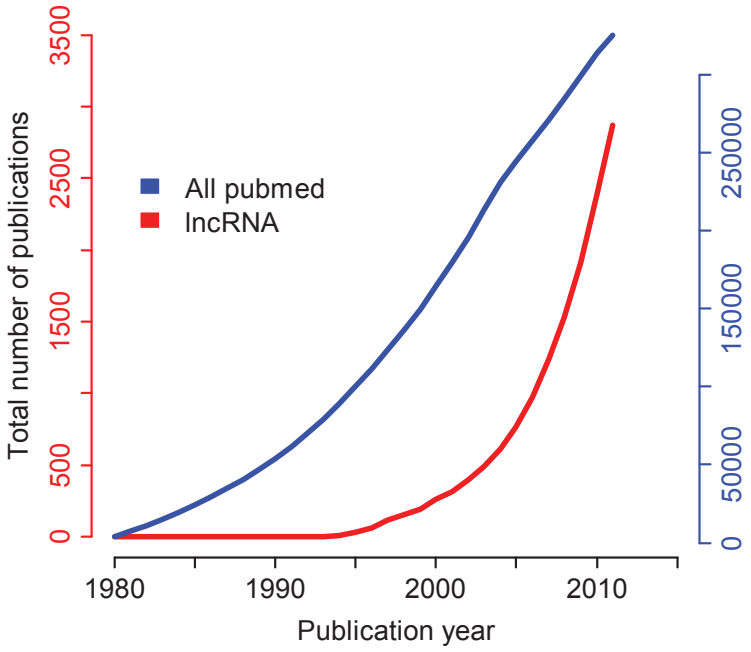

Figure 1.3: Research on lncRNAs is rapidly increasing.

The total number of publication entries in PubMed (blue line and axis) and of entries related to long non-coding RNAs (red line and axis). Adapted from Atkinson et al. (2012).

1.5.1 Guides for chromatin modifying complexes

Several studies have described a role for lncRNAs in binding chromatin-modifying complexes and guiding them to appropriate genomic locations. These studies have led to the suggestion that lncRNAs recruit chromatin-modifying complexes to specific genomic loci (Koziol and Rinn 2010). Such a function helps solve the apparent paradox of how chromatin-remodeling complexes, with little DNA sequence specificity but often with RNA-binding domains, are able to control complex chromatin modifications at specific genomic loci.

The example of lincRNAs, such as HOTAIR, which bind and recruit the PRC2 chromatin-modifying complex has already been discussed (see Section 1.3.4). Such lincRNAs are thought to largely act *in trans* with these lincRNAs controlling chromatin states at loci in a distinct location from those from which they are transcribed. In contrast, a recent study describes an example where *cis*-acting antisense lncRNAs act to recruit a chromatin-modifying complex at their site of

transcription. Thus, quiescence-induced lncRNAs are transcribed antisense to ribosomal DNA (rDNA) in murine tissue culture (Bierhoff et al. 2014). Such lncRNAs recruit a H4K20 methyltransferase to rDNA, thereby leading to chromatin compaction and subsequent repression of ribosomal genes upon growth factor deprivation or terminal differentiation. Importantly, this mechanism of heterochromatin formation is distinct from heterochromatin formation at pericentric and telomeric regions in that it requires neither H3K9 trimethylation nor interaction with heterochromatin protein 1 (HP1) to recruit H4K20 methyltransferases.

Analogous to this recruitment of chromatin-modifying complexes *in cis*, lncRNAs emerging from promoters or enhancers have been proposed to act as scaffolds by recruiting and possibly coordinating transcriptional activators and repressors (Mercer et al. 2009, Wang K. C. and Chang 2011).

1.5.2 Evictors

While some lncRNAs have been shown to recruit proteins to specific genomic loci, others may evict factors from chromatin. This has been demonstrated during the onset of X-chromosome inactivation, which requires expression of the Xist RNA. Thus, upon onset of X-chromosome inactivation, an X-encoded lncRNA, transcribed from the *Jpx* gene ~10kb upstream of the *Xist* gene, is up-regulated. This lncRNA binds and titrates away the repressive CTCF protein from the Xist promoter, thus enabling its transcriptional activation (Sun S. et al. 2013b).

A similar evictor role has also been recently described for lncRNAs transcribed from the pericentromeric heterochromatin borders in *S. pombe* (Keller et al. 2013). Such lncRNAs, termed BORDERLINE, compete with methylated histone H3K9 (a hallmark of heterochromatin) for binding the HP1 protein Swi6. Thus Swi6 is evicted from heterochromatin at these sites, and the spreading of heterochromatin regions into neighbouring euchromatic regions is prevented.

Intriguingly, BORDERLINE lncRNAs can be replaced with heterologous sequence without compromising boundary activity, demonstrating that these RNAs act in a sequence-independent but locus-dependent manner. Thus, an interesting mechanistic

difference between the eviction of CTCF and Swi6 is that while Jpx RNA binds CTCF with sequence specificity, any RNA will bind the hinge regions of Swi6 (most likely via positively charged residues in the hinge domain) and cause a conformational change of the chromodomain, resulting in its loss of affinity for H3K9-methylated nucleosomes.

1.5.3 Sponges and decoys

Several studies have reported 'decoy' roles for lncRNAs in sequestering miRNAs or transcription factors, thereby regulating gene expression. For example, lncRNAs arising from transcription of PTEN and KRAS pseudogenes regulate expression of their corresponding protein-coding genes by competing for regulatory miRNA binding (Poliseno et al. 2010, Tay et al. 2011). Such lncRNAs have been termed competitive endogenous RNAs (ceRNAs).

A recent study (Guo et al. 2014) identified ~3000 transcribed human pseudogenes that could produce lncRNAs in a manner of low abundance but high tissue specificity. Analysis of RNA-seq data revealed that many of these transcribed pseudogenes could produce small RNAs (sRNAs). Such pseudogene-derived sRNAs were associated with H3K9Me3 enrichment at both the pseudogene loci and their adjacent regions. This suggests that, in addition to acting as ceRNAs, lncRNAs derived from pseudogene transcription may play a role in regional chromatin repression, and raises the possibility that lncRNAs may exert function by acting as precursors for small regulatory ncRNAs. Such processing to smaller RNAs is reminiscent of the previously described BORDERLINE lncRNAs, which are processed by dicer into small RNAs termed brdrRNAs (Keller et al. 2013).

ceRNAs are similar in function to circular RNAs (circRNAs). Two recent studies have described circRNAs acting as stable miRNA reservoirs, thereby relieving suppression of mRNA targets (Hansen et al. 2013, Memczak et al. 2013). Thousands are expressed in a tissue- and developmental-specific manner and many derive from head-to-tail splicing of coding exons.

1.5.4 Co-transcriptional functions

Several studies have suggested that the process of lncRNA transcription, rather than the lncRNA product itself, may be functional.

Non-coding transcription may impact gene regulation by contributing to nucleosome positioning. Several studies have described such a mechanism of action. For example, lncRNA transcription has been shown to facilitate an open chromatin structure at protein-coding promoters, thereby increasing access to transcriptional activators and to Pol II. Such a mechanism has been shown to function at the *S. pombe fbp1*$^{+}$ locus, where a cascade of several species of lncRNAs are transcribed through the *fbp1*$^{+}$ promoter, progressively altering chromatin structure and making the DNA accessible to transcriptional activators and Pol II (Hirota et al. 2008). Interestingly, promoter proximal non-coding transcription may also have a negative impact on protein-DNA interactions. For example, during *S. cerevisiae* growth on rich media, the serine biosynthesis gene *SER3* is tightly repressed. An upstream lncRNA, termed *SRG1*, has been shown to be transcribed under such conditions, and overlaps the *SER3* promoter. *SRG1* transcription represses the downstream *SER3* gene by enhancing nucleosome assembly over the *SER3* promoter, thereby limiting transcription factor access (Hainer et al. 2011, Martens et al. 2004).

It has also been reported that lncRNA transcription can introduce co-transcriptional regulatory histone marks at loci they overlap. Two recent studies in *S. cerevisiae* have shown that overlapping non-coding transcription can regulate gene expression by creating a repressive chromatin landscape of histone methylation and deacetylation. In haploid cells, expression of *IME1*, the central inducer of gametogenesis, is inhibited by transcription of lncRNA *IRT1* in the *IME1* promoter. Overlapping *IRT1* transcription recruits Set2 histone methyltransferase and Set3 histone deacetylase to establish repressive chromatin at the *IME1* promoter (van Werven et al. 2012). Kim T. et al. (2012) performed a genome-wide study to determine how Set3 affects gene expression. In line with the findings of van Wervern et al., they found that the majority of Set3-affected genes had overlapping non-coding transcription. Thus overlapping non-coding transcription may be a more widespread mechanism to fine-tune gene expression by depositing H3K4Me2, which in turn recruits Set3 HDAC.

Similar mechanisms of non-coding transcription recruiting histone methylation and deacetylation activities have been described for the repression of both the *GAL1-GAL10* locus (Houseley et al. 2008) and the *PHO84* gene (Camblong et al. 2007) by transcription of lncRNAs antisense to these loci (see Section 1.3.1). Recent evidence suggests that these antisense lncRNA effects may be mediated co-transcriptionally. For example, single molecule RNA FISH (fluorescence *in situ* hybridization) analyses revealed that sense and antisense *PHO84* RNA never coexist in the *PHO84* locus, but instead *PHO84* antisense (AS) lncRNA is efficiently exported to the cytoplasm (Castelnuovo et al. 2013). Furthermore, rather than stabilising *PHO84* AS-lncRNA (Camblong et al. 2007), loss of Rrp6 reduces *PHO84* AS-lncRNA early termination by NNS. These observations suggest *PHO84* AS-lncRNA effects are likely mediated via its transcription rather than by static lncRNA accumulation.

Such functional mechanisms, which are dependent on transcription of lncRNAs rather than the lncRNA product itself, may help reconcile poor sequence conservation and low steady-state levels with functional capacity (Section 1.2.2).

In summary, molecular mechanisms are emerging for lncRNA function, but those lncRNAs that have had a functional mode of action elucidated are still relatively few. Since lncRNAs represent widespread pervasive transcripts, determining whether more global mechanisms of action exist is of intense interest. Techniques such as Chromatin Isolation by RNA Purification (ChIRP)-seq are likely to be instrumental in providing such global insights into function. Chromatin is cross-linked to lncRNAs and biotinylated oligonucelotide probes are subsequently used to retrieve specific lncRNAs, together with bound DNA sequences which can then be interrogated by NGS (Chu et al. 2011). Just as ChIP-seq has greatly improved our understanding of protein-DNA interactions on a genomic scale, ChIRP-seq has the potential to map lncRNA:chromatin interactions *in vivo*, genome-wide and at high resolution. A more recently described variation of this technique – domain-specific chromatin isolation by RNA purification (dChIRP) – uses biotinylated oligonucleotide probes to recover chromatin fragments containing specific lncRNA domains of interest (Quinn et al. 2014). The RNA, DNA and protein components associated with the lncRNA domains can then be analysed separately. Characterising lncRNAs at the domain level may provide insights into to how the

modularity of specific domains could contribute to the diverse functions of lncRNAs in gene regulation.

1.6 Aims of this thesis

Unlike microarray-based approaches, RNA-seq does not depend on available genome annotation or sequence, can accurately detect expression levels over a wide dynamic range, and can reveal entire transcriptomes with high sensitivity and to nucleotide resolution (Marguerat and Bahler 2010, Ozsolak and Milos 2011, Wang Z. et al. 2009). Thus, RNA-seq allows the global detection and quantification of lncRNAs, revealing lncRNA genomic origins and expression regulation, and thereby providing insight into functions on a genome-wide scale.

Larger genomes have more complex transcriptomes, and therefore require a greater sequencing depth for adequate coverage. Given sequencing costs, simpler genomes are more amenable to high coverage to reveal their entire transcript repertoire. Therefore simple model organisms with smaller, more compact genomes, such as yeasts, are instrumental in revealing the nature of often lowly expressed, overlapping and complex lncRNA transcripts.

So far lncRNAs have been more extensively studied in in the budding yeast *S. cerevisiae*, but the existence of RNA metabolism pathways more akin to higher eukaryotes, such as interference (RNAi) and splicing, in the fission yeast *S. pombe*, makes this yeast an important complementary model in which to study lncRNAs.

A much larger proportion of the *S. pombe* genome is transcribed than can be accounted for by currently annotated genes (Dutrow et al. 2008, Wilhelm et al. 2008). This additional transcription likely comprises extensions of existing annotated transcripts, biological and experimental noise, as well as novel ncRNAs.

Two RNA-sequencing studies have together annotated approximately 1500 lncRNAs in wild-type (WT) *S. pombe* (Rhind et al. 2011, Wilhelm et al. 2008), the nature and biological function of which remain unclear. However, many more unstable lncRNAs, such as CUTs and XUTs, which are only detectable when stablilised by impairing RNA degradation machineries, have now been described in *S. cerevisiae*.

This suggests that lncRNAs identified by *S. pombe* RNA-seq studies to date may reflect only partially the rich complexity of the transcriptome. I therefore used strand-specific RNA-seq of a comprehensive panel of *S. pombe* RNA processing mutants to identify and analyse different classes of unstable lncRNAs.

Possible regulatory functions can be imagined for lncRNAs degraded by RNA-processing pathways during vegetative growth but stabilised under physiologically relevant growth conditions. I therefore also employed RNA-seq analysis of nitrogen-starved cells, glucose-starved cells and meiotic cells to address whether the identified classes of unstable lncRNAs likely possess any biological function.

Chapter 2 Materials and Methods

Unless acknowledged as otherwise in the text, I have carried out all experimental work and computational analyses reported in this thesis.

2.1 *S. pombe* strains and growth conditions

All strains used for RNA-seq in the current study are detailed in Tables 2.1 and 2.2.

RNA processing mutants used in the current study are detailed in Table 2.1. All RNA processing mutant cell cultures were harvested at mid-log phase (optical density, OD_{595nm} = 0.5). With the exception of conditional mutants (*rrp6-ts, dis3-54, mtr4*), all RNA processing mutant cultures were grown at 32°C in rich medium (YES) (see Table 2.1). For more details of the relevance of these RNA processing mutants to the current study see Chapter 3.

Different environmental perturbations/physiological conditions used in the current study are detailed in Table 2.2. For more details of the relevance of these growth conditions to the current study see Chapter 3.

Meiotic time-course samples were prepared by Dr Cristina Cotobal (Mata laboratory, University of Cambridge). Briefly, *S. pombe pat1-114* cells were grown to mid-log phase before being shifted to Edinburgh minimal media (EMM) without nitrogen. Cells were incubated at 25°C overnight to synchronise them in G1 phase. Meiosis was induced by addition of NH_4Cl to final concentration of 0.5g/L and incubation at 34°C. Cells were harvested and RNA extracted hourly after induction of meiosis. A pool of meiotic RNA was created by pooling equal amounts from all time-points (0, 1, 2, 3, 4, 5, 6, 7, 8h) (Schlackow et al. 2013).

For glucose starvation experiments, WT cells were grown in EMM at 32°C, 2% glucose. A sample representing 100% survival was harvested when cultures reached a stable maximal density. Measurements of Colony Forming Units (CFUs) were conducted every 24 hours after this initial time-point, and another sample harvested when cultures reached 50% survival. *pka1* samples under glucose starvation were prepared in the same way, with an additional sample taken when cultures reached

87% survival (this is the same time-point at which a WT culture run in parallel reached 50% survival). *pka1* glucose starvation samples were prepared by Dr Charalampos Rallis (Bähler laboratory).

For nitrogen starvation experiments, cells were grown in EMM at 32°C to a stable maximal density. Cells were then washed twice in EMM without nitrogen source (NH_4Cl) and cultured in EMM without nitrogen at 32°C. Cells were harvested at 24 hours and 7 days after nitrogen removal.

The number of biological replicates used for RNA-seq for each sample is stated in Table 2.1 and Table 2.2. In the current study, biological replicates are cultures grown from independent single colonies.

Cells were harvested by centrifugation of 50ml of culture at 2300rpm for 3 minutes. Pellets were snap frozen and stored at -80°C prior to RNA extraction.

Table 2.1: Table of mutants

Name	Information	RNA processing pathway impaired	Genotype	Growth conditions	Number of biological replicates	Strain reference
rrp6-ts	3'-5' exonuclease; nuclear-specific exosome component	Nuclear exosome	*h90 ade6-M216 leu1 rrp6::rrp6.9-GFP kan*	YES; heat-shocked at 36°C for 2.5 hours	2	Harigaya et al. (2006)
rrp6	3'-5' exonuclease; nuclear-specific exosome component	Nuclear exosome	*h+ ade6-M216 leu1-32 ura4-D18 his3-D1 rrp6::URA4*	YES 32°C	2	Lemay et al. (2014) (in press)
pab2	Poly(A)-binding protein; targets RNAs for exosomal degradation	Nuclear exosome	*h+ ade6M216 leu1-32 ura4D18 his3D1 pab2::kanMX6*	YES 32°C	2	Lemieux et al. (2011)
mtr4	RNA helicase; component of the TRAMP complex which acts as an exosome cofactor targeting RNAs to the exosome	Nuclear exosome	*h+ ade6-M216 leu1-32 ura4-D18 his3-D1 p81nmt1-kanMX6::mtr4*	EMM + 60μM thiamine for 15h	2	Lemay et al. (2014) (in press)
cid14	Poly(A)-polymerase; TRAMP complex component	Nuclear exosome	*h- ade6-M210 leu1-32 ura4-D18 his3-D1 cid14::URA4*	YES 32°C	2	Lemay et al. (2010)

Name	Information	RNA processing pathway impaired	Genotype	Growth conditions	Number of biological replicates	Strain reference
dis3-54	Dis3 is a 3'-5' exonuclease and endonuclease; core exosome component. Dis3-54 strain contains an amino acid substitution in the exonuclease domain that impairs its catalytic activity. Endonucleolytic activity is retained in this strain.	Core exosome	*h+ leu1 ura4 his2 dis3-54*	YES 30°C; cold-sensitive	2	Wang S. W. et al. (2008)
ski7	GTPase; cytoplasmic exosome cofactor connecting exosome to Ski complex	Cytoplasmic exosome	*h+ ade6M210 leu1-32 ura4Δ18 his3Δ1 ski7::kanMX6*	YES 32°C	2	Francois Bachand, unpublished
ski2	Cytoplasmic helicase; member of cytoplasmic Ski complex which connects the exosome with its cytoplasmic substrates	Cytoplasmic exosome	*h- ski2::hph*	YES 32°C	2	Malecki et al. (2013)
exo2	Cytoplasmic 5'-3' exonuclease	Cytoplasmic 5'-3' exonuclease	*h- exo2::kanMX6 ade6-216*	YES 32°C	2	Samuel Marguerat, unpublished

Name	Information	RNA processing pathway impaired	Genotype	Growth conditions	Number of biological replicates	Strain reference
lsm1	Member of lsm1-7 complex which binds the 3' ends of transcripts, protecting them from 3' trimming and activating decapping and 5'-3'degradation	Decapping, cytoplasmic 5'-3' degradation	*h- lsm1::hph*	YES 32°C	2	Malecki et al. (2013)
dis3l2	Cytoplasmic 3'-5' exonuclease	Cytoplasmic 3'-5' exonuclease	*h- dis3l2::hph*	YES 32°C	2	Malecki et al. (2013)
pan2	Deadenylase	Cytoplasmic deadenylation	*h- pan2::natMX6 ade6-216*	YES 32°C	2	Bähler lab, unpublished
dcr1	Endoribonuclease cleaving dsRNA	RNAi	*h- Δdcr1::kanMX6, otr1R(Sph1)::ura4, ura4-DS/E, ade6-M210, leu1-32, his7-366*	YES 32°C	2	Volpe et al. (2002)
ago1	Member of the RITS (RNA-induced transcriptional gene silencing) effector complex	RNAi	*h- Δago1::kanMX6, otr1R(Sph1)::ura4, ura4-DS/E, ade6-210, leu1-32, his7-366*	YES 32°C	2	Volpe et al. (2002)

Name	Information	RNA processing pathway impaired	Genotype	Growth conditions	Number of biological replicates	Strain reference
rdp1	RNA-dependent RNA polymerase	RNAi	*h- Δrdp1::kanMX6, otr1R(Sph1)::ura4, ura4-DS/E, ade6-M216, leu1-32, his7-366*	YES 32°C	2	Volpe et al. (2002)
upf1	ATP-dependent RNA helicase	Nonsense mediated decay	*h- upf1::kanMX6*	YES 32°C	2	Rodriguez-Gabriel et al. (2006)
nab2	Poly(A)-binding protein	Competes with pab2	*h+ ade6M210 leu1-32 ura4Δ18 his3Δ1 nab2::kanMX6*	YES 32°C	2	Grenier St-Sauveur et al. (2013)
rrp6/dcr1	See *rrp6* and *dcr1*	Nuclear exosome; RNAi	*h- rrp6::URA4 dcr1::natMX6 ura4-D18*	YES 32°C	2	Sophie Atkinson, this study
exo2/dcr1	See *exo2* and *dcr1*	Cytoplasmic 5'-3' exonuclease; RNAi	*h+ exo2::kanMX6 rrp6::URA4 ura4-D18*	YES 32°C	2	Sophie Atkinson, this study

Name	Information	RNA processing pathway impaired	Genotype	Growth conditions	Number of biological replicates	Strain reference
dis3l2/lsm1	See *dis3l2* and *lsm1*	Cytoplasmic 3'-5' exonuclease; decapping, cytoplasmic 5'-3' degradation	*h- dis3l2::KanMS6 lsm1::hph*	YES 32°C	2	Malecki et al. (2013)
dis3l2/ski2	See *dis3l2* and *ski2*	Cytoplasmic 3'-5' exonuclease; cytoplasmic exosome	*h- dis3l2::KanMS6 ski2::hph*	YES 32°C	2	Malecki et al. (2013)
*dis3**	3'-5' exonuclease and endonuclease; core catalytic exosome component	Exosome	*h+ ade6 leu1-32 ura4-D18 his3-D1 p81nmt1-kanMX6::dis3*	EMM + 60µM thiamine for 15h	2	Lemay et al. (2014) (in press)
*rrp41**	Structural component of core exosome	Exosome	*h+ ade6 leu1-32 ura4-D18 his3-D1 p81nmt1-kanMX6::rrp41*	EMM + 60µM thiamine for 15h	2	Lemay et al. (2014) (in press)
*dhp1**	Nuclear 5'-3' exonuclease; plays a role in transcription termination	Nuclear 5'-3' exonuclease	*h+ ade6-M216 ura4-D18 leu1-32 dhp1-1::ura4+*	YES; heat-shocked at 36°C for 4 hours	2	Shobuike et al. (2001)

Name	Information	RNA processing pathway impaired	Genotype	Growth conditions	Number of biological replicates	Strain reference
*seb1**	RNA-binding protein; homologue of *S. cerevisiae* Nrd1; Nrd1-Nab3-Sen1 (NNS) complex involved in transcription termination and targeting of ncRNAs to nuclear exosome in *S. cerevisiae*	Homologue of *S. cerevisiae* NNS complex component; interacts with exosome	*h+ ade6 leu1-32 ura4-D18 his3-D1 p81nmt1-kanMX6::seb1*	EMM + 60μM thiamine for 15h	2	Francois Bachand, unpublished

* These samples were prepared in collaboration with the Francois Bachand lab, and have contributed to the following publication: *Lemay et al., 2014 (in press). The RNA exosome promotes transcription termination of backtracked RNA polymerase II. Nature Structural and Molecular Biology*. These samples display a global transcriptional read-through phenotype, which complicates the detection of distinct lncRNA transcripts from the RNA-seq data. These samples will therefore not be discussed further in this thesis.

Table 2.2: Table of physiological conditions/environmental perturbations

Name	Details	RNA processing pathway impaired	Genotype	Growth conditions	Number of biological replicates
WT	Mid-log phase WT cells	None	*h⁻ 972* (3 repeats) *h⁺ ade6-M216 leu1-32 ura4-D18 his3-D1* (3 repeats)	EMM 32°C (2 repeats) YES 32°C (3 repeats) EMM + 60μM thiamine for 15h (1 repeat)	6
Meiotic pool	A pool of different meiotic time-points (0, 1, 2, 3, 4, 5, 6, 7h) after induction of meiosis	None	*pat1-114/pat1-114 ade6-M210/ade6-M216* h^+/h^+	See Section 2.1 for details	1
Meiosis_0h	Cells harvested 0 hours after induction of meiosis	None	*pat1-114/pat1-114 ade6-M210/ade6-M216* h^+/h	See Section 2.1 for details	2

Name	Details	RNA processing pathway impaired	Genotype	Growth conditions	Number of biological replicates
Meiosis_2h	Cells harvested 2 hours after induction of meiosis	None	*pat1-114/pat1-114 ade6-M210/ade6-M216* h^+/h	See Section 2.1 for details	2
Meiosis_4h	Cells harvested 4 hours after induction of meiosis	None	*pat1-114/pat1-114 ade6-M210/ade6-M216* h^+/h	See Section 2.1 for details	2
Meiosis_6h	Cells harvested 6 hours after induction of meiosis	None	*pat1-114/pat1-114 ade6-M210/ade6-M216* h^+/h	See Section 2.1 for details	2
Meiosis_8h	Cells harvested 8 hours after induction of meiosis	None	*pat1-114/pat1-114 ade6-M210/ade6-M216* h^+/h	See Section 2.1 for details	2
N_24h	Nitrogen starvation for 24 hours	None	h^- *972*	See Section 2.1 for details	2
N_7d	Nitrogen starvation for 7 days	None	h^- *972*	See Section 2.1 for details	2

Name	**Details**	**RNA processing pathway impaired**	**Genotype**	**Growth conditions**	**Number of biological replicates**
WT_100	WT quiescent cells under glucose starvation at 100% survival	None	*h⁻ 972*	See Section 2.1 for details	2
WT_50	WT quiescent cells under glucose starvation at 50% survival	None	*h⁻ 972*	See Section 2.1 for details	2
pka1_100	Long-lived *pka1* mutant under glucose starvation at 100% survival	None	*h- pka1::kanMX4* (Rallis et al. 2013)	See Section 2.1 for details	2
pka1_87	Long-lived *pka1* mutant under glucose starvation at 87% survival (same time-point as when WT_50 sample was taken)	None	*h- pka1::kanMX4*	See Section 2.1 for details	2

Name	Details	RNA processing pathway impaired	Genotype	Growth conditions	Number of biological replicates
pka1_50	Long-lived *pka1* mutant under glucose starvation at 50% survival	None	*h- pka1::kanMX4*	See Section 2.1 for details	2

2.2 Transcriptomics

RNA was extracted from harvested cells using the hot-phenol technique as described previously (Lyne et al. 2003). The quality of total extracted RNA was assessed on a Bioanalyser instrument (Agilent).

Strand-specific RNA-seq libraries were prepared from poly(A)-enriched (poly(A)+) RNA using an unreleased early version of the Illumina TruSeq Small RNA Sample Prep Kit. Briefly, 10μg of total RNA was used as starting material and poly(A)+ RNA was enriched by two rounds of poly(dT) Sera-Mag magnetic bead purification. RNA was fragmented to an average size of ~200 nucleotides (nt). Fragmented RNA was 3' de-phosphorylated with Antarctic phosphatase and 5' phosphorylated with polynucleotide kinase; this treatment prepared RNA fragments for subsequent ligation of Illumina RNA adaptors to their 5' and 3' ends using a 3' RNA ligase and a T4 RNA ligase respectively. First-strand cDNA was produced using a primer specific for the Illumina 3' adaptor. The library was amplified with 15 PCR cycles using primers specific for the Illumina adaptors and purified using SPRI-beads (Agencourt, Beckman Coulter). Library size distributions and concentrations were determined on a Bioanalyser (Agilent).

RNA-seq libraries were sequenced on an Illumina HiSeq 2000 instrument, using single end runs with 50 base pair (bp) reads (Wei Chen laboratory, The Berlin Institute for Medical Systems Biology, Max-Delbrück-Centrum für Molekulare Medizin, Germany).

Reads were aligned to the fission yeast genome with the exonerate software (Slater and Birney 2005) and reads matching to multiple locations in the genome were assigned at random to only one of these locations. Reads containing up to 5 mismatches (not clustered at read ends) were kept for further analysis ("Genomic set"). The remaining sequencing reads were then mapped against fission yeast spliced exons and filtered as above ("Spliced set"). Pools of both sets of reads were used for further analyses. Between 20 and 50 million mappable reads were obtained for each library (~80-85% of total reads were mappable).

Relative sequencing expression scores were calculated for annotated features (spliced transcripts and introns), using the genome annotation available in GeneDB (http://old.genedb.org/, now PomBase http://www.pombase.org/) on 9th May 2011. The number of reads mapping to each feature was divided by the length of the feature per kilobase and read depth per million mapped reads (reads per kilobase of transcript per million reads mapped, RPKM). RPKMs for annotated features correlated strongly between biological replicates ($r_{pearson}$ >0.98).

Mapping and expression score pipelines were provided by Dr Samuel Marguerat (Bähler laboratory).

2.3 Segmentation of sequencing data

A crude heuristic was designed to detect candidate novel lncRNAs from RNA-seq data. Briefly, RNA-seq data from initial sequencing runs, which comprised the following samples, were pooled together (2 biological repeats of each): *rrp6-ts, exo2, dis3-54, pab2, mtr4, ago1, rdp1, dcr1, pan2, upf1*, WT_100, WT_50, -N_7d, -N_24h, meiotic pool, mid-log phase WT cells grown in YES, mid-log-phase WT cells grown in EMM (see Tables 2.1 and 2.2).

Segments were delimited from the pooled data using a 10 hits/bp cut-off. Segments <100bp apart and differing in pooled read density (i.e. average hits/bp of segments) by <10-fold were joined together. Using the genome annotation available in GeneDB (http://old.genedb.org/, now PomBase http://www.pombase.org/) on 9th May 2011, segments overlapping annotations on the same strand (including untranslated regions, UTRs) were removed.

lncRNAs have an arbitrary minimal length cut-off of 200 nucleotides, due mainly to RNA-seq library preparation protocols that exclude small RNAs (Perkel 2013). Therefore segments <200bp were discarded, and remaining consecutive segments longer than 200bp were retained as final transcripts corresponding to 5775 candidate novel lncRNAs.

The described segmentation heuristic was optimised for its ability to detect the 1557 lncRNAs already annotated in *S. pombe* (SPNCRNAs), and validated by visual inspection of RNA-sequencing data (see Chapter 3 for details).

Custom scripts for segmentation of RNA-seq data were written in R (http://www.r-project.org/) and Perl.

2.4 Identifying differentially expressed transcripts

Candidate novel lncRNAs identified by the segmentation process described above, together with all annotated transcripts, were analysed using the DESeq package (Anders and Huber 2010) (available from Bioconductor, http://www.bioconductor.org/) to determine whether their expression level was significantly up- or down-regulated in each sample (Tables 2.1, 2.2) compared to WT cells in mid-log phase growth. Using the DESeq package, for each transcript a p-value was computed based on a negative binomial distribution. P-values were corrected for multiple testing, and transcripts with an adjusted p-value <0.01 were considered to be significantly differentially expressed.

Expression profiles of transcripts significantly differentially expressed relative to WT vegetative cells (DESeq; p <0.01) were analysed by hierarchical clustering of their fold changes relative to WT vegetative cells in GeneSpring GX7 (Agilent). For visualisation purposes, any transcripts with normalized read counts <1 had their read counts rounded up to 1 to avoid visualizing artificially high or low fold changes. Transcripts (rows) were clustered using the Pearson correlation, and columns (mutants and environmental perturbations) were clustered using the Up-regulated correlation.

2.5 lncRNA location analysis

2.5.1 Categorising lncRNAs by location

Based on their location relative to mRNAs, several mutually exclusive categories for the full set of 7332 lncRNAs (5775 novel lncRNAs identified in the current study; 1557 already annotated lncRNAs i.e. SPNCRNAs) were defined (see Chapter 4).

For the set of lncRNAs differentially expressed (p <0.01, DESeq) in each sample (Table 2.1, 2.2) the observed distribution amongst these defined categories was calculated. The expected distribution amongst the same categories was calculated based on full set of 7332 lncRNAs. Expected and observed distributions were compared using a chi-squared test and p-values corrected for multiple testing (Bonferroni method).

Custom scripts were written in R.

2.5.2 Distribution of lncRNAs with respect to mRNA TSSs and TTSs

The location of lncRNAs was analysed relative to mRNA transcription start sites (TSSs) and transcription termination sites (TTSs).

mRNA TSSs and TTSs were taken as the most 5' and 3' annotation available respectively using the genome annotation available in GeneDB (http://old.genedb.org/, now PomBase http://www.pombase.org/) on 9th May 2011.

Briefly, for each mRNA TSS, a window 1kb upstream and downstream of the TSS was created. To avoid consideration of lncRNAs associated with other mRNA TSSs or TTSs within this window, if the mRNA under consideration was <1kb in length, the downstream window only extended to the end of the mRNA. Similarly, if the distance between the mRNA TSS and another mRNA in the upstream direction was <1kb, then the window in the upstream direction only extended to the end of the intergenic region. Distances between the mRNA TSS and any lncRNA TSSs occurring within this window were calculated. The distribution of these distances on either DNA strand, and centred around mRNA TSSs, was plotted.

A similar analysis was performed for distances between mRNA TTSs and lncRNA TSSs.

For the set of lncRNAs differentially expressed (p <0.01, DESeq) in each sample (Table 2.1, 2.2) the observed distribution with respect to mRNA TSSs and TTSs was calculated. The expected distribution (antisense strand only) with respect to mRNA TSSs and TTSs was calculated based on full set of 7332 lncRNAs. Expected and

observed distributions were compared using a chi-squared test and p-values corrected for multiple testing (Bonferroni method).

Custom scripts were written in R.

2.6 Expression correlation analysis

Based on their location relative to mRNAs, lncRNA:mRNA pairs were defined as described in Chapter 4.

2.6.1 Fisher tests

For a given set of lncRNA:mRNA pairs, Fisher exact tests were used to determine whether there is an association between the lncRNA transcript being significantly differentially expressed (DESeq, $p<0.01$) and the mRNA transcript being significantly differentially expressed, relative to WT cells in log-phase growth. Similar analyses were performed to test for associations in lncRNA:mRNA pairs for both transcripts to be significantly up- or down-regulated (positively correlating pairs), or for one to be up-regulated while the other is down-regulated (anti-correlating pairs). P-values were corrected for multiple testing (Bonferroni method). Custom scripts were written in R.

2.6.2 Correlation coefficients

Using normalised read counts from DESeq, an association between the expression of lncRNA and mRNA transcripts in specific types of pairs was tested using Pearson's product moment correlation coefficient (CC). For each transcript, the difference in its normalised read counts in a given sample and WT log phase growth, was calculated on a log2 scale. Correlation coefficients for these expression values were then calculated for specific types of mRNA:lncRNA pairs. Associated p-values were corrected for multiple testing (Bonferroni method). Transcripts with zero read counts in either one or both of the WT and sample cannot be analysed in this way, and pairs containing such transcripts were excluded from the correlation analysis**.** Note that if for an individual correlation test $p<2.2\text{x}10^{-16}$, the adjusted p-value has

been set to 10^{-16}. All p-values corrected for multiple testing. Custom scripts were written in R.

2.7 Functional enrichment analyses

All analyses for functional enrichments were performed using a set of gene ontology (GO) lists based on fission and budding yeast GOslim annotations (release September 2011). This set was complemented with a series of lists based on gene expression. Five lists of cell-cycle regulated genes contain either all periodic genes, or periodic genes peaking in G1, S, M, or G2 phases of the cell cycle (Rustici et al. 2004). Two lists contained genes of the core environmental stress response (CESR), either induced ("stress-related") or repressed ("growth-related") during stress (Chen D. et al. 2003). Four lists contained genes regulated upon nitrogen removal or during early, middle, or late meiosis (Mata et al. 2002). Finally four lists were computed in the Bähler laboratory and contained the 10% shortest and the 10% longest mRNAs, a list of transcription factors, and a list of proteins containing RNA-recognition (RRM) motifs (based on annotation available in PomBase).

The level and significance of the overlap between a list of mRNAs of interest, and specific functional categories/gene lists was calculated using Fisher's exact test. P-values were corrected for multiple testing (Bonferroni method).

2.8 PCR-based gene deletions

The one step PCR-based approach as described previously (Bahler et al. 1998) was used for deletion of *dcr1* in *h- ura4-D18*. Sequences of the 100bp primers used for the gene-specific deletion of *dcr1* are given below:

Forward primer (5' to 3'):

ATAGCTTAGGATTCATTATTTTTTAAGAGACAAATTTCTCGTCAATTGAATGAAA
CCTTCCGCCTTTATTTTCTTTTTGACGGATCCCCGGGTTAATTAA

Reverse primer (5' to 3'):

GCTTTGGAGACCCAAATTGAAAGTTTGAAAAGTTACAAGGGCCGCGGTCATAAA
AAATGAAATACTGTATATTTCAAGTCGAATTCGAGCTCGTTTAAAC

Both left and right junctions arising from this gene-specific deletion were PCR verified using primers in Table 2.3 (Dcr1-L-F, Dcr1-L-R, Dcr1-R-F, Dcr1-R-R), and observed fragment sizes were as expected (Table 2.4).

2.9 Random spore analysis

Double mutants of the *dcr1*, *rrp6* and *exo2* genes were created by crossing single mutants. Genotypes of single mutant strains used for these matings are listed below and were derived from the single mutant strains listed in Table 2.1:

h+ rrp6::URA ura-
h- dcr1::NAT ura-
h+ exo2::KAN ura-
h- exo2::KAN ura-

h- dcr1::NAT ura- was created using one step PCR-based approach (Bahler et al. 1998) to delete *dcr1* in *h- ura4-D18* as described in Section 2.8.

Random spore analysis (Sabatinos and Forsburg 2010) was used to create the remaining single mutant strains: *h+ rrp6::URA ura-* ; *h+ exo2::KAN ura-* and *h- exo2::KAN ura-*. Briefly, strains were crossed and incubated on malt extract agar (MEA) plates for 2-3 days at 25°C. Tetrads were treated with zymolyase (0.5 mg/ml, MP Biomedicals Europe) and incubated at 37°C for a minimum of 4 hours to release spores. Spores were germinated on rich (YES) agar plates before being replica plated to selective plates as appropriate. All deletion junctions were PCR verified (see Table 2.3 for primers, and Table 2.4 for expected fragment sizes).

Crosses and selection by random spore analysis were as follows: *h+ ade6-M216 leu1-32 ura4-D18 his3-D1 rrp6::URA4* (Table 2.1) was crossed with *h- ura4-D18* with selection on EMM agar plates to create *h+ rrp6::URA ura-* ; h^{-} *exo2::kanMX6*

ade6-216 (Table 2.1) was crossed with *h+ ura4*-D18 with selection on YES + kanamycin agar plates, as well as EMM agar plates with or without uracil, to select for selected *h+ exo2::KAN ura-* and *h- exo2::KAN ura-*.

2.10 Tetrad dissection

Tetrad analysis was used to analyse the meiotic products resulting from the crossing in all combinations of the single mutant strains:

h+ rrp6::URA ura-
h- dcr1::NAT ura-
h+ exo2::KAN ura-
h- exo2::KAN ura-

Strains were crossed and incubated on MEA plates for 2-3 days at 25°C. Resulting tetrads were picked and dissected using a micromanipulator (Singer Instruments) and spores germinated on rich media (YES) plates. Plates were imaged after 5 days growth.

Haploid colonies arising from germinated spores were then streaked to selective plates to test for KAN, NAT and URA markers.

All deletion junctions in doubly resistant colonies were PCR verified using primers in Table 2.3, and observed fragment sizes were as expected (Table 2.4).

Table 2.3: Primers used to check left and right deletion junctions

Primer Name	Sequence (5' to 3')
Exo2-L-F	CGCTATACTGCTGTCGATTCG
Exo2-L-R	TGTTGATGTTCCATTTGATGGT
Exo2-R-F	ATTTTCCCGAGGAGCAATTT
Exo2-R-R	ACGTCGTAGAAAGCGCTGAT
Dcr1-L-F	CGCTATACTGCTGTCGATTCG
Dcr1-L-R	AAAGAGTGGCATTCCTAAGCTG
Dcr1-R-F	GAAAAATTGCCGACCTTGAA
Dcr1-R-R	CAATTGTGATTCCGCAGATG
Rrp6-L-F	GATGCCGACGAAGCATAGTT
Rrp6-L-R	CTCGATGATCCAACTGATTTGA
Rrp6-R-F	CCAGGGAGTTTGTGTGGTCT
Rrp6-R-R	ACGGAATCTCTCCAATCGTG

Table 2.4: Expected fragment sizes

Primers used	Fragment size (bp)
Exo2-L-F, Exo2-L-R	459
Exo2-R-F, Exo2-R-R	905
Dcr1-L-F, Dcr1-L-R	330
Dcr1-R-F, Dcr1-R-R	902
Rrp6-L-F, Rrp6-L-R	800
Rrp6-R-F, Rrp6-R-R	934

2.11 Phenotyping mutants

2.11.1 Live cell imaging

Log-phase cells in rich medium were imaged using phase-contrast microscopy (Hamamatsu digital camera C4742-95 fitted to a Zeiss Axioskop microscope). One ml of culture was pelleted at 6000rpm for 30 seconds and 2µl of pelleted cells were mounted on a microscope slide. Images were captured using 5ms exposure and a 63x oil objective.

2.11.2 Growth analysis

2.11.2.1 BioLector microfermentation

The growth of mutant strains in rich media (YES) was profiled using the BioLector micro-fermentation system (mp2p-labs). Based on light scattering (LS) technology, the BioLector system records biomass values at 620 nm.

Cells of each strain were grown for 36 hours in pre-cultures. The pre-cultures were used to inoculate microtiter plates with 48 'flower'-shaped wells (m2p-labs) filled with YES medium. Compared to round or square wells, 'flower'-shaped wells provide improved mixing and oxygen transfer within the culture. The cell density in each well of the plate was adjusted to an OD of 0.15, with the final culture volume being 1.5ml. Gas permeable adhesive seals (Thermo Fisher Scientific) were used to cover the wells of the plate.

Triplicate cultures were prepared for each strain tested. Micro-fermentations were performed with the following settings: temperature was set at 32°C, humidity at 99%, shaking at 1000 rpm and LS measurements were retrieved every three minutes over a 48 hour period.

Growth parameters were extracted from the automatically recorded growth curves. Calculation of the maximum slope within the exponential phase provided an estimation of the maximum growth rate, while the average growth rate was calculated as the average biomass change per unit time during exponential phase.

2.11.2.2 Spot assay

For a semi-quantitative analysis of growth of mutant strains on rich media, the spot assay was employed as described previously (Tsutsui et al. 2000). Briefly, for each strain, 5µl of four serial (tenfold) dilutions of log-phase cells were spotted onto a YES agar plate. The plate was imaged after 3 days of growth at 32°C.

Chapter 3 Identifying novel lncRNAs

3.1 Introduction

This Chapter describes how, using strand-specific RNA-sequencing (RNA-seq), I have identified thousands of novel long noncoding RNAs (lncRNAs) in *S. pombe* that are regulated in response to a range of genetic and environmental perturbations.

While functions have been ascribed to several lncRNAs, and themes in their mechanisms of action are emerging (see Chapter 1), the function, if any, of the vast majority of identified lncRNAs remains unclear. With its unrivalled sensitivity and resolution, RNA-seq allows the global detection and quantification of lncRNAs, revealing their genomic origins and expression, and thereby providing insights into functions on a genome-wide scale.

Uncontrolled accumulation of noncoding RNA can adversely affect genome stability (Li and Manley 2006), and so eukaryotes have evolved multiple strategies to regulate noncoding RNA expression. These pathways often comprise RNA processing and degradation (see Chapter 1).

As such, transcriptomic studies in *S. cerevisiae* have revealed classes of lncRNAs, such as CUTs (Neil et al. 2009, Xu et al. 2009) and XUTs (van Dijk et al. 2011), which are stabilised upon impairment of the RNA degradation pathways of the nuclear exosome or cytoplasmic 5'-3' exonuclease respectively. There is evidence that such unstable or cryptic lncRNAs, normally targeted for rapid degradation, are stabilised and functional under relevant growth conditions. For example, *S. cerevisiae* antisense *PHO84* and *GAL1-10* transcripts are both regulated by exosomal degradation. These transcripts are stabilised during chronological ageing or repressive nutrient conditions, respectively, leading to repression of their respective sense loci via the recruitment of histone deacetylases (Camblong et al. 2007, Houseley et al. 2008). Similarly, it has been shown that XUTs accumulate in lithium-containing media indicating a possible role in adaptive responses to changing growth conditions.

These studies illustrate that systematic RNA-seq analysis of mutants defective for RNA degradation/processing pathways may represent a powerful approach to reveal novel, possibly regulatory, classes of lncRNAs.

In *S. pombe* a much larger proportion of the genome is transcribed than can be accounted for by currently annotated genes (Dutrow et al. 2008, Wilhelm et al. 2008). This additional transcription likely comprises extensions of existing annotated transcripts, biological and experimental noise, as well as novel lncRNAs.

Two RNA-sequencing studies have together annotated approximately 1500 lncRNAs in *S. pombe* (Rhind et al. 2011, Wilhelm et al. 2008), the nature and biological function of which remain unclear. However, since many more unstable lncRNAs, such as CUTs and XUTs, have been described in *S. cerevisiae*, lncRNAs identified by *S. pombe* RNA-seq studies to date may reflect only partially the rich complexity of the transcriptome.

Additionally, *S. pombe* appears to share more conservation in RNA metabolism with higher eukaryotes than *S. cerevisiae*. For example, RNA interference (Moazed 2009), transcript uridylation (Schmidt et al. 2011) and the PABPN1 pathway (Lemay et al. 2010, Lemieux et al. 2011) are all conserved from humans to *S. pombe*, but not conserved in *S. cerevisiae*. This makes *S. pombe* a meaningful model in which to study lncRNA processing and degradation.

In the current study, strand-specific RNA-seq of *S. pombe* RNA processing mutants was used to identify and analyse different classes of unstable lncRNAs. Possible regulatory functions can be imagined for lncRNAs degraded by RNA-processing pathways during vegetative growth but stabilised under physiologically relevant growth conditions. RNA-seq datasets for nitrogen-starved cells, glucose-starved cells and meiotic cells were therefore used to address whether the identified classes of unstable lncRNAs likely possess any biological function.

3.2 RNA-seq samples

Since many lncRNAs are Pol II-transcribed, capped and polyadenylated (see Chapter 1), all RNA-seq samples were poly(A)-enriched (see Chapter 2 for details).

All RNA-seq samples of RNA processing mutants used in the current study are detailed in Table 2.1. All RNA-seq samples of different physiological conditions/environmental perturbations used in the current study are detailed in Table 2.2.

For all samples, cells were grown, RNA harvested and strand-specific cDNA libraries prepared by myself (see Chapter 2). This is with the exception of: meiotic time-course samples, for which cells were grown and RNA prepared by Dr Cristina Cotobal (Mata laboratory, University of Cambridge); *pka1* samples under glucose starvation, for which cells were grown by Dr Charalampos Rallis (Bähler laboratory).

3.2.1 RNA processing/ degradation mutants

An overview of RNA processing/degradation pathways has been set out in Section 1.4, Chapter 1.

A panel of RNA processing mutants was chosen to cover key aspects of RNA processing and degradation (see Table 2.1). This panel includes: exonucleases in both the nucleus and cytoplasm; cofactors of these exonucleases; factors involved in deadenylation; factors that target decapping substrates; poly(A)-binding proteins; RNA interference (RNAi) factors; and factors involved in nonsense mediated decay (NMD).

The exosome contributes to the processing, quality control and turnover of a large number of cellular RNAs in both the nucleus and the cytoplasm where it degrades RNAs in a 3'-5' direction (see Section 1.4, Chapter 1). The exosome consists of nine inactive subunits plus an active ribonuclease, Dis3, which has both 3'-5' exonucleolytic as well as endonucleolytic activity. Thus the mutant *dis3-54* was chosen for the current study to reveal the role of the core exosome on lncRNA regulation.

Rrp6 confers nuclear specificity to the exosome, and both a temperature-sensitive mutant (*rrp6-ts*) and a null-mutant (*rrp6*) were included in the current study. In addition, null-mutants for Pab2, and TRAMP complex components Cid14 and Mtr4,

were chosen based on their known interactions with the nuclear exosome in *S. pombe*. A null mutant of the poly(A)-binding protein Nab2 was also included in the current study since Nab2 has been proposed to play a role in differentiating between mRNAs and lncRNAs (Tuck and Tollervey 2013), and has also been shown to impede Pab2/Rrp6-mediated decay by competing with Pab2 for binding poly(A) tails (Grenier St-Sauveur et al. 2013).

In the cytoplasm of *S. pombe*, three key pathways contribute to RNA degradation: the exosome which degrades in a 3'-5' manner and is connected to its cytoplasmic substrates via the SKI complex; the 5'-3' exonuclease Exo2; and the recently identified 3'-5' exonuclease Dis3L2 (Malecki et al. 2013). Mutants of all of these factors were therefore included in the current study.

The RNAi pathway in *S. pombe* is emerging as having roles in gene expression beyond heterochromatin formation at centromeres and telomeres, with several recent studies revealing roles for RNAi in transcriptional and post-transcriptional regulation of euchromatic loci (Cawley et al. 2004, Smialowska et al. 2014, Woolcock et al. 2011, Woolcock et al. 2012, Yamanaka et al. 2013). Moreover, connections between the RNA processing activities of the exosome and the RNAi pathway are emerging as playing a role in the regulation of lncRNAs (Gullerova and Proudfoot 2008, Lee et al. 2013, Shah et al. 2014, Zofall et al. 2009). Thus, mutants of the core RNAi components Dicer (*dcr1*), Argonaute (*ago1*) and RNA-dependent RNA polymerase (*rdp1*) were included in the current study.

Several double mutants for these key pathways were included to elucidate interactions between pathways in regulating lncRNA expression (see Table 2.1). The *rrp6/dcr1* and *exo2/dcr1* mutants will be discussed in more detail in Chapter 5.

3.2.2 Environmental perturbations/physiological conditions

Different physiologically relevant growth conditions (referred to as 'growth conditions') were chosen to reflect key physiologically relevant states for *S. pombe*.

In fission yeast, meiosis is accompanied by a complex gene expression programme in which more than 50% of the genome is regulated. Moreover, microarray studies

have shown that changes in gene expression occur in successive expression waves throughout the course of meiosis (Mata et al. 2007, Mata et al. 2002), and RNA-seq studies have revealed the expression of several meiosis-specific lncRNAs (Bitton et al. 2011, Rhind et al. 2011, Wilhelm et al. 2008). Cells harvested at time-points throughout the course of meiosis were therefore included in the current study.

Fission yeast is emerging as a complementary model system to budding yeast for studying the molecular mechanisms for cellular ageing (Rallis et al. 2013). *S. pombe* undergoes a progressive decline in viability after entering a quiescent stationary phase, a phenomenon known as chronological ageing, and which is often modeled by survival in glucose starvation (Roux et al. 2006). A nutrient-signaling pathway, which includes the serine/threonine cAMP-activated protein kinase Pka1, is known to regulate this process. *Pka1* is a long-lived mutant that accumulates fewer reactive oxygen species and has delayed initiation of apoptosis compared with WT cells (Roux et al. 2006).

Thus stationary WT cells at 100% and 50% survival in glucose starvation were included in the current study to determine lncRNA expression during chronological ageing. *pka1* cells at 100% and 50% survival were also included to determine lncRNA expression during chronological ageing in a long-lived mutant. When WT cells were at 50% survival, *pka1* cells were at 87% survival. This *pka1* time-point was included in the current study and is of interest because it may help determine whether changes in lncRNA expression during chronological ageing are a function of time in stationary phase, or whether they are related to the underlying molecular mechanisms of chronological ageing.

The cellular response of fission yeast to nitrogen starvation has been extensively studied. Following the withdrawal of the nitrogen source (NH_4Cl) from the culture medium, *S. pombe* cells exit the cell cycle, inhibit cellular growth and enter the G0 phase, thus providing an excellent model to study cellular quiescence (Yanagida 2009). The adaptation to nitrogen starvation occurs in several stages. In the initial stage (within the first ~8h), the cell size is reduced by two subsequent cell divisions, consequently forming small round cells (Shimanuki et al. 2007). Following these two rounds of division, the cell cycle is arrested in an 'uncommitted' G1 phase.

During this period, cells may undergo meiosis and produce spores, providing that a partner of the opposite mating type is available. In the absence of a mating partner, the cells lose the ability to mate after 12h and commit to the G0 phase (Yanagida et al. 2011). Within 24h of nitrogen starvation, the cellular volume, mRNA and protein content are reduced to about 55%, 20%, and 50% of the vegetative cell content, respectively (Marguerat et al. 2012). Fully adapted G0 cells can survive for months in the nitrogen starvation culture medium, exhibiting increased stress resistance. Thus nitrogen starved cells at 24h and 7 days were included in the current study to reflect distinct stages of adaptation to nitrogen starvation.

3.3 Method to identify novel lncRNAs and its validation

A crude heuristic was designed to detect candidate novel lncRNAs from RNA-seq data.

Briefly, RNA-seq data from initial sequencing runs, which comprised the following samples, were pooled together (two biological repeats of each): *rrp6-ts, exo2, dis3-54, pab2, mtr4, ago1, rdp1, dcr1, pan2, upf1*, WT_100, WT_50, N_7d, N_24h, meiotic pool, mid-log phase WT cells grown in YES, mid-log-phase WT cells grown in EMM.

Segments were delimited from the pooled data using a 10 hits/bp cut-off. Segments <100bp apart and differing in pooled read density (i.e. average hits/bp of segments) by <10-fold were joined together. Using the genome annotation available in GeneDB (http://old.genedb.org/, now PomBase http://www.pombase.org/) on 9th May 2011, segments overlapping annotations on the same strand (including untranslated regions, UTRs) were removed.

lncRNAs have an arbitrary minimal length cut-off of 200 nucleotides, due mainly to RNA-seq library preparation protocols that exclude small RNAs (Perkel 2013). Therefore segments <200bp were discarded, and remaining consecutive segments longer than 200bp were retained as final transcripts corresponding to 5775 candidate novel lncRNAs.

The described segmentation heuristic was optimised on its ability to detect the 1557 lncRNAs already annotated in *S. pombe* (SPNCRNAs). The ability to detect SPNCRNAs was judged on: the percentage coverage of each segment overlapping an SPNCRNA; the percentage coverage of each SPNCRNA overlapping a segment (as illustrated in Figure 3.1D). To optimise these two values the following parameters were varied: hits/bp cut-off; fusing distance; imposing a rule on whether or not to fuse based on fold difference in expression of consecutive segments; varying this fold difference in expression at which consecutive segments are fused (Figure 3.1A-C).

The optimisation strategy aimed to be as inclusive as possible. Thus an initial hits/bp cut-off of 2 hits/bp was used, and segments <20 nts apart were fused (Figure 3.1A). Imposition of the criterion that consecutive segments should only be fused if their fold difference in expression meets a certain threshold aimed to detect lncRNAs which are near mRNAs but distinct from mRNAs, while discarding data which likely represents misannotated untranslated regions (UTRs) (Figure 3.1B-C).

With the optimized segmentation procedure described, SPNCRNAs were covered at 92% by the detected segments, and the segments intersecting SPNCRNAs were covered at 62% by the SPNCRNAs (Figure 3.1C-D).

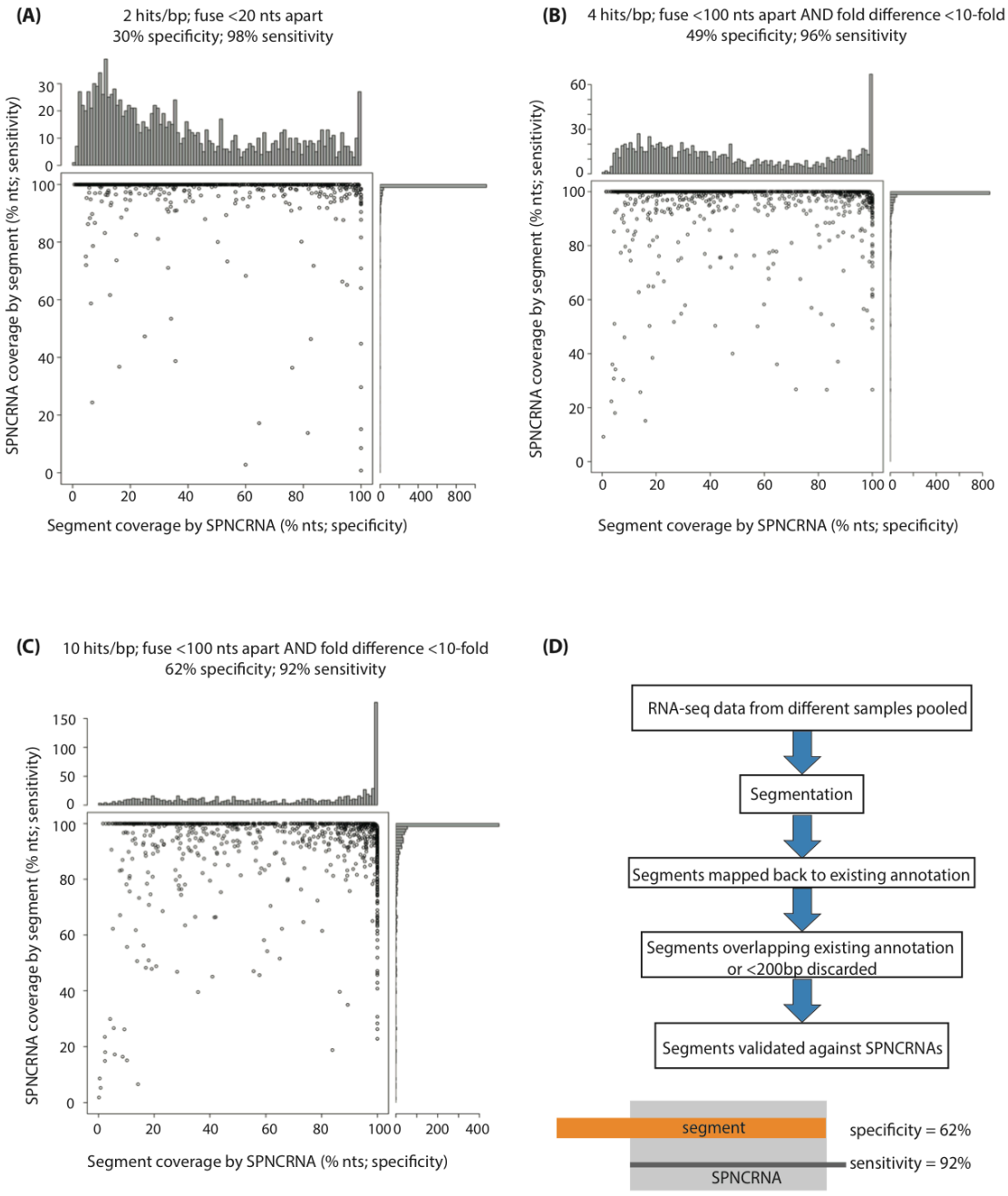

Figure 3.1: Optimisation of the segmentation process.

Most SPNCRNAs are efficiently detected as unique putative transcripts. A–C: Sensitivity and specificity of the segmentation method; each SPNCRNA is represented by a point; in abscissa the percentage of the nucleotide coverage of each segment that intersects a SPNCRNA; in ordinate the percentage of the nucleotide coverage of each SPNCRNA that is intersected by a segment. Segmentation parameters as indicated on figures and in text. D: Schematic of the segmentation process.

Finally, visual inspection was used to check the segmentation procedure (examples in Figure 3.2). From such visual inspection it was clear that novel lncRNAs are more lowly expressed than mRNAs (Figure 3.2A-E) but are comparable to already annotated SPNCRNAs in expression level (Figure 3.2E). Moreover, it was striking that lncRNA expression appeared to be more sample specific than that of either mRNAs or SPNCRNAs (Figure 3.2).

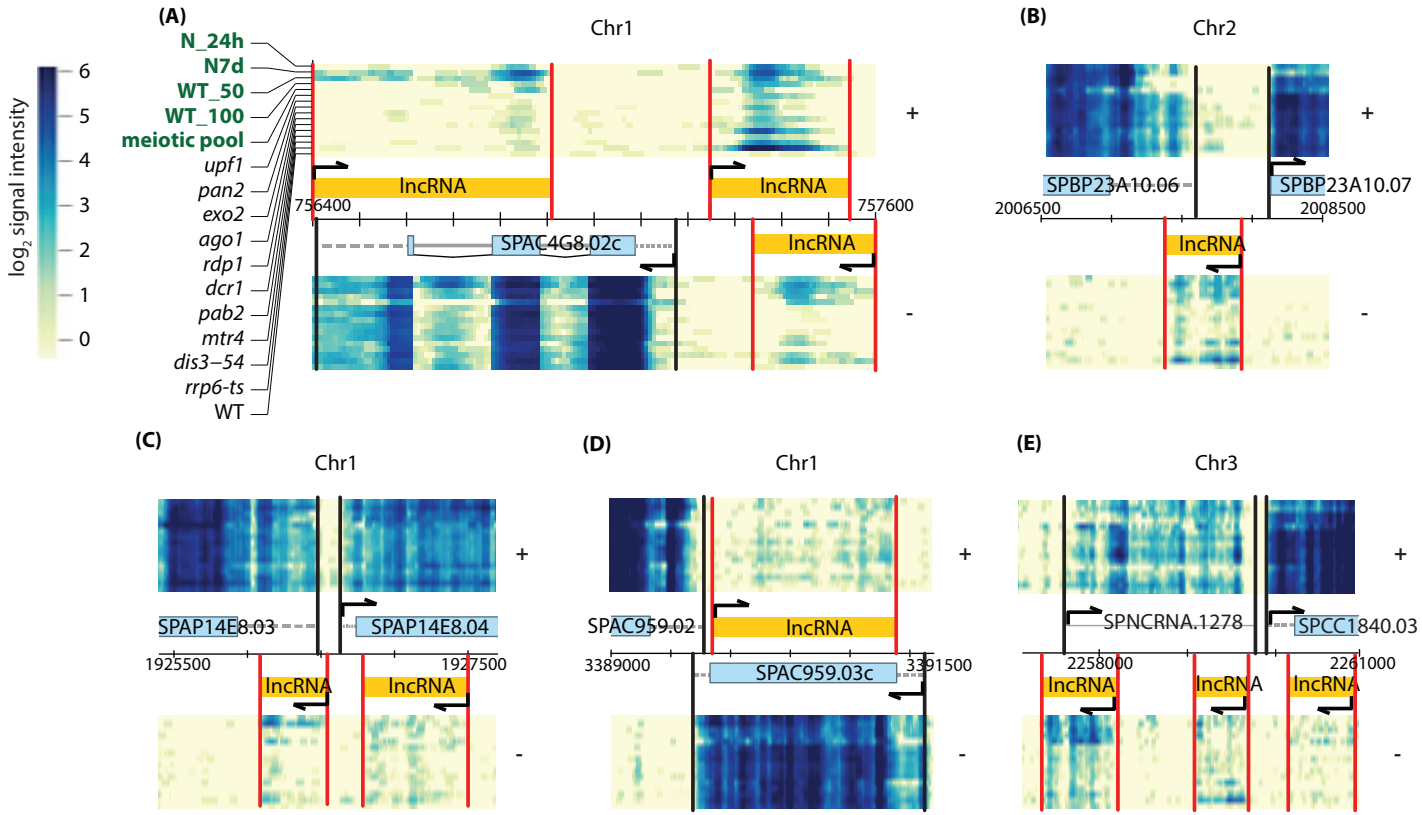

Figure 3.2: Examples of novel lncRNAs.

A–E: Normalised signal intensities (log_2 hits per base divided by depth) are shown on the y-axis. Expression data for plus and minus strands as indicated; chromosome number as indicated; vertical lines represent transcript boundaries; coordinates are indicated in base pairs. Annotated ORFs (blue boxes) and their mapped UTRs (dashed grey lines), novel lncRNAs (orange boxes) and transcript start sites (arrows) are as indicated.

3.4 Identifying differentially expressed transcripts

To identify novel lncRNAs normally degraded in WT cells but stabilised under the genetic and environmental perturbations tested, the DESeq package was used to determine significantly differentially expressed (DE) transcripts (see Chapter 2 for details).

For all samples, differential expression was determined relative to the 6 WT samples detailed in Table 2.2. This panel of WT samples was chosen to represent the main genetic backgrounds of the samples analysed, as well as the key growth media used (rich media, minimal media), thereby reducing the chances that DE transcripts could be attributed to differences in genetic background or growth media. Moreover, very few DE transcripts were found when these 6 WT samples were analysed against each other in DESeq, in all possible combinations (~30 DE transcripts when two different WT strains analysed against each other; ~40 DE transcripts when WT growth media analysed against each other), and none of these DE transcripts were lncRNAs.

3.5 Expression profiles of differentially expressed transcripts

Expression profiles of significantly differentially expressed mRNAs, SPNCRNAs and novel lncRNAs were examined by clustering their fold changes in expression relative to WT vegetative cells across all samples (Figure 3.3; see Chapter 2 for details).

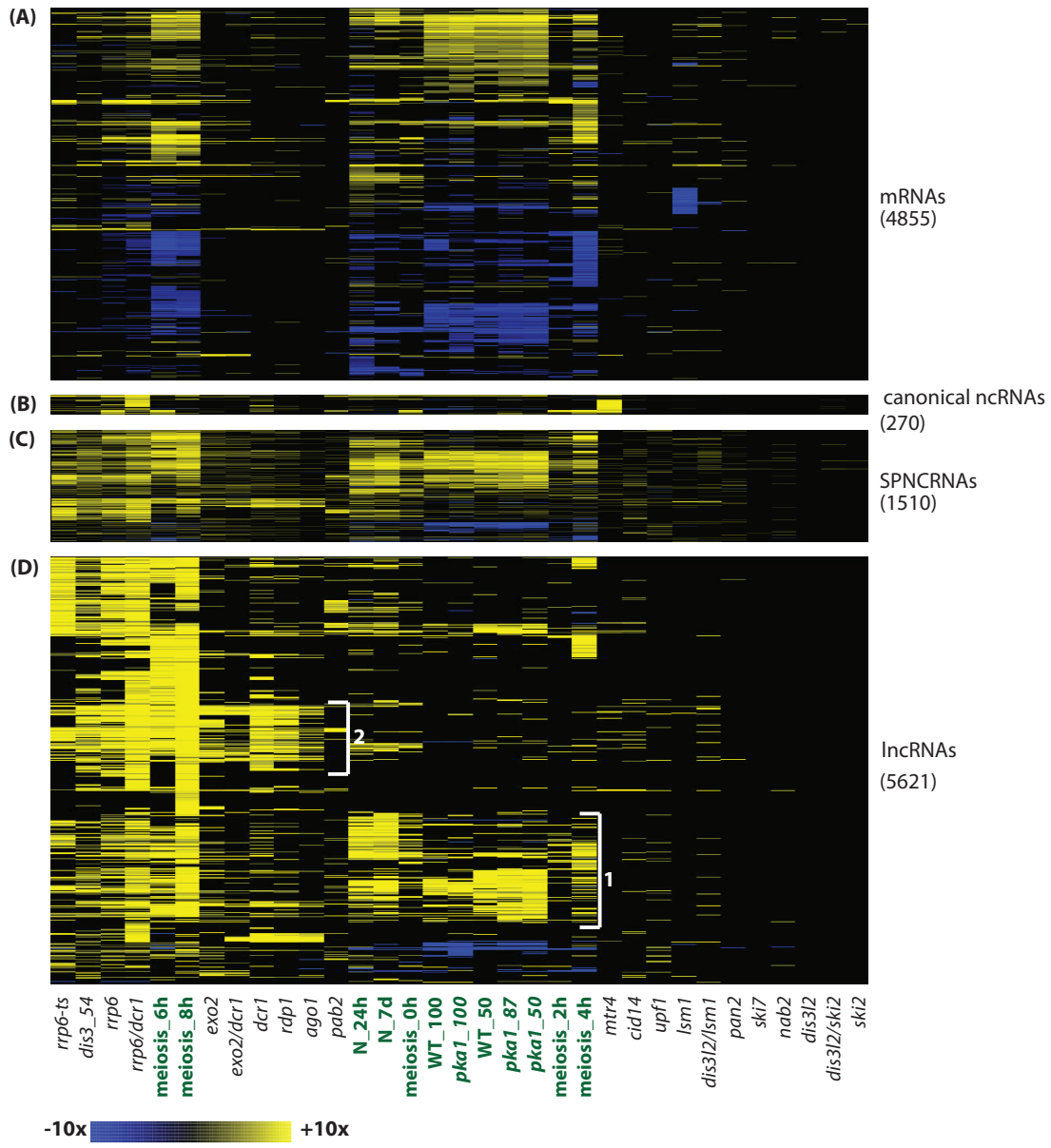

Figure 3.3: Expression profiles of significantly differentially expressed transcripts.

Fold changes relative to WT vegetative cells clustered in GeneSpring GX7 (Agilent); colour coded as indicated. Transcripts not significantly differentially expressed relative to WT (DESeq; p-value <0.01) have fold change set to 1 (appear black in heatmaps; see Chapter 2 for details). Transcripts with a fold change of 1 across all samples do not appear in these heatmaps. Columns represent different samples and rows represent transcripts; rows clustered using the Pearson correlation. Columns clustered using the up-regulated correlation based on lncRNAs only. (A) rows represent mRNAs (B) rows represent canonical ncRNAs (C) rows represent annotated SPNCRNAs (D) rows represent novel lncRNAs; white brackets highlight clusters of lncRNAs described in text. Number of transcripts in each heatmap is as indicated.

3.5.1 mRNA expression signatures

Expression signatures of mRNAs revealed a large number of DE mRNAs under environmental perturbations and relatively fewer under genetic perturbations (Figure 3.3A and Figure 3.5). Nitrogen starved, glucose starved and meiotic samples each have ~1000 mRNAs significantly differentially up-regulated or down-regulated relative to WT vegetative cells. This is in contrast to the RNA processing mutants for which an average of only 300 mRNAs are up-regulated and 30 mRNAs down-regulated.

Gene ontology (GO) analysis of up-regulated and down-regulated mRNAs reveals up-regulation of the stress module across all samples, and down-regulation of the growth module predominantly in growth conditions (Figure 3.4).

Figure 3.4: Gene ontology analysis of significantly differentially expressed mRNAs.

Significantly enriched GO-terms for mRNAs up-regulated (beige) or down-regulated (red) in each sample. GO-terms (rows) clustered by the "maximum" method in heatmap.2 function of R.

Consistent with known roles for the exosome, the poly(A)-binding protein Pab2 (St-Andre et al. 2010, Yamanaka et al. 2010), and the RNAi machinery (Lee et al. 2013, Yamanaka et al. 2013) in meiotic gene silencing, mRNAs up-regulated in *rrp6, dis3-54, pab2, ago1, dcr1* and *rdp1* mutants are enriched for meiotic GO-terms

(Figure 3.4). Notably, mRNAs up-regulated in the *exo2* mutant are also strongly and specifically enriched for middle meiotic genes. This implicates a role for the 5'-3' cytoplasmic RNA exonuclease Exo2 in control of meiotic gene expression. Although no such role has been described for Exo2, it has been observed that Exo2 expression is induced during meiosis in *S. pombe* (Amorim et al. 2010).

Although there are relatively fewer mRNAs significantly up- or down-regulated in the RNA processing mutants compared to the physiological conditions, a cluster of ~350 mRNAs is down-regulated in *lsm1* (cluster indicated by '4' on Figure 3.5). This group of down-regulated mRNAs is significantly enriched for the following GO-terms ($p < 10^{-5}$), a signature which is possibly indicative of rRNA metabolism, and is consistent with the slow growth phenotype reported for *lsm1* (Malecki et al. 2013):

nucleus
RNA metabolic process
organelle organization
chromosome organization
chromosome segregation
nucleolus
nucleotidyltransferase activity
DNA metabolic process

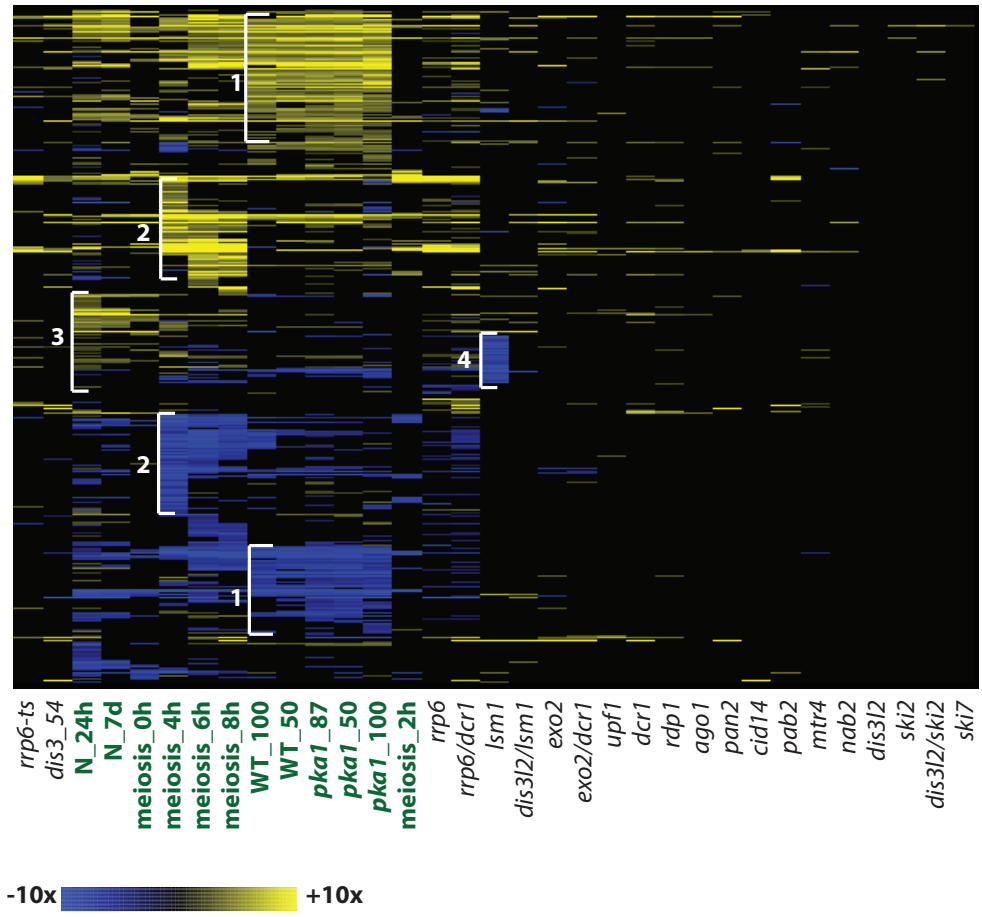

Figure 3.5: mRNA expression signatures.

See legend for Figure 3.3. Columns represent different samples and rows represent mRNAs; rows clustered using the Pearson correlation. Columns clustered using the up-regulated correlation. White brackets highlight clusters of mRNAs described in text.

Some mutants (*ski2, ski7, pan2, upf1, dis3l2, cid14*) have no GO-terms enriched amongst up or down-regulated mRNAs (Figure 3.4), which is reflective of the little to no phenotype observed for these mutants.

Distinct clusters of up- and down-regulated mRNAs were observed, which seemed to define to one of the following three processes: glucose starvation (clusters indicated by '1' on Figure 3.5); meiosis (clusters indicated by '2' on Figure 3.5); or nitrogen starvation (cluster indicated by '3' on Figure 3.5). Such groups of mRNAs may confer specificity to these phenotypes. In addition, mRNA expression signatures appear to be distinct at different meiotic time-points, while there is comparatively

little change in mRNA expression signatures at differing time-points for nitrogen or glucose starved cells (Figure 3.5).

Distinct mRNA expression signatures at the transcriptional level for glucose versus nitrogen starvation are in line with preliminary translational profiling data from *S. pombe* cells starved for either glucose or nitrogen. Such data reveal two apparently different modes of starvation at the translational level – while translation is completely shut down during glucose starvation, during nitrogen starvation many mRNAs are still associated with polysomal fractions and cells remain metabolically active, surviving for more extended time periods (Daniel Lackner PhD thesis, Bähler laboratory).

3.5.2 Canonical ncRNA expression signatures

The expression signatures of canonical ncRNAs in Figure 3.3B and Figure 3.6 represent 270 ncRNAs comprising:

164 tRNA (61%)
50 snoRNA (19%)
49 rRNA (18%)
7 snRNA (2%)

These canonical ncRNAs are non-polyadenylated and so, while this poly(A)-enriched RNA-seq data cannot be used in a quantifiable manner to analyse canonical ncRNA expression, certain qualitative features of their expression signatures are of note.

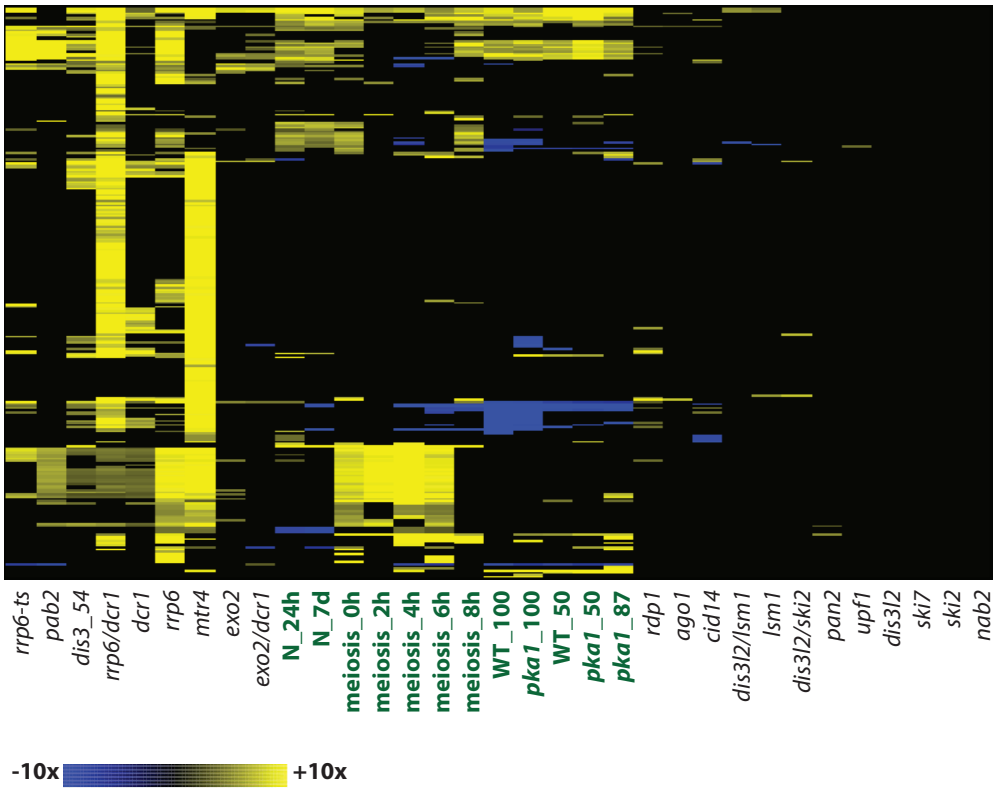

Figure 3.6: Canonical ncRNA expression signatures.

See legend for Figure 3.3. Columns represent different samples and rows represent canonical ncRNAs; rows clustered using the Pearson correlation. Columns clustered using the up-regulated correlation.

The up-regulation of canonical ncRNAs in the exosome mutants, *mtr4* and *pab2* is consistent with known roles of the exosome, Pab2 and the TRAMP complex in negatively controlling snoRNA expression (Larochelle et al. 2012, Lemay et al. 2010), as well as pre-tRNAs and rRNAs known as being exosome targets (Gudipati et al. 2012, Schneider et al. 2012).

It is striking that the *rrp6/dcr1* double mutant has a large number of canonical ncRNAs strongly up-regulated – more than either single mutant (Figure 3.6) - hinting at a possible cooperative role between the exosome and RNAi in regulating expression of these canonical ncRNAs. Also of note is that *dcr1* has a stronger phenotype in terms of canonical ncRNA up-regulation than either *rdp1* or *ago1*. However, since these canonical non-coding RNAs are not polyadenylated, the

possibility exists that these observations are simply due to a difference in poly(A)-enrichment efficiency.

3.5.3 SPNCRNA expression signatures

Expression profiles of annotated lncRNAs (SPNCRNA) revealed SPNCRNAs to be predominantly up-regulated under nitrogen starvation, glucose starvation or meiosis, and largely unchanged in RNA processing mutants (Figure 3.3C).

Notable exceptions to this are the exosome mutants *rrp6, rrp6-ts* and *dis3*. This indicates that some SPNCRNAs may represent cryptic lncRNAs analogous to *S. cerevisiae* CUTs, which are subject to exosomal degradation (Neil et al. 2009, Xu et al. 2009).

Although SPNCRNAs are predominantly up-regulated across samples for different growth conditions, there is a small cluster of down-regulated SPNCRNAs in glucose starvation samples. GO-analysis of mRNAs these SPNCRNAs are associated with, in terms of location (i.e. mRNAs these SPNCRNAs are transcribed divergently from, or antisense to), revealed no enrichment of any GO-terms (lncRNA:mRNA associations by location discussed further in Chapter 4).

3.5.4 Novel lncRNA expression signatures

Expression profiles of differentially expressed novel lncRNAs reveal extensive DE under genetic and environmental perturbations (Figure 3.3D). There is much more up-regulation across samples than seen with either mRNAs or SPNCRNAs, indicating that these novel lncRNAs comprise many cryptic lncRNAs specifically degraded by RNA-processing pathways during vegetative growth, and up-regulated under physiologically relevant growth conditions.

3.5.4.1 Novel lncRNA expression signatures across RNA-processing mutants

The nuclear-specific 3'-5' exosome component Rrp6, the core exosome component Dis3 and, to a lesser extent, the poly(A)-binding protein Pab2 act in concert to degrade a common set of novel lncRNAs (Figure 3.3D). Such cryptic lncRNAs resemble *S. cerevisiae* CUTs (Neil et al. 2009, Xu et al. 2009), which are targeted for

degradation by the nuclear exosome. The involvement of Pab2 in their degradation is consistent with the cooperation between Pab2 and the exosome in degrading meiotic transcripts in *S. pombe* (St-Andre et al. 2010, Yamanaka et al. 2010) and snoRNAs (Lemay et al. 2010). Additionally, the human nuclear poly(A)-binding protein, PABPN1, has been shown to mediate the exosomal degradation of a group of ~60 lncRNAs (Beaulieu et al. 2012). The weak lncRNA signatures observed for mutants of the TRAMP complex *mtr4* and *cid14* suggest a possible role for TRAMP-independent exosome degradation of lncRNAs (see Chapter 6).

The cytoplasmic 5'-3' RNA exonuclease Xrn1 has been shown to degrade a class of lncRNAs in *S. cerevisiae* referred to as XUTs (van Dijk et al. 2011). The current study reveals that Exo2, the *S. pombe* orthologue of Xrn1, similarly acts to degrade a set of novel lncRNAs (Figure 3.3D). It is striking that mutants for cytoplasmic degradation pathways other than Exo2 show very little phenotype in terms of lncRNA expression (*ski2, ski7, dis3l2, lsm1*). Even the double cytoplasmic degradation mutants *dis3l2/ski2* and *dis3l2/lsm1* show little phenotype in terms of lncRNA expression. This suggests that Exo2 represents the main pathway controlling cytoplasmic lncRNA degradation, with other cytoplasmic degradation factors playing minor or redundant roles.

The RNAi proteins Ago1, Dcr1 and Rdp1 act in concert to degrade a common set of novel lncRNAs (Figure 3.3D). This set of novel lncRNAs partially overlaps those targeted for degradation by the exosome. Such cooperation between RNAi and the exosome has already been reported in *S. pombe* in the silencing of centromeric transcripts (Buhler et al. 2007, Reyes-Turcu et al. 2011) and the degradation of some antisense ncRNAs arising from read-through at convergent genes (Zhang et al. 2011, Zofall et al. 2009). Moreover, Rdp1 and Dcr1 physically associate with several euchromatic loci, including 30 SPNCRNAs, at which the TRAMP component Cid14 is also enriched (Woolcock et al. 2011, Woolcock et al. 2012). This suggests a cooperative role for RNAi and the exosome in silencing lncRNAs, a conclusion which is corroborated by the current study (see Chapter 6).

Extensive overlap between novel lncRNAs degraded by the exosome, by RNAi and by Exo2, reflects known interactions between the exosome and RNAi. It also

possibly hints at a novel cooperation between cytoplasmic Exo2 activity and the nuclear processing activities of RNAi and the exosome.

In summary, the novel lncRNAs identified in the current study resemble classes of cryptic lncRNAs, such as CUTs and XUTs, already described in *S. cerevisiae* (see Chapter 5). Moreover, the current study reveals overlapping and specific roles for the nuclear exosome, the cytoplasmic exonuclease Exo2 and the RNAi pathway, in degrading novel lncRNAs.

Double and triple mutants were created to assess interactions between these three key pathways, and will be discussed in detail in Chapter 5. Briefly, it is of note that the *rrp6/dcr1* double mutant has more up-regulated lncRNAs than either single mutant, while the *exo2/dcr1* double mutant has fewer up-regulated lncRNAs than *exo2* or *dcr1* (Figure 3.3D).

3.5.4.2 Novel lncRNA expression signatures across environmental perturbations

Consistent with a functional role for the novel lncRNAs identified in the current study, their expression appears to be regulated under physiologically relevant growth conditions (Figure 3.3D). Two key clusters of lncRNAs emerge from the expression profile in Figure 3.3D: firstly, lncRNAs up-regulated under nitrogen or glucose starvation (and early meiosis) cluster together and are regulated predominantly by the nuclear exosome (cluster indicated by '1' on Figure 3.3D); secondly, the RNAi machinery and Exo2, together with the nuclear exosome, appear to have greater involvement in the regulation of lncRNAs up-regulated in later meiotic time-points (cluster indicated by '2' on Figure 3.3D).

In *S. pombe* the exosome is known to silence meiotic gene expression (St-Andre et al. 2010, Yamanaka et al. 2010). Moreover, Yamanaka et al. have recently described interaction between RNAi and the exosome in silencing meiotic genes (Yamanaka et al. 2013). While a handful of meiosis-specific non-coding RNAs are known to exist in *S. pombe* (Bitton et al. 2011, Ding et al. 2012, Watanabe T. et al. 2001), in *S. cerevisiae* a class of Rrp6-regulated lncRNAs, up-regulated on entry into meiosis, have been described (Lardenois et al. 2011). Consistent with the above, I observe a group of novel lncRNAs cooperatively regulated by the exosome, RNAi and Exo2,

and up-regulated in later meiotic time-points (cluster indicated by '2' on Figure 3.3D). Importantly, the stabilisation of Exo2-degraded lncRNAs in meiosis further implicates a novel role for Exo2 in meiosis.

Notably, regulation of meiotic lncRNAs by the RNAi machinery seems to be largely Ago1-independent (Figure 3.3D; RNAi mutants and meiotic samples were also re-clustered on their own - data not shown). This is reminiscent of the recently described heterochromatin-independent RNAi silencing of LTRs and Atf1-bound stress response genes (Woolcock et al. 2011, Woolcock et al. 2012). Silencing of these loci requires physical association with Rdp1 and Dcr1, and such associations are unaffected in *ago1Δ* cells.

In contrast to lncRNAs up-regulated in later meiotic time-points, those up-regulated in nitrogen or glucose starvation and early meiotic time-points appear to be largely unaffected by RNAi and Exo2 (cluster indicated by '1' on Figure 3.3D). Instead, these lncRNAs appear to be predominantly regulated by the nuclear exosome.

While no direct role for the exosome in nitrogen or glucose starvation has been reported, the exosome has been implicated in *S. cerevisiae* chronological ageing, which is often modeled by survival in quiescence. Weaker association of Rrp6 with the *S. cerevisiae PHO84* locus upon chronological ageing leads to up-regulation of *PHO84* antisense transcripts, and subsequent repression of the *PHO84* locus by recruitment of histone deacetylases (Camblong et al. 2007). Moreover, a recent study has revealed a role for quiescence-induced lncRNAs in recruitment of a H4K20 methyltransferase to ribosomal DNA in murine tissue culture (Bierhoff et al. 2014). This leads to chromatin compaction and subsequent repression of ribosomal genes upon growth factor deprivation or terminal differentiation.

While many lncRNAs up-regulated in nitrogen or glucose starvation clearly overlap (cluster indicated by '1' on Figure 3.3D), there also appears to be a distinct cluster of lncRNAs up-regulated under nitrogen starvation but not glucose starvation. Interestingly, differences between glucose-starved WT and *pka1* cells were less pronounced, and these samples cluster together based on time spent in glucose starvation, rather than by percentage survival or genetic background. Although

lncRNAs are predominantly up-regulated across physiological condition samples, there is a small cluster of down-regulated lncRNAs in glucose starvation samples. GO-analysis of mRNAs these lncRNAs are associated with, in terms of location (i.e. mRNAs these lncRNAs are transcribed divergently from, or antisense to), revealed no enrichment of any GO-terms (lncRNA:mRNA associations by location discussed further in Chapter 4).

Finally, it is interesting to note that more lncRNAs are up-regulated at progressively later meiotic time-points (Figure 3.3D). A similar trend is observed at progressively later nitrogen or glucose starvation time-points. In addition, for both WT and *pka1* cells under glucose starvation, this trend appeared to be related to time spent in stationary phase rather than genetic background. Thus when WT cells were at 50% survival, *pka1* cells were at 87% survival, and similar numbers of up-regulated lncRNAs were observed in the two samples. At the later time-point, when *pka1* cells were at 50% survival, a greater number of up-regulated lncRNAs were observed (Figure 3.3D).

In summary, I observe distinct groups of lncRNAs degraded by RNA-processing pathways during vegetative growth but stabilised under physiologically relevant growth conditions. Quiescent lncRNAs (nitrogen or glucose starvation and early meiosis) are regulated predominantly by the exosome. In contrast, the RNAi pathway and the cytoplasmic exonuclease Exo2 appear to play a greater role in the regulation of lncRNAs up-regulated in later meiotic time-points. A more extensive analysis of the overlap between these groups is discussed in Chapter 5.

3.6 Summary of main conclusions

A method for detecting novel lncRNAs from strand-specific RNA-seq data has been successfully developed, implemented and validated.

The main pathways regulating lncRNA post-transcriptional expression and degradation are: RNAi, nuclear exosome, and the cytoplasmic exonuclease Exo2. The exosome is the main player in the nucleus but not in the cytoplasm where Exo2

likely takes over. Other cytoplasmic degradation pathways play little or no role in lncRNA degradation.

lncRNAs are stabilised under physiologically relevant growth conditions. Quiescent lncRNAs (nitrogen or glucose starvation and early meiosis) are regulated predominantly by the exosome. In contrast, the RNAi pathway and the cytoplasmic exonuclease Exo2 appear to play a greater role in the regulation of lncRNAs up-regulated in later meiotic time-points.

A more extensive analysis of the overlap between key clusters of lncRNAs, and inferences that can be made about overlap between the RNA processing pathways examined, are discussed in results Chapter 5.

Chapter 4 lncRNA location & expression in relation to mRNAs

4.1 Introduction

Studies of lncRNAs in other organisms have shown a strong association between lncRNAs and other transcribed units, in terms of both their location and their expression.

For example, CUTs in *S. cerevisiae* predominantly arise within nucleosome-free regions (NFRs) that correspond to promoter regions of known genes, and tend to positively correlate in expression with the gene they are transcribed divergently from (Neil et al. 2009, Xu et al. 2009). A more recent study has revealed that another class of cryptic lncRNAs, NUTs, chiefly originate from both 5'- and 3'-NFRs flanking known genes (Schulz et al. 2013). Those NUTs originating in 3'-NFRs, and running antisense to known genes, are proposed to mediate a repressive effect on the genes they run antisense to. Similarly, many *S. cerevisiae* XUTs originate antisense to open reading frames, and have been proposed to have an inhibitory effect on expression of their sense loci (van Dijk et al. 2011).

In this Chapter, I analyse the 5775 lncRNAs identified in the current study (Chapter 3), together with the 1557 already annotated SPNCRNAs, in terms of both their location and expression with respect to 'other transcribed units'. 'Other transcribed units' were taken to be any annotated non-lncRNA transcript (using the genome annotation available in GeneDB, http://old.genedb.org/, now PomBase http://www.pombase.org/, on 9th May 2011). As such, there are 5716 'other transcribed units' comprising:

5175 mRNA
7 snRNA
53 snoRNA
195 tRNA
49 rRNA
237 LTR

Transcribed units run from the most 5' annotation to the most 3' annotation i.e. untranslated regions (UTRs) are included where these have been annotated. For ease

of reference, ‘other transcribed units’ will be referred to as mRNAs in the remainder of this Chapter.

4.2 Location of lncRNAs relative to mRNAs

To analyse the location of lncRNAs with respect to mRNAs, I developed two complementary methods: firstly, lncRNAs were categorized into mutually exclusive groups based on their location with respect to mRNAs (see below and Section 2.5, Chapter 2); secondly, the distribution of lncRNA transcription start sites (TSSs) with respect to mRNA TSSs and mRNA transcription termination sites (TTSs) was analysed (see below and Section 2.5, Chapter 2).

4.2.1 Categorising lncRNAs by location

Based on their location relative to mRNAs, several mutually exclusive categories for lncRNAs were defined (Figure 4.1A), and are described below:

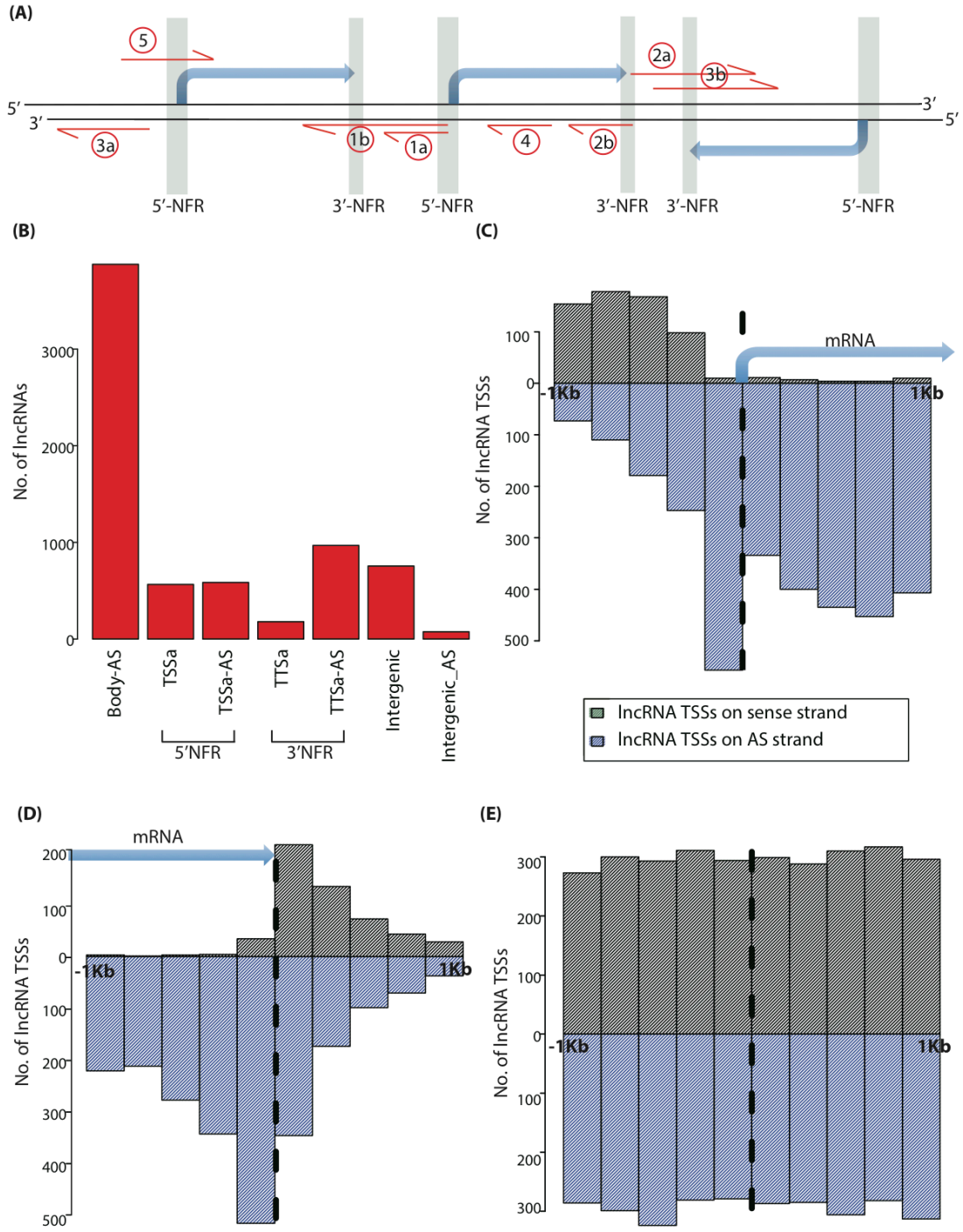

Figure 4.1: Location of full set of lncRNAs relative to mRNAs.

(A) Schematic of the different lncRNA locations relative to mRNAs. mRNAs represented in blue; lncRNAs in red; 3'- and 5'-NFRs shaded in grey. 1a: TSSa; 1b: TSSa-AS; 2a: convergent read-through; 2b: TTSa-AS-own; 3a: intergenic; 3b: intergenic-AS; 4: body; 5: SPNCRNA overlapping an mRNA on the sense strand (overlapping). (B) Number of each type of lncRNA categorised by location ("overlapping" omitted). (C) Distribution of lncRNAs relative to mRNA TSSs. Dotted vertical line indicates reference mRNA TSS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. (D) Distribution of lncRNAs relative to mRNA TTSs. Dotted vertical line indicates reference mRNA TTS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. (E) Distribution of

lncRNAs relative to random coordinates. Dotted vertical line indicates reference random coordinate; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand.

Transcription start site associated (TSSa): the lncRNA originates upstream and antisense to an mRNA transcription start site (TSS). These lncRNAs arise from bidirectional transcription, i.e. lncRNAs transcribed divergently from mRNA TSSs. TSSa lncRNAs are considered to originate within the 5'-(nucleosome free region)-NFR associated with an mRNA TSS (represented by 1a on Figure 4.1A) and, as such, a maximum distance of 300bp between the mRNA TSS and lncRNA TSS is allowed (see note later in this Section). Some TSSa lncRNAs additionally run antisense (AS) to an upstream gene (TSSa-AS; represented by 1b on Figure 4.1A).

Transcription termination site associated (TTSa): the lncRNA originates within the 3'-(nucleosome free region)-NFR associated with an mRNA transcription termination site (TTS). A maximum distance of 300bp between the mRNA TTS and the lncRNA TSS is allowed. The vast majority of these lncRNAs run antisense to an adjacent gene (TTSa-AS). A TTSa-AS lncRNA may run antisense to either the gene from whose 3'-NFR it originates (referred to as 'TTSa-AS-own'; represented by 2b on Figure 4.1A), or antisense to a downstream gene in a convergent orientation. This latter class is referred to as 'convergent read-through' (represented by 2a on Figure 4.1A) since, although these lncRNAs have been annotated as distinct transcripts in the current study (Chapter 3), it is possible that such non-coding transcription in fact arises from continued action of Pol II at the 3' ends of annotated genes in a convergent orientation. Such a phenomenon has already been described in *S. pombe* (Bitton et al. 2011, Gullerova and Proudfoot 2008, Leong et al. 2014). Most TTSa-AS lncRNAs arise from convergent read-through (~75%).

Body antisense (body-AS): the lncRNA originates and runs antisense to the body of an mRNA. These are represented by 4 on Figure 4.1A.

Intergenic: the lncRNA originates in an intergenic region (no mRNA annotation on either strand, and not sufficiently close to an mRNA TSS or TTS to be considered as originating from a 3'- or 5'-NFR). These are represented by 3a on Figure 4.1A.

Some intergenic lncRNAs additionally run antisense (AS) to an adjacent gene (intergenic-AS; represented by 3b on Figure 4.1A).

Overlapping: the lncRNA overlaps an mRNA on the same strand (represented by 5 on Figure 4.1A). The 321 lncRNAs that can be assigned to this class are all SPNCRNAs i.e. already annotated lncRNAs. The current study aims to identify distinct lncRNA transcripts, and is therefore biased against non-coding transcription that overlaps coding genes on the sense strand (Chapter 3). Although such non-coding transcription has been suggested to play regulatory roles in *S. pombe* (Shah et al. 2014), since the algorithm developed in the current study (Section 3.3, Chapter 3) cannot identify these lncRNAs, they cannot be fully addressed and therefore this class will not be considered further in this Chapter.

As described above, TSSa and TTSa lncRNAs are assumed to originate within 5'-NFRs and 3'-NFRs respectively. For these lncRNA locations, a maximum distance of 300bp between the lncRNA TSS and the mRNA TSS (or TTS) was allowed in the current study. This is based on *in vivo* genome-wide maps of nucleosomes in *S. pombe* revealing that nucleosome occupancy at 5'-NFRs appears to follow a bimodal distribution with 5'-NFR width being ~70bp or ~150bp, and the maximum 5'-NFR width ~200bp (Moyle-Heyrman et al. 2013). Assuming a model of a single NFR surrounded by two regions inside the flanking nucleosomes from which divergent transcripts initiate, a non-conservative cut-off of 300bp was chosen as the maximum distance between an mRNA TSS and lncRNA TSS that putatively originate within the same 5'-NFR. For *S. cerevisiae* divergent transcript pairs, the modal 5'-NFR length is 131bp, with a mode TSS:TSS distance of 180bp (Xu et al. 2009). Although Xu et al. found TSS:TSS distances of divergent transcript pairs range up to 500bp in *S. cerevisiae*, by far the majority are <300bp. The same distance constraint of 300bp was applied to define an mRNA TTS proximal region akin to a 3'-NFR.

It is also worth noting that, for a lncRNA originating in the intergenic region between two tandem mRNAs which are less than 300bp apart, the lncRNA could be considered as originating in the 5'-NFR of the downstream mRNA or the 3'-NFR of

the upstream mRNA. In these cases, the 5'-NFR is given precedence, and the lncRNA is classed as originating in the 5'-NFR. This is in line with the finding that 5'-NFRs are more pronounced and better defined in *S. pombe* than 3'-NFRs (Moyle-Heyrman et al. 2013).

The percentages of the total set of lncRNAs, that each of the categories described above represents, are given below (also see Figure 4.1B):

53% body-AS
16% TSSa i.e. bidirectional
16% TTSa
11% from intergenic regions
4% overlapping

To some extent, the above numbers are simply a reflection of the relative genomic space in which these transcripts can be initiated. The compact *S. pombe* genome makes it no surprise that approximately half of all lncRNAs originate antisense to the body of mRNAs. Overall, 18% of the *S. pombe* genome is intergenic (no annotation on either strand), roughly correlating with the 11% of lncRNAs that originate in intergenic regions.

4.2.2 Distribution of lncRNAs with respect to mRNA TSSs and TTSs

To assess the location of lncRNAs with respect to mRNAs, a complementary approach of examining the distribution of lncRNA TSSs within a 1kb window upstream and downstream of mRNA TSSs and TTSs was used (Figure 4.1C,D; see Chapter 2, Section 2.5 for details).

The average length (both mean and median) of mRNAs is ~2kb, and the average length of intergenic regions is ~300-500bp (median and mean respectively). Therefore, examination of both windows (mRNA TSS ± 1kb and mRNA TTS ± 1kb) should provide information not only on the distribution of lncRNAs immediately surrounding mRNA TSSs and TTSs, but also on their distribution relative to the body of the mRNA and any associated intergenic region(s).

Figure 4.1C shows a peak of antisense lncRNAs and depletion of sense lncRNAs initiating 0-200bp upstream of mRNA TSSs. Such a peak is indicative of lncRNAs transcribed divergently from mRNA TSSs, and corresponds to the TSSa classes in Figure 4.1B.

From both Figure 4.1C and Figure 4.1D, lncRNA transcription appears to be able to initiate along length of mRNAs (corresponds to body-AS in Figure 4.1B). However, there is a peak of antisense lncRNA transcription at the 3' end of mRNAs, 0-200bp upstream of mRNA TTSs (Figure 4.1D). Since 3'-UTRs are more highly variable than 5'UTRs (Mata 2013, Schlackow et al. 2013), there is likely more error in mRNA TTS annotation. As such it is possible that this peak in fact represents lncRNA transcription initiating in 3'-NFRs. Figure 4.1D additionally shows a peak of lncRNA transcription initiating 0-200bp downstream and antisense to mRNA TTSs, which likely represents lncRNAs initiating in 3'-NFRs and running antisense to the mRNA from whose 3'-NFR they initiate (2b on Figure 4.1A).

As a control, the distribution of lncRNAs was assessed relative to random coordinates (Figure 4.1E). For each chromosome, the TSS and TTS coordinates of each mRNA were replaced with random coordinates, which could be sampled from the entire length of the chromosome. Whether each mRNA is transcribed from the plus or minus strand was also randomised. The distribution of all lncRNAs relative to these random coordinates was a uniform distribution, with roughly equal numbers of lncRNAs being found at any point within the 1kb window upstream or downstream of the random coordinate on either strand (Figure 4.1E).

4.2.3 Enrichment and depletion in different samples

Different groups of lncRNAs are differentially expressed (DE) upon differing genetic or environmental perturbations (Chapter 3). Differential expression is relative to expression in WT log phase growth, and was calculated using DESeq ($p < 0.01$; see Chapter 2, Section 2.4). I analysed whether such groups are enriched for lncRNAs with a particular location with respect to mRNAs (Figure 4.2-4.5).

Firstly, for the set of lncRNAs differentially expressed in each sample (Table 2.1, 2.2) the observed distribution amongst the location categories described in Section 4.2.1 was calculated. The expected distribution amongst the same categories was calculated based on full set of 7332 lncRNAs (Figure 4.1B). Expected and observed distributions were compared using a chi-squared test and p-values corrected for multiple testing (Bonferroni method).

Secondly, for the set of lncRNAs differentially expressed in each sample (Table 2.1, 2.2) the observed distribution with respect to mRNA TSSs and TTSs was calculated. The expected distribution (antisense strand only) with respect to mRNA TSSs and TTSs was calculated based on full set of 7332 lncRNAs (Figure 4.1C-D). Expected and observed distributions were compared using a chi-squared test and p-values corrected for multiple testing (Bonferroni method).

4.2.3.1 Exosome mutants and nitrogen and glucose starvation

Differentially expressed (DE) lncRNAs in mutants of the exosome (*rrp6, rrp6-ts, dis3-54*), or known exosome-interacting factors (*cid14, pab2*), showed similar location patterns (Figure 4.2). Interestingly, lncRNAs differentially expressed upon nitrogen and glucose starvation displayed the same location pattern as these mutants (Figure 4.3).

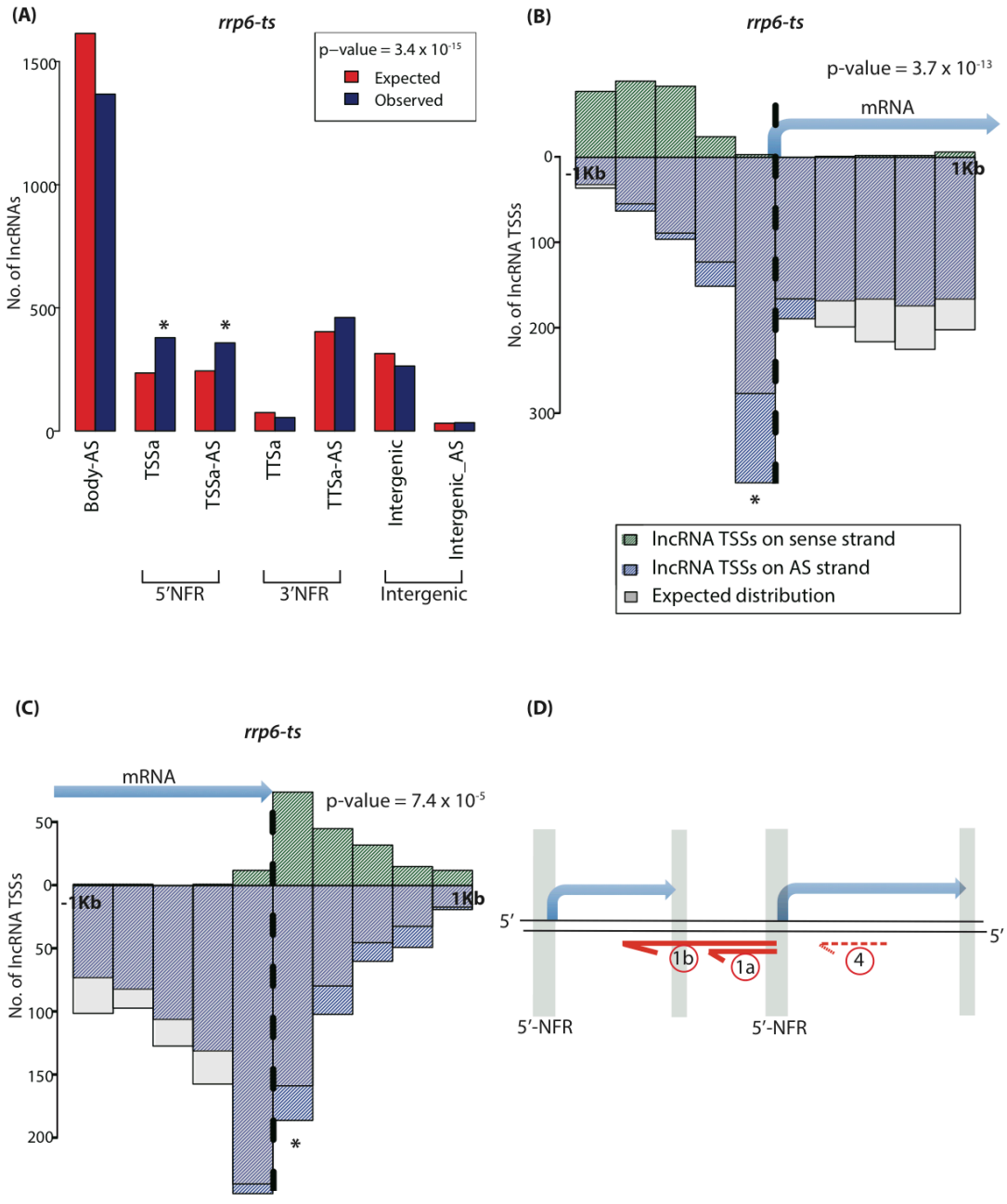

Figure 4.2: Location of lncRNAs in exosome mutants.

rrp6-ts used as representative of exosome mutants in A-C; astrices indicate key enrichments. (A) Distribution by type of lncRNAs differentially expressed (DE) in *rrp6-ts* (red bars) as compared to the distribution expected from the full set of lncRNAs (blue bars). Chi-squared test used to compare expected and observed distributions. (B) Distribution of lncRNAs DE in *rrp6-ts* relative to mRNA TSSs. Dotted vertical line indicates reference mRNA TSS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. Distribution is compared to that expected from the full set of lncRNAs on the antisense strand (light grey bars) using a chi-squared test. (C) Distribution of lncRNAs DE in *rrp6-ts* relative to mRNA TTSs. Dotted vertical line indicates reference mRNA TTS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. Distribution is compared to that expected from the full set of lncRNAs on the antisense strand (light grey bars) using a chi-squared test. (D) Schematic of the locations of lncRNAs enriched (solid red arrows) and depleted (dashed red arrow) in exosome mutants. Numbering as for Figure 4.1A.

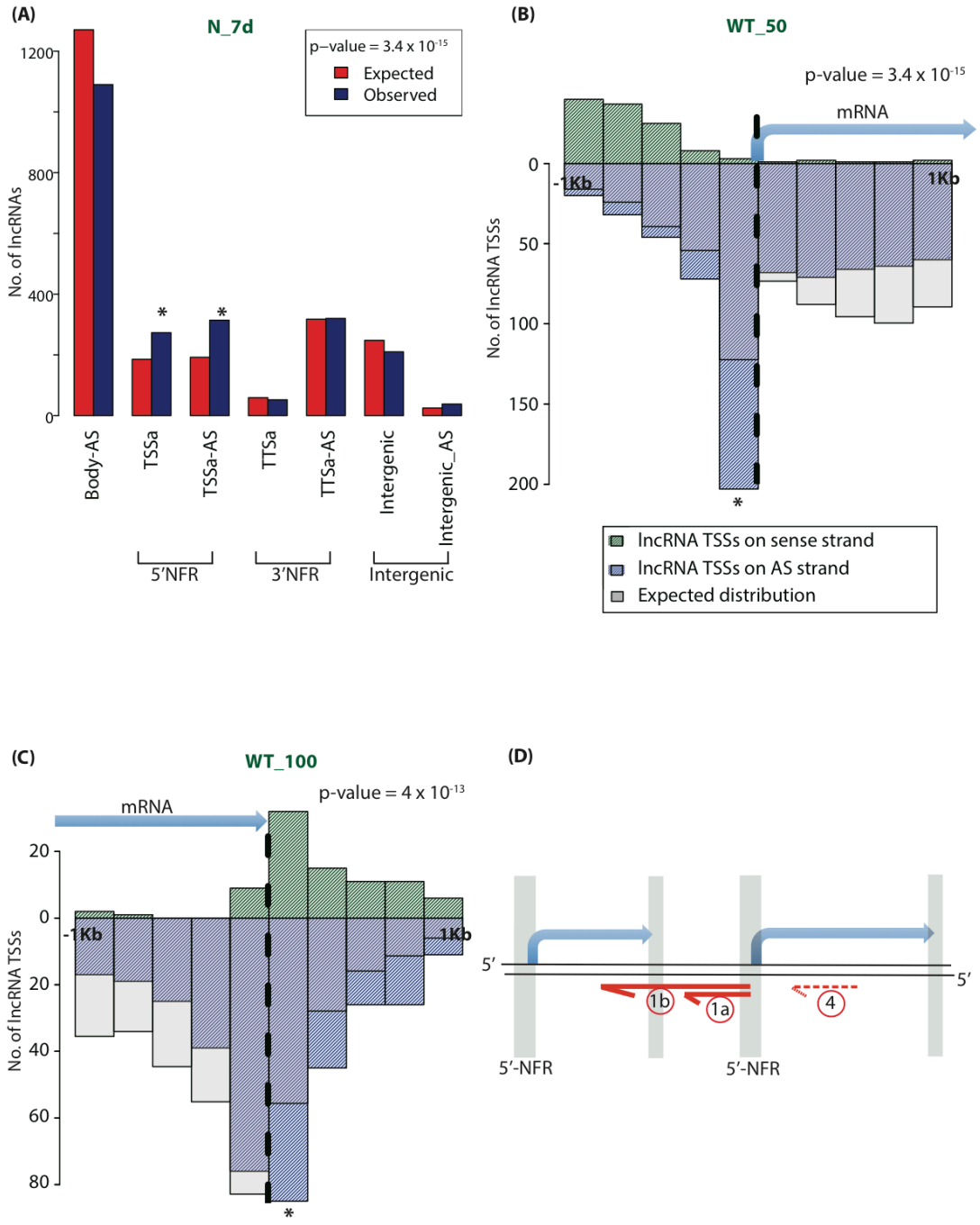

Figure 4.3: Location of lncRNAs under nitrogen and glucose starvation.

Astrices indicate key enrichments. (A) Distribution by type of lncRNAs DE in starvation samples (red bars; nitrogen starvation at 7 days used as representative) as compared to the distribution expected from the full set of lncRNAs (blue bars). Chi-squared test used to compare expected and observed distributions. (B) Distribution of lncRNAs DE in WT-50 (WT cells glucose starved at 50% survival) relative to mRNA TSSs. Dotted vertical line indicates reference mRNA TSS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. Distribution is compared to that expected from the full set of lncRNAs on the antisense strand (light grey bars) using a chi-squared test. (C) Distribution of lncRNAs DE in in WT-100 (glucose starved WT cells at 100% survival) relative to mRNA TTSs. Dotted vertical line indicates reference mRNA TTS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. Distribution is compared to that expected from the full set of lncRNAs on the

antisense strand (light grey bars) using a chi-squared test. (D) Schematic of the locations of lncRNAs enriched (solid red arrows) and depleted (dashed red arrow) in starvation samples. Numbering as for Figure 4.1A.

Most striking is the enrichment of TSSa lncRNAs, and depletion of lncRNAs originating antisense to the body of mRNAs (Figure 4.2A, 4.3A). Examining the distributions relative to mRNA TSSs in these samples supports this, revealing an enriched peak of divergent lncRNA transcription initiation accompanied by fewer lncRNAs initiating antisense to the body of mRNAs (Figure 4.2B, 4.3B).

Similarly, distributions relative to mRNA TTSs reveal a depletion of lncRNAs initiating antisense to the body of mRNAs (Figure 4.2C, 4.3C). Additionally, there is an enriched peak originating 0-200bp downstream and antisense to mRNA TTSs. This peak is in fact reflective of increased non-coding transcription from 5'-NFRs, since it corresponds to lncRNAs originating from the antisense strand in intergenic regions between two tandem mRNAs that are closer together than 300bp. Such lncRNAs could be considered as originating from the 3'-NFR of the upstream mRNA or the 5'-NFR of the downstream lncRNA. As already described (see Section 4.2.1), such lncRNAs will contribute to enrichment in TSSa location categories, but they will show both as an enriched bidirectional peak when examining the distribution relative to mRNA TSSs, as well as an enriched peak antisense and downstream of mRNA TTSs.

In summary, lncRNAs differentially expressed in exosome mutants and under glucose or nitrogen starvation are enriched for being transcribed divergently from mRNA TSSs (Figure 4.2D, 4.3D).

4.2.3.2 RNAi and exo2 mutants

RNAi, *exo2*, and *exo2/dcr1* mutants all show enrichments in the TTSa-AS location category (Figure 4.4A-C). This category is also enriched in exosome mutants but to a lesser extent (Figure 4.2A). The enrichment in this class arises from increased convergent read-through in these mutants.

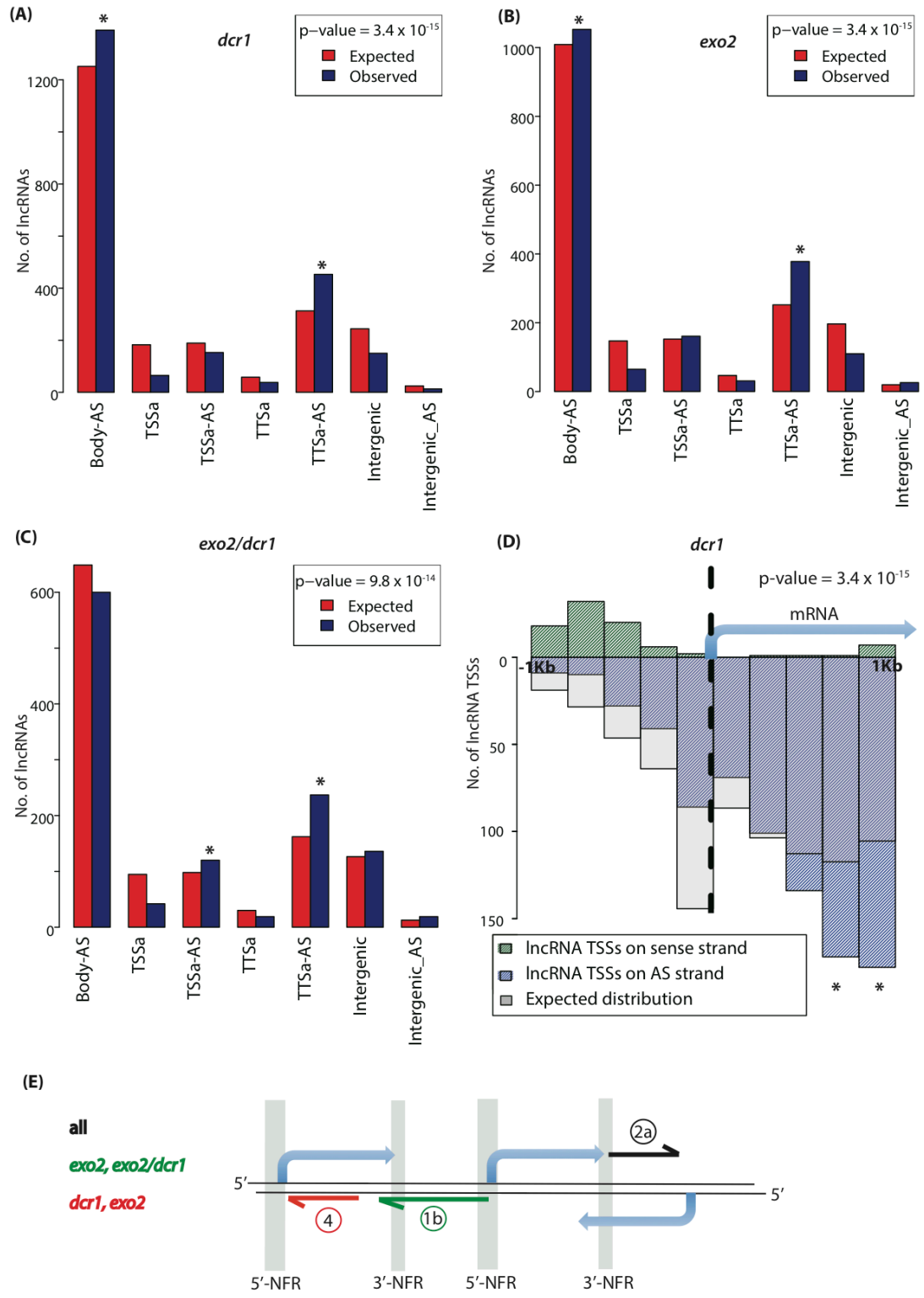

Figure 4.4: Location of lncRNAs in RNAi and *exo2* mutants.

Astrices indicate key enrichments. (A) Distribution by type of lncRNAs DE in RNAi samples (red bars; *dcr1* used as representative) as compared to the distribution expected from the full set of lncRNAs (blue bars). Chi-squared test used to compare expected and observed distributions. (B) Distribution by type of lncRNAs DE in *exo2* (red bars) as compared to the distribution expected from the full set of lncRNAs (blue bars). Chi-squared test used to compare expected and observed distributions. (C) Distribution by type of lncRNAs DE in *exo2/dcr1* (red bars) as compared to the

distribution expected from the full set of lncRNAs (blue bars). Chi-squared test used to compare expected and observed distributions. (D) Distribution of lncRNAs DE in *dcr1* relative to mRNA TSSs. Dotted vertical line indicates reference mRNA TSS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. Distribution is compared to that expected from the full set of lncRNAs on the antisense strand (light grey bars) using a chi-squared test. (E) Schematic of the locations of lncRNAs enriched RNAi and *exo2* mutants. Numbering as for Figure 4.1A.

However, apart from *dcr1* and *rdp1*, the distribution of lncRNAs around mRNA TTSs is not significantly different to what would be expected based on the full set of lncRNAs (not shown). This is because TTSa-AS lncRNAs arising from convergent read-through will not be represented on these plots if the lncRNA doesn't originate in a 'clean' intergenic region (see Section 2.5, Chapter 2; an example is represented by '2a' in Figure 4.4E). *dcr1* and *rdp1* show a significantly different distribution around mRNA TTSs as compared to the distribution of the full set of lncRNAs ($p < 0.05$; plots not shown). This is due to a depleted peak 0-200bp downstream and antisense to mRNA TTSs. This is consistent with the finding that enrichment in the TTSa-AS location category is due to increased convergent read-through, rather than enrichment of lncRNAs running antisense to the gene from whose 3'-NFR they originate.

Furthermore, RNAi, *exo2*, and *exo2/dcr1* mutants all show depletion in the TTSa location category (no antisense overlap) (Figure 4.4A-C) which suggests that, in these mutants, non-coding read-through which specifically results in sense:antisense overlap is up-regulated, as opposed to read-through *per se*.

For *dcr1* and *exo2* mutants, the distributions of lncRNAs relative to mRNA TSSs show enriched non-coding transcription initiating antisense to the body of mRNAs, which is in agreement with the enrichment in the 'body-AS' class seen in these mutants (Figure 4.4A-B, D).

In summary, as depicted in Figure 4.4E, RNAi, *exo2*, and *exo2/dcr1* mutants all show increased TTSa-AS non-coding transcription, arising from increased convergent read-through. *exo2* and *exo2/dcr1* mutants additionally display a slight enrichment of TSSa-AS lncRNAs. Interestingly, while lncRNAs initiating antisense

to the body of mRNAs are enriched in *exo2* and *dcr1* mutants, this class of lncRNA is in fact depleted in the double mutant.

4.2.3.3 Meiotic samples

Early meiotic time-points (0h, 2h, 4h) all show enrichments in the TSSa-AS location category, but depletion in the body-AS class (Figure 4.5A). This is supported by the distribution of lncRNAs relative to mRNA TSSs (Figure 4.5C). Such a distribution of lncRNAs is similar to that seen in the exosome mutants (Figure 4.2) and is also reminiscent of *exo2*, which shows a slight but specific enrichment of the TSSa-AS lncRNAs (Figure 4.4B).

In contrast, later meiotic time-points (6h, 8h) show enrichments in the body-AS and TTSa-AS classes (Figure 4.5B, D). Such a distribution of lncRNAs is reminiscent of that seen for both RNAi and *exo2* mutants (Figure 4.4).

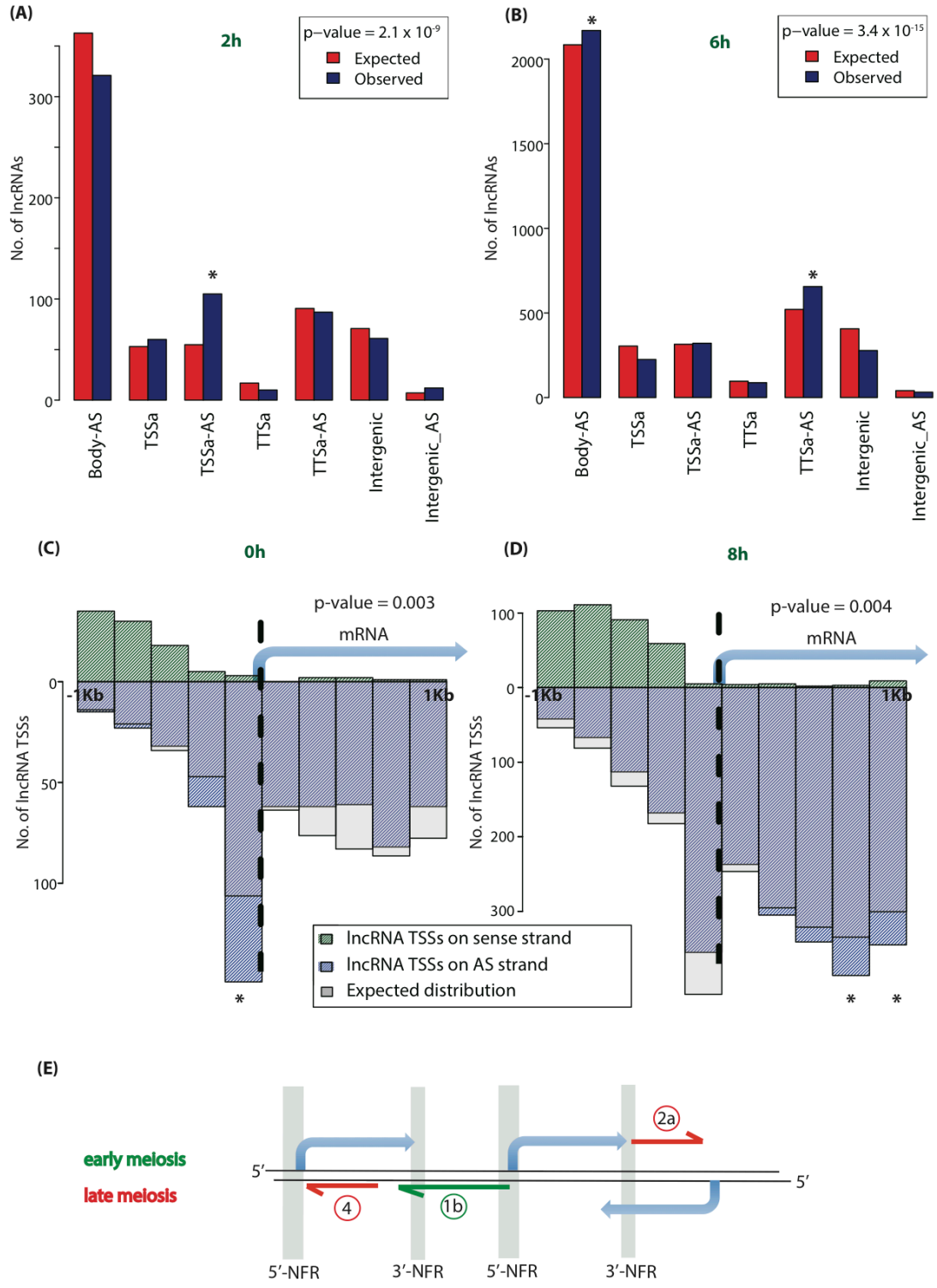

Figure 4.5: Location of lncRNAs in meiotic samples.

Astrices indicate key enrichments. (A) Distribution by type of lncRNAs DE in early meiosis (red bars; time-point at 2 hours used as representative) as compared to the distribution expected from the full set of lncRNAs (blue bars). Chi-squared test used to compare expected and observed distributions. (B) Distribution by type of lncRNAs DE in late meiosis (red bars; time-point at 6 hours used as representative) as compared to the distribution expected from the full set of lncRNAs (blue bars). Chi-squared test used to compare expected and observed distributions. (C) Distribution of

lncRNAs DE in early meiosis (time-point at 0 hours used as representative) relative to mRNA TSSs. Dotted vertical line indicates reference mRNA TSS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. Distribution is compared to that expected from the full set of lncRNAs on the antisense strand (light grey bars) using a chi-squared test. (D) Distribution of lncRNAs DE in late meiosis (time-point at 8 hours used as representative) relative to mRNA TSSs. Dotted vertical line indicates reference mRNA TSS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. Distribution is compared to that expected from the full set of lncRNAs on the antisense strand (light grey bars) using a chi-squared test. (E) Schematic of the locations of lncRNAs enriched at different meiotic time-points. Numbering as for Figure 4.1A.

In summary, lncRNAs up-regulated during early meiosis are enriched for being transcribed divergently from mRNAs and running antisense to upstream mRNAs. In contrast, antisense lncRNAs arising from convergent read-through, or initiating antisense to the body of mRNAs, are enriched during later meiotic time-points (Figure 4.5E). The similarity of these location profiles to those seen for *exo2*, as well as exosome and RNAi mutants, could point to differing roles of these RNA processing factors in lncRNA regulation throughout the course of meiosis.

4.3 Expression correlation of lncRNAs and mRNAs

It is clear that lncRNAs are not randomly distributed throughout the genome, but can instead be described by where they initiate relative to mRNAs (Section 4.2). Moreover, lncRNAs differentially expressed (DE) under differing genetic and environmental perturbations display distinct location profiles (Section 4.2). Since many lncRNAs are proposed to exert a regulatory effect on mRNA expression *in cis* (see Chapter 1, Section 1.5), it was of interest to examine whether lncRNA:mRNA transcript pairs, which are related to each other in terms of their location, are also related in terms of their expression. How lncRNA:mRNA pairs relate to each other, in terms of expression, was assessed by (a) testing for association between lncRNA differential expression and mRNA differential expression in a given set of mRNA:lncRNA pairs (Fisher tests); (b) calculating the Pearson's product moment correlation coefficient (CC) for normalised lncRNA and mRNA read counts (see Chapter 2, Section 2.6 for further details of these methods).

4.3.1 lncRNA:mRNA pairs

All lncRNAs, apart from those that are intergenic, can be associated with at least one mRNA. Broadly speaking, lncRNAs will form a pair with an mRNA in one of two ways: the lncRNA and mRNA are transcribed divergently from each other (bidirectional pairs); the lncRNA and mRNA have antisense overlap (antisense pairs).

TSSa lncRNAs are transcribed divergently from mRNAs, and each TSSa lncRNA will form a bidirectional lncRNA:mRNA pair.

Antisense lncRNAs will form an antisense lncRNA:mRNA pair with at least one mRNA. As described in Section 4.2, antisense lncRNAs can originate antisense to the body of mRNAs, from intergenic regions, or from 5'- or 3'-NFRs.

For most RNA-processing mutants, out of all possible lncRNA:mRNA pairs, pairs in which both the lncRNA and the mRNA are DE are *not* enriched over pairs in which only one or neither is DE (Figure 4.6A). Thus, DE lncRNAs are equally likely to be associated with a DE mRNA or a non-DE mRNA. For example, for the *rrp6-ts* mutant, of those pairs where the mRNA is DE, ~50% are associated with a DE lncRNA. Similarly, of those pairs where the mRNA is non-DE, ~50% are associated with a DE lncRNA. A lack of association between differential expression of lncRNAs and differential expression of mRNAs, likely reflects the fact that in these mutants lncRNAs are selectively post-transcriptionally stabilised, i.e. these factors act to selectively degrade lncRNAs but not mRNAs.

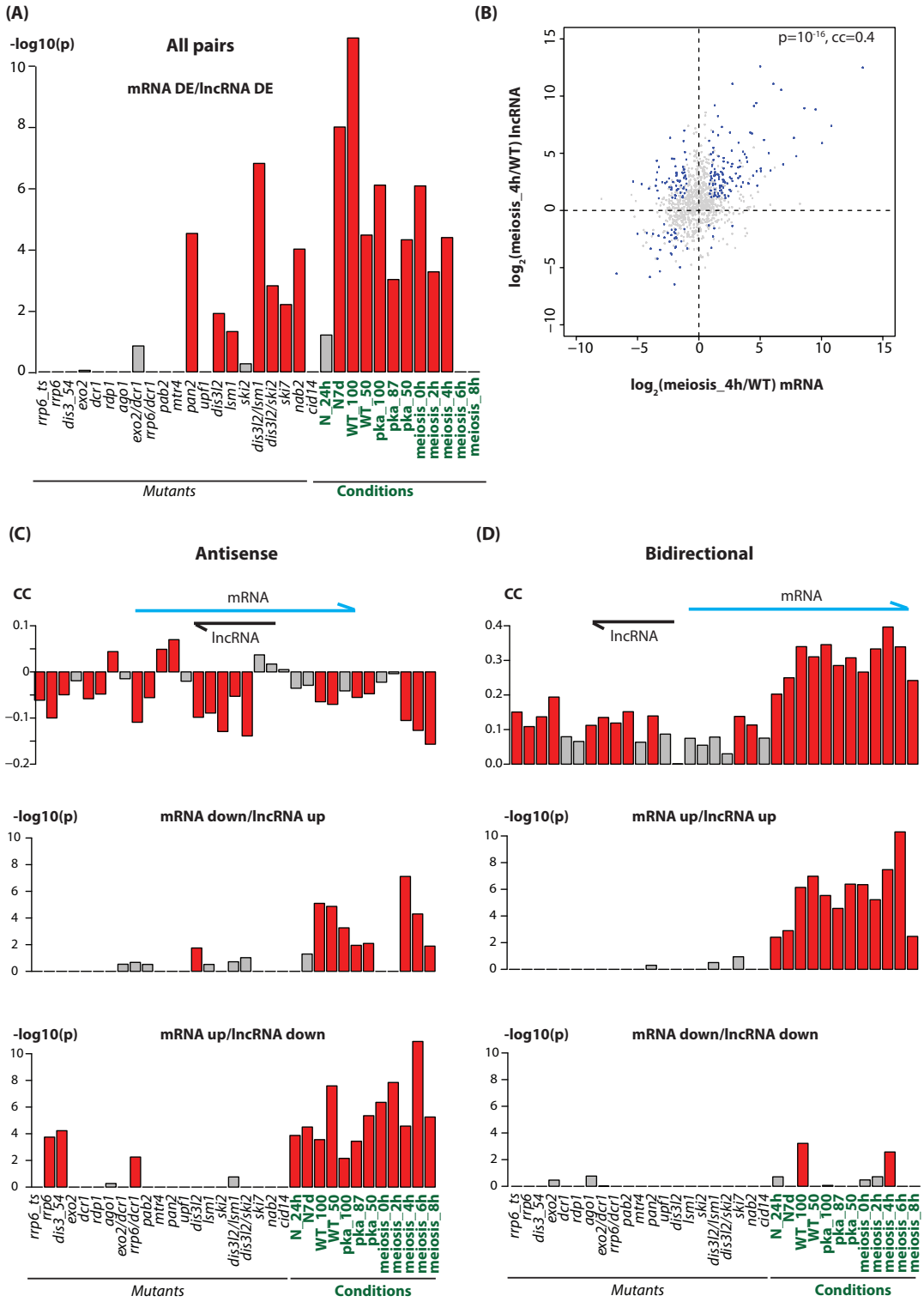

Figure 4.6: Expression correlation of lncRNA:mRNA pairs.

Red indicates p <0.05. (A) Fisher tests for enrichment of pairs in which both transcripts are DE; all lncRNA:mRNA pairs considered. (B) Scatterplot of normalised read counts for divergent lncRNA:mRNA transcript pairs for the comparison of conditions meiosis_4h/WT. Blue indicates transcript pairs in which both the lncRNA and mRNA are significantly DE (DESeq). CC = Pearson's correlation coefficient and associated p-value. (C) Antisense lncRNA:mRNA pairs: top panel- Pearson's correlation coefficients; middle and bottom panel - Fisher tests for enrichment of anti-

correlating pairs. (D) Bidirectional lncRNA:mRNA pairs: top panel - Pearson's correlation coefficients; middle and bottom panel - Fisher tests for enrichment of positively correlating pairs.

RNA-processing mutants that do not show this relationship are: *pan2*, *dis3l2*, *lsm1*, *ski7*, *nab2,* as well as the double mutants *dis3l2/lsm1* and *dis3l2/ski2* (Figure 4.6A). All of these mutants have relatively low numbers of DE lncRNAs and DE mRNAs (Chapter 3). Nonetheless, DE lncRNAs in these mutants are significantly more likely to be associated with DE mRNAs (Figure 4.6A). This may reflect the fact that these factors play little or no role in selectively targeting lncRNAs for degradation.

For WT cells under different physiologically relevant growth conditions, out of all possible lncRNA:mRNA pairs, pairs in which both the lncRNA and the mRNA are DE *are* significantly enriched over pairs in which only one or neither is DE (Figure 4.6A). Thus, DE lncRNAs are more likely to be associated with a DE mRNA than with a non-DE mRNA. This suggests that, in these samples, differential expression of lncRNAs is significantly associated with differential expression of the mRNAs that they are related to in terms of location. Exceptions to this are the meiotic samples at 6h and 8h. The lack of significant association of DE mRNAs and DE lncRNAs in these two samples may suggest that some lncRNAs are differentially expressed independently of their associated mRNAs at these later time-points of meiosis. Selective post-transcriptional stabilisation of lncRNAs by global down-regulation of RNA processing factors may account for this (see Chapter 5). In addition, after the conservative Bonferroni correction for multiple testing, the p-value for the N_24h sample only just misses the cut-off of $p < 0.05$ ($p = 0.06$).

In summary, in mutants where large numbers of lncRNAs are selectively stabilised, post-transcriptional stabilisation of lncRNAs is not associated with differential expression of mRNAs. In contrast, under physiologically relevant growth conditions, differential expression of lncRNAs and mRNAs are associated. This may indicate that lncRNA expression may have a regulatory effect on mRNA expression under such conditions. DE of lncRNAs under different growth conditions may be due to increased or decreased transcription rather than post-transcriptional stabilisation. If this is the case, the fact that differential lncRNA transcription is

associated with differential mRNA expression, while post-transcriptional stabilisation of lncRNAs is not, suggests that any functions of these lncRNAs may be mediated through the act of them being transcribed, rather than by the lncRNA transcript itself.

To verify that the relationships observed are a consequence of the locational constraints placed on the lncRNA:mRNA pairs analysed, the same Fisher tests were performed on the same number of randomly chosen lncRNA:mRNA pairs (control pairs). For these pairs, no significant association between mRNA differential expression and lncRNA differential expression was observed across all the samples.

Finally, 1334 (~25%) mRNAs are not associated with any lncRNA. These are enriched for the GO-terms 'shortest mRNAs' and 'RNA binding' ($p < 10^{-9}$).

4.3.2 Bidirectional pairs

Bidirectional transcript pairs have significantly positive expression correlation coefficients in WT cells under physiologically relevant growth conditions (upper panel of Figure 4.6D; Figure 4.6B). Some, but not all, mutants also have significant positive expression correlations amongst these pairs, but the correlations are generally weaker and less significant (not shown) than in WT cells under different growth conditions.

In support of this, amongst bidirectional pairs, there is enrichment for pairs in which both the lncRNA and mRNA are significantly up-regulated in WT cells under different growth conditions but not in mutants (middle panel of Figure 4.6D). Enrichment of pairs in which both mRNA and lncRNA are significantly down-regulated is observed for WT_100 and meiosis_4h samples (bottom panel of Figure 4.6D). Fisher tests show no enrichment of anti-correlating DE mRNA and DE lncRNA amongst these pairs (not shown).

Together, these results suggest that bidirectional lncRNA:mRNA pairs positively correlate in expression. This positive correlation is more pronounced under environmental perturbations than under the genetic perturbations examined, possibly for the same reasons discussed in Section 4.3.1.

To verify that the relationships observed are a consequence of the locational constraints placed on the bidirectional lncRNA:mRNA pairs analysed, the same number of randomly chosen lncRNA:mRNA pairs were analysed (control pairs). For these pairs, no significant association between mRNA up- or down-regulation and lncRNA up- or down-regulation was observed across all the samples. In addition, correlation coefficients for these random pairs were extremely weak, positive or negative, and not significant ($p < 0.05$).

DE mRNAs associated with DE bidirectional lncRNA are enriched for the following GO-terms:

Stress module (Chen D. et al. 2003); $p = 0.007$

No GO-term enrichment was observed amongst the same number of control pairs ($p < 0.05$).

4.3.3 Antisense pairs

Antisense transcript pairs have weakly negative, but significant, expression correlation coefficients in most samples (Figure 4.6C upper panel). Notably, expression correlations are stronger and more significant at progressively later meiotic time-points (4h, 6h, 8h). In contrast, weaker and less significant correlation coefficients are observed amongst nitrogen and glucose starved samples.

In support of this, amongst antisense pairs, Fisher tests reveal an enrichment of pairs in which both the lncRNA and mRNA are significantly DE and anti-correlate (pairs where the lncRNA is significantly up-regulated while mRNA is significantly down-regulated, or vice versa; Figure 4.6C middle and bottom panels). Such enrichments are observed mainly in WT cells under different growth conditions but not in mutants, possibly for the same reasons discussed in Section 4.3.1. Fisher tests show no enrichment of DE positively correlating pairs in any samples (not shown).

Together, these results suggest that antisense lncRNA:mRNA pairs generally anti-correlate in expression.

To verify that the relationships observed are a consequence of the locational constraints placed on the antisense lncRNA:mRNA pairs analysed, the same number of randomly chosen lncRNA:mRNA pairs were analysed (control pairs). For these pairs, no significant association between mRNA up- or down-regulation and lncRNA up- or down-regulation was observed across all samples. In addition correlation coefficients for control pairs were close to zero and not significant ($p< 0.05$).

DE mRNAs associated with DE antisense lncRNA are enriched for the following GO-terms ($p <10^{-7}$):

longest mRNAs (Bähler lab)
middle meiotic genes (Mata et al. 2002)
hydrolase activity
stress module (Chen D. et al. 2003)

No GO-term enrichment was observed amongst the same number of control pairs ($p <0.05$).

4.3.4 Antisense lncRNAs originating in NFRs

As described in Section 4.2, most antisense lncRNAs originate antisense to the body of mRNAs, or from 5'- or 3'-NFRs. Expression correlation analyses of these different types of antisense lncRNA:mRNA pairs were performed.

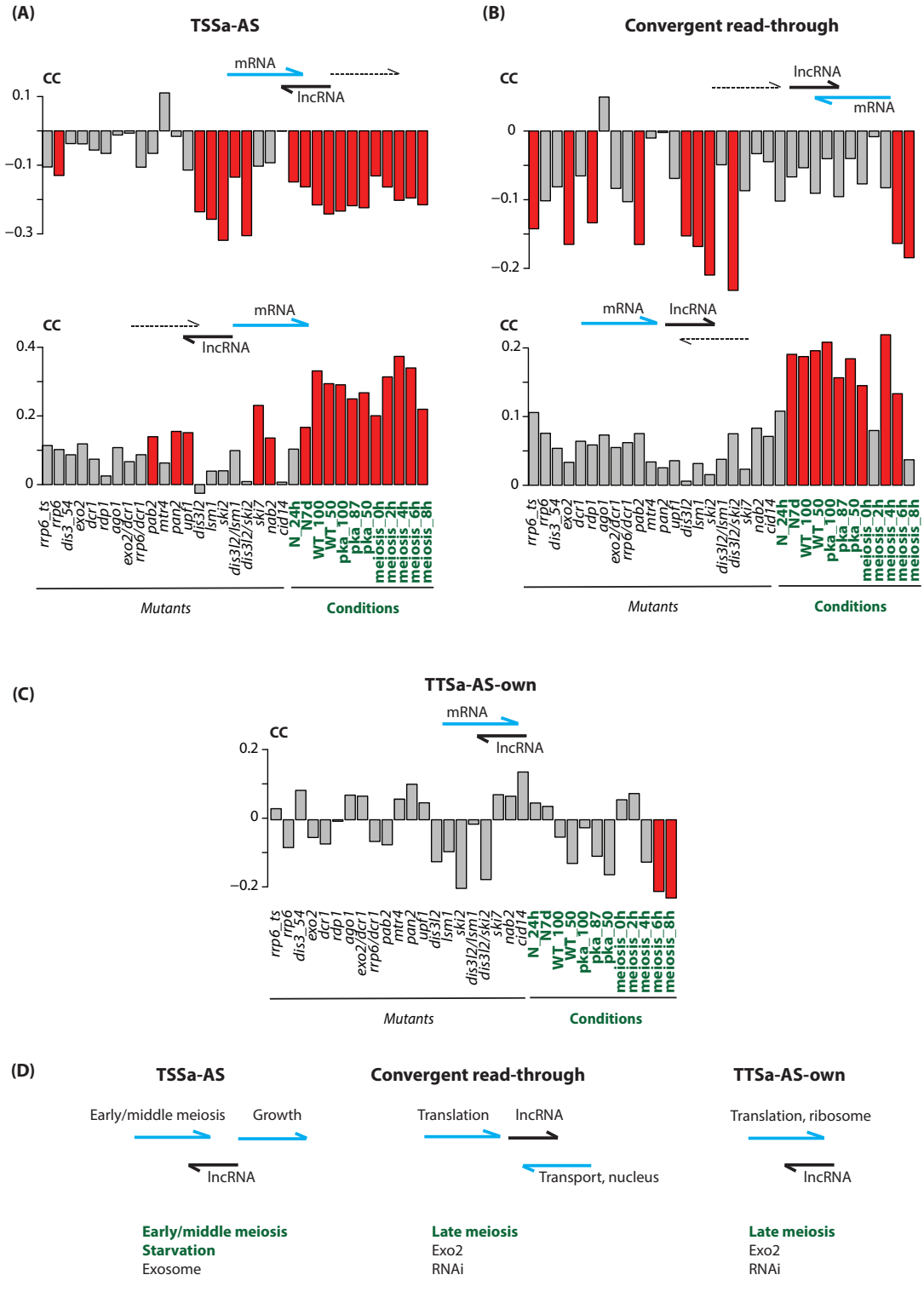

Figure 4.7: Expression correlation of antisense lncRNA:mRNA pairs in which the lncRNA originates in an NFR.

Red indicates p <0.05. (A) Pearson's correlation coefficients for pairs involving TSSa-AS lncRNAs as indicated in schematic. (B) Pearson's correlation coefficients for pairs involving a convergent read-through lncRNA as indicated in schematic. (C) Pearson's correlation coefficients for pairs involving a TTSa-AS-own lncRNA as indicated in schematic. (D) Schematics of lncRNA:mRNA

relationships that are enriched in the samples indicated. Blue indicates mRNA and the GO-terms these mRNAs are enriched for are indicated; black indicates lncRNA.

For antisense lncRNA:mRNA pairs in which the lncRNA originates in a 5'-NFR (i.e. a TSSa lncRNA which additionally runs antisense to an upstream gene), correlation coefficients are more strongly negative and more significant than for the full set of antisense lncRNA:mRNA pairs (Figure 4.7A, top panel; compared to Figure 4.6C, top panel). Correlation coefficients are significant and negative for all conditions, as well as the following mutants: *rrp6, dis3l2*, *lsm1*, *ski2*, and the double mutants *dis3l2/lsm1* and *dis3l2/ski2*. Additionally, amongst these antisense pairs, Fisher tests show enrichments for differentially expressed anti-correlating pairs for all glucose starvation samples as well as meiosis_4h (not shown).

Amongst these pairs, the following GO-terms are enriched ($p < 0.001$):

early meiotic genes (Mata et al. 2002)
middle meiotic genes (Mata et al. 2002)

For the same number of control pairs there is no enrichment of any GO-terms ($p<0.05$).

Antisense lncRNAs originating in a 5'-NFR have the potential to link the expression of two mRNAs – the mRNA they are transcribed divergently from, and the mRNA they run antisense to. Such lncRNAs anti-correlate in expression with their antisense mRNA (see above), but positively correlate in expression with the mRNA they are transcribed divergently from, predominantly under different growth conditions (Figure 4.7A, bottom panel). Moreover, mRNAs these lncRNAs are transcribed divergently from, are enriched for the following GO-terms ($p < 0.01$):

growth module (Chen D. et al. 2003)

For the same number of control pairs there is no enrichment of any GO-terms ($p < 0.05$).

In summary, lncRNAs transcribed divergently from mRNAs and running antisense to upstream mRNAs are enriched under nitrogen or glucose starvation, early/middle meiosis, and are targeted for degradation predominantly by the exosome (Section 4.2.3). Such lncRNAs anti-correlate with mRNAs they run antisense to (Figure 4.7A top panel), which are enriched for early/middle meiotic GO-terms. Additionally, such lncRNAs positively correlate with the mRNA they are transcribed divergently from (Figure 4.7A bottom panel), which are enriched for growth terms. Thus such lncRNAs may represent a mechanism to coordinate the expression of growth and meiotic genes (Figure 4.7D).

In contrast, antisense lncRNAs originating in 3'-NFRs (TTSa) are enriched in late meiosis, and targeted for degradation by RNAi factors, Exo2 and, to a lesser extent, by the exosome (Section 4.2.3). Such lncRNAs may run antisense to either the gene from whose 3'-NFR they originate (TTSa-AS-own; represented by 2b on Figure 4.1A) or antisense to a downstream gene in a convergent orientation (convergent read-through; represented by 2a on Figure 4.1A). Most TTSa-AS lncRNAs arise from convergent read-through (~75%).

For antisense lncRNA:mRNA pairs arising from convergent read-through, negative correlation coefficients are seen for *exo2*, RNAi (*rdp1*), late meiotic time-points (6h and 8h), and to a lesser extent for the exosome-related mutants (*rrp6*, *pab2*) (Figure 4.7B top panel). Additionally, *dis3l2*, *lsm1* and *ski2* mutants also show significant negative correlation coefficients for these pairs. mRNAs in these antisense lncRNA:mRNA pairs are enriched for the following GO-terms ($p < 10^{-3}$):

nucleus
vesicle-mediated transport
transport

For the same number of control pairs there is no enrichment of any GO-terms ($p<0.05$).

Antisense lncRNAs arising from convergent read-through have the potential to link the expression of two mRNAs – the mRNA from which they originate, and the mRNA they run antisense to. mRNAs which give rise to such lncRNAs are enriched

for the GO-term ‘cytoplasmic translation’ ($p < 0.001$), and such lncRNA:mRNA pairs positively correlate in expression across most conditions (Figure 4.7B bottom panel; Figure 4.7D).

lncRNAs which run antisense to the gene from whose 3’-NFR they originate, anti-correlate in expression with their antisense mRNA at late meiotic time-points (6h and 8h) (Figure 4.7C). mRNAs in such pairs are enriched for the following GO-terms ($p < 0.01$) (Figure 4.7D):

ribosome
translation
cytosolic large ribosomal subunit

For the same number of control pairs there is no enrichment of any GO-terms ($p<0.05$).

4.3.5 Antisense lncRNAs originating antisense to the body of mRNAs

lncRNAs originating antisense to the body of mRNAs have weaker and less significant correlation coefficients that than for antisense lncRNAs originating in NFRs (Figure 4.9A compared to Figure 4.7). The median percentage of mRNA covered (coding coverage) by body-AS lncRNAs is 20%. By contrast, antisense lncRNAs originating in NFRs have a higher median coding coverage of ~60%. AS-body lncRNAs were therefore binned by coding coverage to see if this affected the degree of correlation (Figure 4.8).

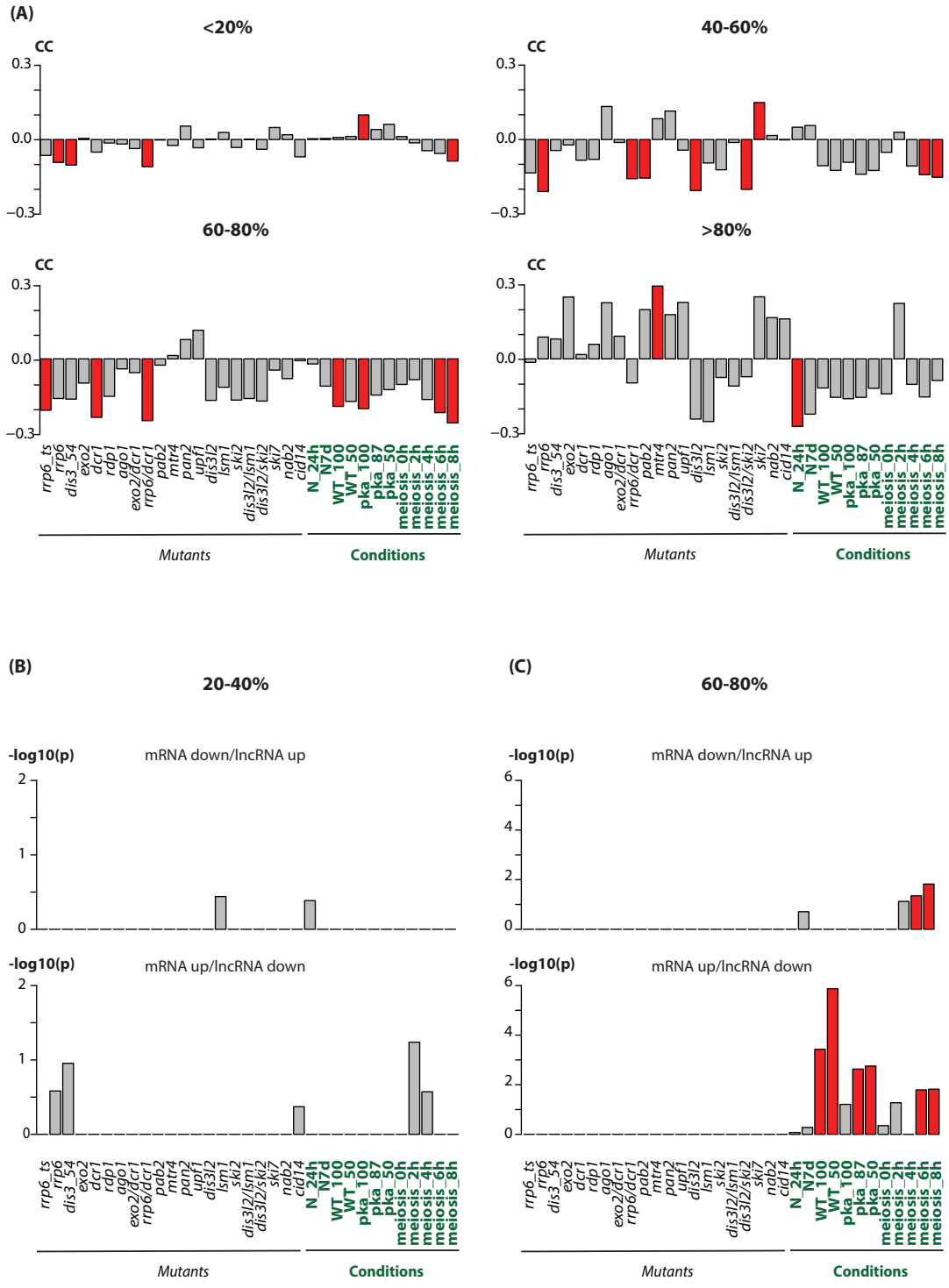

Figure 4.8: Expression correlation of 'body-AS' lncRNA:mRNA pairs by coding coverage.

Red indicates p <0.05. (A) Pearson's correlation coefficients for 'body' antisense lncRNA:mRNA pairs binned by coding coverage as indicated. (B) 20-40% coding coverage: top panel – Fisher tests for enrichment of anti-correlating pairs. (C) 60-80% coding coverage: top panel – Fisher tests for enrichment of anti-correlating pairs.

lncRNAs with a coding coverage of <20% and 20-40% have very weak correlation coefficients that are significant and negative in only four samples (Figure 4.8A). Additionally Fisher tests show no enrichment of anti-correlating pairs in either of these bins (Figure 4.8B).

In contrast, for 40-60% and 60-80% bins, correlation coefficients are more strongly negative and significant in more samples (Figure 4.8A), with Fisher tests revealing an enrichment of anti-correlating pairs in which both the lncRNA and mRNA are differentially expressed across most physiologically relevant growth conditions (Figure 4.8C).

Interestingly, increasing coding coverage to >80% produces a significantly negative correlation coefficient for only one sample (Figure 4.8A). Thus, although increasing coding coverage seems to result in stronger and more significant negative correlation coefficients, this relationship doesn't appear to hold beyond 80% coverage.

lncRNAs originating antisense to the body of mRNAs can either overlap the 5' end of mRNAs (Figure 4.9B) or neither end (Figure 4.9C). Only two lncRNAs cover the entire length of mRNAs (100% coding coverage) and therefore overlap both ends.

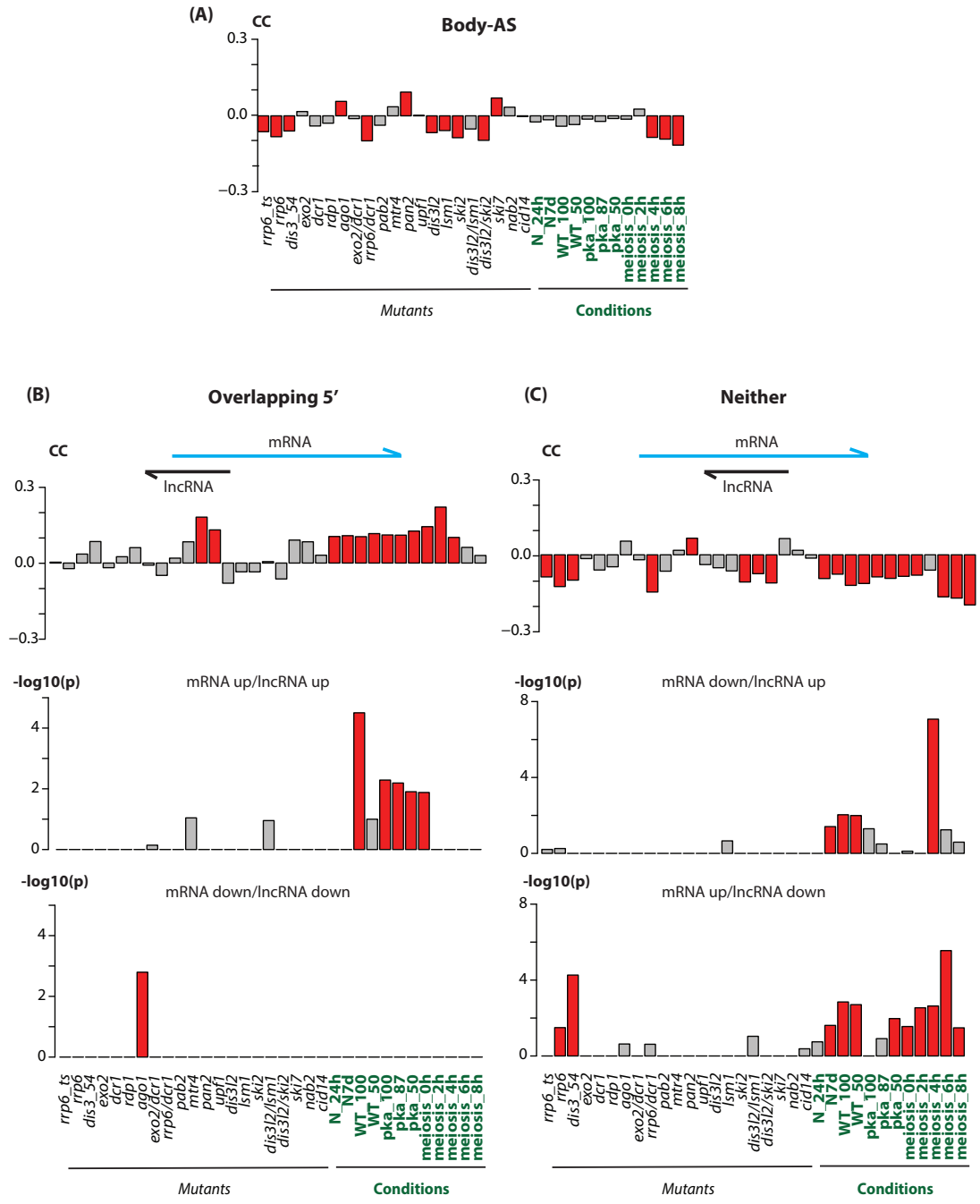

Figure 4.9: Expression correlation of 'body-AS' lncRNA:mRNA pairs by where the lncRNA overlaps its mRNA partner.

Red indicates p <0.05. (A) Pearson's correlation coefficients for all 'body' antisense lncRNA:mRNA pairs. (B) 'body' lncRNA:mRNA pairs which overlap 5' of mRNA as indicated on schematic; top panel - Pearson's correlation coefficients; middle and bottom panels – Fisher tests for enrichment of positively correlating pairs. (C) 'body' lncRNA:mRNA pairs which overlap neither end of the mRNA as indicated on schematic: top panel - Pearson's correlation coefficients; middle and bottom panels – Fisher tests for enrichment of anti-correlating pairs.

Interestingly, antisense lncRNAs which overlap the 5' end of mRNAs positively correlate with the mRNA they are antisense to (Figure 4.9B), while those that overlap neither end anti-correlate with the mRNA they're antisense to (Figure 4.9C). The correlation coefficients in this class of lncRNAs which overlap neither end are more strongly negative and significant in more samples than for the entire set of body-AS lncRNAs (Figure 4.9A compared to Figure 4.9C). This is despite a median coding coverage of 20%, which suggests that it may be where the antisense lncRNA lies relative to the mRNA, rather than the percentage of mRNA it overlaps, which is responsible for mediating an inhibitory effect on sense mRNA expression.

4.4 Summary of main conclusions

lncRNAs can be described by where they originate relative to mRNAs. Apart from intergenic lncRNAs, all lncRNAs can be paired with at least one mRNA. Amongst these pairs there is a significant association between differentially expressed (DE) mRNAs and DE lncRNAs in physiological conditions/environmental perturbations but not in mutants. This relationship does not hold for mutants that play little or no role in selectively degrading lncRNAs.

Bidirectional lncRNAs are enriched in nitrogen and glucose starvation samples and appear to be targeted for degradation by the exosome. These generally positively correlate in expression with the mRNA they are transcribed divergently from.

Bidirectional lncRNAs that additionally run antisense to an upstream mRNA are enriched at early meiotic time-points. They may act to couple the transcriptional regulation of the gene they are transcribed divergently from (positively correlate; 'growth' GO-terms enriched in these mRNAs) and the gene they run antisense to (anti-correlate; 'early/middle meiosis' GO-terms enriched in these mRNAs).

lncRNAs originating in 3'-NFRs (mRNA TTS associated) and running antisense to mRNAs are enriched in late meiotic time-points and appear to be targeted for degradation by RNAi factors, Exo2, and to a lesser extent by the exosome. lncRNAs originating in 3'-NFRs and running antisense to a downstream gene in a convergent orientation may act to couple the transcriptional regulation of the gene whose 3'-

NFR they are transcribed from (positively correlate; ‘growth’ GO-terms enriched in these mRNAs) and the gene they run antisense to (anti-correlate; ‘transport’ GO-terms enriched in these mRNAs). lncRNAs originating in 3’-NFRs and running antisense to the mRNA from whose 3’-NFR they originate may also down-regulate ‘growth’ genes at later meiotic time-points.

Antisense (AS) lncRNA:mRNA pairs generally anti-correlate in expression. The degree of anti-correlation is stronger for AS lncRNAs originating in NFRs, or for lncRNAs which originate antisense to the body of lncRNAs but overlap neither end of the mRNA. lncRNAs which originate antisense to the body of mRNAs, and overlap the 5’ end of the mRNA, actually positively correlate in expression.

Chapter 5 Classification and regulation of *S. pombe* lncRNAs

5.1 Introduction

In *S. cerevisiae* lncRNAs have been classified by the RNA processing pathways regulating their expression. Thus CUTs (cryptic unstable transcripts) are a group of lncRNAs that are targeted for degradation by the nuclear exosome and stabilised upon deletion of the nuclear-specific exosome subunit Rrp6 (Neil et al. 2009, Xu et al. 2009). XUTs (Xrn1-dependent unstable transcripts) are a lncRNA group which are targeted for degradation by the cytoplasmic exonuclease Xrn1 (van Dijk et al. 2011), while NUTs (Nrd1-unterminated transcripts) are a group of lncRNAs detected upon nuclear depletion of Nrd1 – a member of the Nrd1-Nab3-Sen1 (NNS) complex which interacts with Pol II to specifically promote their termination (Schulz et al. 2013).

These lncRNAs identified in *S. cerevisiae* represent vast and pervasive groups of transcripts. 925 CUTs were identified by tiling array analysis of an Rrp6 deletion mutant (Xu et al. 2009). A second study identified 1496 CUTs using a 3' long-SAGE (serial analysis of gene expression) approach adapted to deep sequencing of an *S. cerevisiae* strain defective for both the nuclear specific exosome component Rrp6, as well as the TRAMP complex component Trf4 (Neil et al. 2009). A third study identified 1600 CUTs using tiling array analysis of a mutant with a fully inactivated exosome (a triple mutant defective for Rrp6 as well as both the exonucleolytic and endonucleolytic functions of the core exosome catalytic subunit Dis3) (Gudipati et al. 2012). Similar numbers of XUTs (1658) (van Dijk et al. 2011), and NUTs (1526) (Schulz et al. 2013) have been identified in *S. cerevisiae*.

However, these classes of lncRNAs are not distinct but overlap substantially, indicating that complex and partially redundant mechanisms likely regulate their expression. For example, ~600 NUTs substantially overlap annotated CUTs, and a further, but overlapping, group of ~600 NUTs are also annotated as XUTs (Schulz et al. 2013).

In this Chapter, I perform a similar classification of *S. pombe* lncRNAs – both those that are already annotated, as well as the extensive number identified in the current

study. I then consider the overlap between these classes, the differential expression of these classes under different physiological conditions, and the mechanisms underlying their observed regulated expression.

5.2 *S. pombe* CUTs, XUTs, RUTs

From the panel of RNA processing mutants analysed in Chapter 3, it is clear that there are three main pathways regulating lncRNA post-transcriptional expression and degradation in *S. pombe*: the nuclear exosome, the RNAi machinery, and the cytoplasmic exonuclease Exo2. Other cytoplasmic degradation pathways seem to play little or no role in lncRNA degradation Given that Exo2 is generally presumed to function in the cytoplasm, the exosome appears to be the main player mediating lncRNA degradation in the nucleus, but not in the cytoplasm where Exo2 takes over.

However, there is extensive overlap in the lncRNAs regulated by these three pathways (Figure 3.3, Chapter 3). To further probe the biological mechanisms underlying this observed overlap, I constructed double and triple mutants of three key genes (*dcr1*, *rrp6* and *exo2*) to determine their genetic interactions. The *dcr1* null mutant has the strongest lncRNA phenotype amongst the three RNAi mutants (*dcr1, rdp1, ago1*; Chapter 3), while Rrp6 is the nuclear-specific catalytic exosome component.

5.2.1 Creation of double and triple mutants

Double mutants of the *dcr1*, *rrp6* and *exo2* genes were created by crossing single mutants, and analysing the resulting tetrads by tetrad dissection (see Chapter 2 for details). Genotypes of single mutant strains used for matings are listed below and were derived from the single mutant strains listed in Table 2.1, Chapter 2:

h+ rrp6::URA ura-
h- dcr1::NAT ura-
h+ exo2::KAN ura-
h- exo2::KAN ura-

Tetrad dissection results are given in Tables 5.1-5.4 below, and example tetrad dissections are shown in Figure 5.1.

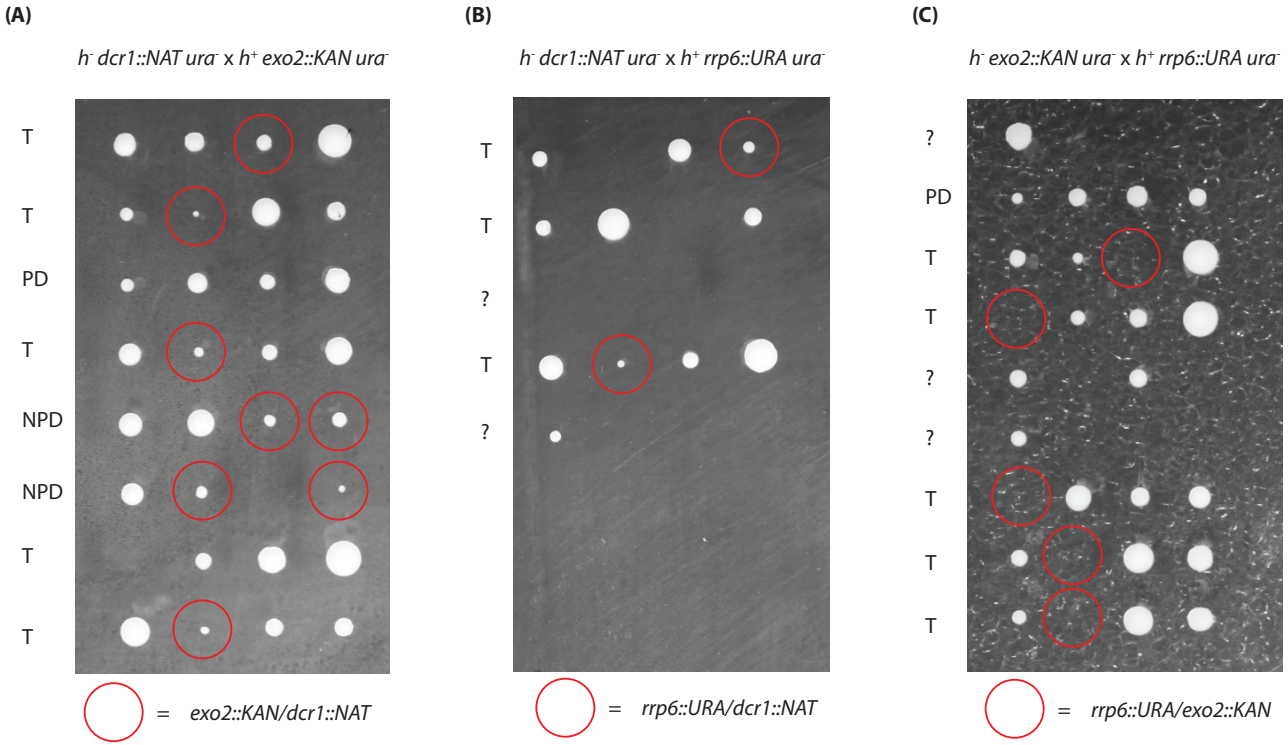

Figure 5.1: Tetrad analysis of double mutants.

(A)–(C): The indicated strains were crossed and the resulting tetrads dissected (see Chapter 2 for details). Each row represents a single tetrad dissection; plates were imaged after 5 days growth on rich medium; T = tetratype; PD = parental ditype; NPD = non-parental ditype; ? = indeterminable tetrad type due to insufficient spore survival in tetrad. Colonies arising from double mutant spores are as indicated.

h+ exo2::KAN ura- and *h- dcr1::NAT ura-* were crossed and 24 resulting tetrads analysed by tetrad dissection (Figure 5.1A, Table 5.1). Spore viability amongst these 24 tetrads was 85%, and there was no significant difference between observed and expected frequencies (calculated based on standard Mendelian ratios) of WT, single and double mutant spores (chi-squared test, p=0.93; calculated based on the number of surviving spores). Additionally, tetratype (T), non-parental ditype (NPD) and parental ditype (PD) tetrads were observed in the expected ratio 4:1:1. Thus the *exo2/dcr1* double mutant is viable, although the colony sizes arising from germinated haploid spores were consistently smaller for the *exo2/dcr1* double mutant than for either single mutant, or for WT spores (Figure 5.1A).

Table 5.1: Tetrad analysis of h- dcr1::NAT ura- x h+ exo2::KAN ura-

(n=96; 24 tetrads)

	Observed	**Expected**
Dead	14	0
WT	22	24
exo2::KAN	19	24
dcr1::NAT	19	24
exo2::KAN/dcr1::NAT	22	24

h+ rrp6::URA ura and *h- dcr1::NAT ura-* were crossed and 22 resulting tetrads analysed by tetrad dissection (Figure 5.1B, Table 5.2). Spore viability amongst these 22 tetrads was only 65% (lower than for the *h- dcr1::NAT ura-* x *h+ exo2::KAN ura-* cross), and there was no significant difference between observed and expected frequencies of WT, single and double mutant spores (chi-squared test, p=0.93; calculated based on the number of surviving spores). Additionally, tetratype, non-parental ditype and parental ditype tetrads were observed in the expected ratio 4:1:1. Thus the *rrp6/dcr1* double mutant is viable, although the colony sizes arising from germinated haploid spores were consistently smaller for the double mutant than for either single mutant, or for WT spores (Figure 5.1B). Additionally, there seems to be lower spore viability for this cross with the *rrp6* mutant.

Table 5.2: Tetrad analysis of h- dcr1::NAT ura- x h+ rrp6::URA ura-

(n=88; 22 tetrads)

	Observed	**Expected**
Dead	32	0
WT	14	22
dcr1::NAT	15	22
rrp6::URA	12	22
rrp6::URA/dcr1::NAT	15	22

h+ rrp6::URA ura and *h- exo2::KAN ura*-were crossed and 24 resulting tetrads analysed by tetrad dissection (Figure 5.1C, Table 5.3). Spore viability amongst these 24 tetrads was only 61%. Amongst the tetrads analysed, there were many examples where the only non-surviving spore was the *rrp6::URA exo2::KAN* double mutant (Figure 5.1C). Thus, there is a significant difference between observed and expected frequencies of WT, single and double mutant spores (chi-squared test, $p=10^{-4}$; calculated based on the number of surviving spores). Additionally, tetratype, non-parental ditype and parental ditype tetrads were observed in the expected ratio of 4:1:1. Thus the *rrp6/exo2* double mutant is most likely inviable. Spore viability was low and comparable to the spore viability seen in the *dcr1*x *rrp6* cross, thereby suggesting that deletion of *rrp6* adversely affects spore viability.

Table 5.3: Tetrad analysis of h+ rrp6::URA ura- x h- exo2::KAN ura-

(n=96; 24 tetrads)

	Observed	**Expected**
Dead	37	0
WT	21	24
exo2::KAN	18	24
rrp6::URA	20	24
exo2::KAN/rrp6::URA	0	24

Although the *rrp6/exo2* double mutant is inviable, the two viable double mutants created from the crosses described above (*h- rrp6::URA dcr1::NAT ura-, h+ exo2::KAN dcr1::NAT ura-*) were mated to determine whether any rescue phenotype could be observed which may result in a viable triple mutant *rrp6/exo2/dcr1*. Overall, 24 tetrads from the *h- rrp6::URA dcr1::NAT ura-* x *h+ exo2::KAN dcr1::NAT ura-* cross were analysed by tetrad dissection (Table 5.4). Spore viability amongst these 24 tetrads was only 17%, and no triple mutant spores were observed. Thus, there is a significant difference between observed and expected frequencies of single, double and triple mutant spores (chi-squared test, $p=10^{-4}$; calculated based on the number of surviving spores), and the triple mutant is most likely inviable. Additionally the much lower spore viability observed in this cross compared to all previously described crosses, suggests that *dcr1* and *exo2* may interact with *rrp6* to regulate expression of meiotic genes (*dcr1* x *exo2* cross has a normal spore viability of 85%; *rrp6* x *dcr1* and *rrp6* x *exo2* crosses have a lower spore viability of ~60%, consistent with known roles for *rrp6* in meiotic gene expression; *rrp6/dcr1* x *exo2/dcr1* cross has a spore viability of 17%, much lower than that observed in crosses involving a single *rrp6* mutant).

Table 5.4: Tetrad analysis of h- rrp6::URA dcr1::NAT ura- x h+ exo2::KAN dcr1::NAT ura- (n=96; 24 tetrads)

	Observed	**Expected**
Dead	80	0
dcr1::NAT	11	24
dcr1:: NAT rrp6::URA	5	24
dcr1::NAT exo2::KAN	0	24
dcr1::NAT exo2::KAN rrp6::URA	0	24

The growth rates of the *rrp6/dcr1* and *exo2/dcr1* double mutants were assessed by a spot assay (Figure 5.2B) and growth monitoring in a BioLector (Figure 5.2C, Table 5.5). From both these assays, it is clear that all single and double mutants have slow growth phenotypes. *dcr1* has a growth rate only slightly impaired relative to WT, while *rrp6/dcr1* has the most severe slow growth phenotype. Surprisingly the

exo2/dcr1 double mutant seems to grow faster than the *exo2* single mutant (Figure 5.2C) suggesting a possible suppression phenotype, which is reflected in the lncRNA phenotype described below in Section 5.2.2. However, the slow growth phenotype of these mutants is not directly related to the number of lncRNAs up-regulated in each mutant since *dcr1* has more lncRNAs up-regulated than *exo2* and yet grows faster.

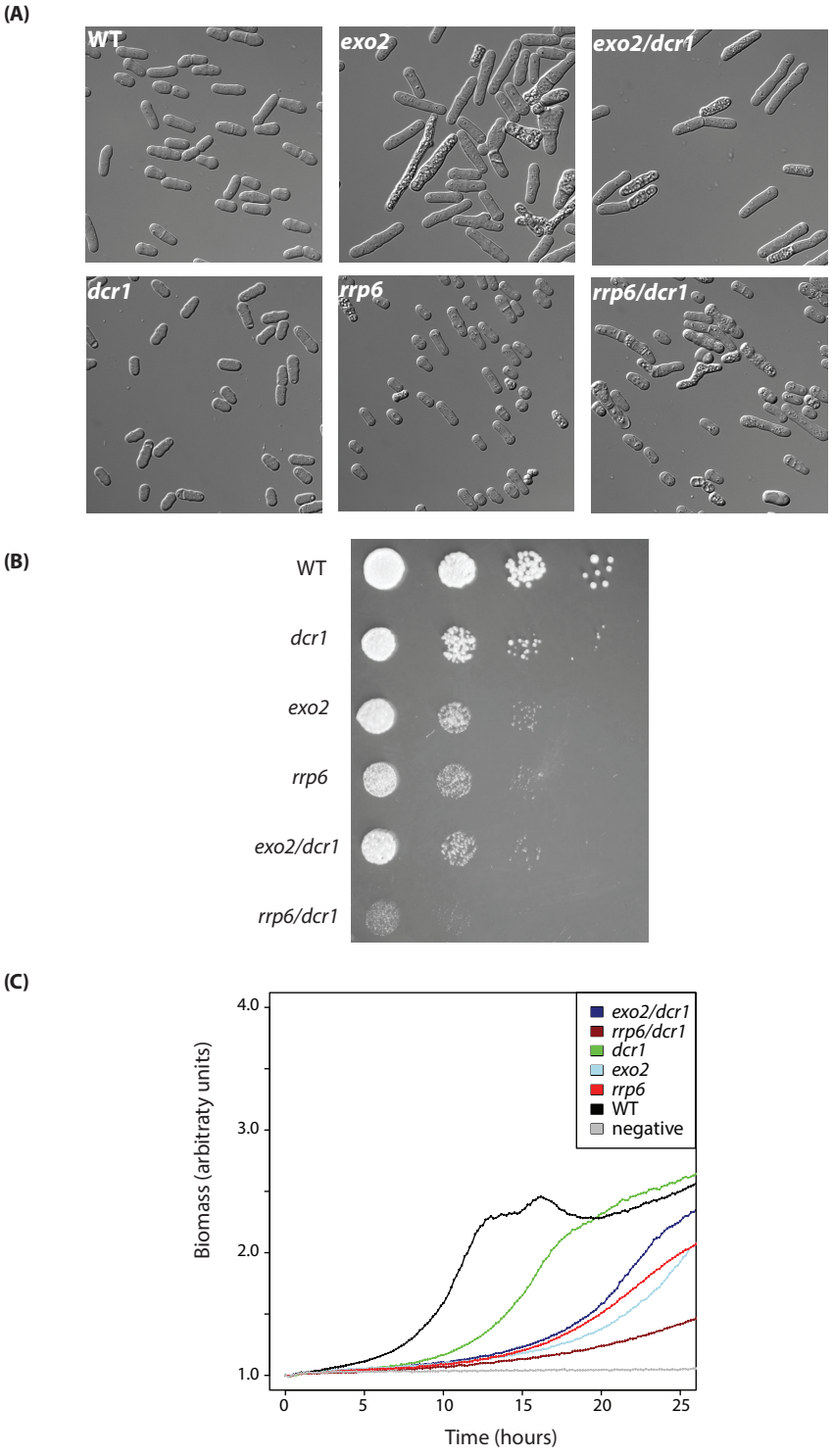

Figure 5.2: Phenotyping double mutants.

(A) Morphology: live cell imaging of mutants as indicated (see Chapter 2 for details); all cells were imaged in mid-log phase OD = 0.5. (B) Spot test for growth: four serial (tenfold) dilutions of cells were plated on rich medium and photographed after 3 days of growth at 32°C (see Chapter 2 for details). (C) Growth curves: strains were grown in triplicate in a BioLector and their biomass monitored by light spectroscopy (see Chapter 2 for details). Negative is media alone (no cells); for each strain the averages of three technical repeats is shown, and biomass is normalized to the 0h time-point.

Table 5.5: BioLector analysis of double mutants h- rrp6::URA dcr1::NAT ura- and h+ exo2::KAN dcr1::NAT ura-

Maximum growth rate = maximum biomass change in unit time during exponential phase; average growth rate = average biomass change in unit time during exponential phase. These measurements were taken over a 48-hour period.

Sample	**Maximum growth rate (Arbitrary units)**	**Average growth rate (Arbitrary units)**
WT control	0.29	0.11
dcr1	0.20	0.09
exo2/dcr1	0.18	0.08
exo2	0.17	0.07
rrp6	0.11	0.06
rrp6/dcr1	0.10	0.05
Negative control	0.02	0.00

Live cell imaging was used to assess the morphology of the *rrp6/dcr1* and *exo2/dcr1* double mutants (Figure 5.2A). The *exo2* and *exo2/dcr1* cells were long, consistent with the previously described increased cell size of the *exo2* mutant (Szankasi and Smith 1996), implicating a requirement for Exo2 in normal mitotic growth. The *rrp6* and *rrp6/dcr1* cells were small, with *rrp6/dcr1* cells exhibiting the most severe phenotype with many aberrant shapes and high cell death.

5.2.2 lncRNA phenotypes of double mutants

RNA-seq was used to determine the lncRNA phenotypes of the *rrp6/dcr1* and *exo2/dcr1* double mutants. Expression profiles of significantly differentially expressed (DE) mRNAs, SPNCRNAs and novel lncRNAs were examined by clustering their fold changes in expression relative to WT vegetative cells (Figure 5.3A; see Chapter 2 for details).

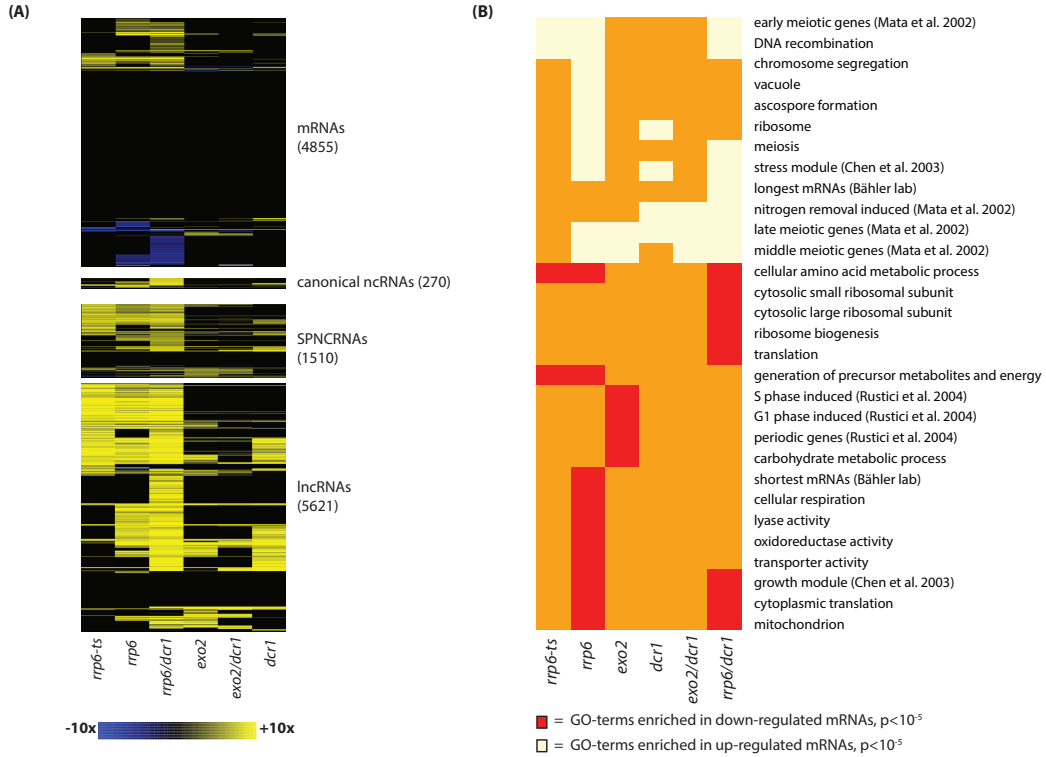

Figure 5.3: lncRNA phenotypes of double mutants.
(A) Expression profiles of significantly differentially expressed (DE) transcripts. Fold changes of transcripts relative to WT vegetative cells clustered in GeneSpring GX7 (Agilent); colour coded as indicated. Transcripts not significantly differentially expressed relative to WT (DESeq; p-value <0.01) have fold change set to 1 (appear black in heatmaps; see Chapter 2 for details). Transcripts with a fold change of 1 across all samples do not appear in these heatmaps. Columns represent different samples and rows represent transcripts; rows clustered using the Pearson correlation. Columns clustered using the up-regulated correlation based on lncRNAs only. (B) Gene ontology analysis of significantly differentially expressed mRNAs: significantly enriched GO-terms for mRNAs up-regulated (beige) or down-regulated (red) in each sample. GO-terms (rows) clustered by the "maximum" method in heatmap.2 function of R.

The *rrp6/dcr1* mutant shows a greater number of up-regulated lncRNAs than either single mutant (Figure 5.3A), indicating that *rrp6* and *dcr1* may act cooperatively. Such cooperation between RNAi and the exosome in lncRNA regulation has already been suggested by several studies in *S. pombe* (Gullerova and Proudfoot 2008, Lee et al. 2013, Shah et al. 2014, Zofall et al. 2009).

Surprisingly, the *exo2/dcr1* double mutant has fewer up-regulated lncRNAs than either single mutant (Figure 5.3A). This is consistent with the mild rescue phenotype in growth observed for the *exo2/dcr1* double mutant when compared to the *exo2* single mutant (Figure 5.2C). Since *rrp6*, *exo2* and *dcr1* seem to be key players in lncRNA regulation, a possible explanation for this observed lncRNA phenotype is that if the nuclear RNAi pathway is impaired, lncRNAs are channeled to the nuclear exosome. Thus, in the *rrp6/dcr1* double mutant, where the nuclear exosome pathway is also impaired, a greater number of lncRNAs are up-regulated than in either *dcr1* or *rrp6*. However, in the *exo2/dcr1* mutant, where the nuclear exosome is intact, defects in RNAi may efficiently channel lncRNAs to the nuclear exosome. This results in fewer lncRNAs reaching the cytoplasm and defective Exo2 therefore has less impact on lncRNA expression, leading to the observed rescue phenotype in *exo2/dcr1*.

The fact that impairing key nuclear lncRNA degradation pathways (either alone, *rrp6* and *dcr1*, or in combination, *rrp6/dcr1*) results in more severe lncRNA phenotypes than when the main cytoplasmic pathway is impaired (either alone, *exo2*, or in combination with a nuclear pathway *exo2/dcr1*) is consistent with a greater degree of lncRNA regulation occurring in the nucleus than in the cytoplasm.

Finally, the lethality of the *rrp6/exo2* mutant may suggest that while Exo2 is the key cytoplasmic player regulating lncRNAs, Rrp6 is the main nuclear player. Thus, although *rrp6* and *dcr1* appear to act cooperatively, impairing *dcr1* in an *exo2*-background produces a much less severe phenotype than impairing *rrp6* in an *exo2*-background, which is synthetically lethal.

Already annotated lncRNAs (SPNCRNAs) have similar but less pronounced expression signatures in the *exo2/dcr1* and *rrp6/dcr1* double mutants (Figure 5.3A),

when compared to lncRNAs identified in the current study. This is consistent with the observation in Chapter 3 that fewer SPNCRNAs are substrates of these RNA degradation/processing pathways.

In fact the most striking changes in expression signatures of the *exo2/dcr1* and *rrp6/dcr1* mutants, compared to the single mutants, appear to be amongst lncRNAs (Figure 5.3A). The relatively small changes in mRNA expression signatures upon comparison of single and double mutants, supports the notion that these pathways act specifically on lncRNAs. It is notable that the strongest mRNA expression signature phenotypes are observed for *rrp6* and *dcr1/rrp6*, mutants that also have the strongest lncRNA phenotypes.

Gene ontology (GO) analysis of differentially expressed mRNAs reveals up-regulation of the stress module, and down-regulation of the growth module (Figure 5.3B). This is consistent with GO-enrichments amongst differentially expressed mRNAs in single mutants (Chapter 3).

mRNAs up-regulated in *rrp6/dcr1* but not in either single mutant are enriched for the following GO-terms ($p < 10^{-6}$):

RNA binding
cytoplasm
longest mRNAs (Bähler lab)
nitrogen removal induced (Mata et al. 2002)
stress module (Chen D. et al. 2003)

mRNAs down-regulated in *rrp6/dcr1*, but not in either single mutant, are much more strongly enriched for the following GO-terms ($p < 10^{-16}$):

cytoplasmic translation ($p < 10^{-75}$)
translation ($p < 10^{-57}$)
cytosolic large ribosomal subunit ($p < 10^{-47}$)
growth module (Chen D. et al. 2003) ($p < 10^{-45}$)
cytosolic small ribosomal subunit ($p < 10^{-36}$)
ribosome biogenesis ($p < 10^{-16}$)

This indicates a strong down-regulation of growth-related functions in the *rrp6/dcr1* mutant and fits with the severe slow growth phenotype observed when compared to either single mutant (Figure 5.2C).

There are no GO-terms enriched amongst mRNAs up-regulated in *exo2/dcr1* but not in either single mutant. However, mRNAs down-regulated in *exo2/dcr1* but not in either single mutant are mildly enriched for the following GO-terms ($p < 10^{-3}$):

cellular amino acid metabolic process
oxidoreductase activity

Finally, it is striking that the *rrp6/dcr1* double mutant shows a large number of canonical ncRNAs (tRNAs, rRNA, snRNAs, snoRNAs) strongly up-regulated – more than either single mutant (Figure 5.3A) - hinting at a possible cooperative role between the exosome and RNAi in regulating expression of these canonical ncRNAs (see Chapter 3).

5.2.3 Defining CUTs, RUTs, XUTs

Given that the three main pathways regulating lncRNA expression in *S. pombe* appear to be the nuclear exosome, the RNAi pathway and the cytoplasmic exonuclease Exo2, three groups of lncRNAs, termed CUTs, RUTs and XUTs, were defined in *S. pombe*.

Thus, in analogy to *S. cerevisiae* CUTs, *S. pombe* CUTs (cryptic unstable transcripts) were defined to be those lncRNAs up-regulated upon nuclear exosome impairment. CUTs were taken to be the group of lncRNAs up-regulated in both the *rrp6* and *rrp6-ts* mutant. The *rrp6* mutant has a stronger lncRNA phenotype than the *rrp6-ts* mutant (3891 lncRNAs up-regulated in *rrp6*, compared to 3042 in *rrp6-ts*, most likely due to incomplete inactivation of Rrp6 activity in the *rrp6-ts*). There are 2506 *S. pombe* CUTs (both novel and already annotated lncRNAs).

RUTs (RNAi-dependent unstable transcripts) were defined to be those lncRNAs up-regulated in the *dcr1* mutant. There are 2343 *S. pombe* RUTs (both novel and

already annotated lncRNAs). Unsurprinsgly, no analogous group of lncRNAs has yet been described in *S. cerevisiae* as the RNAi pathway is missing in this yeast.

Finally XUTs (Xrn1-dependent unstable transcripts) were defined to be those lncRNAs up-regulated in the *exo2* mutant. This group is analogous to *S. cerevisiae* XUTs, which are targeted for degradation by cytoplasmic exonuclease Xrn1, which is the homologue of *S. pombe* Exo2. There are 1888 *S. pombe* XUTs (both novel and already annotated lncRNAs).

There is extensive overlap amongst these 3 groups of *S. pombe* CUTs, RUTs and XUTs, with ~600 lncRNAs belonging to all 3 classes (Figure 5.4A). Notably, the least degree of overlap occurs between CUTs and XUTs. This suggests that the nuclear exosome and the cytoplasmic exonuclease Exo2 may represent the most distinct pathways, with Exo2 being the key cytoplasmic player regulating lncRNAs, and Rrp6 the main nuclear player. This finding is consistent with the synthetic lethality observed for the double mutant *rrp6/exo2*.

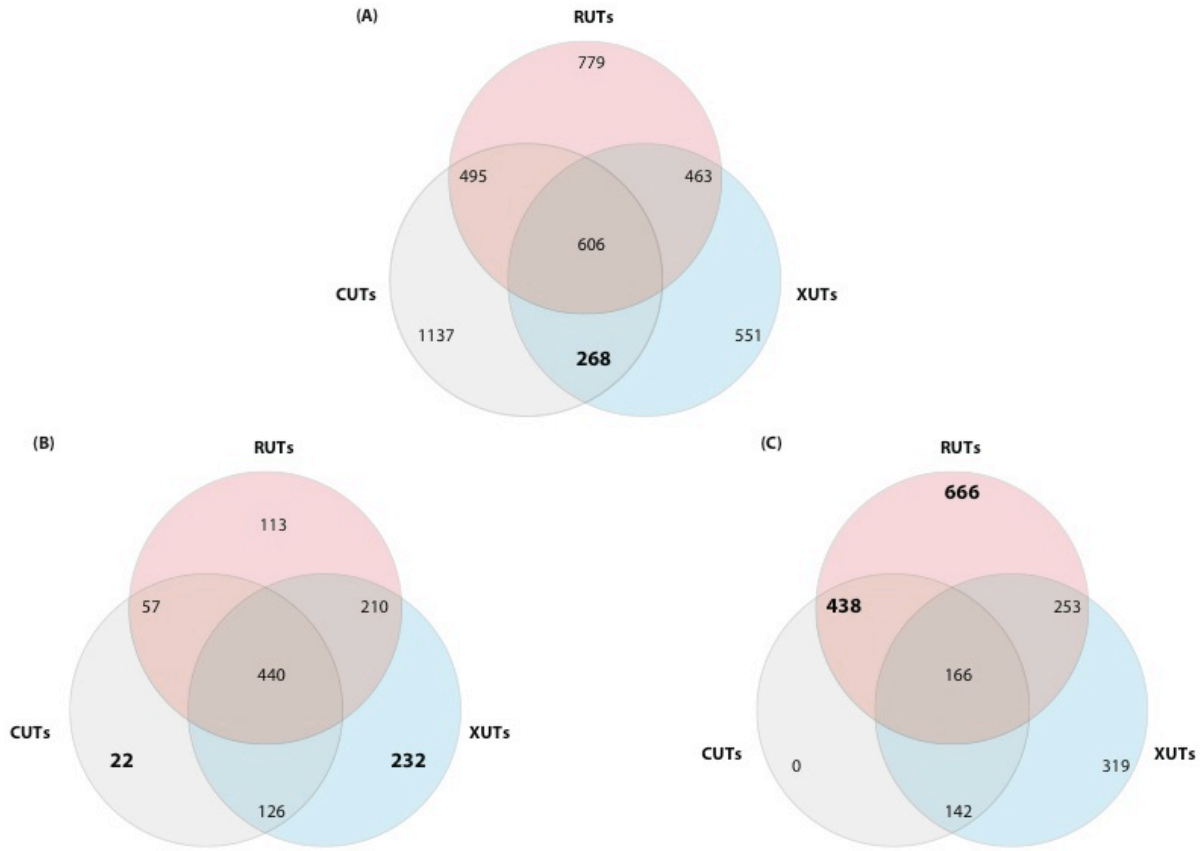

Figure 5.4: Overlap of CUTs, RUTs and XUTs.
Venn diagrams representing the overlap amongst CUTs, RUTs and XUTs. Areas of circles are not proportional to the number of lncRNAs indicated. Highlighted numbers in bold are discussed in the text. (A) Overlap of CUTs, RUTs and XUTs. (B) Overlap of CUTs, RUTs, and XUTs up-regulated in the *exo2/dcr1* mutant. (C) Overlap of XUTs and RUTs which are not up-regulated in *exo2/dcr1*.

As would be expected, the vast majority (98%) of lncRNAs up-regulated in *exo2/dcr1* are either XUTs and/or RUTs (Figure 5.4B), with a slight over-representation of XUTs. As described above, the *exo2/dcr1* mutant has fewer lncRNAs up-regulated than either single mutant. Thus ~2000 lncRNAs are XUTs and/or RUTs but are no longer up-regulated upon double deletion. These ~2000 XUTs and/or RUTs which are no longer up-regulated in *exo2/dcr1* are predominantly RUTs (Figure 5.3A and Figure 5.4C). This is consistent with the idea that, in the *exo2/dcr1* mutant, lncRNAs that may otherwise have been targeted to the RNAi pathway are channeled to the nuclear exosome.

In summary, the nuclear exosome, the RNAi pathway and the cytoplasmic exonuclease Exo2 represent three key pathways regulating lncRNAs in *S. pombe*, defining the overlapping classes of CUTs, RUTs and XUTs, respectively. The nuclear exosome and the RNAi pathway act cooperatively to control nuclear lncRNA expression, while the cytoplasmic Exo2 pathway is more distinct.

5.3 Regulation of CUTs, XUTs, RUTs in response to external stimuli

5.3.1 Late meiotic, early meiotic and quiescent CUTs, RUTs and XUTs

An emerging theme amongst lncRNAs is their regulated expression in response to environmental cues (Bierhoff et al. 2014, Lardenois et al. 2011, Leong et al. 2014).

Expression profiles of lncRNAs across both genetic and environmental perturbations (Figure 3.3D, Chapter 3) revealed that lncRNAs are stabilised under physiologically relevant growth conditions. Furthermore, these profiles indicated that quiescent lncRNAs (nitrogen or glucose starvation and early meiosis) are regulated predominantly by the exosome. In contrast, the RNAi pathway and the cytoplasmic exonuclease Exo2 appear to play a greater role in the regulation of lncRNAs up-regulated in later meiotic time-points.

Fisher tests were therefore performed to determine whether, for each physiological condition, there is a statistically significant association between lncRNAs being up-regulated in a given condition and being a CUT, XUT or RUT (Figure 5.5A).

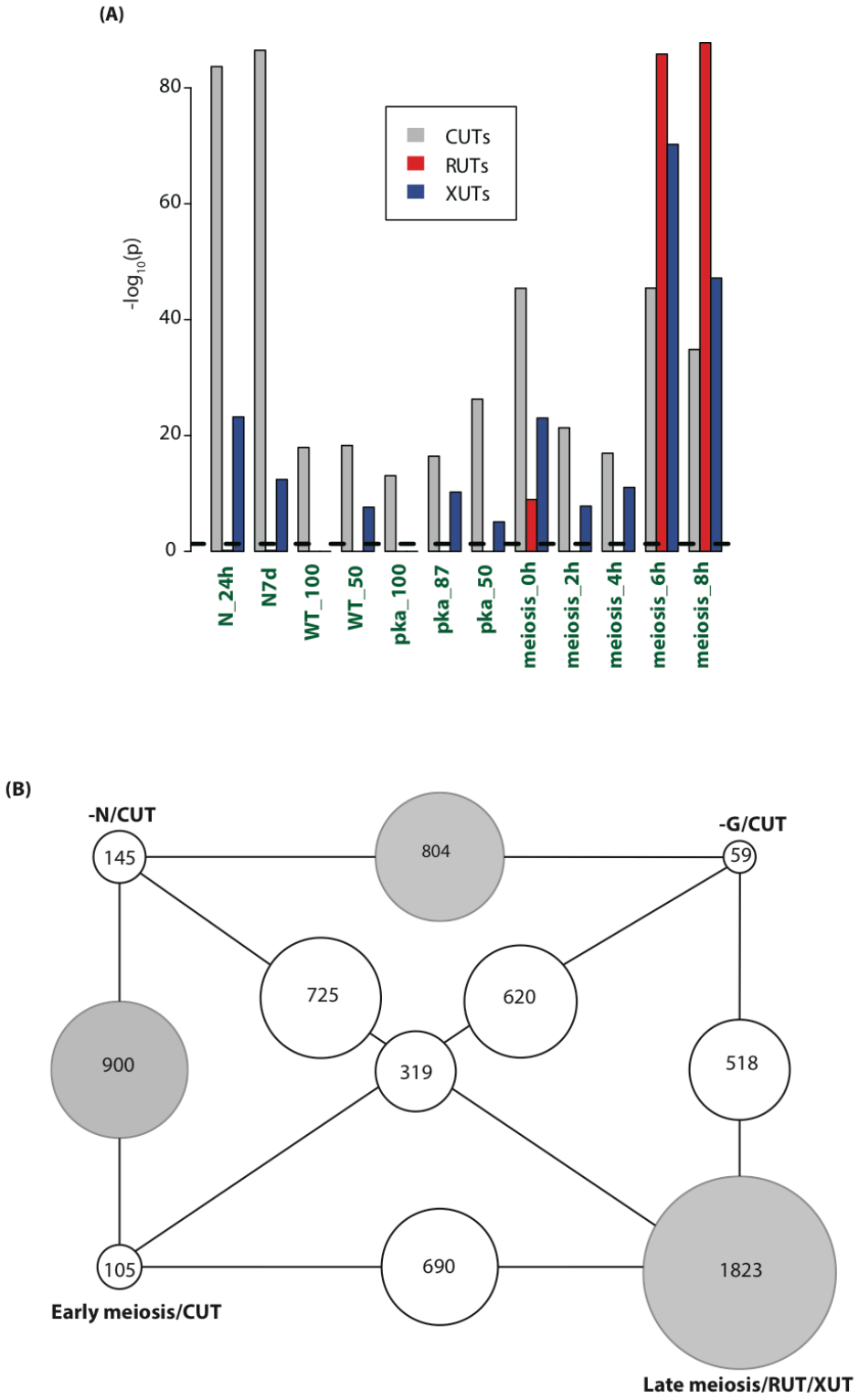

Figure 5.5: Regulation of CUTs, RUTs and XUTs in response to external stimuli.

(A) Enrichments of CUTs, RUTs and XUTs under different physiological growth conditions. Dotted line is $-\log_{10}(p = 0.05)$. (B) Overlap of –N/CUTs, -G/CUTs, early meiotic CUTs and late meiotic RUT/XUTs (see text for details). Area of circles is proportional to number of lncRNAs indicated.

These Fisher tests revealed CUTs to be the most significantly enriched class under all conditions except the late meiotic time-points of 6h and 8h. In contrast, RUTs are strongly significantly enriched only at these late meiotic time-points of 6h and 8h. XUTs are enriched in most conditions, although to a lesser extent than CUTs, except for the late meiotic time-points (6h and 8h) where their enrichment exceeds that of CUTs (Figure 5.5A).

These enrichments of CUTs, XUTs and RUTs under different growth conditions reinforce that quiescent lncRNAs (nitrogen or glucose starvation and early meiosis) are regulated predominantly by the exosome, while the RNAi pathway and the cytoplasmic exonuclease Exo2 play a greater role in the regulation of lncRNAs up-regulated in later meiotic time-points.

Four groups of lncRNAs were therefore defined:

RUTs and/or XUTs up-regulated in late meiosis (6h and/or 8h)
CUTs up-regulated in early meiosis (at least one of 0h, 2h and 4h)
CUTs up-regulated in nitrogen starvation (24h and/or 7 days)
CUTs up-regulated in glucose starvation (at least one time-point)

From the overlap of these 4 groups (Figure 5.5B) it is clear that late meiotic RUTs/XUTs represent the most distinct group, with 1823 lncRNAs specific to this group only – this group of 1823 late meiotic RUTs/XUTs was therefore retained for further analysis.

Amongst CUTs up-regulated in early meiosis (early meiosis/CUT), nitrogen starvation (-N/CUT), and glucose starvation (-G/CUT), early meiotic CUTs and glucose starvation CUTs have the least overlap (Figure 5.5B). Quiescent-specific CUTs were therefore taken to be the 804 lncRNAs present in both –N/CUT and -G/CUT groups. Early meiosis specific CUTs were taken to be the 900 lncRNAs present in both early meiosis/CUT and –N/CUT groups (Figure 5.5B).

The expression profiles of these 3 groups (late meiotic RUTs/XUTs; early meiotic CUTs; quiescent CUTs) are shown in Figure 5.6B-D, along with the expression profile of the full set of novel lncRNAs (Figure 5.6A). Late meiotic RUTs/XUTs (Figure 5.6B) overlap strongly with cluster ‘2’ picked out by eye in Chapter 3

(Figure 3.3D). Quiescent CUTs (Figure 5.6D) and early meiotic CUTs (Figure 5.6C) broadly overlap with cluster '1' picked out by eye in Chapter 3 (Figure 3.3D), but are less well defined by this cluster.

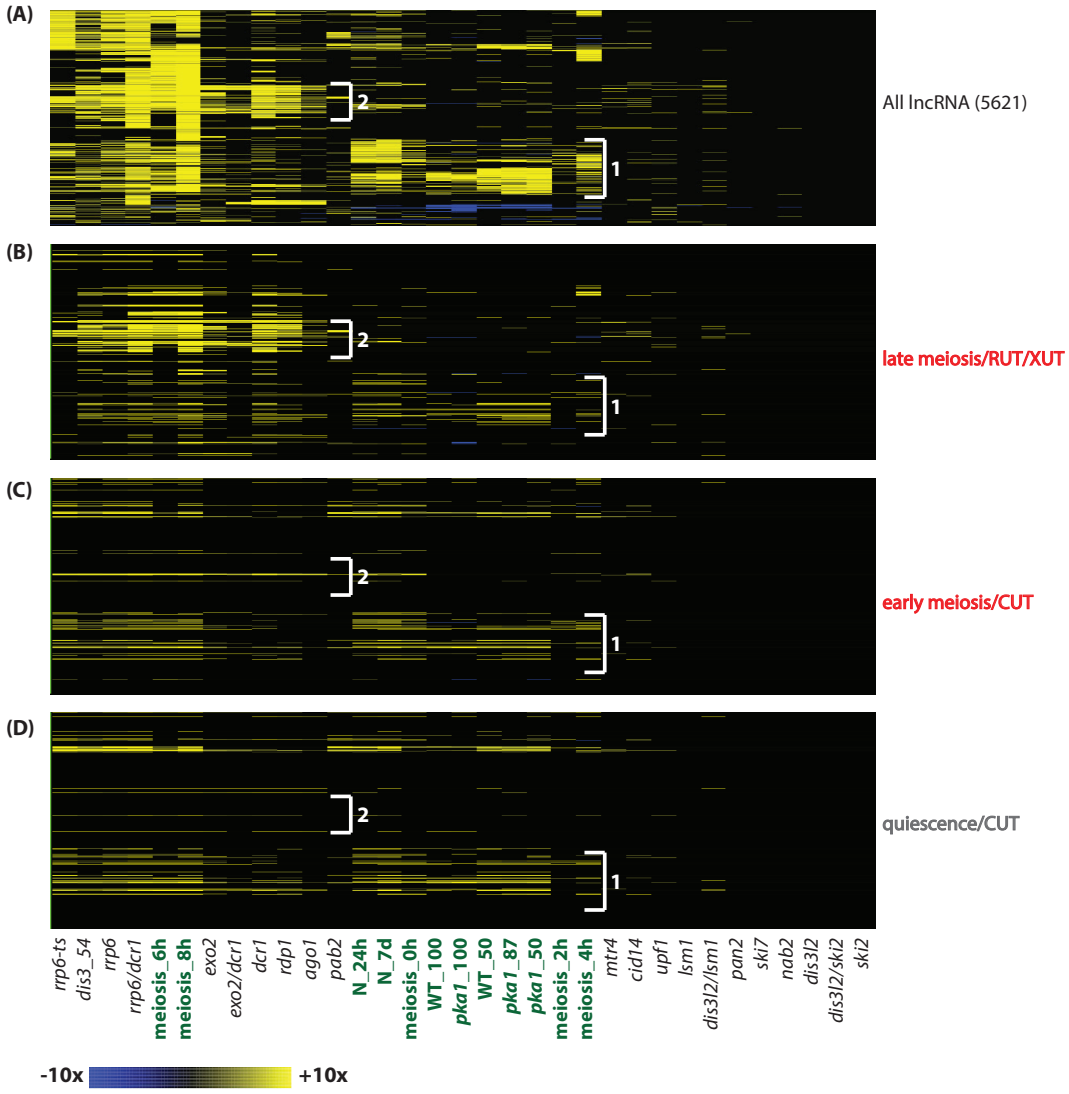

Figure 5.6: Expression profiles of late meiotic RUTs/XUTs, early meiotic CUTs and quiescent CUTs.

See legend for Figure 5.3A. White brackets highlight clusters of lncRNAs described in the text.

5.3.2 Location analysis of three groups

Analyses were performed to determine whether these three groups (late meiotic RUTs/XUTs; early meiotic CUTs; quiescent CUTs) are enriched for lncRNAs with a particular location with respect to mRNAs (Figure 5.7 and Figure 5.8).

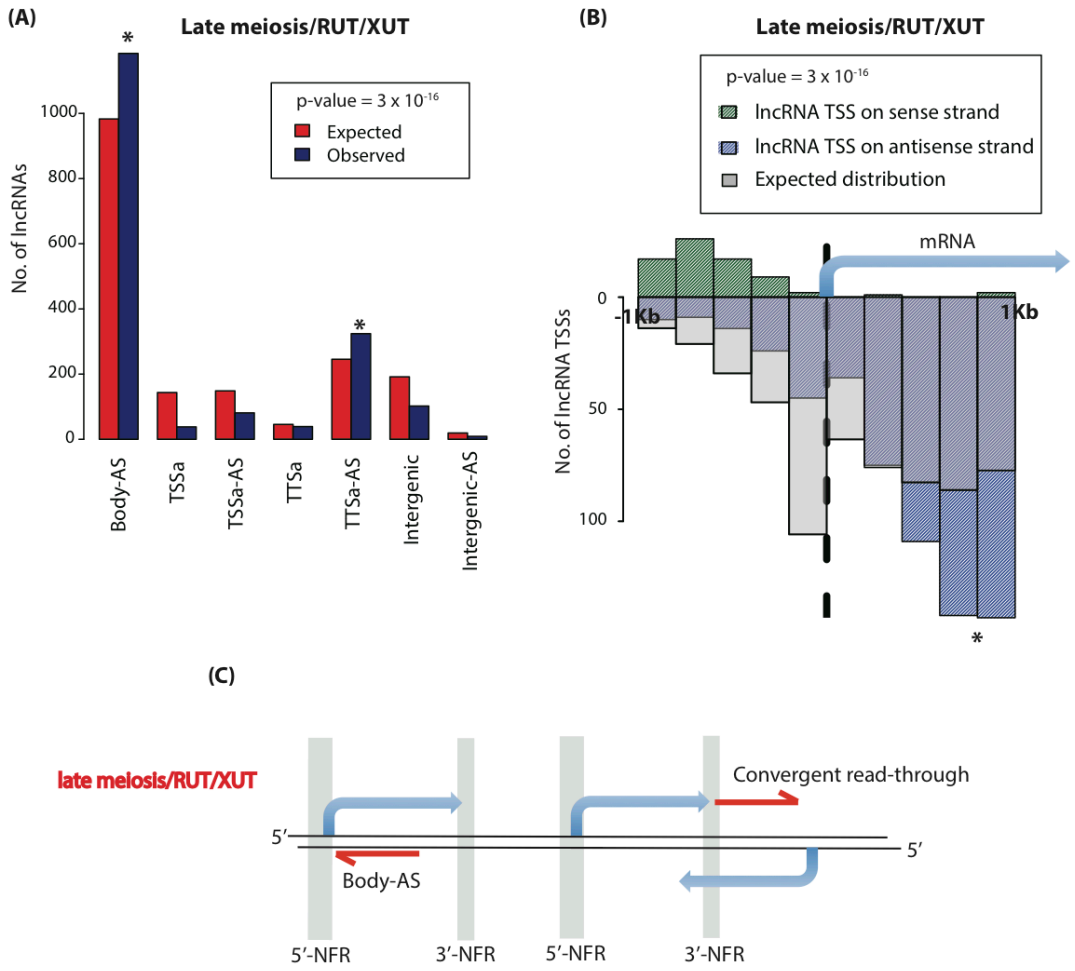

Figure 5.7: Location of late meiotic RUTs/XUTs.

(A) Distribution by type of late meiotic RUT/XUTs (red bars) as compared to the distribution expected from the full set of lncRNAs (blue bars). Chi-squared test used to compare expected and observed distributions. Astrices indicate key enrichments. (B) Distribution of late meiotic RUT/XUTs relative to mRNA TTSs. Dotted vertical line indicates reference mRNA TTS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. Distribution is compared to that expected from the full set of lncRNAs on the antisense strand (light grey bars) using a chi-squared test. (C) Schematic of the location of late meiotic RUT/XUTs.

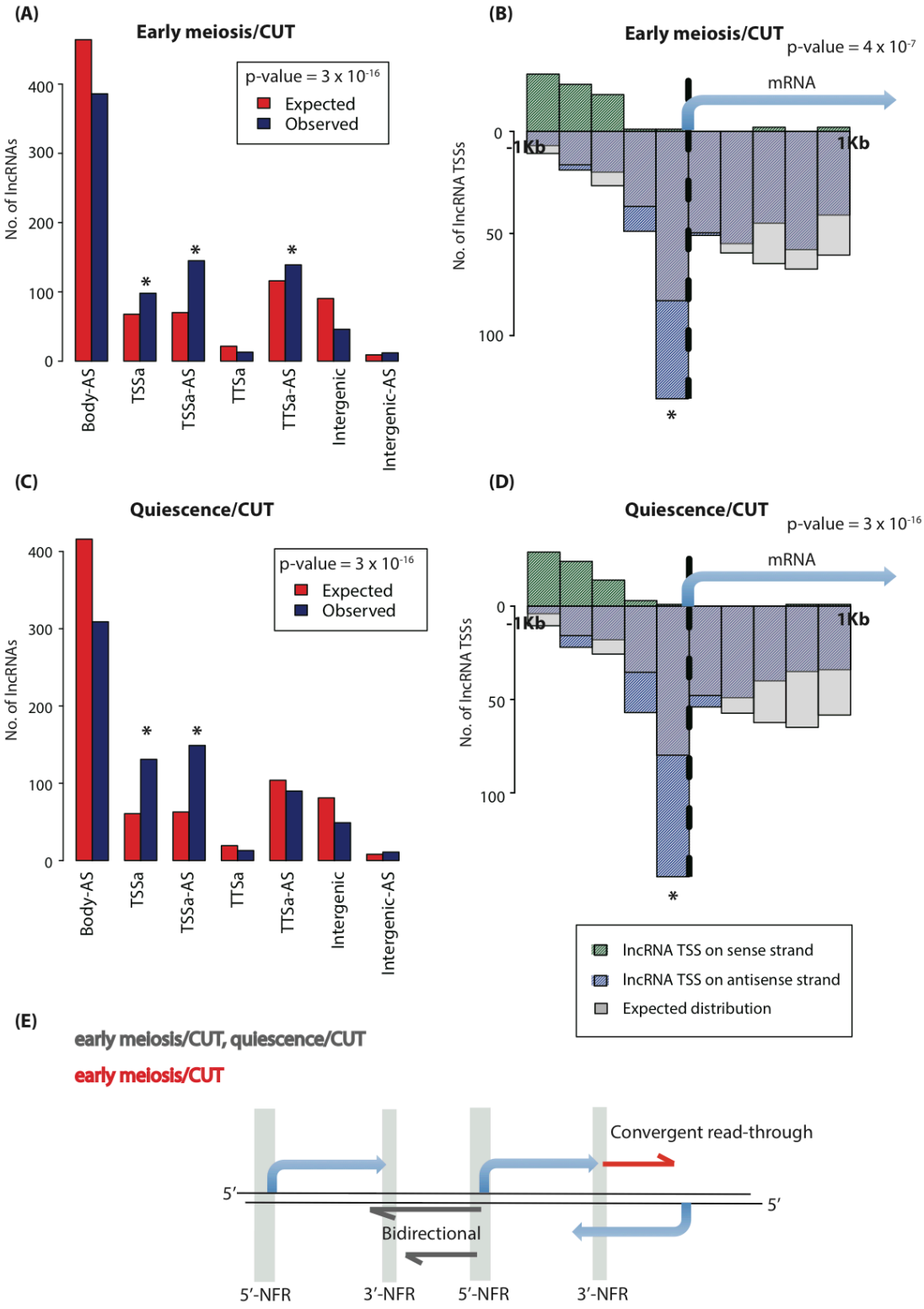

Figure 5.8: Location of early meiotic CUTs and quiescent CUTs.

(A) Distribution by type of early meiotic CUTs (red bars) as compared to the distribution expected from the full set of lncRNAs (blue bars). Chi-squared test used to compare expected and observed distributions. Astrices indicate key enrichments. (B) Distribution of early meiotic CUTs relative to mRNA TTSs. Dotted vertical line indicates reference mRNA TTS; black bars are number of lncRNA TSSs on sense strand; blue bars are number of lncRNA TSSs on antisense strand. Distribution is

compared to that expected from the full set of lncRNAs on the antisense strand (light grey bars) using a chi-squared test. (C) As for (A) but for quiescent CUTs. (D) As for (B) but for quiescent CUTs. (E) Schematic of the location of early meiotic CUTs and quiescent CUTs.

Based on their location relative to mRNAs, mutually exclusive categories for lncRNAs were defined in Chapter 4. These include:

Transcription start site associated (TSSa): the lncRNA originates upstream and antisense to an mRNA transcription start site. These can be considered to be bidirectional lncRNAs, i.e. lncRNAs transcribed divergently from mRNA TSSs (Figure 5.8E).

Transcription start site associated antisense (TSSa-AS): as for TSSa but the lncRNA additionally runs antisense to an upstream gene (Figure 5.8E). These can also be considered to be bidirectional lncRNAs (Figure 5.8E).

Transcription termination site associated (TTSa): the lncRNA originates within the 3'-(nucleosome free region)-NFR associated with an mRNA transcription termination site (TTS).

Transcription termination site associated antisense (TTSa-AS): as for TTSa but the lncRNA additionally runs antisense to either the gene from whose 3'-NFR it originates (TTSa-AS-own) or antisense to a downstream gene in a convergent orientation (convergent read-through; Figure 5.7C).

Body antisense (body-AS): the lncRNA originates and runs antisense to the body of an mRNA (Figure 5.7C).

Intergenic: the lncRNA originates in an intergenic region (no mRNA annotation on either strand, and not sufficiently close to an mRNA TSS or TTS to be considered as originating from a 3'- or 5'-NFR).

Intergenic-antisense(AS): as for intergenic but the lncRNA additionally runs antisense to an adjacent gene.

Late meiotic RUTs/XUTs, early meiotic CUTs and quiescent CUTs were classified according to the definitions above, and their distribution amongst these classes compared to the expected distribution derived from the full set of 7332 lncRNAs. Additionally, the distribution of these groups of CUTs, RUTs and XUTs relative to mRNA TSSs was compared to the expected distributions derived from the full set of 7332 lncRNAs (Figure 5.7 and Figure 5.8; see Chapters 2 and 4 for further for details of methods).

Late meiotic RUTs/XUTs are enriched for antisense lncRNAs, specifically those originating antisense to the body of mRNAs (body-AS) and those arising from convergent read-through (TTSa-AS; 90% of late meiotic XUTs/RUTs in this category are from convergent read-through as opposed to TTSa-AS-own) (Figure 5.7A, C). Consistent with this, the distribution of late meiotic RUTs/XUTs around mRNA TSSs shows enrichments of lncRNAs initiating antisense to the body of mRNAs, and a depletion of lncRNAs originating antisense and 0-200bp upstream of mRNA TSSs (these can be considered as arising from divergent transcription) (Figure 5.7B). Notably, all these enrichments are stronger and more significant than for the full set of lncRNAs up-regulated in later meiotic time-points (Chapter 4, Figure 4.5). This suggests that late meiotic lncRNAs regulated by RNAi and/or Exo2 have a highly specific location with respect to mRNAs, more so than for the full set of late meiotic lncRNAs.

Quiescent CUTs are enriched for bidirectional lncRNAs (TSSa and TSSa-AS) (Figure 5.8C,E). Consistent with this, the distribution of quiescent CUTs around mRNA TSSs shows an enrichment of lncRNAs originating antisense and 0-200bp upstream of mRNA TSSs (these can be considered as arising from divergent transcription) and depletion of lncRNAs initiating antisense to the body of mRNAs (Figure 5.8D). Notably, all these enrichments are stronger and more significant than for the full set of lncRNAs up-regulated at any nitrogen or glucose starvation time-point (Chapter 4, Figure 4.3). This suggests that quiescent lncRNAs regulated by the nuclear exosome have a highly specific location with respect to mRNAs, more so than for the full set of quiescent lncRNAs.

Similar to quiescent CUTs, early meiotic CUTs are also enriched for bidirectional lncRNAs (Figure 5.8A-B, E), and, to a lesser extent, for antisense lncRNAs arising from convergent read-through (TTSa-AS; 70% of early meiotic CUTs in this category are from convergent read-through as opposed to TTSa-AS-own) (Figure 5.8A). This is supported by the distribution of early meiotic CUTs around mRNA TSSs, which shows an enrichment of divergent lncRNAs (Figure 5.8B). This distribution also shows a depletion of lncRNAs initiating antisense to the body of mRNAs, but this antisense depletion is less pronounced than for quiescent CUTs (Figure 5.8B,D). Once again, these enrichments are stronger and more significant than for the full set of lncRNAs up-regulated at early meiotic time-points (Chapter 4, Figure 4.5). This suggests that early meiotic lncRNAs regulated by the nuclear exosome have a highly specific location with respect to mRNAs, more so that for the full set of early meiotic lncRNAs.

In summary, late meiotic RUT/XUTs are enriched for antisense lncRNAs originating either antisense to the body of mRNAs or arising from convergent read-through; early meiotic CUTs are enriched for bidirectional lncRNAs and, to a lesser extent, for antisense lncRNAs arising from convergent read-through; quiescent CUTs are enriched for bidirectional lncRNAs. Interestingly, an enrichment of convergent read-through antisense lncRNAs seems to distinguish meiosis from quiescence.

It was of interest to determine how these lncRNAs correlate in expression with their associated mRNAs. Analyses were performed as in Chapter 4, and expression correlations followed a similar trend as in Chapter 4 – bidirectional lncRNA:mRNA pairs positively correlated in expression, while antisense lncRNA:mRNA pairs negatively correlated in expression.

However, these trends were no stronger or more significant for these sub-groups than for the location groups as a whole (for example, the correlation coefficient for bidirectional lncRNA:mRNA pairs involving quiescent CUTs was of a similar strength and significance as that for bidirectional lncRNA:mRNA pairs involving any lncRNA). This is likely due to a reduction in power of statistical testing accompanying the smaller sample sizes of late meiotic RUTs/XUTs, early meiotic CUTs and quiescent CUTs. Therefore, instead of comparing to the whole set of

lncRNA:mRNA pairs, a more appropriate control was a random set of the same number. So, for example, the expression correlation of bidirectional lncRNA:mRNA pairs involving quiescent CUTs was compared to a random set *of the same number* of bidirectional lncRNA:mRNA pairs involving any lncRNA. Consistently, the trends seen for the specific groups were stronger and more significant than those seen for the random group of the same number.

This finding indicates that sub-setting lncRNAs in this way (late meiotic RUTs/XUTs, early meiotic CUTs and quiescent CUTs) identifies lncRNAs with specific relationships (in terms of location and expression) with respect to mRNAs. Such lncRNA groups represent candidates for experimental functional follow-up.

5.4 Expression of RNA processing genes in physiological conditions

Since CUTs, XUTs and RUTs are up-regulated in *S. pombe* upon impairment of the nuclear exosome, the cytoplasmic exonuclease Exo2 and the RNAi pathway respectively, it was of interest to determine whether their up-regulation in different physiological conditions is due to a similar impairment of such RNA processing pathways.

5.4.1 Expression of rrp6, dcr1 and exo2 in physiological conditions

To determine whether transcriptional down-regulation of *rrp6*, *dcr1* and *exo2* accompanies enrichment of CUTs, RUTs and XUTs in physiological conditions, the expression profiles of these three genes across all samples were examined (Figure 5.9A). These profiles showed *rrp6*, *dcr1* and *exo2* transcripts to be knocked-out in these mutants as expected. It is of note that in the *mtr4* mutant *dcr1* is up-regulated, while in the *lsm1* mutant *rrp6* is down-regulated. This reflects possible interactions between RNA processing pathways and will be discussed in more detail in Section 5.4.2.

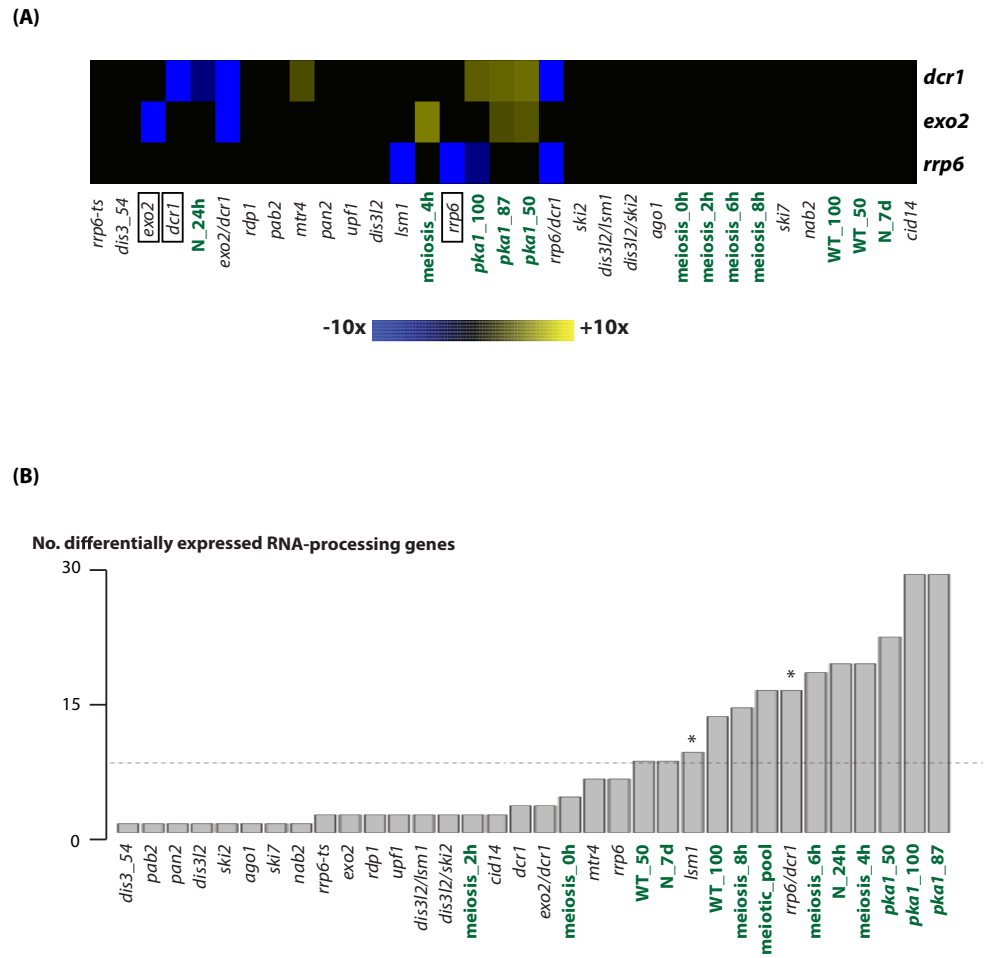

Figure 5.9: Expression of RNA processing genes in physiological conditions.

(A) Expression profiles of *dcr1*, *exo2* and *rrp6* across all samples. See legend for Figure 5.3A for details. (B) Barplot of the number of significantly differentially expressed RNA processing genes (from Table 5.6) across all samples.

However, Figure 5.9A shows relatively little significant differential expression of *rrp6*, *dcr1* and *exo2* in physiologically relevant growth conditions. It may be expected that significant down-regulation of the transcripts of these three genes would accompany the observed up-regulation of CUTs, XUTs and RUTs in physiological conditions. Contrary to this, significant up-regulation of *exo2* was seen in meiosis_4h, *pka1*_87, *pka1*_50; and significant up-regulation of *dcr1* was seen in all glucose starvation time-points of the *pka1* mutant (Figure 5.9A). The only significant down-regulation of these three RNA processing genes was: down-regulation of *dcr1* in N_24h; down-regulation of *rrp6* in *pka1*_100 (Figure 5.9A).

Transcriptional down-regulation of exosome components at progressive time-points of meiosis has been observed (Danny Bitton, personal communication). Since the expression profile of Figure 5.9A only considers transcripts which are called as significantly differentially expressed relative to WT log-phase cells (using the DESeq package; see Chapter 2 for details), it is possible that *dcr1*, *rrp6* and *exo2* are down-regulated in physiological conditions, but this down-regulation is too subtle to be called significant by DESeq. To explore this possibility, normalized read counts of these three transcripts in different physiological conditions were plotted (Figure 5.10).

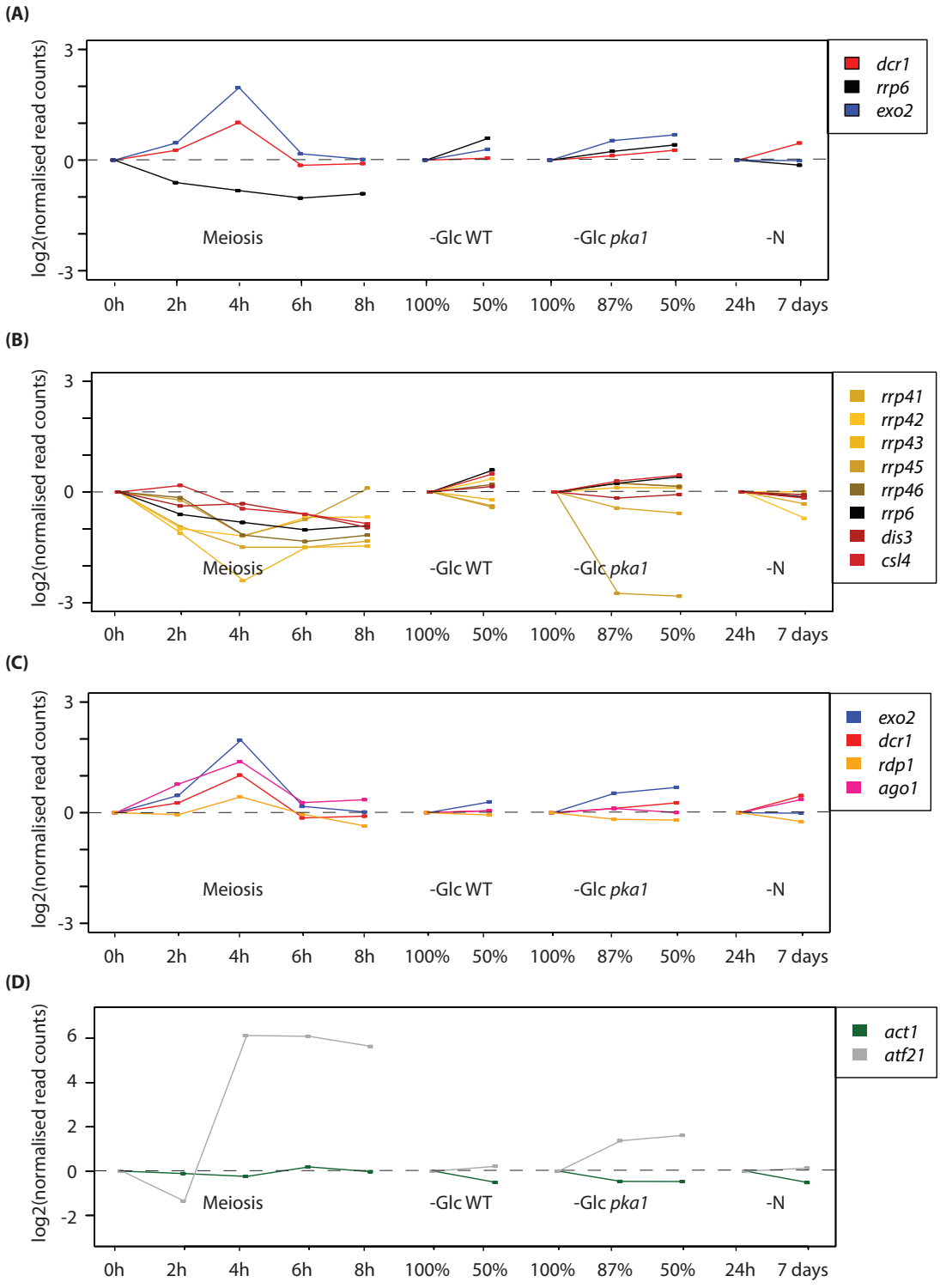

Figure 5.10: Expression profiles of key transcripts as a function of meiotic progression, or survival under glucose or nitrogen starvation.

(A)–(D) Profiles of transcripts as indicated. Normalised read counts from DESeq are plotted; mean expression of two biological replicates is shown at each time-point; for each profile, each time-point is normalized to the first time-point. –Glc = glucose starvation; -N = nitrogen starvation.

Progressive down-regulation of *rrp6* occurs throughout the course of meiosis (Figure 5.10A). While this transcriptional down-regulation is too subtle to be called as statistically significant by DESeq, other exosome components follow a similar trend (Figure 5.10B). Such transcriptional down-regulation of the exosome may explain the observed enrichment of CUTs in meiosis (Figure 5.5A). Interestingly, *exo2* and *dcr1* transcripts appear to be up-regulated during early meiosis, peak in expression at 4h and then rapidly drop in expression (Figure 5.10A). Again, although these changes in expression are not all called as significant by DESeq, other RNAi components follow a similar trend (Figure 5.10C), suggesting that these changes in expression may be regulated. Such a drop in *exo2* and RNAi machinery expression at later meiotic time-points may explain the strong enrichments of XUTs and RUTs at these stages of meiosis (Figure 5.5A). The patterns of expression of *rrp6* (and other exosome components), *dcr1* (and other RNAi components) and *exo2* during meiosis are distinct from the housekeeping gene *act1*, whose expression remains relatively constant throughout the course of meiosis, as well as being distinct from the meiotic transcript *atf21* whose expression rises rapidly during meiosis and remains high (Figure 5.10D).

Patterns of expression of *rrp6* (and other exosome components), *dcr1* (and other RNAi components) and *exo2* at progressive nitrogen and glucose starvation time-points are much less coherent and pronounced (Figure 5.10A-C), and are similar to the fluctuations seen for the house-keeping control *act1*, whose expression would be expected to remain relatively constant during nitrogen and glucose starvation (Figure 5.10D). This may suggest that a complex interplay of altered RNA processing activities underlies the observed enrichments, particularly of CUTs, in nitrogen and glucose starvation (Figure 5.5A).

5.4.2 Expression of a panel of RNA processing genes in physiological conditions

As described in Section 5.4.1, exosome, RNAi component, and *exo2* transcripts may be subtly down-regulated at appropriate meiotic time-points leading to appearance of CUTs, RUTs and XUTs. However, any altered RNA processing activities of these factors underlying the observed enrichment of CUTs during glucose and nitrogen starvation is less clear.

Therefore, the expression profiles across all samples of a comprehensive list of 60 genes belonging to key RNA processing pathways (see Table 5.6 at the end of this Chapter) were examined (Figure 5.11; *rrp47, lsm5, pop2, mot2, not3, not2* and *pac1* are not significantly differentially expressed in any samples and therefore do not appear in Figure 5.11).

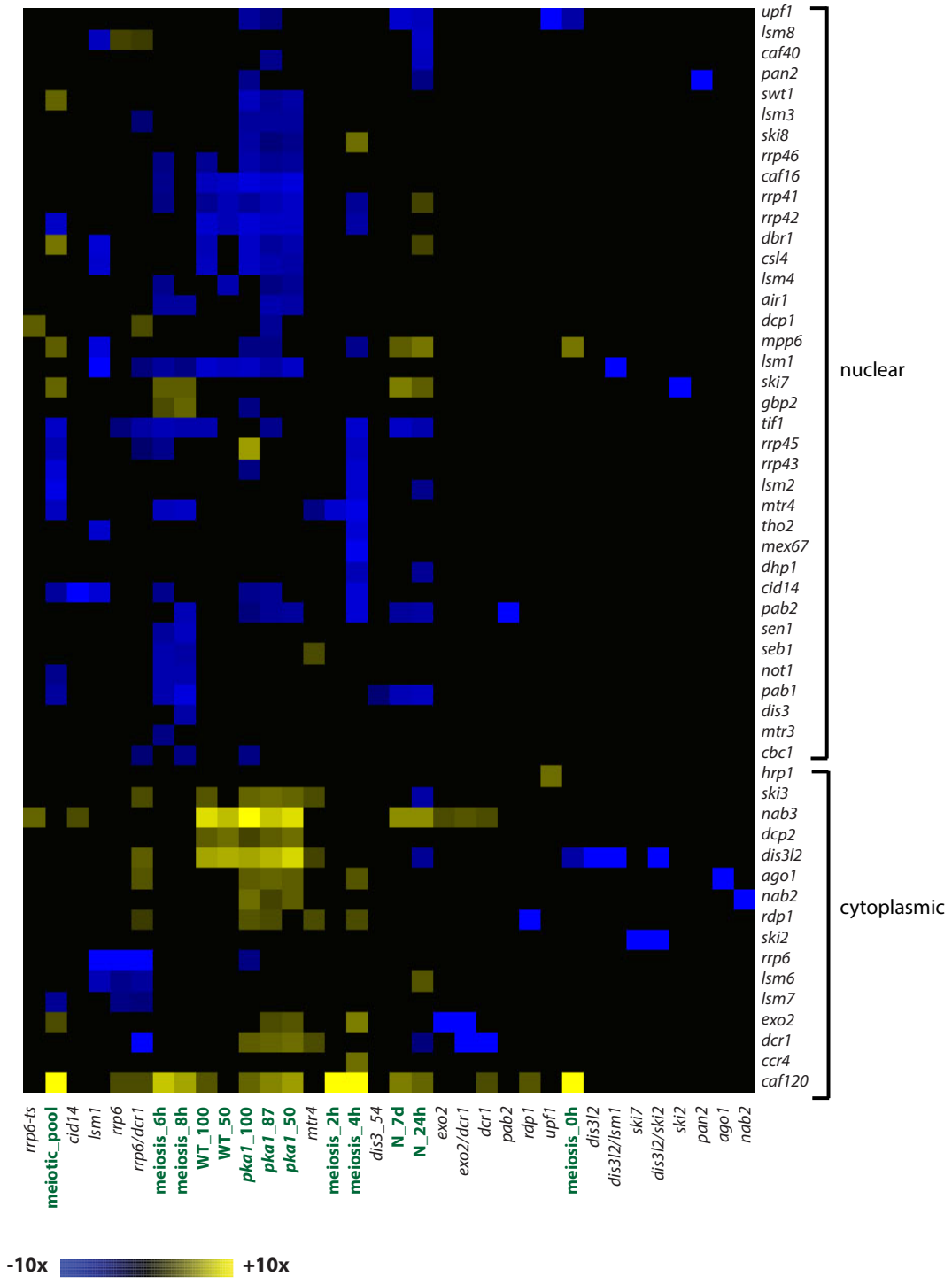

Figure 5.11: Expression profiles of a panel of 60 RNA processing genes.

See legend for Figure 5.3A. Columns clustered using the up-regulated correlation for the transcripts indicated.

For each RNA processing mutant, the transcript of the corresponding gene was strongly down-regulated relative to WT, confirming successful knock-out of these genes (Figure 5.11). Notable exceptions are *rrp6-ts* and *dis3-54*, for which down-regulation of activity of the corresponding gene occurs at the post-translational, rather than transcriptional, level.

As with the full set of mRNA transcripts (Figure 3.3A, Chapter 3), differential expression of this panel of RNA processing genes is more pronounced in WT cells under different growth conditions than in mutants (Figure 5.11, 5.9B). However, there is significantly more down-regulation amongst this set of RNA-processing genes than would be expected based on down-regulation observed across the full set of mRNAs (chi-squared test, $p = 2 \times 10^{-6}$). This suggests that global down-regulation of a complex interplay of RNA-processing pathways may help mediate the observed enrichments of CUTs, RUTs and XUTs in WT cells under different growth conditions.

RNA-processing genes were categorised based on whether their transcript was predominantly down-regulated across samples (must be down-regulated in at least 50% of the samples where it is differentially expressed) or predominantly up-regulated (must be up-regulated in at least 50% of samples where it is differentially expressed). Gene ontology (GO) analysis revealed that RNA-processing genes which are predominantly down-regulated are significantly enriched for nucleus and nucleolus GO-terms ($p = 3 \times 10^{-6}$ and $p = 2 \times 10^{-6}$ respectively). In contrast, RNA-processing genes which are significantly up-regulated are not enriched for nucleus or nucleolus GO-terms ($p = 1$). This suggests that down-regulation of nuclear RNA-processing activities may help mediate the observed up-regulation of lncRNA transcripts, whose location and site of any action, is presumably also nuclear. Since 50% of RNA-processing genes that are predominantly up-regulated belong to the 'cytoplasm' GO-term, a down-regulation of nuclear RNA-processing activities may be accompanied by a concomitant up-regulation of cytoplasmic RNA-processing activities (Figure 5.11).

While there are generally more DE RNA-processing genes in WT cells under different growth conditions than in the mutants (Figure 5.9B), two notable

exceptions to this are *lsm1* and *rrp6/dcr1*, both of which have a greater number of DE RNA processing genes than the mean for all samples (dotted line on Figure 5.9B). This could implicate Lsm1, Rrp6 and Dcr1 as hubs in RNA-processing networks. It is interesting to note that in *rrp6/dcr1* both *ago1* and *rdp1* are significantly up-regulated. No such similar up-regulation of RNAi components is observed when any of *ago1*, *dcr1* or *rdp1* is deleted alone. Such a result may indicate a tight functional link or redundancy between the nuclear exosome and the RNAi pathway.

In summary, global down-regulation of a complex interplay of RNA-processing pathways may help mediate the observed enrichments of CUTs, RUTs and XUTs in WT cells under different growth conditions. With the RNA-seq data available, it is only possible to assess the extent to which RNA processing pathways are likely to be transcriptionally down-regulated, and the possibility remains that these pathways may be post-translationally inhibited. Finally, expression correlations of mRNA:lncRNA pairs in RNA processing mutants as compared to WT cells under different growth conditions indicated that increased transcription, rather than post-transcriptional stabilisation, may account for increased lncRNA expression under different physiological conditions (Chapter 4). There is a growing body of evidence to suggest that RNA processing factors likely regulate transcript levels via a dynamic interplay of synthesis as well as degradation (Chapter 1). Thus, enrichments of CUTs, RUTs, and XUTs in different growth conditions may occur via a regulated interplay of synthesis and degradation, both of which are coordinated by RNA processing factors. Amongst such RNA processing factors, the nuclear exosome, RNAi and Exo2 are likely to play key roles.

5.5 Summary of main conclusions

The nuclear exosome, the RNAi pathway and the cytoplasmic exonuclease Exo2 represent three key pathways regulating lncRNAs in *S. pombe*, defining the overlapping classes of CUTs, RUTs and XUTs, respectively. The nuclear exosome and the RNAi pathway act cooperatively to control nuclear lncRNA expression, while the cytoplasmic Exo2 pathway is more distinct. Impairing both the nuclear exosome and the cytoplasmic exonuclease Exo2 is lethal in *S. pombe*.

CUTs, RUTs and XUTs are differentially enriched in WT cells under physiologically relevant growth conditions. Three key groups emerge: late meiotic RUTs/XUTs, early meiotic CUTs and quiescent CUTs.

Late meiotic RUTs/XUTs tend to be antisense (originating either antisense to the body of mRNAs or from convergent read-through). Both early meiotic and quiescent CUTs tend to be bidirectional, although early meiotic CUTs are also slightly enriched for antisense lncRNAs arising from convergent read-through.

Global down-regulation of a complex interplay of RNA-processing pathways may help mediate the observed enrichments of CUTs, RUTs and XUTs in WT cells under different growth conditions. Amongst such RNA processing factors, the nuclear exosome, RNAi and Exo2 are likely to play key roles.

Table 5.6: A panel of RNA processing genes

Information taken from Houseley and Tollervey (2009), Tuck and Tollervey (2013) and PomBase http://www.pombase.org/

RNA processing pathway	Gene	Information
5' end processing	*dcp1*	Member of decapping complex with *dcp2*
	dcp2	Catalytic pyrophosphatase subunit of decapping complex
	dhp1	Nuclear 5' exonuclease
Core exosome	*dis3*	Catalytic component of core exosome; 3' exonuclease; also endonuclease activity
	rrp41	Member of core exosome
	rrp42	Member of core exosome
	rrp43	Member of core exosome
	rrp45	Member of core exosome
	rrp46	Member of core exosome
	mtr3	Member of core exosome
	csl4	Member of core exosome
Nuclear exosome associated factors	*rrp6*	Nuclear specific catalytic exosome component; 3' exonuclease
	rrp47	Nuclear exosome cofactor
	mpp6	Nuclear exosome cofactor

RNA processing pathway	Gene	Information
Nuclear exosome associated factors (continued)	*mtr4*	Helicase; TRAMP-complex component (an exosome cofactor which targets nuclear RNAs to the exosome)
	cid14	Nuclear poly(A) polymerase; TRAMP-complex component
	air1	TRAMP complex component
Sen1-Seb1-Nab3 complex (involved in PolII termination and canonical non-coding RNA processing)	*sen1*	helicase
	seb1	RNA-binding protein
	nab3	RNA binding protein
Cleavage, polyadenlylation and poly(A) binding	*hrp1*	Cleavage and polyadenylation factor
	pab1	Poly(A)-binding protein
	pab2	Poly(A)-binding protein; targets RNAs for exosomal degradation; interacts with nuclear exosome
	nab2	Poly(A)-binding protein; competes with pab2
Lsm complexes (7 member ring-shaped RNA chaperone complexes; Lsm2-8 complex is nuclear and involved in pre-mRNA splicing, mRNA decay and processing of canonical non-coding RNAs; Lsm1-7 is cytoplasmic and promotes decapping)	*lsm1*	Member of cytoplasmic 1-7 complex
	lsm2	Member of both nuclear and cytoplasmic complexes
	lsm3	Member of both nuclear and cytoplasmic complexes

RNA processing pathway	Gene	Information
Lsm complexes (continued)	*lsm4*	Member of both nuclear and cytoplasmic complexes
	lsm5	Member of both nuclear and cytoplasmic complexes
	lsm6	Member of both nuclear and cytoplasmic complexes
	lsm7	Member of both nuclear and cytoplasmic complexes
	lsm8	Member of nuclear 2-8 complex
Nuclear export	*cbc1*	Nuclear cap binding complex subunit; mRNA stabilisation, processing, export and decay
	tho2	TREX complex component; regulates nuclear export of mRNAs
	gbp2	TREX complex component; regulates nuclear export of mRNAs
	mex67	mRNA export receptor
Cytoplasmic deadenylation	*pan2*	deadenylase
	ccr4	Member of Ccr4-NOT complex
	pop2	Member of Ccr4-NOT complex
	not1	Member of Ccr4-NOT complex
	not2	Member of Ccr4-NOT complex
	not3	Member of Ccr4-NOT complex

RNA processing pathway	Gene	Information
Cytoplasmic deadenylation (continued)	*mot2*	Member of Ccr4-NOT complex
	caf16	Member of Ccr4-NOT complex
	caf40	Member of Ccr4-NOT complex
	caf120	Member of Ccr4-NOT complex
Cytoplasmic exonucleases	*exo2*	Cytoplasmic 5' exonuclease
	dis3l2	Cytoplasmic 3' exonuclease
	ski7	Cytoplasmic exosome cofactor; connects exosome to Ski complex (a cytoplasmic exosome cofactor complex)
	ski2	Cytoplasmic helicase; member of exosome cofactor Ski complex
	ski3	member of exosome cofactor Ski complex
	ski8	member of exosome cofactor Ski complex
RNAi pathway (well-established role in heterochromatin assembly and gene silencing; small-interfering RNAs, siRNAs, are products and mediators of the RNAi pathway)	*ago1*	Member of the RITS (RNA-induced transcriptional gene silencing) effector complex
	dcr1	Endoribonuclease cleaving dsRNA
	rdp1	RNA-dependent RNA polymerase
Miscellaneous	*dbr1*	Debranches intron lariat structures
	pac1	Endonuclease involved in termination and processing of non-coding RNAs

RNA processing pathway	Gene	Information
Miscellaneous (continued)	*swt1*	Endonuclease involved in perinuclear mRNP surveillance
	upf1	ATP-dependent RNA helicase involved in nonsense mediated decay
	tif1	Helicase within the cytoplasmic cap-binding complex; promotes translation

Chapter 6 Discussion

6.1 *S. pombe* CUTs, XUTs and RUTs

6.1.1 Model emerging from the current study

In the current study I have developed, validated and implemented a method to detect novel lncRNAs from strand-specific RNA-seq data (Chapter 3). Using this method, I have described thousands of novel lncRNAs in *S. pombe*, providing a rich and comprehensive resource for further studies of lncRNA function.

Approximately 1500 lncRNAs have been previously annotated in *S. pombe* (Rhind et al. 2011, Wilhelm et al. 2008), and are referred to as SPNCRNAs. The method I have developed to identify novel lncRNAs has been optimised for its ability to detect SPNCRNAs from RNA-seq data. Nonetheless, several important differences between SPNCRNAs, and the lncRNAs identified in the current study, are apparent. Firstly, lncRNAs identified in the current study are more numerous and more pervasively transcribed than SPNCRNAs. Secondly, they are more variable across samples than SPNCRNAs, suggesting they are more subject to differential regulation.

Analysis of the expression of lncRNAs across a comprehensive panel of RNA processing mutants has indicated that the nuclear exosome, the RNAi pathway and the cytoplasmic exonuclease Exo2 are three key pathways regulating lncRNAs in *S. pombe*. Regulation by the nuclear exosome, RNAi and Exo2 has been used to define CUTs, RUTs and XUTs respectively (Chapter 5). These classes are not entirely distinct, but overlap to some degree. Interactions between these three key pathways will be discussed further in Section 6.2.

Importantly, despite their other functions in RNA metabolism (see Chapter 1), the nuclear exosome, the RNAi pathway and the cytoplasmic exonuclease Exo2 all appear to selectively target lncRNAs for degradation. Thus, the same expression signatures and extent of up-regulation are not seen in the corresponding mutants for mRNAs or canonical ncRNAs (tRNAs, rRNAs, sno/sn-RNAs). This is supported by the observed decoupling of expression correlation between bidirectional or antisense

mRNA:lncRNA pairs in RNA-processing mutants (Chapter 4). In such mutants, lncRNAs are specifically stabilised, and therefore a concomitant effect on their partner mRNA is not seen, resulting in a lack of expression correlation.

lncRNAs are differentially expressed under physiologically relevant growth conditions, implicating them as having possible functions under such conditions (Chapter 5). Quiescent and early meiotic lncRNAs are regulated predominantly by the exosome (strong enrichment of CUTs in quiescence and early meiosis), while RNAi and Exo2 appear to play a greater role in the regulation of lncRNAs up-regulated during late meiosis (strong enrichment of RUTs and XUTs in late meiosis). Such specific repertoires of lncRNAs being stabilised under differing physiological perturbations argues in favour of such lncRNAs having a function under these conditions, rather than simply representing non-specific transcriptional noise (see Section 6.3).

Based on their regulation by differing RNA processing pathways, together with their enrichments under the physiological conditions considered, three key groups of lncRNAs emerge: late meiotic RUTs/XUTs, early meiotic CUTs and quiescent CUTs (Chapter 5; Figure 6.1 below). Early meiotic and quiescent CUTs tend to be divergently transcribed from mRNAs and positively correlate in expression with the mRNAs they are divergently associated with. In contrast, late meiotic XUTs/RUTs are transcribed predominantly antisense to mRNAs. Interestingly, there is an enrichment for such antisense transcription to arise from read-through at convergent mRNAs, and these antisense-lncRNA:mRNA pairs tend to anti-correlate in expression. This suggests such antisense lncRNAs may exert inhibitory regulatory effects on their sense loci during the later stages of meiosis, and will be discussed further in Section 6.3.

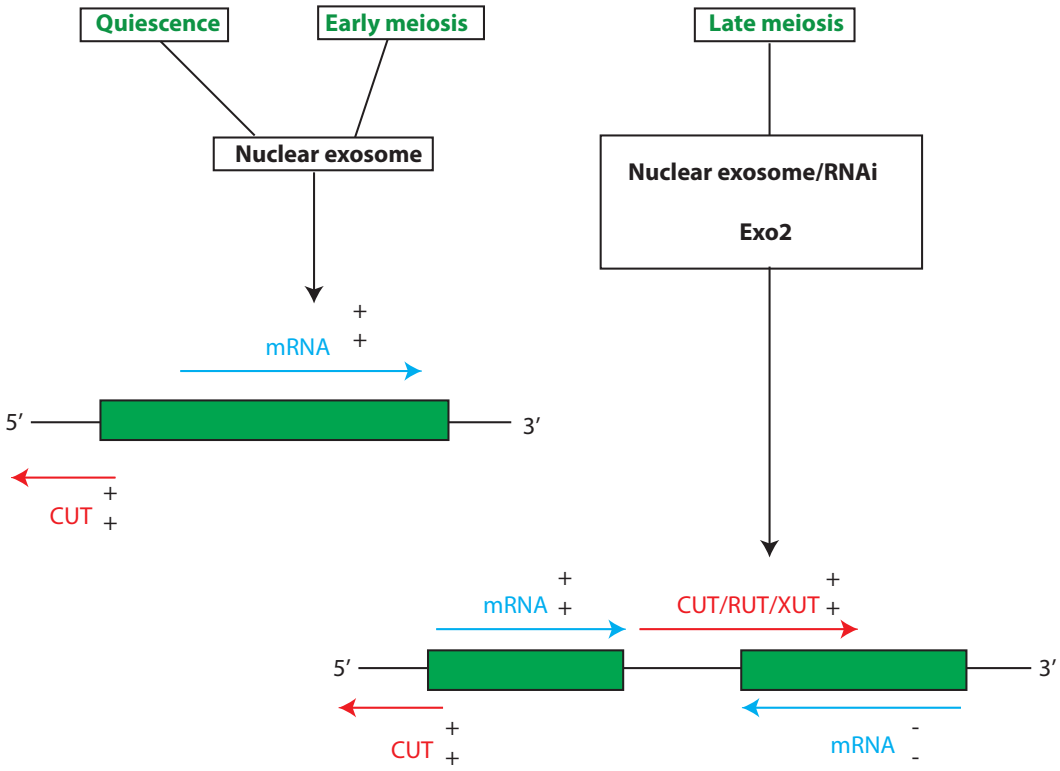

Figure 6.1: Schematic of CUTs, RUTs and XUTs enriched under different physiologically relevant perturbations due to underlying changes in synthesis and degradation mediated by the nuclear exosome, RNAi and Exo2.

CUTs, RUTs and XUTs have distinct locations and relationships with the mRNAs they are associated with. Green boxes represent mRNA genes; plus and minus symbols represent up-regulation or down-regulation of RNA levels respectively.

The molecular mechanisms underlying enrichment of CUTs, RUTs and XUTs under physiologically relevant growth conditions remain unclear, since *rrp6*, *dcr1* and *exo2* transcript levels are not significantly differentially expressed under such conditions. This will be discussed further in Section 6.2.

Unstable lncRNAs have been most extensively described at a genome-wide level in the budding yeast *S. cerevisiae*. I will briefly compare the findings in the current study to what has been described for *S. cerevisiae*, and the relevance of these yeast model systems to lncRNA regulation in higher eukaryotes.

6.1.2 S. pombe lncRNAs compared to other species

The CUTs and XUTs identified in the current study are analogous to similar groups of unstable lncRNAs identified in *S. cerevisiae*.

In *S. cerevisiae*, CUTs are a group of lncRNAs targeted for degradation by the nuclear exosome and stabilised upon deletion of the nuclear-specific exosome subunit Rrp6 (Neil et al. 2009, Xu et al. 2009). Approximately 1500 CUTs were identified genome-wide by Neil et al. (2009), and ~1000 identified in an independent study by Xu et al. (2009). Approximately 56% of the lncRNAs identified by Xu et al. (2009) were also identified by Neil et al. (2009), meaning that ~2000 CUTs have been described in *S. cerevisiae*, a comparable number to the 2500 CUTs identified in *S. pombe* in the current study.

Although a few *S. cerevisiae* CUTs have had functions ascribed (Camblong et al. 2007, Hainer et al. 2011, Houseley et al. 2008, Martens et al. 2004), it is unknown whether the majority has any function. Importantly, as shown for *S. pombe* CUTs in the current study, *S. cerevisiae* CUTs tend to be transcribed divergently from mRNAs, and positively correlate in expression with the mRNAs they are divergently transcribed from (Neil et al. 2009, Xu et al. 2009). A key difference between *S. pombe* and *S. cerevisiae* CUTs, however, is that loss of Trf4 in *S. cerevisiae*, a key component of the exosome-targeting TRAMP complex, causes dramatic stabilisation of CUTs, with many CUTs known to be targeted by the TRAMP complex (Frenk et al. 2014, Wlotzka et al. 2011, Wyers et al. 2005). The same phenotype is not seen in the current study upon deletion of the *S. pombe* Trf4 homologue Cid14, raising the possibility of a TRAMP-independent mechanism of exosomal degradation of CUTs in *S. pombe* (see Section 6.2).

Importantly, exosome depletion in mammalian cells has revealed lncRNAs mapping to promoter regions of known protein-coding genes (Core et al. 2008, Preker et al. 2008, Seila et al. 2008). Such lncRNAs are reminiscent of *S. cerevisiae* and *S. pombe* CUTs. Whether the evolutionary conservation of this transcriptional pattern of lncRNAs reflects functional importance of CUTs, or simply that they are non-functional by-products of the basic mechanics of transcription, remains an open

question. Nonetheless, it is clear that pervasive non-coding transcription is a universal feature of transcriptomes, and that such transcription is not random but has clear patterns in terms of its genomic location and the pathways regulating its expression.

In *S. cerevisiae*, XUTs (Xrn1-dependent unstable transcripts) are a group of ~1700 lncRNAs which are targeted for degradation by the cytoplasmic exonuclease Xrn1 (homologue of *S. pombe* Exo2) (van Dijk et al. 2011). As such, *S. cerevisiae* XUTs are analogous to the similar number (1888) of *S. pombe* XUTs identified in the current study. Importantly, the majority of *S. cerevisiae* XUTs were shown to be antisense to mRNAs, with Pol II ChIP-seq revealing that accumulation of antisense XUTs correlates with decreased transcription of their sense loci (van Dijk et al. 2011). This is in agreement with the enrichment of antisense lncRNAs amongst *S. pombe* XUTs identified in the current study, which tend to anti-correlate in expression with their sense mRNA transcripts. The targeting of XUTs by a cytoplasmic exonuclease implies their efficient export to the cytoplasm. However, their proposed inhibitory functions on coding transcription are likely to be mediated co-transcriptionally in the nucleus, and so the functional importance of their cytoplasmic export is unclear (Jensen et al. 2013). Approximately 20% of *S. cerevisiae* CUTs were found to be Xrn1-sensitive (van Dijk et al. 2011). The overlap observed in the current study is comparable but slightly higher, with 35% of CUTs also classed as XUTs. Such overlap between CUTs and XUTs hints at partially redundant mechanisms of regulating lncRNA expression, and will be discussed further in Section 6.2. It will be of interest to determine whether an analogous group of Xrn1-sensitive lncRNAs exists in multicellular eukaryotes.

There is no analogous group in *S. cerevisiae* to the *S. pombe* RUTs identified in the current study, since the RNAi pathway is missing in *S. cerevisiae*. The identification of RUTs as an important class of lncRNA in the current study, and the conservation of RNAi from *S. pombe* to higher eukaryotes, makes *S. pombe* an attractive complementary model system in which to study lncRNAs. The possible functional importance of the RNAi system in lncRNA mechanisms of action will be discussed Section 6.2.

The median lengths of the CUTs, RUTs and XUTs identified in the current study are 864bp, 908bp, and 1074bp respectively (for comparison, the median length of *S. pombe* mRNAs is 1.9kb), and the distribution of lengths is progressively higher for CUTs, RUTs, and then XUTs (Figure 6.2).

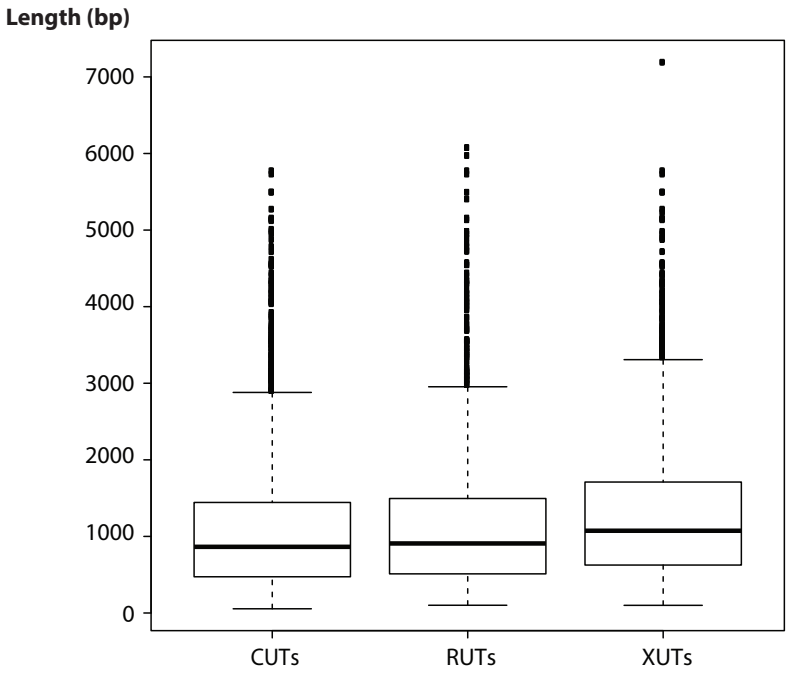

Figure 6.2: Lengths in base-pairs (bp) of CUTs, RUTs and XUTs.

It is of note that the longer classes of RUTs and XUTs are also enriched for being antisense lncRNAs. CUTs identified in *S. cerevisiae* by Xu et al. (2009) have a shorter median length of 440bp. The longer length of CUTs identified in the current study could be due to differences in the methods used to profile transcriptomes. For example, the high depth of RNA-seq used in the current study is a higher resolution and more sensitive technique than the tiling arrays used by Xu et al. (2009), and as such may identify the full length of lncRNAs. Indeed, the length of *S. pombe* XUTs identified in the current study is comparable to the *S. cerevisiae* XUTs identified by

strand-specific RNA-seq (van Dijk et al. 2011).

A recent genome-wide study has revealed another class of unstable lncRNAs in *S. cerevisiae*, which are detected upon nuclear depletion of Nrd1, and are referred to as Nrd1-unterminated transcripts (NUTs) (Schulz et al. 2013). Nrd1 is a member of the Nrd1-Nab3-Sen1 (NNS) complex, which is thought to interact with Pol II to specifically promote lncRNA termination in *S. cerevisiae*. Approximately ~1500 NUTs have been described in *S. cerevisiae*, and these overlap substantially with CUTs and XUTs. Thus ~600 NUTs overlap annotated CUTs, and a further, but overlapping, group of ~600 NUTs are also annotated as XUTs. Importantly, Nrd1 and Nab3 binding motifs are depleted in mRNAs but enriched in NUTs, indicating that NNS selectively terminates this class of lncRNAs. This is in contrast to *S. pombe,* where a similar motif bias is not seen (Aylin Cakiroglu, personal communication). Moreover, RNA-seq of the *S. pombe seb1* mutant (homologue of Nrd1) reveals a global transcriptional read-through phenotype (Chapter 2, Table 2.1). Thus it seems unlikely that Seb1 acts to selectively terminate lncRNA transcription in *S. pombe*, and a group of lncRNAs analogous to NUTs is unlikely to exist in *S. pombe*. The NNS pathway in *S. pombe* likely acts in a dissimilar manner to the NNS pathway in *S. cerevisiae*. This is discussed further in Section 6.2.3.

Finally, stable unannotated transcripts (SUTs) are a class of ~850 *S. cerevisiae* lncRNAs which, unlike CUTs and XUTs, are detectable in WT vegetative cells (Xu et al. 2009). Tuck and Tollervey (2013) analysed the transcriptome-wide targets of 13 RNA processing, export and turnover factors in *S. cerevisiae*. Clustering of the binding profiles of these factors to different transcripts revealed CUTs and SUTs to cluster largely separately from each other, suggesting that they do indeed represent predominantly distinct transcript classes. *S. cerevisiae* SUTs could be considered to be analogous to the ~1500 already annotated SPNCRNAs in *S. pombe*, which have been detected predominantly by studies in WT vegetative cells (Rhind et al. 2011, Wilhelm et al. 2008), and whose expression is less variable across samples than the novel lncRNAs annotated in the current study (Chapter 3).

6.2 Regulation of lncRNAs

6.2.1 Overlap of CUTs, RUTs and XUTs

There is extensive overlap amongst the CUTs, RUTs and XUTs identified in the current study, with ~600 lncRNAs belonging to all three classes (Chapter 5). Such overlap could be indicative of redundancy and coordination between degradation pathways, i.e. in the absence of one RNA degradation pathway, lncRNAs are simply turned over by another.

Notably, the least degree of overlap occurs between CUTs and XUTs. This suggests that the nuclear exosome and the cytoplasmic exonuclease Exo2 may represent the most distinct pathways, with Exo2 being the key cytoplasmic player regulating lncRNAs, and Rrp6 the main nuclear player. This finding is consistent with the synthetic lethality observed for the double mutant *rrp6/exo2*. While these pathways function in other aspects of RNA metabolism, which may contribute to this observed synthetic lethality, this result may also indicate the functional importance of regulating lncRNA expression. In contrast, the greater number of lncRNAs up-regulated in the *rrp6/dcr1* double mutant when compared to either single mutant, suggests that the nuclear exosome and the RNAi pathway may act cooperatively to control nuclear lncRNA expression (see Section 6.2.2).

Surprisingly, the *exo2/dcr1* double mutant has *fewer* up-regulated lncRNAs than either single mutant (Chapter 5). A possible explanation for this observed lncRNA phenotype is that if the nuclear RNAi pathway is impaired, lncRNAs are channeled to the nuclear exosome. Thus, in the *rrp6/dcr1* double mutant, where the nuclear exosome pathway is also impaired, a greater number of lncRNAs are up-regulated than in either *dcr1* or *rrp6*. However, in the *exo2/dcr1* mutant, where the nuclear exosome is intact, defects in RNAi may efficiently channel lncRNAs to the nuclear exosome. This results in fewer lncRNAs reaching the cytoplasm and defective Exo2 therefore has less impact on lncRNA expression, leading to the observed rescue phenotype in *exo2/dcr1*.

As described in Section 6.1, a similar overlap between classes of *S. cerevisiae*

unstable lncRNAs has been reported. It has been suggested that this redundancy may be a result of inefficient transcription termination (Jensen et al. 2013). Accordingly, it has been reported that, in *S. cerevisiae,* CUTs are terminated co-transcriptionally by the Nrd1-Nab3-Sen1 (NNS) pathway (Arigo et al. 2006, Schulz et al. 2013, Thiebaut et al. 2006). In contrast, since XUTs are presumably exported to the cytoplasm (they are targeted for degradation by cytoplasmic Xrn1), they are likely terminated further downstream at a polyadenylation signal by the canonical CPF (cleavage and polyadenylation factor) pathway. However, a proportion of CUTs may escape early transcription termination by the NNS pathway, and subsequent degradation by the nuclear exosome. Instead, such elongated CUTs ("eCUTs") may terminate further downstream near a polyadenylation site. As a result, they may be treated more like mRNAs, and exported to the cytoplasm, where they are degraded by Xrn1. Consequently lncRNAs with a common transcription origin may be classified as both CUTs and XUTs (Marquardt et al. 2011).

6.2.2 Interplay between RNAi and the nuclear exosome

An intriguing finding from the current study is that the exosome and the RNAi pathway appear to act cooperatively to regulate lncRNA expression. Thus, there is a high degree of overlap between CUTs and RUTs, and the *rrp6/dcr1* double mutant displays a stronger lncRNA expression than either single mutant (Chapter 5).

Connections between the RNA processing activities of the exosome and the RNAi pathway are emerging as playing a role in the regulation of lncRNAs in *S. pombe* (Gullerova and Proudfoot 2008, Lee et al. 2013, Shah et al. 2014, Woolcock et al. 2011, Woolcock et al. 2012, Zofall et al. 2009). Intriguingly, a recent study has identified a lncRNA, named *prt*, which is transcribed upstream of the *pho1* gene, and overlaps the *pho1* promoter (Shah et al. 2014). The *pho1* gene encodes a phosphatase required for uptake of extracellular phosphate, and is induced during phosphate starvation. Shah et al. (2014) show that, in the presence of extracellular phosphate, *prt* transcription is induced, leading to transient RNAi-dependent deposition of H3K9Me2 across the *pho1* locus, and subsequent *pho1* repression. RNAi-recruitment by *prt1* is dependent on DSR (determinant of selective removal) motifs in *prt1*, which bind the RNA-binding protein Mmi1. Moreover, the same

DSR-Mmi1 system also recruits the exosome, which selectively targets *prt1* transcript termination and degradation.

A similar DSR-Mmi1 system has been shown to result in both exosomal targeting of meiotic transcripts during mitotic growth (Harigaya et al. 2006, Sugiyama and Sugioka-Sugiyama 2011), as well as formation of transient heterochromatin at these loci (Tashiro et al. 2013, Zofall et al. 2012). Such a role for the exosome and RNAi acting in parallel at meiotic loci is consistent with the strong enrichment of RUTs and CUTs at late meiotic time-points in the current study. This result implicates RNAi/exosome-mediated lncRNA regulation as playing an important physiological role during late meiosis.

Furthermore, late meiotic RUTs in the current study are enriched for antisense transcripts, arising to a large extent from convergent read-through, and anti-correlate in expression with the mRNAs they run antisense to. This observation is consistent with a study showing that antisense transcription, arising in part from read-through at convergent genes, is increased during meiosis. Moreover, this antisense transcription exerts an inhibitory effect on sense loci, which appears to be RNAi-dependent (Bitton et al. 2011). Although a subsequent study (Chen H. M. et al. 2012) showed antisense-mediated repression of three mid-meiotic genes to be RNAi-*in*dependent, this cannot be extended to suggest that the RNAi machinery plays no role in antisense repression during meiosis at a global level. Finally, a role for convergent read-through and the RNAi-machinery in gene regulation is reminiscent of reports that double-stranded RNA arising from defective transcription termination at convergent genes can lead to RNAi-dependent transient heterochromatin and gene silencing (Gullerova and Proudfoot 2008, Gullerova et al. 2011).

In summary, although further studies are required to understand the interplay between the exosome and RNAi in the regulation of gene silencing, it is becoming increasingly clear that these two pathways interact to mediate the effects, and control the expression of, regulatory lncRNAs in *S. pombe*. It is plausible that transient heterochromatin formation lies downstream of the RNAi/exosome/lncRNA-mediated mechanism of gene regulation suggested in the current study. In fact, lncRNA-mediated transient heterochromatin formation may represent a general mechanism

for achieving an effective response to environmental stimuli.

H3K9Me2 ChIP-seq in vegetative and meiotic cells, both in WT and in mutants defective for RNAi and exosome components, will be instructive in elucidating how widespread putative RNAi/exosome/lncRNA-mediated formation of transient heterochromatin is in fission yeast. RNA immunoprecipitation (RIP)-seq of RNAi-components may help determine whether late meiotic RUTs are involved in targeting the RNAi machinery to specific loci, while determining whether such lncRNAs are enriched for DSR motifs will be of interest.

6.2.3 TRAMP-independent exosome degradation

The weak lncRNA phenotypes observed for mutants of the TRAMP complex *mtr4* and *cid14* (Chapter 3) suggest a possible role for TRAMP-independent exosome degradation of lncRNAs. Such TRAMP-independent exosome degradation has been reported for a group of lncRNAs in human cells (Beaulieu et al. 2012). This group of lncRNAs are instead targeted for exosomal degradation predominantly by the poly(A)-binding protein PABPN1 (homologue of *S. pombe* Pab2). A similar Pab2-Rrp6 mechanism has been described for targeting degradation of *S. pombe* meiotic genes during vegetative growth (St-Andre et al. 2010). Such a Pab2-Rrp6 TRAMP-independent degradation mechanism for *S. pombe* lncRNA degradation is consistent with the stronger lncRNA phenotype observed in the current study for the *pab2* mutant, when compared to *cid14* or *mtr4* mutants. Moreover, although weaker than the lncRNA phenotype observed for exosome mutants, lncRNAs up-regulated in *pab2* strongly overlap with those up-regulated in exosome mutants (Chapter 3).

In contrast to *S. pombe*, loss of Trf4 in *S. cerevisiae* (homologue of *S. pombe* Cid14) causes dramatic stabilisation of CUTs, with many CUTs known to be targeted by the TRAMP complex (Frenk et al. 2014, Wlotzka et al. 2011, Wyers et al. 2005). This suggests that exosomal targeting of lncRNAs in *S. cerevisiae* is TRAMP-dependent, consistent with literature describing an important role for NNS termination of lncRNAs in *S. cerevisiae* (Arigo et al. 2006, Schulz et al. 2013, Thiebaut et al. 2006). NNS termination likely occurs co-transcriptionally with TRAMP-dependent

exosomal targeting (Tudek et al. 2014).

It is of note that the NNS pathway in *S. pombe* likely acts in a dissimilar manner to the NNS pathway in *S. cerevisiae*. Thus, Sen1 is essential in *S. cerevisiae* but not in *S. pombe*. Intriguingly, Seb1 (*S. pombe* Nrd1 homologue) is unable to complement the *S. cerevisiae NRD1* mutant and vice versa, and Seb-1-Nab3-Sen1 *S. pombe* homologues do not co-purify (Francois Bachand – personal communication). Once again, given that there are no human homologs for Nrd1 and Nab3, *S. pombe* may present a model system in which the mechanisms of lncRNA degradation are more akin to those in multicellular eukaryotes.

In summary, differences between the findings in the current study and those described for *S. cerevisiae* could indicate that NNS-TRAMP-dependent exosome degradation is a predominant pathway for lncRNA degradation in *S. cerevisiae* but not in *S. pombe*. In contrast, in *S. pombe* it may be that most lncRNAs are cleaved and polyadenylated by the canonical CPF pathway, and are therefore subject to TRAMP-independent exosome degradation, possibly mediated by Pab2. It would be of interest to determine how prevalent canonical polyadenylation sequence elements (for example the hexanucleotide AAUAAA followed by a downstream G/U-rich sequence) are in the lncRNAs in the current study.

It cannot be excluded, however, that the minimal role observed for the TRAMP complex in lncRNA degradation in the current study may be an artefact of the poly(A)-enrichment employed for RNA-seq sample preparation. RNAs polyadenylated by the canonical 3' end processing machinery have generally long poly(A) tails. In contrast, shorter oligo(A) tails (4-5 nucleotides) arise from the adenylation activity of the TRAMP complex. It is possible that the poly(A) enrichment used to prepare samples for RNA-seq in the current study, biases against the shorter oligo(A) tails, thereby biasing against lncRNAs that are TRAMP-targeted.

6.2.4 Stabilisation of lncRNAs – synthesis vs. degradation

CUTs, RUTs and XUTs have been defined in the current study from *rrp6*, *dcr1* and *exo2* deletion mutants respectively. A key finding is that specific repertoires of CUTs, RUTs and XUTs are up-regulated under physiologically relevant conditions. However, since only subtle changes in transcript levels of *rrp6*, *dcr1* and *exo2* were observed in WT cells under different physiologically relevant growth conditions, it is unclear what underlies CUT, RUT and XUT stabilisation under such conditions.

While changes in *rrp6*, *dcr1* and *exo2* transcripts were largely not significant under different growth conditions, they appeared to follow coherent patterns, particularly throughout the course of meiosis, thereby suggesting that these changes in transcript levels may still be of functional relevance in controlling CUT, RUT and XUT expression (Chapter 5). Analysing the differential expression of a panel of 60 RNA processing factors revealed that a complex interplay of altered RNA processing activities may underlie the observed enrichments of CUTs, RUTs and XUTs (Chapter 5).

In addition, it is possible that the activity of RNA processing factors could be translationally regulated, and Western blot analysis of Rrp6, Dcr1 and Exo2 under physiologically relevant growth conditions would be required to determine this. Furthermore, activities of RNA processing factors may be post-translationally down-regulated. For example, a weaker association of Rrp6 with the *S. cerevisiae PHO84* locus upon chronological ageing has been reported to result in stabilisation of lncRNAs antisense to this locus (Camblong et al. 2007). ChIP-seq of Rrp6 and Dcr1 in WT cells under different growth conditions may provide some insight into the association of these factors with different loci, and whether this may account for the observed patterns of CUT and RUT expression under these conditions. Although Exo2 is largely presumed to be cytoplasmic, Xrn1 (*S. cerevisiae* homologue of Exo2) has recently been shown to associated with chromatin (Haimovich et al. 2013). ChIP-seq of Exo2 may therefore also be informative.

Increased lncRNA expression may occur not only from post-transcriptional stabilisation (altered RNA degradation activities) but also from increased

transcription. For example, Castelnuovo et al. (2014) recently reported increased H3K4Me3 marks (marking transcription initiation) at the 3' ends of antisense-lncRNA producing genes in a *S. cerevisiae rrp6* mutant. This indicates that lncRNA accumulation upon loss of exosome activity may be due not only to stabilisation, but also to increased transcriptional activity. Indeed, there is a growing body of evidence to suggest that RNA processing factors likely regulate transcript levels via a dynamic interplay of synthesis as well as degradation, and may be more appropriately referred to as synthedegradases (Haimovich et al. 2013, Medina et al. 2014, Sun M. et al. 2013a, Sun M. et al. 2012).

Thus, enrichments of CUTs, RUTs, and XUTs under different physiologically relevant perturbations may occur via a regulated interplay of synthesis and degradation, both of which are coordinated by RNA processing factors. Amongst such RNA processing factors, the nuclear exosome, RNAi and Exo2 are likely to play key roles (Figure 6.1).

A key limitation of the current study is that RNA-seq looks at steady state transcript levels, which are a combined outcome of degradation and synthesis. Thus the weak lncRNA phenotypes seen in cytoplasmic RNA processing mutants, other than *exo2*, has led us to conclude that such factors play only minor or redundant roles in lncRNA degradation, and that Exo2 is the key cytoplasmic player. However, recent studies point to key role of Xrn1 (Exo2) in buffering RNA levels, enabling an increased degradation rate to be accompanied by an increased synthesis rate, and vice versa (Haimovich et al. 2013, Sun M. et al. 2013a). Thus, weak lncRNA phenotypes seen in cytoplasmic RNA degradation mutants, other than *exo2*, could be due to RNA levels being efficiently buffered in these mutants. In contrast, in the *exo2* mutant this buffering system is impaired. Thus a decreased degradation rate in the *exo2* mutant is accompanied by an increased synthesis rate, and thus a stronger lncRNA phenotype.

It would therefore be of interest to measure both synthesis and degradation rates of lncRNAs in the mutants and WT cells under different growth conditions used in the current study. While Pol II ChIP-seq has been criticised as not giving a direct measure of synthesis rates, comparative dynamic transcriptome analysis (cDTA) uses

non-perturbing metabolic labelling of newly transcribed RNA to directly measure RNA synthesis (Sun M. et al. 2012).

6.3 lncRNA functions

A major question arising from the current study is how many of the thousands of newly identified lncRNAs are functional, and how many are simply by-products of transcription with no obvious function.

6.3.1 Do expression correlations indicate functionality?

lncRNAs can be related to their mRNA partners in one of two key ways – divergently (TSSa) or antisense (body, TSSa-AS, TTSa-AS) (Chapter 4). How these correlate in expression with each other enables some speculation as to functionality.

6.3.1.1 Bidirectional mRNA:lncRNA pairs

Consistent with the findings of the current study, bidirectional non-coding transcription from mRNA promoters is conserved in *S. cerevisiae* (Neil et al. 2009, Xu et al. 2009) and mammalian cells (Core et al. 2008, Preker et al. 2008, Seila et al. 2008), and across these species mRNAs and their associated divergent transcripts tend to be co-regulated.

While such co-regulation could imply that bidirectional lncRNAs are simply by-products of transcription, various functions can nonetheless be imagined for divergent lncRNAs that positively correlate in expression with their mRNA partners. They may act to fine-tune regulation of their associated genes for example by modifying local chromatin or by recruiting transcriptional activators/repressors (see Chapter 1).

It has been postulated that transcription of divergent lncRNAs is a noise-reduction, expression-priming mechanism (Wang G. Z. et al. 2011). As such, divergent lncRNAs are more associated with essential genes and less associated with stress-response genes. However, no enrichment of GO-terms was seen amongst mRNAs with divergent lncRNAs in the current study. The reason for this discrepancy could be that lncRNAs annotated in the current study (which represent by far the majority

of lncRNAs) have been defined as those which are differentially expressed in mutants and conditions relative to WT vegetative cells. Therefore this set of lncRNAs is biased against constitutively expressed lncRNAs, which are perhaps more likely to be associated with essential genes.

6.3.1.2 Antisense mRNA:lncRNA pairs

Antisense (AS) lncRNA:mRNA pairs were found to generally anti-correlate in expression (Chapter 4). This is in agreement with previous studies describing repressive effects of AS lncRNAs in *S. pombe* (Bitton et al. 2011, Chen H. M. et al. 2012, Leong et al. 2014, Ni et al. 2010).

The degree of anti-correlation amongst AS lncRNA:mRNA pairs was found to be stronger for AS lncRNAs originating in either the 5'- or 3'- NFR, or for lncRNAs which originate antisense to the body of mRNAs but overlap neither end of the mRNA. lncRNAs that originate antisense to the body of mRNAs and overlap the 5' end of the mRNA were found to actually co-correlate in expression with their sense loci. This is in contradiction to recent findings in *S. cerevisiae*, which suggest that a distinguishing feature of AS lncRNAs with inhibitory regulatory potential, is that the lncRNA extends into the promoter of its sense mRNA (Castelnuovo et al. 2014).

In *S. cerevisiae*, genes that respond in a switch-like manner, such as stress-response and environment specific genes, are enriched for antisense expression (Xu et al. 2011). Such antisense non-coding transcription has been suggested to 'switch off' sense loci expression when low levels of expression of the sense mRNA are required. In agreement with this, I find that this pattern is conserved in *S. pombe*, with mRNAs with an antisense lncRNA enriched for meiotic and stress module GO-terms (Chapter 4). This is also consistent with Marguerat et al. (2012), which shows that the most highly expressed antisense SPNCRNAs in vegetative *S. pombe* cells are antisense to meiotic differentiation loci that are tightly repressed during vegetative growth.

While expression correlations of associated mRNAs and lncRNAs provide some speculation as to possible functions, the possibility that observed correlations are simply an effect of basic transcriptional mechanics cannot be excluded. Thus any

pair of antisense transcripts may be inversely correlated in expression due to an inability of Pol II to concomitantly transcribe from both strands of DNA. Similarly any pair of divergently oriented transcripts may be positively correlated in expression due to shared upstream regulation, or proximal transcriptional activity in surrounding open chromatin. To probe this possibility further it would be of interest to determine whether mRNA:mRNA divergent and antisense pairs display the same patterns of expression correlation as lncRNA:mRNA pairs. This would help address whether expression correlation between neighbouring genes is simply a positional effect as opposed to being indicative of active regulation.

6.3.2 Conservation of lncRNAs

Another parameter that could be used to determine the likelihood of lncRNA functionality is the degree of sequence conservation.

By sequencing the genomes of 161 natural isolate *S. pombe* strains, a recent study has analysed genetic diversity in different regions of the *S. pombe* genome (Jeffares et al., manuscript under review). Jeffares et al. show that already annotated intergenic lncRNAs in *S. pombe* appear to be subject to little or no purifying selection, with their genetic diversity being comparable to 4-fold degenerate sites. However, there is evidence for purifying selection amongst the 20% most highly expressed intergenic lncRNAs.

Using a whole-genome alignment across the four species of the fission yeast clade (*S. pombe, S. octosporus, S. cryophilus, S. japonicus*), sequence conservation and selective constraints in the fission yeast genome have been assessed (Victor Sojo, unpublished data). Preliminary analyses have revealed that the vast majority of already annotated *S. pombe* lncRNAs, as well as the lncRNAs identified in the current study, show significantly less constraint than 4-fold degenerate sites (regions of lncRNAs overlapping mRNAs, in either a sense or antisense orientation, were not considered in the analysis).

Thus, in agreement with lncRNAs described in other species, the sequence conservation of the majority of *S. pombe* lncRNAs is poor. However, it has recently

been suggested that despite their poor sequence conservation, many lncRNAs may be conserved in secondary structure, thereby providing insights into the likelihood and possible mechanisms of function (Smith et al. 2013).

Moreover, if one considers that lncRNAs may exert their function through the act of being transcribed, rather than transcript itself (see Section 6.3.5), it is perhaps more meaningful to determine the extent to which transcription of lncRNA loci is retained across evolutionary lineages, rather than DNA sequence conservation. Transcriptional conservation of lncRNAs has recently been assessed using RNA-seq and H3K4Me3 ChIP-seq data from closely related rodent species (Kutter et al. 2012). Kutter et al. (2012) show that lncRNA transcription undergoes much greater gain and loss during rodent evolution compared with protein-coding genes, and such rapid transcriptional turnover of lncRNAs has been suggested to contribute to the evolution of lineage-specific gene expression. In addition, by comparing the genomes and transcriptomes across the fission yeast clade, Rhind et al. (2011) have shown that ~450 already annotated SPNCRNAs in *S. pombe* are conserved in location, but not necessarily sequence, in two or more species of this clade.

6.3.3 Subcellular locations of lncRNAs

Where lncRNAs reside in the cell can provide some insight into their possible functions. Although RNA-seq of the lncRNA content of subcellular fractions has not been performed in the current study, their low level of expression possibly suggests a nuclear location, as described below.

It has been reported that most annotated lncRNAs in *S. pombe* (SPNCRNAs) are expressed below 1 copy/cell (Marguerat et al. 2012). To analyse the expression of CUTs, RUTs and XUTs identified in the current study in similar terms, their expression was compared to 'zone 1' mRNAs (low abundance mRNAs detected at <0.5 copies/cell) from Marguerat et al. (2012). These mRNAs are enriched for meiotic differentiation functions, and it is well known that genes induced during meiosis are tightly repressed during proliferation. During log-phase growth in WT cells, the RPKMs of CUTs, RUTs and XUTs are much lower than for zone 1 mRNAs (Figure 6.3A). Although CUT, RUT and XUT expression is significantly

increased in *rrp6*, *dcr1* and *xrn1* mutants, the expression levels of these lncRNA classes are still lower than for zone 1 mRNAs (Figure 6.3B). Thus, most lncRNAs are lowly expressed, likely below 1 copy per cell. This suggests that such lncRNAs are unlikely to act as diffusible *trans*-acting factors under the conditions studied, and instead any function they have will likely be exerted via *cis*-acting mechanisms.

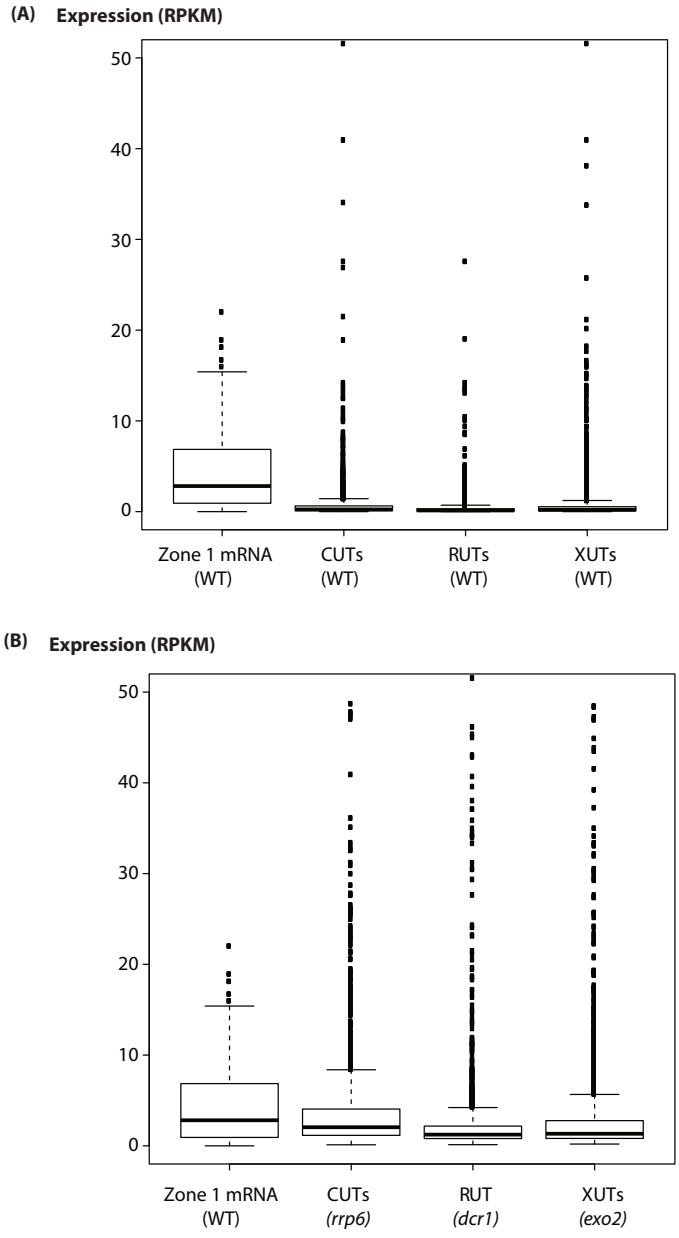

Figure 6.3: Expression of CUTs, RUTs and XUTs.

Expression in RPKM of CUTs, RUTs and XUTs in (A) vegetative WT cells, and (B) mutants, as indicated.

Thus, the majority of lncRNAs likely exert any effects co-transcriptionally in the nucleus. This is in agreement with the fact that impairing key nuclear lncRNA degradation pathways (either alone, *rrp6* and *dcr1*, or in combination, *rrp6/dcr1*) results in more severe lncRNA phenotypes than when the main cytoplasmic pathway is impaired (either alone, *exo2*, or in combination with a nuclear pathway *exo2/dcr1*). Furthermore, although it has been assumed that *S. cerevisiae* XUTs are exported to the cytoplasm where they are targeted by Xrn1 (van Dijk et al. 2011), their function could still be nuclear as recently suggested by Castelnuovo et al. (2013). Furthermore, by assessing the binding of different transcript classes to nuclear and cytoplasmic surveillance factors it has been shown that *S. cerevisiae* lncRNAs are bound predominantly by nuclear processing factors, while mRNAs are bound more extensively by nuclear export and cytoplasmic factors (Tuck and Tollervey 2013). This suggests that *S. cerevisiae* lncRNAs are likely largely confined to the nucleus.

Many lncRNAs whose mechanisms of action are well-characterised, such as HOTAIR (Rinn et al. 2007), use chromatin as a substrate to execute their biological function. It has therefore been postulated that lncRNAs associated with chromatin may be more likely to have biological functions. RNA-seq of chromatin fractions may therefore be an insightful approach.

6.3.4 lncRNA functions in meiosis and quiescence

A key finding of the current study is that expression profiles of lncRNAs vary with external stimuli. Although part of this is likely to be indirectly due to overall changes in protein coding-gene activities, it may also imply a widespread role of lncRNA transcription in gene regulation. Thus, although lncRNA differential expression is associated with mRNA differential expression, lncRNA expression profiles do not exactly mirror mRNA expression profiles. This suggests lncRNAs may exert some regulatory effect on mRNAs rather than representing transcriptional by-products. Secondly, CUTs, RUTs and XUTs are differentially enriched under different physiological conditions. For example, lncRNAs up-regulated at early meiotic time-points are predominantly CUTs and transcribed divergently from mRNAs, while those up-regulated at later meiotic time-points are enriched for RUTs and XUTs and transcribed antisense to mRNAs. Such distinct lncRNA

transcriptional profiles, with distinct genomic locations and targeted by different degradation pathways, argue against such lncRNA transcription representing random transcriptional noise.

The observed up-regulation of lncRNAs in quiescence is in agreement with a recent study describing a set of quiescence-induced lncRNAs in murine tissue culture (Bierhoff et al. 2014). Such lncRNAs have been shown to recruit H4K20 methyltransferase to ribosomal DNA, leading to chromatin compaction and subsequent repression of ribosomal genes upon growth factor deprivation or terminal differentiation.

Similarly the observed up-regulation of lncRNAs during meiosis is reminiscent of a class of budding yeast meiotic lncRNAs that are actively degraded during the mitotic cell-cycle by the nuclear exosome and stabilised as the cell proceeds into meiotic differentiation (Lardenois et al. 2011). Moreover, a well-studied example of a meiotic lncRNA in *S. pombe* is the meiRNA whose expression is increased upon entry into meiosis (Watanabe Y. and Yamamoto 1994). This lncRNA is transcribed from the *sme2* locus where it forms a nuclear dot structure together with its protein-binding partner Mei2. This dot structure promotes the progression of meiosis by sequestering and thereby inhibiting Mmi1, which is involved in the selective elimination of meiotic transcripts during vegetative growth (Shichino et al. 2014). Furthermore, expression of *sme2* lncRNA, and its retention at the *sme2* locus, has been shown to help mediate robust pairing of homologous chromosomes during meiosis (Ding et al. 2012).

Bidirectional lncRNAs which run antisense to upstream mRNAs were found to be enriched at early meiotic time-points, where they may act to couple the transcriptional regulation of the gene they are transcribed divergently from and the gene they run antisense to (Chapter 4). For mRNAs in such a set up, divergent mRNAs were enriched for growth GO-terms, while antisense mRNAs were enriched for meiotic GO-terms. Thus, such a lncRNA-mRNA arrangement may represent a mechanism to coordinate expression of growth and meiotic/stress genes. It is also in agreement with previous suggestions that antisense expression initiated from bidirectional promoters enables the spreading of regulatory signals from one locus to

neighbouring genes (Xu et al. 2011, Xu et al. 2009).

It may be of interest to extend the current study to other physiologically relevant cell states, for example lncRNA expression throughout the cell cycle, or lncRNA expression in response to environmental stresses such as heat-shock, oxidative stress or DNA damage. Such physiological perturbations are of relevance in understanding the mechanisms underlying disease progression in higher eukaryotes. The role of lncRNAs in human health and disease is of intense interest since a number of genome-wide association (GWA) studies have shown that variations associated with complex disorders often map to non-coding regions of the genome (Wapinski and Chang 2011).

6.3.5 Can pervasive transcription be functional?

From the current study, lncRNAs are pervasive throughout the *S. pombe* genome (85% of mRNAs can be associated with at least one lncRNA), lowly expressed (Section 6.3.3), poorly conserved (6.3.2), and are targeted by RNA surveillance machineries. Such findings could imply that most lncRNAs are by-products of transcription.

However, poorly conserved, low level, pervasive non-coding transcription is not necessarily incompatible with being functional, especially if one considers that any function may lie in the act of being transcribed rather than the transcript products themselves. Several studies have suggested that the process of lncRNA transcription, rather than the lncRNA product itself, may be functional. For example, lncRNA transcription may facilitate an open chromatin structure at protein-coding promoters, thereby increasing access to transcriptional activators and to Pol II (Hirota et al. 2008). It has also been reported that the lncRNA transcription can introduce co-transcriptional regulatory histone marks at loci they overlap (Kim T. et al. 2012, van Werven et al. 2012). In such cases, lncRNA expression level, how rapidly they are targeted for degradation, and their sequence conservation, are not necessarily indicators of functionality.

Moreover, even if few individual lncRNA transcripts are found to possess any

obvious regulatory function, a broad expression of genomes may hold advantages over an investment in fail-safe transcription initiation control (Jensen et al. 2013). For example, preventing large genomic regions from being completely silenced may prevent the formation of too tightly compacted chromatin domains, which would otherwise be hard to reopen. Secondly, pervasive RNAs, or their transcription, may aid the 3D shaping of chromosomes into active and repressive domains. Finally, it has also recently been suggested that low-level pervasive transcription may be part of a defence mechanism capable of directing an RNAi-dependent silencing response against transposable elements amplifying within in the genome (Cruz and Houseley 2014).

In summary, the current study provides an in-depth description of the lncRNA repertoire of *S. pombe*, and the pathways that regulate their expression. It seems likely that any regulatory functions mediated by most of these lncRNAs are in *cis*, nuclear and co-transcriptional. The current study provides a framework for the further functional characterisation of lncRNAs, and I will briefly outline future work that may help to achieve this.

6.4 Future Work

6.4.1 Further genome-wide data sets

6.4.1.1 RNA-seq

As already discussed, it may be of interest to extend the current study to other physiologically relevant cell states, for example lncRNA expression throughout the cell cycle, or lncRNA expression in response to environmental stresses such as heat-shock, oxidative stress or DNA damage.

In addition, given the findings of the current study point to a nuclear co-transcriptional function of lncRNAs, RNA-seq of the RNA content associated with chromatin fractions may present a method of enriching for lncRNAs which likely possess regulatory functions.

Finally, it is increasingly appreciated that gene expression in individual cells deviates significantly from the average behaviour of cell populations (Kaufmann and van

Oudenaarden 2007, Saliba et al. 2014). Thus, single-cell transcriptome profiling could revolutionise our understanding of genome regulation beyond population averages. Technologies for single-cell RNA-seq profiling are rapidly emerging, and their application to study lncRNA expression may be of interest.

6.4.1.2 ChIP-seq

Condition-specific transcription factor (TF) binding sites have been used as an indicator of active regulation of lncRNAs, which argues against lncRNAs being by-products of non-specific transcription in open chromatin regions. For example, Necsulea et al. (2014) have shown that in 11 tetrapod species homeobox TFs preferentially bind to lncRNA promoters, suggesting that such lncRNAs function in embryonic development. Thus analysing TF binding sites at promoter regions of lncRNAs, computationally by motif searching and experimentally by ChIP-seq, could represent an approach to identify actively regulated, and therefore potentially functional, lncRNAs.

As discussed above, ChIP-seq of H3K4Me3 and Pol II may provide some insight into the extent to which lncRNAs are transcriptionally regulated (Section 6.2.4). Similarly H3K9Me2 ChIP-seq will be of interest in determining whether lncRNA transcription is associated with formation of transient heterochromatin (Section 6.2.2).

6.4.1.3 Proteomics

Recently, Leong et al. (2014) have integrated global proteomics and RNA-seq data from *S. pombe* cells exposed to osmotic stress. Such data revealed elevated antisense lncRNA levels, predominantly arising from read-through at convergent genes, to be associated with a reduction in protein levels of these genes. Genes with reduced antisense levels and increased protein production were enriched for core environmental stress response (CESR) genes.

It would therefore be of interest to combine the transcriptomics data in the current study with global proteomics data. Firstly this would help elucidate whether lncRNA-mediated inhibitory effects on mRNA expression result in changes at the protein level. Secondly it would be informative as to whether any lncRNAs are

translated into small peptides (Duncan and Mata 2014, Ingolia et al. 2011, Ingolia et al. 2014, Wilson and Masel 2011).

6.4.2 Deletion and over-expression screens

Ultimately further work to determine functionality of lncRNAs lies in their systematic deletion and/or over-expression, and the effects such manipulations have on phenotypes such as cell growth and morphology. The current study serves as a basis for identifying lncRNAs most amenable to deletion or over-expression, and such experiments are already underway in the Bähler laboratory (Dr Michal Malecki, Dr Maria Rodriguez).

Follow-up studies to determine mechanisms underlying lncRNA functionality, such as ChIRP-seq (see Chapter 1) or RIP-seq to identify DNA and protein interacting partners, may then be considered.

References

Adachi N, Lieber MR. 2002. Bidirectional gene organization: a common architectural feature of the human genome. Cell 109: 807-809.

Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA. 2013. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature 499: 360-363.

Amberg DC, Goldstein AL, Cole CN. 1992. Isolation and characterization of RAT1: an essential gene of Saccharomyces cerevisiae required for the efficient nucleocytoplasmic trafficking of mRNA. Genes Dev 6: 1173-1189.

Amorim MJ, Cotobal C, Duncan C, Mata J. 2010. Global coordination of transcriptional control and mRNA decay during cellular differentiation. Mol Sys Biol 6: 380.

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

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

Atkinson SR, Marguerat S, Bahler J. 2012. Exploring long non-coding RNAs through sequencing. Semin Cell Dev Biol 23: 200-205.

Bahler J, Wu JQ, Longtine MS, Shah NG, McKenzie A, 3rd, Steever AB, Wach A, Philippsen P, Pringle JR. 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. Yeast 14: 943-951.

Beaulieu YB, Kleinman CL, Landry-Voyer AM, Majewski J, Bachand F. 2012. Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. PLoS Genet 8: e1003078.

Bierhoff H, Dammert MA, Brocks D, Dambacher S, Schotta G, Grummt I. 2014. Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing. Mol Cell 54: 675-682.

Birney E, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799-816.

Bitton DA, Grallert A, Scutt PJ, Yates T, Li Y, Bradford JR, Hey Y, Pepper SD, Hagan IM, Miller CJ. 2011. Programmed fluctuations in sense/antisense transcript ratios drive sexual differentiation in S. pombe. Mol Sys Biol 7: 559.

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

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25: 1915-1927.

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

Carninci P, et al. 2005. The transcriptional landscape of the mammalian genome. Science 309: 1559-1563.

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

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

Cawley S, et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116: 499-509.

Chang CT, Bercovich N, Loh B, Jonas S, Izaurralde E. 2014. The activation of the decapping enzyme DCP2 by DCP1 occurs on the EDC4 scaffold and involves a conserved loop in DCP1. Nucleic Acids Res 42: 5217-5233.

Chen D, Toone WM, Mata J, Lyne R, Burns G, Kivinen K, Brazma A, Jones N, Bahler J. 2003. Global transcriptional responses of fission yeast to environmental stress. Molecular biology of the cell 14: 214-229.

Chen HM, Rosebrock AP, Khan SR, Futcher B, Leatherwood JK. 2012. Repression of meiotic genes by antisense transcription and by Fkh2 transcription factor in Schizosaccharomyces pombe. PLoS One 7: e29917.

Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. 2011. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Mol Cell.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322: 1845-1848.

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

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

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol 8: e1000384.

Ding DQ, Okamasa K, Yamane M, Tsutsumi C, Haraguchi T, Yamamoto M, Hiraoka Y. 2012. Meiosis-specific noncoding RNA mediates robust pairing of homologous chromosomes in meiosis. Science 336: 732-736.

Duncan CD, Mata J. 2014. The translational landscape of fission-yeast meiosis and sporulation. Nat Struct Mol Biol 21: 641-647.

Dutrow N, Nix DA, Holt D, Milash B, Dalley B, Westbroek E, Parnell TJ, Cairns BR. 2008. Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. Nat Genet 40: 977-986.

Faulkner GJ, Carninci P. 2009. Altruistic functions for selfish DNA. Cell Cycle 8: 2895-2900.

Faulkner GJ, et al. 2009. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet 41: 563-571.

Frenk S, Oxley D, Houseley J. 2014. The Nuclear Exosome Is Active and Important during Budding Yeast Meiosis. PLoS One 9: e107648.

Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. Nat Rev Genet 10: 94-108.

Grenier St-Sauveur V, Soucek S, Corbett AH, Bachand F. 2013. Poly(A) tail-mediated gene regulation by opposing roles of Nab2 and Pab2 nuclear poly(A)-binding proteins in pre-mRNA decay. Molecular and cellular biology 33: 4718-4731.

Gudipati RK, Xu Z, Lebreton A, Seraphin B, Steinmetz LM, Jacquier A, Libri D. 2012. Extensive degradation of RNA precursors by the exosome in wild-type cells. Mol Cell 48: 409-421.

Gullerova M, Proudfoot NJ. 2008. Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. Cell 132: 983-995.

Gullerova M, Moazed D, Proudfoot NJ. 2011. Autoregulation of convergent RNAi genes in fission yeast. Genes Dev 25: 556-568.

Guo X, Lin M, Rockowitz S, Lachman HM, Zheng D. 2014. Characterization of human pseudogene-derived non-coding RNAs for functional potential. PloS one 9: e93972.

Guttman M, et al. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28: 503-510.

Guttman M, et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458: 223-227.

Haimovich G, Medina DA, Causse SZ, Garber M, Millan-Zambrano G, Barkai O, Chavez S, Perez-Ortin JE, Darzacq X, Choder M. 2013. Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. Cell 153: 1000-1011.

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

Halbach F, Reichelt P, Rode M, Conti E. 2013. The yeast ski complex: crystal structure and RNA channeling to the exosome complex. Cell 154: 814-826.

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013. Natural RNA circles function as efficient microRNA sponges. Nature 495: 384-388.

Harigaya Y, Tanaka H, Yamanaka S, Tanaka K, Watanabe Y, Tsutsumi C, Chikashige Y, Hiraoka Y, Yamashita A, Yamamoto M. 2006. Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature 442: 45-50.

He W, Parker R. 2000. Functions of Lsm proteins in mRNA degradation and splicing. Curr Opin Cell Biol 12: 346-350.

Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K. 2008. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature 456: 130-134.

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

Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M. 2008. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol Cell 32: 685-695.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of Mammalian proteomes. Cell 147: 789-802.

Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJ, Jackson SE, Wills MR, Weissman JS. 2014. Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. Cell Rep.

Jensen TH, Jacquier A, Libri D. 2013. Dealing with pervasive transcription. Mol Cell 52: 473-484.

Kaufmann BB, van Oudenaarden A. 2007. Stochastic gene expression: from single molecules to the proteome. Curr Opin Genet Dev 17: 107-112.

Keller C, Kulasegaran-Shylini R, Shimada Y, Hotz HR, Buhler M. 2013. Noncoding RNAs prevent spreading of a repressive histone mark. Nat Struct Mol Biol 20: 994-1000.

Khalil AM, et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 106: 11667-11672.

Kim T, Xu Z, Clauder-Munster S, Steinmetz LM, Buratowski S. 2012. Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. Cell 150: 1158-1169.

Kim TK, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. Nature 465: 182-187.

Koziol MJ, Rinn JL. 2010. RNA traffic control of chromatin complexes. Curr Opin Genet Dev 20: 142-148.

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

LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell 121: 713-724.

Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, Chu A, Davis RW, Steinmetz LM, Primig M. 2011. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proceedings of the National Academy of Sciences of the United States of America 108: 1058-1063.

Larochelle M, Lemay JF, Bachand F. 2012. The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. Nucleic Acids Res 40: 10240-10253.

Lee NN, et al. 2013. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. Cell 155: 1061-1074.

Lemay JF, D'Amours A, Lemieux C, Lackner DH, St-Sauveur VG, Bahler J, Bachand F. 2010. The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs. Mol Cell 37: 34-45.

Lemieux C, Marguerat S, Lafontaine J, Barbezier N, Bahler J, Bachand F. 2011. A Pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein. Mol Cell 44: 108-119.

Leong HS, Dawson K, Wirth C, Li Y, Connolly Y, Smith DL, Wilkinson CR, Miller CJ. 2014. A global non-coding RNA system modulates fission yeast protein levels in response to stress. Nat Commun 5: 3947.

Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods 7: 709-715.

Li X, Manley JL. 2006. Cotranscriptional processes and their influence on genome stability. Genes Dev 20: 1838-1847.

Ling J, Baibakov B, Pi W, Emerson BM, Tuan D. 2005. The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a cis-linked globin promoter. J Mol Biol 350: 883-896.

Lubas M, Damgaard CK, Tomecki R, Cysewski D, Jensen TH, Dziembowski A. 2013. Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation pathway of human cytoplasmic mRNA. EMBO J 32: 1855-1868.

Lyne R, Burns G, Mata J, Penkett CJ, Rustici G, Chen D, Langford C, Vetrie D, Bahler J. 2003. Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility, and processing of array data. BMC Genomics 4: 27.

Malecki M, Viegas SC, Carneiro T, Golik P, Dressaire C, Ferreira MG, Arraiano CM. 2013. The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. EMBO J 32: 1842-1854.

Marguerat S, Bahler J. 2010. RNA-seq: from technology to biology. Cell Mol Life Sci 67: 569-579.

Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R, Bahler J. 2012. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell 151: 671-683.

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

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

Mata J. 2013. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. RNA Biol 10: 1407-1414.

Mata J, Wilbrey A, Bahler J. 2007. Transcriptional regulatory network for sexual differentiation in fission yeast. Genome Biol 8: R217.

Mata J, Lyne R, Burns G, Bahler J. 2002. The transcriptional program of meiosis and sporulation in fission yeast. Nat Genet 32: 143-147.

Mattick JS. 2011. The central role of RNA in human development and cognition. FEBS Lett 585: 1600-1616.

Medina DA, Jordan-Pla A, Millan-Zambrano G, Chavez S, Choder M, Perez-Ortin JE. 2014. Cytoplasmic 5'-3' exonuclease Xrn1p is also a genome-wide transcription factor in yeast. Front Genet 5: 1.

Memczak S, et al. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495: 333-338.

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

Mitchell P. 2014. Exosome substrate targeting: the long and short of it. Biochem Soc Trans 42: 1129-1134.

Moazed D. 2009. Small RNAs in transcriptional gene silencing and genome defence. Nature 457: 413-420.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5: 621-628.

Moyle-Heyrman G, Zaichuk T, Xi L, Zhang Q, Uhlenbeck OC, Holmgren R, Widom J, Wang JP. 2013. Chemical map of Schizosaccharomyces pombe reveals species-specific features in nucleosome positioning. Proc Natl Acad Sci U S A 110: 20158-20163.

Muhlrad D, Decker CJ, Parker R. 1994. Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. Genes Dev 8: 855-866.

Nadal-Ribelles M, Sole C, Xu Z, Steinmetz LM, de Nadal E, Posas F. 2014. Control of Cdc28 CDK1 by a stress-induced lncRNA. Mol Cell 53: 549-561.

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

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

Ni T, Tu K, Wang Z, Song S, Wu H, Xie B, Scott KC, Grewal SI, Gao Y, Zhu J. 2010. The prevalence and regulation of antisense transcripts in Schizosaccharomyces pombe. PLoS One 5: e15271.

Okazaki Y, et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420: 563-573.

Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12: 87-98.

Perkel JM. 2013. Visiting "noncodarnia". Biotechniques 54: 301, 303-304.

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465: 1033-1038.

Ponjavic J, Oliver PL, Lunter G, Ponting CP. 2009. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLoS Genet 5: e1000617.

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. Science 322: 1851-1854.

Quinn JJ, Ilik IA, Qu K, Georgiev P, Chu C, Akhtar A, Chang HY. 2014. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. Nat Biotechnol.

Rallis C, Codlin S, Bahler J. 2013. TORC1 signaling inhibition by rapamycin and caffeine affect lifespan, global gene expression, and cell proliferation of fission yeast. Ageing Cell 12: 563-573.

Rands CM, Meader S, Ponting CP, 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: e1004525.

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: 1132-1138.

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

Rhind N, et al. 2011. Comparative functional genomics of the fission yeasts. Science 332: 930-936.

Rinn JL, et al. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129: 1311-1323.

Rodriguez-Gabriel MA, Watt S, Bahler J, Russell P. 2006. Upf1, an RNA helicase required for nonsense-mediated mRNA decay, modulates the transcriptional response to oxidative stress in fission yeast. Molecular and cellular biology 26: 6347-6356.

Roux AE, Quissac A, Chartrand P, Ferbeyre G, Rokeach LA. 2006. Regulation of chronological ageing in Schizosaccharomyces pombe by the protein kinases Pka1 and Sck2. Ageing Cell 5: 345-357.

Rustici G, Mata J, Kivinen K, Lio P, Penkett CJ, Burns G, Hayles J, Brazma A, Nurse P, Bahler J. 2004. Periodic gene expression program of the fission yeast cell cycle. Nat Genet 36: 809-817.

Sabatinos SA, Forsburg SL. 2010. Molecular genetics of Schizosaccharomyces pombe. Methods Enzymol 470: 759-795.

Saliba AE, Westermann AJ, Gorski SA, Vogel J. 2014. Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res 42: 8845-8860.

Schlackow M, Marguerat S, Proudfoot NJ, Bahler J, Erban R, Gullerova M. 2013. Genome-wide analysis of poly(A) site selection in Schizosaccharomyces pombe. RNA 19: 1617-1631.

Schmidt MJ, West S, Norbury CJ. 2011. The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. RNA 17: 39-44.

Schneider C, Kudla G, Wlotzka W, Tuck A, Tollervey D. 2012. Transcriptome-wide analysis of exosome targets. Mol Cell 48: 422-433.

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

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. 2008. Divergent transcription from active promoters. Science 322: 1849-1851.

Shah S, Wittmann S, Kilchert C, 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: 231-244.

Shichino Y, Yamashita A, 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: 140022.

Shimanuki M, Chung SY, Chikashige Y, Kawasaki Y, Uehara L, Tsutsumi C, Hatanaka M, Hiraoka Y, Nagao K, Yanagida M. 2007. Two-step, extensive alterations in the transcriptome from G0 arrest to cell division in Schizosaccharomyces pombe. Genes Cells 12: 677-692.

Shobuike T, Tatebayashi K, Tani T, Sugano S, Ikeda H. 2001. The dhp1(+) gene, encoding a putative nuclear 5'-->3' exoribonuclease, is required for proper chromosome segregation in fission yeast. Nucleic Acids Res 29: 1326-1333.

Shoemaker CJ, Green R. 2012. Translation drives mRNA quality control. Nat Struct Mol Biol 19: 594-601.

Skruzny M, Schneider C, Racz A, Weng J, Tollervey D, Hurt E. 2009. An endoribonuclease functionally linked to perinuclear mRNP quality control associates with the nuclear pore complexes. PLoS Biol 7: e8.

Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6: 31.

Smialowska A, Djupedal I, Wang J, Kylsten P, Swoboda P, Ekwall K. 2014. RNAi mediates post-transcriptional repression of gene expression in fission yeast Schizosaccharomyces pombe. Biochem Biophys Res Commun 444: 254-259.

Smith MA, Gesell T, Stadler PF, Mattick JS. 2013. Widespread purifying selection on RNA structure in mammals. Nucleic Acids Res 41: 8220-8236.

St-Andre O, Lemieux C, Perreault A, Lackner DH, Bahler J, Bachand F. 2010. Negative regulation of meiotic gene expression by the nuclear poly(A)-binding protein in fission yeast. The Journal of biological chemistry 285: 27859-27868.

Steinmetz EJ, Conrad NK, Brow DA, Corden JL. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. Nature 413: 327-331.

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

Sugiyama T, Sugioka-Sugiyama R. 2011. Red1 promotes the elimination of meiosis-specific mRNAs in vegetatively growing fission yeast. EMBO J 30: 1027-1039.

Sun M, Schwalb B, Pirkl N, Maier KC, Schenk A, Failmezger H, Tresch A, Cramer P. 2013a. Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol Cell 52: 52-62.

Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Lariviere L, Maier KC, Seizl M, Tresch A, Cramer P. 2012. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res 22: 1350-1359.

Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT. 2013b. Jpx RNA activates Xist by evicting CTCF. Cell 153: 1537-1551.

Szankasi P, Smith GR. 1996. Requirement of S. pombe exonuclease II, a homologue of S. cerevisiae Sep1, for normal mitotic growth and viability. Curr Genet 30: 284-293.

Tashiro S, Asano T, Kanoh J, Ishikawa F. 2013. Transcription-induced chromatin association of RNA surveillance factors mediates facultative heterochromatin formation in fission yeast. Genes Cells 18: 327-339.

Tay Y, et al. 2011. Coding-Independent Regulation of the Tumor Suppressor PTEN by Competing Endogenous mRNAs. Cell 147: 344-357.

Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D. 2006. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol Cell 23: 853-864.

Tsutsui Y, Morishita T, Iwasaki H, Toh H, Shinagawa H. 2000. A recombination repair gene of Schizosaccharomyces pombe, rhp57, is a functional homolog of the Saccharomyces cerevisiae RAD57 gene and is phylogenetically related to the human XRCC3 gene. Genetics 154: 1451-1461.

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

Tudek A, et al. 2014. Molecular basis for coordinating transcription termination with noncoding RNA degradation. Mol Cell 55: 467-481.

van Dijk EL, et al. 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature 475: 114-117.

van Werven FJ, Neuert G, Hendrick N, Lardenois A, Buratowski S, van Oudenaarden A, Primig M, Amon A. 2012. Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. Cell 150: 1170-1181.

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

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297: 1833-1837.

Wang GZ, Lercher MJ, Hurst LD. 2011. Transcriptional coupling of neighboring genes and gene expression noise: evidence that gene orientation and noncoding transcripts are modulators of noise. Genome Biol Evol 3: 320-331.

Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. Mol Cell 43: 904-914.

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

Wang SW, Stevenson AL, Kearsey SE, Watt S, Bahler J. 2008. Global role for polyadenylation-assisted nuclear RNA degradation in posttranscriptional gene silencing. Molecular and cellular biology 28: 656-665.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57-63.

Wapinski O, Chang HY. 2011. Long noncoding RNAs and human disease. Trends Cell Biol 21: 354-361.

Watanabe T, Miyashita K, Saito TT, Yoneki T, Kakihara Y, Nabeshima K, Kishi YA, Shimoda C, Nojima H. 2001. Comprehensive isolation of meiosis-specific genes identifies novel proteins and unusual non-coding transcripts in Schizosaccharomyces pombe. Nucleic Acids Res 29: 2327-2337.

Watanabe Y, Yamamoto M. 1994. S. pombe mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. Cell 78: 487-498.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453: 1239-1243.

Wilson BA, Masel J. 2011. Putatively noncoding transcripts show extensive association with ribosomes. Genome Biol Evol 3: 1245-1252.

Wlotzka W, Kudla G, Granneman S, Tollervey D. 2011. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J 30: 1790-1803.

Woolcock KJ, Gaidatzis D, Punga T, Buhler M. 2011. Dicer associates with chromatin to repress genome activity in Schizosaccharomyces pombe. Nat Struct Mol Biol 18: 94-99.

Woolcock KJ, Stunnenberg R, Gaidatzis D, Hotz HR, Emmerth S, Barraud P, Buhler M. 2012. RNAi keeps Atf1-bound stress response genes in check at nuclear pores. Genes Dev 26: 683-692.

Wu X, Sharp PA. 2013. Divergent transcription: a driving force for new gene origination? Cell 155: 990-996.

Wyers F, et al. 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121: 725-737.

Xu Z, Wei W, Gagneur J, Clauder-Munster S, Smolik M, Huber W, Steinmetz LM. 2011. Antisense expression increases gene expression variability and locus interdependency. Mol Syst Biol 7: 468.

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

Xue Z, Ye Q, Anson SR, Yang J, Xiao G, Kowbel D, Glass NL, Crosthwaite SK, Liu Y. 2014. Transcriptional interference by antisense RNA is required for circadian clock function. Nature doi:10.1038/nature13671.

Yamanaka S, Yamashita A, Harigaya Y, Iwata R, Yamamoto M. 2010. Importance of polyadenylation in the selective elimination of meiotic mRNAs in growing S. pombe cells. The EMBO journal 29: 2173-2181.

Yamanaka S, Mehta S, Reyes-Turcu FE, Zhuang F, Fuchs RT, Rong Y, Robb GB, Grewal SI. 2013. RNAi triggered by specialized machinery silences developmental genes and retrotransposons. Nature 493: 557-560.

Yamashita A, Chang TC, Yamashita Y, Zhu W, Zhong Z, Chen CY, Shyu AB. 2005. Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol 12: 1054-1063.

Yanagida M. 2009. Cellular quiescence: are controlling genes conserved? Trends Cell Biol 19: 705-715.

Yanagida M, Ikai N, Shimanuki M, Sajiki K. 2011. Nutrient limitations alter cell division control and chromosome segregation through growth-related kinases and phosphatases. Philos Trans R Soc Lond B Biol Sci 366: 3508-3520.

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

Zhang K, Fischer T, Porter RL, Dhakshnamoorthy J, Zofall M, Zhou M, Veenstra T, Grewal SI. 2011. Clr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. Science 331: 1624-1627.

Zofall M, Yamanaka S, Reyes-Turcu FE, Zhang K, Rubin C, Grewal SI. 2012. RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. Science 335: 96-100.

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