THE UNIVERSITY
of EDINBURGH

Genome-wide identification of non-canonical targets of messenger RNA synthesis and turnover factors in *Saccharomyces cerevisiae*

Alex Tuck

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

2012

Declaration

I declare that this thesis was composed by myself. The research in this thesis is my own, unless otherwise stated, and has not been submitted for any other degree or professional qualification.

Alex Tuck, 2013

In loving memory of my grandmother, Margaret Tuck

Acknowledgements

I would first like to thank David for outstanding supervision. It has been an incredible privilege to work in his lab, and he has been a constant source of inspiration. I also thank everyone in the Tollervey lab, past and present, for their advice and support – Sander, Greg, Wiebke, Claudia, Rebecca, Simon, Liz, Laura, Jai, Clementine, Tanya, Konstantin, Dariusz and Aziz.

I would also like to thank my thesis committee, Jean Beggs, Kevin Hardwick and Steve West, for their encouragement and helpful comments. Additionally, I thank David Barrass for help with experiments.

I am very grateful to the Wellcome Trust for funding this 4 year PhD Programme in Cell Biology, and Karen Traill, for her excellent organisation of the programme.

Finally, I would like to thank my family for their unconditional love and support. I am also extremely lucky to have met Ausma while studying in Edinburgh, and am grateful for her support and encouragement during the last four, very happy years.

Contents

Abstract

Pervasive transcription is widespread amongst eukaryotic genomes, and produces long non-coding RNAs (lncRNAs) in addition to classically annotated transcripts such as messenger RNAs (mRNAs). LncRNAs are heterogeneous in length and map to intergenic regions or overlap with annotated genes. Analogous to mRNAs, lncRNAs are transcribed by RNA polymerase II, regulated by common transcription factors, and possess 5' caps and perhaps 3' poly(A) tails. However, lncRNAs perform distinct functions, acting as scaffolds for ribonucleoprotein complexes or directing proteins to nucleic acid targets. The act of transcribing a lncRNA can also affect the local chromatin environment. Furthermore, whereas mRNAs are predominantly turned over in the cytoplasm, both nuclear and cytoplasmic pathways reportedly participate in lncRNA degradation. In this study, I address the question of when and how lncRNAs and mRNAs are distinguished in the cell.

Messenger RNAs interact with a defined series of protein factors governing their production, processing and decay, and I hypothesised that lncRNAs might be similarly regulated. I therefore sought to determine which mRNA-binding proteins, if any, also bind lncRNAs. I reasoned that this would reveal the point at which lncRNAs and mRNAs diverge, and how differences in their biogenesis and turnover equip them for different roles. I selected factors from key stages of mRNA metabolism in *Saccharomyces cerevisiae*, and identified their transcriptome-wide targets using CRAC (crosslinking and analysis of cDNAs). CRAC can detect interactions with low abundance transcripts under physiological conditions, and reveal where within each transcript a protein is bound.

Analyses of binding sites in mature mRNAs and intron-containing pre-mRNAs revealed the order in which the tested factors interact with mRNAs, and which region they bind. The poly(A)-binding protein Nab2 bound throughout mRNAs, consistent with an architectural role, whereas the cytoplasmic decay factors Xrn1 and Ski2 bound to poly(A) tails, which might act as hubs to coordinate turnover. The RNA packaging factors Tho2 and Gbp2, and

nuclear surveillance factors Mtr4 and Trf4 bound abundantly to intron-containing pre-mRNAs, indicating that they act during or shortly after transcription.

The tested factors bound lncRNAs to various extents. LncRNA binding was most abundant for Mtr4 and Trf4, moderate for Tho2, Gbp2, the cap binding complex component Sto1, and the 3' end processing factors Nab2, Hrp1 and Pab1, and lowest for Xrn1, Ski2 and the export receptor Mex67. This suggests that early events in lncRNA and mRNA biogenesis are similar, but unlike mRNAs, most lncRNAs are retained and degraded in the nucleus. Analyses of two documented classes of lncRNA, cryptic unstable transcripts (CUTs) and stable unannotated transcripts (SUTs), revealed some differences. SUTs were most similar to mRNAs, with canonical cleavage and polyadenylation signals flanking their 3' ends, and poly(A) tails bound by the poly(A)-binding protein Pab1. CUTs lacked these characteristics, and in comparison to SUTs bound more abundantly to Mtr4 and Trf4 and less so to Ski2, Xrn1 and Mex67. Furthermore, CUTs accumulated upon Hrp1 depletion, suggesting that Hrp1 functions non-canonically to promote CUT turnover.

Mtr4, Trf4 and Nab2 also bound abundantly to promoter-proximal RNA fragments generated from ~1000 protein coding genes. These fragments possessed short oligo(A) tails (hallmarks of nuclear surveillance substrates), were not bound to cytoplasmic factors, and apparently correspond to a population of ~150-200 nt promoter-proximal lncRNAs. Notably, CRAC analyses of Mtr4 and Sto1 targets in yeast subjected to a media shift revealed widespread changes in the abundance and surveillance of mRNAs, promoter-proximal transcripts and CUTs, which at many loci were arranged in a complex transcriptional architecture.

Overall, the transcriptome-wide binding analyses presented here reveal that lncRNAs diverge from mRNAs prior to export, and are predominantly retained in the nucleus. Transcript fate is apparently determined during 3' end processing, with CUTs diverging from mRNAs early in transcription via a distinct termination pathway coupled to rapid turnover, and SUTs diverging during or shortly after cleavage and polyadenylation, making them more

stable and perhaps prone to escape to the cytoplasm. Promoter-proximal transcripts might arise from termination associated with an early checkpoint in Pol II transcription. The diverse behaviours of lncRNAs arise from their association with distinct subsets of RNA binding proteins, some of which perform different roles when bound to different types of transcript. In conclusion, my results provide the foundation for a mechanistic understanding of how distinct classes of non-coding Pol II transcripts are produced, and how they can perform diverse functions throughout the nucleus.

1: Introduction

1.1 Pervasive transcription

Eukaryotic genomes accommodate a vast amount of information within the DNA sequence of chromosomes. Readout of this information involves transcription, whereby a genomic locus acts as a template for RNA Polymerase I, II or III (Pol I, II or III) to generate an RNA copy (transcript). A "gene" was classically defined as the DNA region directly encoding the transcript, together with adjacent regulatory elements, and transcription was thought to be restricted to protein-coding genes (generating messenger RNAs (mRNAs)) and a small number of genes encoding stable, structural RNAs. These include ribosomal and transfer RNAs (rRNAs and tRNAs), which participate in protein synthesis; small nucleolar RNAs (snoRNAs), which guide modifications of other classes of RNA; and small nuclear RNAs (snRNAs), which participate in splicing.

However, whole-genome microarrays, deep sequencing, and other methods that can detect enormous numbers of transcripts in a single experiment, have recently revealed that transcription is not restricted to annotated genes, but pervades most, if not all, genomic loci (Table 1.1). This generates "non-[protein]-coding RNAs" (ncRNAs), which lack protein-coding capacity and are distinct from well-characterised structural RNAs (rRNAs, tRNAs, snRNAs and snoRNAs). Non-coding RNAs map to regions between genes (intergenic), or overlap partially or fully with one or more annotated genes, and range from <20 nt to >50 kb (Kapranov et al, 2010; Taft et al, 2011). Thus the transcriptome is overwhelmingly complex. Furthermore, most transcripts have alternative 5' and 3' ends (ENCODE, 2007; Ozsolak et al, 2010; Yoon et al, 2010), and "ultra-deep" sequencing is beginning to reveal yet another layer of very rare transcripts (Mercer et al, 2012), suggesting that more transcripts await discovery.

Although ncRNAs were initially dismissed as experimental artefacts or biological noise, functions have now been reported for numerous ncRNAs (Table 1.2). Many ncRNAs are conserved and exhibit expression patterns that are dependent on environmental conditions, cell cycle state or tissue type, and may be correlated with proximal genes (Granovskaia et al, 2010; Lardenois et al, 2011; Ørom et al, 2010; Yassour et al, 2010). Notably, ncRNAs do not just act *in cis*, on the chromatin from which they are transcribed, but can function throughout the nucleus in diverse roles. This suggests that RNA is not just a messenger directing protein synthesis, but is a central, active participant in many cellular functions. The pervasive and interleaved nature of transcription and the plethora of functional ncRNAs refute the mRNA-centric, modular definition of a gene. A better definition is perhaps "a system comprising a genomic region with the corresponding network of control regions and ensemble of transcripts" (Tuck et al, 2011).

In this study, I investigate the biogenesis and properties of non-coding RNAs, to improve our understanding of where and how they function, and how they are regulated.

Technology	Organism	Coverage	Reference
Tiling array	Human	~25%	(Kapranov et al, 2002)
Sequencing of full-length cDNAs or 5'/3' tags	Human	62.5%	(Carninci et al, 2005)
RNA-Seq	*S. pombe*	94 %	(Wilhelm et al, 2008)
RNA-Seq	*S. cerevsiaie*	75 %	(Nagalakshmi et al, 2008)
Tiling array	*S. cerevisiae*	85 %	(David et al, 2006)
RNA-Seq	Fly	75 %	(Graveley et al, 2011)
Tiling array, tag sequencing of cap-selected RNA	Human	93 %	(ENCODE, 2007)

Table 1.1: Recent estimates of the proportion of the genome that is transcribed ("coverage").

Table 1.2: Documented functions of lncRNAs, either direct or arising from the act of lncRNA transcription. *S. cerevisiae* lncRNAs indicated in green, *S. pombe* lncRNAs in orange, and mammalian lncRNAs in red. To date, most documented functions of lncRNAs in *S. cerevisiae* are dependent on the act of lncRNA transcription, rather than the lncRNA transcript itself.

lncRNA	Function	Mechanism	Reference
ZRR1	Represses expression of zinc-dependent factors during zinc deficiency	Transcription through *ADH1* promoter displaces transcriptional activator Rap1	(Bird et al, 2006)
PWR1 and ICR1	Regulate morphological heterogeneity of clonal populations of yeast	Transcription interference between reciprocally expressed lncRNAs at the *FLO11* locus; ICR1 resets chromatin	(Bumgarner et al, 2012)
IGS1-F and ISG2-R ncRNAs	rDNA copy number control	Displace cohesin, which aligns sister chromatids to prevent rDNA copy number change via homologous recombination	(Kobayashi et al, 2005)
GAL10 ncRNA	Reduces expression of galactose utilisation proteins	Set1-dependent Rpd3S recruitment (by H3K4me3, or via transcription-dependent H3K36me3) leads to inhibitory deacetylation, which delays induction by galactose and lowers steady state level of *GAL1* and *GAL10* mRNA	(Houseley et al, 2008; Pinskaya et al, 2009)
SRG1	Suppresses serine biosynthesis in the presence of serine	Nucleosome deposition over *SER3* promoter in the wake of ncRNA transcription suppresses transcription; requires Spt6/16 histone chaperones and the HMG-like protein Spt2	(Hainer et al, 2012; Hainer et al, 2011; Martens et al, 2004; Martens et al, 2005; Pruneski et al, 2011; Thebault et al, 2011; Thiebaut et al, 2006; Thompson et al, 2007)
PHO5 as-ncRNA	Activates transcription in response to phosphate starvation	Nucleosome eviction facilitates induction of *PHO5* (but not steady-state chromatin state)	(Uhler et al, 2007)
SUT719	Represses *SUR7* low-level sense expression	Antisense transcription across *SUR7* TSS	(Xu et al, 2011)
IGS1-R ncRNA	rDNA copy number control	Recruits Trf4 which influences homologous recombination	(Houseley et al, 2007)
IGS ncRNAs	Formation of rDNA 35S and 5S chromatin domains	Stalled Pol II mediates chromatin looping, to confine Pol I and Pol III to 35S and 5S respectively	(Mayan et al, 2010)
KCS1 as-ncRNA	Truncate Kcs1 protein under phosphate starvation, to alter IP7 synthesis	Sense:antisense interaction diverts translation initiation downstream	(Nishizawa et al, 2008)
RME2	Represses entry into meiosis	Antisense transcription through 5' region of ORF inhibits elongation of *IME4*	(Gelfand et al, 2011; Hongay et al, 2006)
IRT1	Represses *IME1* to prevent sporulation in haploid cells	Directs Set1-dependent H3K4 dimethylation (recruits Set3C) and Set2-dependent H3K36 trimethylation (recruits Rpd3S), resulting in histone deacetylation, repressive chromatin (nucleosome deposition) and exclusion of transcription factors such as Pog1	(van Werven et al, 2012)
DCI1 uCUT	Affects kinetics of *DCI1* transient induction in galactose	Upstream CUT that extends across ORF; H3K4me2 leads to Set3C-dependent histone deacetylation	(Kim et al, 2012)
RTL	Suppress Ty1 retrotransposition	Transcriptional effect via H3K4 methylation/histone acetylation, or post-transcriptional interference in VLP assembly	(Berretta et al, 2008; Matsuda et al, 2009)
PHO84 as-ncRNA	Reduces *PHO84* expression during aging	Hda1/2/3 recruitment in cis (elevated by Rrp6 reduction), and silencing in trans (requiring a homologous region)	(Camblong et al, 2009; Camblong et al, 2007)

lncRNA	Function	Mechanism	Reference
IMD2 CUT	Regulates nucleotide biogenesis	Guanine nucleotide deprivation promotes productive initiation from downstream TSS_{mRNA}, rather than the non-productive upstream TSS_{CUT} which is in competition	(Jenks et al, 2008; Kuehner et al, 2008)
ASP3 ncRNA	Contributes to the response to nitrogen limitation	Intragenic, sense lncRNA within *ASP3* ORF is required for efficient *ASP3* transcription initiation, which is induced by GATA TFs in response to nitrogen limitation	(Huang et al, 2010)
XUTs	Antisense repression	Direct Set1-dependent H3K4 dimethylation, which can repress via recruitment of the Set3C HDAC	(Kim et al, 2012; van Dijk et al, 2011)
URA2 CUT	Regulates nucleotide biogenesis	Upstream initiation (TSS_{CUT}) exerts a constant repression on downstream productive initiation (TSS_{mRNA}); uracil deprivation activates TSS_{mRNA} but has no effect on transcription of the CUT	(Thiebaut et al, 2008)
TEL05L	Function unknown	Antisense to putative DNA helicase	(Houseley et al, 2007)
fbp1 ncRNA	Activate transcription	Displace repressors and remodel chromatin	(Hirota et al, 2008)
IGS ncRNAs	Nucleolar detention of proteins	Nucleolar detention motif in proteins binds IGS ncRNAs	(Audas et al, 2012)
pRNAs	rDNA silencing	~200 nt ncRNAs from the mammalian rDNA promoter that form a triplex, which displaces the transcription factor TTF1 and is specifically recognised by Dnmt3b	(Schmitz et al, 2010)
Alpha satellite repeat lncRNAs	Formation of pericentric heterochromatin	lncRNA acts as a scaffold to accumulate and target the repressive protein HP1	(Maison et al, 2011)
Enhancer RNAs	Activate distal transcription	Enhancer-promoter looping and recruitment of transcription factors	(Kim et al, 2010b; Ørom et al, 2010; Wang et al, 2011a)
1/2sbsRNA	Post-transcriptional mRNA downregulation	Staufen 1 recognises dsRNA binding sites comprising a 1/2sbsRNA bound to an Alu element (each contributes half a binding site)	(Gong et al, 2011)
Gas1	Suppresses glucocorticoid induction of anti-apoptotic genes, to sensitise arrested cells to apoptosis	Acts as a decoy response element for the glucocorticoid receptor	(Kino et al, 2010)
NEAT1 (MENε/β)	Paraspeckle formation, involved in nuclear retention of hyperedited Alu-containing RNAs	NEAT1 lncRNAs seed paraspeckle formation; adopt an ordered structure and bind paraspeckle proteins	(Chen et al, 2009; Mao et al, 2011; Murthy et al, 2010; Saha et al, 2006; Souquere et al, 2010)(Clemson *et al*, 2009) (Sunwoo *et al*, 2009) (Sasaki *et al*, 2009)
Linc-MD1	Timing of muscle differentiation	Acts as a microRNA "sponge" for miR-133 and miR-135	(Cesana et al, 2011)
LUST	Prevents premature termination	Hybridises to *RBM5* sense strand mRNA and masks a termination signal	(Rintala-Maki et al, 2009)
SRA	Coactivates nuclear receptors	Highly structured RNA that acts as a scaffold for nuclear receptor complexes	(Novikova et al, 2012)
H19	Regulation of placental growth during early development	microRNA processed from H19 exon 1 is able to regulate genes in trans; HuR regulates this processing	(Keniry et al, 2012)
HAR1F	Associated with neuronal function	Highly structured and rapidly evolving; mechanism unknown	(Beniaminov et al, 2008; Johnson et al, 2010; Pollard et al, 2006)
NRON	Regulate nucleo-cytoplasmic trafficking	Associates with import receptors to prevent NFAT nuclear accumulation	(Willingham et al, 2005)
DHFR	Downregulate DHFR expression	Association with promoter DNA and TFIIB disrupts PIC assembly	(Martianov et al, 2007)

lncRNA	Function	Mechanism	Reference
TERRA	Inhibit telomerase activity and contribute to telomere structure	Binds protein and RNA components of telomerase and inhibits; forms a G-quadruplex that associates with the telomere protein TRF2	(Biffi et al, 2012; Redon et al, 2010)
TUG1	Component of Polycomb bodies; retinal development (photoreceptor gene expression)	Physically recruit genes to PcG bodies via interaction with methylated Pc2 (PRC1); changes Pc2 chromodomain specificity to H3K27me2; interacts w/ repressive proteins	(Ingolia et al, 2011; Khalil et al, 2009; Yang et al, 2011b)
NEAT2 (MALAT1)	Component of interchromatin granules/nuclear speckles	Physically recruit genes to ICGs via interaction with unmethylated Pc2 (PRC1); interact w/ activatory proteins; change Pc2 chromodomain specificity to acetyl marks; also affects genes *in cis*	(Nakagawa et al, 2012; Nakagawa et al, 2011; Yang et al, 2011b; Zhang et al, 2012a)
CCND1 PROMPTs	Gene repression	Allosterically activate TLS, a co-repressor at the CCND1 locus	(Wang et al, 2008)
lincRNA-p21	Repression of hundreds of genes in response to p53 activation; pro-apoptotic	Physical interaction with hnRNP K targets it to promoters; translational repression via dsRNA, recruiting translational repressors or causing ribosome drop-off	(Huarte et al, 2010; Yoon et al, 2012)
PANDA	Anti-apoptotic	Sequesters pro-apoptotic TF NF-YA; expressed from the *CDKN1A* locus alongside lincRNA-p21 and *CDKN1A*	(Hung et al, 2011)
ANRIL	Repression of tumour suppressor genes	In the *INK4b/ARF/INK4a* locus, recruits PRC2 to $p15^{INK4B}$ and PRC1 and PRC2 to $p14^{INK4A}$	(Kotake et al, 2011; Yap et al, 2010)
HOTAIR	Homeotic gene expression	Expressed from *HOXC*; binds and targets LSD1 and PRC2, to nucleate genome-wide PRC2 domains (including within *HOXD*); many binding sites	(Chu et al, 2011; Gupta et al, 2010; Kogo et al, 2011; Tsai et al, 2010)
HOTTIP	Activation of *HOXA* genes	Gene looping targets *HOTTIP*-tethered H3K4 methylation activity to activate *HOXA* promoters	(Wang et al, 2011c)
Bxd ncRNA	Hox gene control	Transcriptional interference	(Brock et al, 2009; Petruk et al, 2006)
AIR	Imprinting at the *Igf2r* locus	Recruits G9a (H3K9 methylation) and interacts with *Slc22a3* promoter for silencing; silences *Igf2r* by an independent mechanism	(Nagano et al, 2008)
Kcnq1ot1	Imprinting at the *Kcnq1* locus	Establishes nuclear silencing domain; recruits G9a and PRC2 to silence eight genes in the *Kcnq1* locus; some evidence suggests a transcript-independent mechanism	(Golding et al, 2011; Mohammad et al, 2012; Pandey et al, 2008; Redrup et al, 2009)
Gtl2 ncRNA	Imprinting of the *Dlk1* gene	Recruits PRC2 to *cis* locus	(Zhao et al, 2010)
roX1 and roX2	Drosophila dosage compensation	Target DCC to many specific sites on X chromosome to promote histone acetylation and increased Pol II initiation; roX2 has many binding sites	(Chu et al, 2011; Conrad et al, 2012a; Conrad et al, 2012b)
Xist	Human X inactivation	Recruits PRC2 (via RepA domains) to inactive X chromosome (Xi) via YY1 (binds scaffold associated regions) and hnRNP U (a transcription factor); coats Xi	(Hasegawa et al, 2010; Jeon et al, 2011; Jeon et al, 2012)
Tsix	Prevents inactivation of Xa	Competes with a short RepA-containing lncRNA for PRC2 binding to suppress Xist promoter	(Zhao et al, 2008)
TERC	Telomere maintenance	Human telomerase RNA provides a scaffold for telomerase proteins, and occupies telomere and Wnt pathway genes	(Chu et al, 2011)

1.2 Origins of non-coding RNAs

Although non-coding RNAs are heterogeneous, distinct classes might exist with common origins, properties and functions. High-throughput studies detect ncRNAs longer or shorter than, but generally not spanning, 200 nt, which has led to the definition of "long" ncRNAs (lncRNAs; >200 nt) and short ncRNAs (<200 nt). Currently, the Argonaute-associated regulatory small ncRNAs are the only well-defined class. These ~20-30 nt RNAs include microRNAs, Piwi-interacting RNAs and endogenous small interfering RNAs, all of which direct Argonaute family proteins to RNA targets via base-pairing (Qureshi et al, 2012). In this study, however, I focus on the numerous non-Argonaute-associated ncRNAs. These are poorly characterised, but generally originate from particular genomic features, such as promoters (Figure 1.1).

Promoter-proximal pausing

Eukaryotic promoters generate several classes of ncRNA, perhaps due to the less condensed packaging of DNA in promoter regions. Generally, DNA wraps around histone proteins to form nucleosomes, the basic unit of chromatin, but promoters contain a nucleosome-free region (NFR), flanked by upstream (-1) and downstream (+1) nucleosomes (Cairns, 2009). This enables binding of a transcription pre-initiation complex (PIC), comprised of the basal transcription factors and Pol II, either upstream of or just within the +1 nucleosome (Rhee et al, 2012). In metazoa, the +1 nucleosome and combined activity of the transcription factors NELF (negative elongation factor) and DSIF (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity-inducing factor) elicit promoter-proximal polymerase pausing ~20-60 nts downstream of the TSS (Core et al, 2008; Peterlin et al, 2006). This leads to the generation of short (~18 to ~90 nt) ncRNAs, including TSSa-RNAs and tiRNAs (Seila et al, 2008; Taft et al, 2009), perhaps involving cleavage by the elongation factor TFIIS (Taft et al, 2011). This mechanism is apparently not conserved in *S. cerevisiae*, which lack a

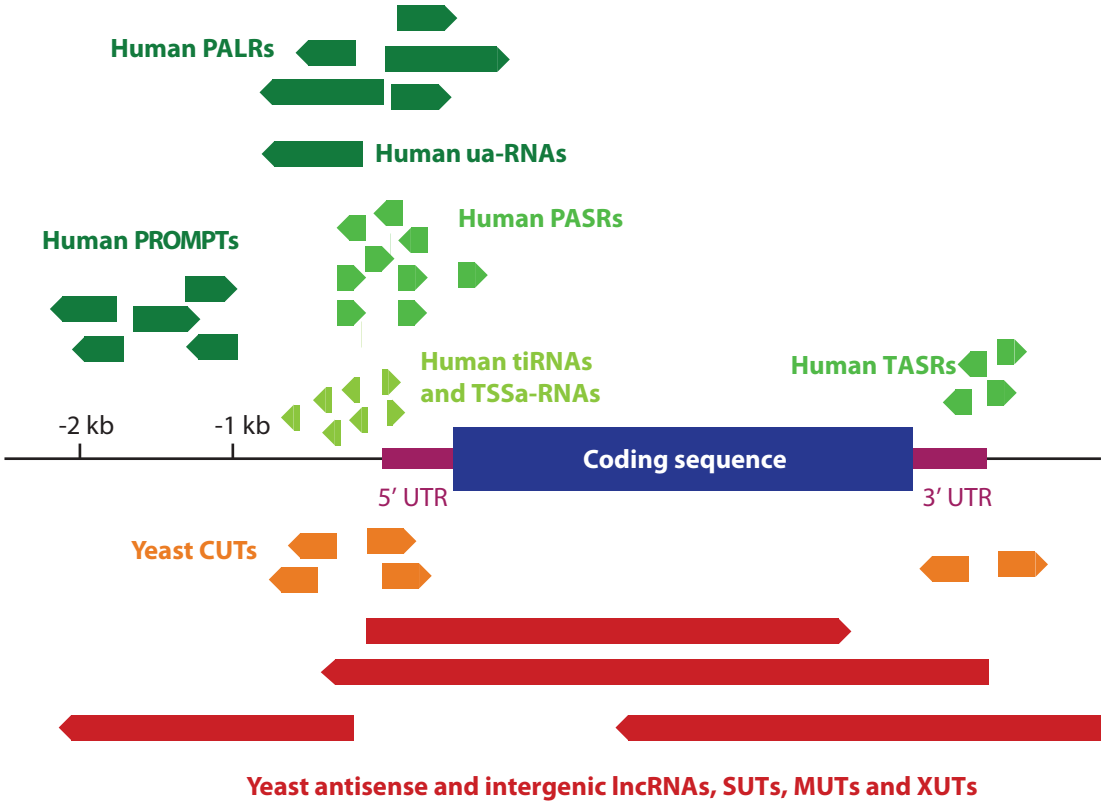


Figure 1.1: Origins of non-coding RNAs. Non-coding RNAs (ncRNAs) of various sizes are transcribed from nucleosome free regions associated with the 5' and 3' ends of protein coding genes. In yeast, these include cryptic unstable transcripts (CUTs) (Neil et al, 2009), which are ~200-500 nt, and longer, more stable ncRNAs, including stable unannotated transcripts (SUTs) (Xu et al, 2009), meiotic unannotated transcripts (MUTs) (Lardenois et al, 2011), Xrn1-sensitive unstable transcripts (XUTs) (van Dijk et al, 2011), and antisense and intergenic lncRNAs (Granovskaia et al, 2010; Yassour et al, 2010). In humans, ncRNAs are transcribed from the 5' end of genes in both directions, including transcription initiation RNAs (tiRNAs) (~18 nt) (Taft et al, 2011), transcription start site associated RNAs (TSSa-RNAs) (~50-90 nt) (Seila et al, 2008), promoter-associated long and short RNAs (PALRs and PASRs) (50-200 nt) (Kanhere et al, 2010; Kapranov et al, 2007), and upstream antisense RNAs (ua-RNAs) (up to 1 kb) (Flynn et al, 2011). Promoter upstream transcripts (PROMPTs) (Preker et al, 2011; Preker et al, 2008) (several hundred nts) are transcribed ~1.5 kb upstream, and termini-associated RNAs (TASRs) (50-200 nt) (Kapranov et al, 2007) at the 3' end of genes.

NELF homologue and do not express short promoter-associated RNAs (Preker et al, 2011). Nonetheless, an accumulation of elongation-competent Pol II ~100-500 bp downstream of many promoters in yeast suggests that pausing does occur (Churchman et al, 2011; McKinlay et al, 2011). A recent high resolution study of Pol II pausing in *Drosophila* found that whereas "focused" pausing arises from interactions between Pol II and the protein complex assembled upon promoter elements, some genes exhibit a more dispersed mode of pausing attributable partly to nucleosomal barriers (Kwak et al, 2013). Pol II pausing in *S. cerevisiae* might resemble this latter mode of pausing in metazoa. Indeed, in yeast nucleosomes pose a significant barrier to elongation and frequently cause Pol II to stall (Churchman et al, 2011). Future studies will help reveal the full spectrum of pause-associated ncRNAs in eukaryotes.

Promoters form a zone of activity

Longer promoter-associated ncRNAs are also documented. Some initiate at or near the TSS, such as PASRs and PALRs (promoter-associated short (~50-200 nt) and long (>1kb) RNAs) (Kanhere et al, 2010; Kapranov et al, 2007). Others initiate just upstream, including yeast ~450 nt cryptic unstable transcripts (CUTs) (David et al, 2006; Neil et al, 2009; Wyers et al, 2005), and ~1 kb stable unannotated transcripts (SUTs), replication/sporulation SUTs (rsSUTs) and meiotic unannotated transcripts (MUTs) (Lardenois et al, 2011; Xu et al, 2009)). This suggests that promoters support initiation from multiple sites within the NFR. Furthermore, transcription frequently initiates in the reverse orientation (Xu et al, 2009), generating antisense versions of these ncRNAs. Most initiate from the upstream border of the NFR (Core et al, 2008), although reverse PASRs/PALRs are transcribed from either side of the canonical TSS (Kapranov et al, 2007). Polymerase pausing is important for the generation of antisense TSSa/tiRNAs, as depletion of pausing factors results in the production of longer upstream antisense RNAs instead (uaRNAs) (Flynn et al, 2011). Thus the antisense activity from the 5' border of the promoter NFR largely mirrors that at the 3'

border, but in most cases only sense transcription generates full-length mRNAs. In yeast, bidirectional promoters can direct transcription of long, stable transcripts in both directions (e.g. divergent mRNAs, or an mRNA and a SUT), but in many cases divergent transcription is suppressed (Churchman et al, 2011).

Non-coding promoter-associated transcription also initiates further upstream, generating PROMPTs (~1 kb upstream) (Preker et al, 2011; Preker et al, 2008) or 50-1500 nt ncRNAs from several kb upstream (Hung et al, 2011). This suggests that promoter-proximal and upstream regions constitute a single domain that is permissive for transcription, and consistently, mRNA and promoter-associated ncRNAs are often co-regulated (Kapranov et al, 2007; Preker et al, 2011; Preker et al, 2008; Seila et al, 2008; Taft et al, 2009; Xu et al, 2009). However, shared regulation does not extend to the most distal ncRNAs (Hung et al, 2011), and in yeast some upstream lncRNAs are independently expressed (Neil et al, 2009), suggesting that mRNA and ncRNA expression can also be individually regulated. Indeed, short promoter-proximal ncRNAs are even generated from "repressed" genes in the apparently complete absence of mRNA expression, although this could reflect regulation at post-initiation steps (Guenther et al, 2007; Kanhere et al, 2010).

The detection of heterogeneous promoter-associated ncRNAs raises the question of how a single promoter directs initiation from multiple sites. In yeast, nucleotide biogenesis genes assemble a single PIC that either initiates proximal to the TATA box or further downstream, resulting in a binary switch between non-productive upstream transcription (in replete conditions) and mRNA production (in conditions of nucleotide shortage) (Jenks et al, 2008; Kuehner et al, 2008; Thiebaut et al, 2008). Thus a single PIC can initiate various transcripts. However, for divergently oriented transcripts in yeast, experimental downregulation of sense transcription (via mutation of promoter elements) can upregulate antisense transcription (Neil et al, 2009), suggesting that divergent initiation events compete for PIC components. Consistently, global mapping of elongating Pol II and PIC components revealed that

divergent transcription is uncoupled, and involves distinct PICs (Churchman et al, 2011; Rhee et al, 2012). Together, these mechanistic studies support a model whereby the heterogeneous transcripts from a single promoter can be variously subject to shared, independent or competitive regulation.

Non-coding transcription is a general property of nucleosome-free regions

The 3' ends of genes resemble promoters in many respects, as they contain a NFR (3' NFR) and are bound by GTFs and Pol II (Murray et al, 2012; Rhee et al, 2012). Indeed, promoters and 3' NFRs are often physically juxtaposed in "gene loops" (Hampsey et al, 2011; Tan-Wong et al, 2012). Furthermore, like promoters, 3' NFRs produce non-coding RNAs. These include termini-associated short RNAs (TASRs) in metazoa, which extend ~300 bp in both orientations (Kapranov et al, 2007), and many classes of lncRNAs in yeast, including CUTs, SUTs, MUTs, Xrn1-sensitive unstable transcripts (XUTs) (van Dijk et al, 2011), and lncRNAs identified in mitotic cycling cells (Granovskaia et al, 2010). Notably, lncRNAs transcribed in the antisense direction from the 3' NFR will overlap part or all of the associated gene, so can potentially influence the expression of their sense partner via mechanisms not available to lncRNAs transcribed in other arrangements. These so-called "antisense ncRNAs" are therefore particularly likely to play functional roles, and are the subject of several dedicated studies (van Dijk et al, 2011; Yassour et al, 2010). Antisense lncRNAs can also be generated via divergent initiation from downstream promoters (Neil et al, 2009; Yassour et al, 2010).

Although many ncRNAs are generated from NFRs associated with the ends of annotated genes, the start sites of ~33 % of antisense ncRNAs in yeast cannot be linked to the 5' or 3' NFR of an adjacent gene (Yassour et al, 2010), and many other studies identify ncRNAs that are completely intergenic. The latter include many yeast lncRNAs (David et al, 2006; Granovskaia et al, 2010; Lardenois et al, 2011; Nagalakshmi et al, 2008; Neil et al, 2009; van Dijk et al, 2011; Xu et al, 2009) and human "large intergenic ncRNAs" (lincRNAs)

(Guttman et al, 2009; Khalil et al, 2009) and "short-lived lncRNAs" (Tani et al, 2012). Indeed, ~30 % of human lncRNAs are intergenic, and can extend for >50 kb ("vlincs") (Kapranov et al, 2010). Non-coding ncRNAs produced independently of annotated genes might arise from NFRs associated with other genomic features, such as enhancers.

Enhancers activate transcription from distal promoters, via long-range chromatin interactions, and possess NFRs. Notably, enhancers are transcribed bidirectionally to generate non-coding enhancer RNAs (De Santa et al, 2010; Kim et al, 2010b; Kowalczyk et al, 2012), perhaps reflecting a mechanism by which Pol II is first recruited to enhancers then transferred to target promoters. Enhancers are associated with a particular post-translational modification within the N-terminal tail of histone H3, H3K4 monomethylation (Heintzman et al, 2009), and bind the transcriptional coactivator p300 (Visel et al, 2009). These hallmarks frequently overlap with non-coding transcription, indicating that enhancer RNAs are prevalent (Kim et al, 2010b). Protein-coding genes are associated with a different chromatin "signature", comprising promoter proximal H3K4 trimethylation (H3K4me3) and downstream H3K36 trimethylation (H3K36me3) marks. Notably, this "K4-K36" signature also occurs in intergenic regions, where it is associated with human and mouse lincRNAs (Guttman et al, 2009; Khalil et al, 2009; Mikkelsen et al, 2007). This suggests that besides "hijacking" pre-existing NFRs associated with genic and enhancer regions, non-coding transcription can also arise independently. Indeed, lncRNAs in yeast even arise from regions of "silenced" chromatin, including centromeres, mating-type cassettes and telomeres (Houseley et al, 2007; Vasiljeva et al, 2008b).

Repression of non-coding transcription

Although non-coding transcription initiation is prevalent, intragenic initiation (within gene bodies) is generally suppressed. This is due to the organisation of intragenic nucleosomes into an arrangement refractory to transcription. Spurious initiation is therefore blocked and, in consequence, elongating polymerases require a cohort of factors to overcome the

nucleosomal barriers (Petesch et al, 2012). In yeast, a cascade of events in the transcription cycle (Owen-Hughes et al, 2012) (Figure 1.2) ensures efficient chromatin disruption ahead of the elongating polymerase, and chromatin reassembly in its wake, illustrating how non-coding transcription initiation can be repressed. These events include post-translational modifications (PTMs) of histone tails and heptad ($Y_1S_2P_3T_4S_5P_6S_7$) repeats within the C-terminal domain of the largest Pol II subunit (Pol II CTD), and the recruitment of regulatory proteins that modify histones, rearrange nucleosomes and/or associate with the Pol II complex. The Pol II and histone PTMs primarily function by recruiting proteins to specific regions of genes (Kouzarides, 2007).

At the promoter, the Pol II CTD is phosphorylated at serine 5 (Ser5) and serine 7 (Ser7) by Kin28 (Akhtar et al, 2009), whereas Bur1 phosphorylates Ser2, Ser5 and Ser7 further downstream (Bataille et al, 2012; Qiu et al, 2009). This establishes high Ser5P/Ser7P and low Ser2P at the 5' end of genes. Diphosphorylated Ser2P,Ser5P or Ser5P,Ser7P recruits the Paf1 elongation complex (Paf1C) (Qiu et al, 2012), which promotes Set1-dependent H3K4 trimethylation and consequent histone acetylation by SAGA within the 5' region (Pascual-Garcia et al, 2008; Pray-Grant et al, 2005). Histone acetylation disrupts inter-nucleosomal or histone:DNA contacts and recruits chromatin modifying factors, facilitating Pol II progression (Kouzarides, 2007). The Paf1 complex also restores nucleosomes to their original state in the wake of Pol II, recruiting the FACT nucleosome reassembly complex and Isw1 and Chd1 remodellers, which regulate nucleosome spacing (Cheung et al, 2008; Mason et al, 2003; Quan et al, 2010; Tirosh et al, 2010). Appropriate spacing suppresses cryptic initiation by stabilising the nucleosome array, preventing nucleosome collisions or providing access to regulatory proteins. Additional factors also suppress cryptic initiation by regulating nucleosome eviction and reassembly, including Spt6, Spt2 and Asf1 (Cheung et al, 2008; Kaplan et al, 2003; Nourani et al, 2006; Schwabish et al, 2006).

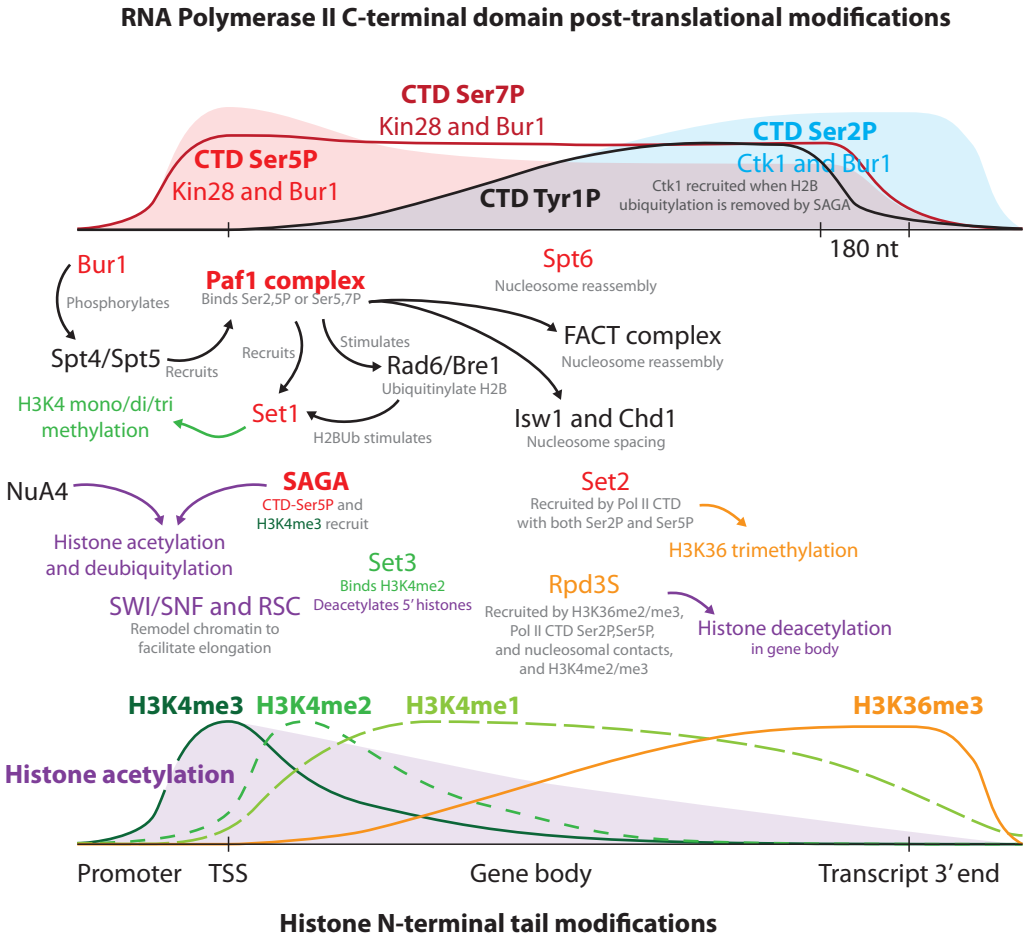


Figure 1.2: The Pol II transcription cycle. Post-translational modifications (PTMs) within Pol II C-terminal domain (CTD) heptad repeats (top graph) and histone N-terminal tails (bottom graph), together with protein factors (central region), coordinate events during transcription. The Pol II CTD is phosphorylated as indicated at serines 2, 5 and/or 7 (Ser2, Ser5 and/or Ser7) by Kin28 at the promoter (Akhtar et al, 2009; Kim et al, 2010a), Bur1 ~450 bp downstream, and Ctk1 yet further downstream (Qiu et al, 2009). The CTD is also phosphorylated at tyrosine 1 (Tyr1P). Pol II CTD modifications recruit proteins, and those interacting with Ser5P are indicated in red. Histone modifications include methylation by Set1 (H3K4) and Set2 (H3K36) and acetylation by SAGA, with the indicated distributions (Kirmizis et al, 2007; Pokholok et al, 2005). At the promoter, the Bur1/Bur2 kinase complex recruited by Pol II Ser5P (Qiu et al, 2009) phosphorylates the Spt4/Spt5 (DSIF) complex, which is partly recruited via interactions with nascent RNA (Missra & Gilmour, 2010). The Paf1 complex (Paf1C) then interacts with Spt4/Spt5 and Pol II Ser5P,Ser7P (Qiu et al, 2012), and directs many events (Jaehning, 2010). These include (i) H2B ubiquitylation by Rad6/Bre1 and downstream H3K4 methylation by Set1, and (ii) nucleosome rearrangements by Isw1, Chd1 and the FACT complex. The SAGA complex is then recruited by CTD Ser5P (Pascual-Garcia et al, 2008) and H3K4me3 (Pray-Grant et al, 2005), resulting in histone acetylation and H2B deubiquitylation (Rodriguez-Navarro, 2009). Additional histone acetyltransferases (HATs) are also recruited, such as NuA4. Histone acetylation facilitates elongation, partly via recruitment of SWI/SNF and RSC chromatin remodelling complexes (Kouzarides, 2007). SAGA association is transient, as H2B deubiquitylation opposes H3K4 methylation by Set1, and also permits Ctk1 recruitment to establish Pol II Ser2P (Wyce et al, 2007). Within the central gene body, Pol II CTD Ser5P and Ser2P coexist, and this bivalent mark recruits Set2, which trimethylates H3K36 (Kizer et al, 2005). Histone acetylation is reversed by histone deacetylases (HDACs), including Set3 towards the 5' end of genes (recruited by H3K4me2 (Kim & Buratowski, 2009)) and Rpd3S within gene bodies (recruited by H3K36me2/3, Ser2P,Ser5P and contacts with dinucleosomes) (Govind et al, 2010; Huh et al, 2012; Li et al, 2007).

In the central and 3' regions of genes, the Ctk1 kinase phosphorylates Ser2P, and diphosphorylated Ser2P,Ser5P CTDs recruit the H3K36 methyltransferase Set2. Consequently, histone acetylation is reduced by the histone deacetylase (HDAC) Rpd3S, which binds Ser5P,Ser2P CTDs, H3K36me3/me2, or non-methylated regions of dinucleosomes (Govind et al, 2010; Huh et al, 2012). Deacetylation suppresses cryptic initiation, and the redundancy in Rpd3S recruitment reflects its importance (Cheung et al, 2008; Li et al, 2009). H3K36 methylation also suppresses co-transcriptional histone exchange, which otherwise incorporates new, acetylated histones (Venkatesh et al, 2012). Although in promoter-proximal regions histone acetylation is more abundant, it is kept in check by the HDACs Set3 (recruited by H3K4me2) (Kim et al, 2009b), Rpd3L (Kim et al, 2009b; Terzi et al, 2011), and perhaps Rpd3S (recruited by H3K4 di/trimethylation under some conditions) (Pinskaya et al, 2009).

Similar mechanisms suppress cryptic initiation outside of gene bodies, for example at promoters and other NFRs. Notably, the Isw2 chromatin remodelling complex slides nucleosomes towards NFRs to repress cryptic transcription (Yadon et al, 2010), and Rpd3S suppresses divergent transcription from promoters (Churchman et al, 2011).

These mechanisms do not completely prohibit the generation of intragenic ncRNAs. For example, enhancer RNAs and longer "multi-exonic eRNAs" (meRNAs) initiate from intragenic enhancers (Kowalczyk et al, 2012), ncRNAs are generated from introns (Kapranov et al, 2010) and the 3' ends of exons (Taft et al, 2011), and many transcripts have 5' ends mapping within gene bodies in yeast (Miura et al, 2006) and other organisms (reviewed in (Tuck et al, 2011)). Although some of these ncRNAs arise from intragenic initiation, others are generated by post-transcriptional processing. For example, some intron-derived lncRNAs are flanked by snoRNA sequences and excised by the snoRNA processing machinery (Yin et al, 2012).

Together, this suggests a model whereby non-coding transcription is unavoidable within active regions of chromatin, as inhibition of non-coding transcription requires a repressive chromatin state. This explains why most, if not all, regions of active chromatin are associated with ncRNAs. Although much lncRNA transcription might arise as a side effect of canonical transcription, some lncRNAs are transcribed from regions of active chromatin that exist primarily for this purpose.

1.3 Direct and indirect functions of lncRNAs

The discovery of abundant non-coding RNAs was accompanied by intense debate as to whether they are functional. It is now apparent that many are, a conclusion supported by genome-wide and individual case studies. Whereas small ncRNAs might predominantly function via Argonaute family proteins, long ncRNAs appear to function by a variety of novel methods, described in detail in Table 1.2. In this study, I therefore focus on lncRNAs. Importantly, although lncRNA transcripts can function directly, in many cases it is actually the act of transcription that is functional. Further investigation (including this study) will reveal more broadly the extent to which functions are dependent on lncRNA transcripts versus the act of transcription.

Indirect evidence that lncRNAs are functional

Several groups have inferred widespread functions for lncRNAs by studying their expression and evolutionary conservation. Protein coding genes exhibit a characteristic pattern of sequence conservation, with variation at three nucleotide intervals due to redundancy in the genetic code, and a deficiency of nonsense or frameshift mutations. LncRNAs exhibit a different pattern of conservation, concentrated at promoters and splice junctions and without periodic bias (Ponjavic et al, 2007). Notably, the non-coding fraction of the genome is proportional to organism complexity (Taft et al, 2007), and the most rapidly evolving region of the human genome (compared to chimpanzees) encodes a brain lncRNA, HAR1F,

downregulated in Huntington's disease (Beniaminov et al, 2008; Johnson et al, 2010; Pollard et al, 2006). Together, these studies suggest that lncRNAs perform important conserved roles, and contribute to species divergence and the evolution of higher neurological functions.

LncRNA expression appears to be regulated, as it varies between normal and disease states and throughout embryonic development, mitosis and meiosis (Granovskaia et al, 2010; Guttman et al, 2009; Hung et al, 2011; Lardenois et al, 2011; Ørom et al, 2010; Yassour et al, 2010), and pluripotency-associated transcription factors bind ~75% of lincRNA promoters. Furthermore, variation in lncRNA expression is correlated with expression changes in particular functional groups of genes (Guttman et al, 2009; Hung et al, 2011), antisense transcription correlates with reduced gene expression (Xu et al, 2011; Xu et al, 2009), and depletion of lincRNAs causes widespread changes in gene expression (Guttman et al, 2011). Together, this suggests that lncRNAs are highly regulated, function in important cellular processes, and influence the expression of other genes.

The most convincing evidence for lncRNAs being functional, however, comes from detailed case studies, which also reveal the molecular mechanisms. I will give an overview of these mechanisms, which are fully described in several recent reviews (Guttman et al, 2012; Kanhere et al, 2012; Magistri et al, 2012; Qureshi et al, 2012; Wang et al, 2011b). Notably, several recent studies have confirmed that lncRNAs do not encode proteins: lncRNAs are seldom represented in the cellular proteome (assessed by mass spectrometry) (Banfai et al, 2012) or bound to ribosomes (Ingolia et al, 2009), and cross-species alignments reveal that lncRNAs possess low coding potential (Lin et al, 2011).

LncRNAs recruit proteins to chromatin

Chromatin modification is dependent on histone modifying enzymes and chromatin remodellers. Many of these lack a DNA binding domain, and lncRNAs might guide them to their targets. Such a role is well characterised for the lncRNA HOTAIR, which is transcribed

from the *HOXC* locus and targets the repressive H3K27 methyltransferase PRC2 and H3K4 demethylase LSD1 to the *HOXD* locus (Tsai et al, 2010). Notably, a genome-wide analysis of lncRNA:chromatin interactions detected 832 HOTAIR occupancy sites, and HOTAIR knockdown disrupts LSD1 and PCR2 recruitment to hundreds of promoters, suggesting that HOTAIR-dependent PRC2/LSD1 targeting is widespread (Chu et al, 2011; Tsai et al, 2010). Furthermore, approximately one third of tested lncRNAs (including lincRNAs, PROMPTs and antisense lncRNAs) bind chromatin modifying complexes (Guttman et al, 2011; Khalil et al, 2009; Zhao et al, 2010), each lncRNA can bind thousands of genomic loci (Chu et al, 2011; Simon et al, 2011), and remodelling complexes such as PRC2 bind thousands of different lncRNAs (Zhao et al, 2010). Thus lncRNAs play a major role in targeting proteins to chromatin, via a complex network of interactions. Furthermore, each lncRNA can bind several chromatin modifying complexes, facilitating cooperation between them (Guttman et al, 2011).

The ability of lncRNAs to bridge proteins and DNA has also been studied at a molecular level. Distinct regions of HOTAIR bind PRC2 and LSD1 (Tsai et al, 2010), and the PRC2 binding region comprises a double hairpin motif (Zhao et al, 2008) present in ncRNAs arising from many PRC2 target genes (Kanhere et al, 2010), indicative of a widespread PRC2 targeting mechanism. Other studies have focused on lncRNA:DNA interactions, and find that HOTAIR and roX2 lncRNA bind particular sequence motifs (Chu et al, 2011). LncRNAs can interact with DNA via several mechanisms, including (i) direct base-pairing, (ii) tethering by ongoing transcription, (iii) chromosome looping bringing lncRNAs into contact with target regions, and (iv) protein bridges.

These mechanisms are illustrated by lncRNAs that contribute to several key biological processes via targeting proteins to DNA:

1. **Tumour suppression**: The INK4b/ARF/INK4a tumour suppressor locus is silenced during normal growth by two Polycomb repressive complexes, PRC1 and PRC2.

The lncRNA ANRIL binds both complexes, tethering them to their target genes and facilitating cooperation between them (Huarte et al, 2010; Kotake et al, 2011).

2. **Anatomic specific expression**: HOX lncRNAs are expressed in distinct anatomic regions and contribute to development. These include HOTAIR and HOTTIP, the latter of which is expressed from the 5' tip of the HOXA cluster and recruits the histone methyltransferase MLL to target genes via chromatin looping (Wang et al, 2011c).
3. **DNA damage response**: p53 is stabilised by DNA damage and activates three transcripts in the *CDKN1A* locus, namely (i) *CDKN1A*, a cell-cycle arrest gene, (ii) the lncRNA PANDA, which sequesters a pro-apoptotic factor (Hung et al, 2011), and (iii) the lncRNA p21, which binds and recruits hnRNP K to hundreds of target promoters (Huarte et al, 2010), explaining how the transcription activator p53 can lead to gene repression.
4. **Genomic stability**: Intergenic lncRNAs from tandem arrays of ribosomal RNA genes recruit Trf4, which prevents recombination-dependent changes in repeat number (Houseley et al, 2007).
5. **Aging**: In aging yeast, an antisense lncRNA at the *PHO84* locus recruits silencing factors to repress PHO84. This is dependent on homology between the lncRNA and *PHO84* upstream region, and the lncRNA can function in *trans*, suggesting that it tethers silencing factors via direct hybridisation to its complementary sequence (Camblong et al, 2009).
6. **Imprinting** (parent-of-origin-specific, monoallelic gene expression): In the imprinted *Igf2r-Slc22a2-Slc22a3* cluster, the paternally expressed lncRNA AIR bridges the *Slc22a3* promoter and histone methyltransferase G9a to establish silencing (Nagano et al, 2008).
7. **Dosage compensation** (counteracts differences in chromosomal copy number to equalise gene expression): *Drosophila* males upregulate expression of genes on their

single X chromosome via deposition of the dosage compensation complex, within which the lncRNA roX2 makes contacts along the X chromosome (Chu et al, 2011; Conrad et al, 2012a; Conrad et al, 2012b). Conversely, mammalian females silence expression from one X chromosome. This is dependent on the lncRNA Xist, within which RepA structural motifs bind and recruit the PRC2 complex for silencing (Zhao et al, 2008), and the RepC localisation domain binds the bridging proteins YY1 and hnRNP U that anchor Xist to specific sequences on the X chromosome (Hasegawa et al, 2010; Jeon et al, 2012).

LncRNAs can also inhibit protein recruitment to chromatin. For example, the lncRNA NRON binds nuclear import factors to exclude the transcriptional activator NFAT from the nucleus (Willingham et al, 2005). Other lncRNAs can act as "sponges" to sequester factors from their targets, such as the lncRNA PANDA which sequesters the transcription factor NF-YA (Hung et al, 2011), or the microRNA target mimic MD1 which sequesters miR-133 and miR-135 (Cesana et al, 2011).

LncRNAs establish nuclear domains

LncRNAs also contribute to the formation of nuclear domains, perhaps via their ability to bind and juxtapose numerous genomic loci. For example, Xist and roX2 bind along the X chromosome, and HOTAIR binds sites genome-wide (Chu et al, 2011; Simon et al, 2011). LncRNAs might bring together different loci by one lncRNA making multiple interactions, or several lncRNAs interacting with a protein hub. Indeed, a chromosome conformation capture (3C) study identified interactions between Polycomb targets, perhaps mediated by Polycomb-binding lncRNAs (Bantignies et al, 2011).

In addition to acting as scaffolds, lncRNAs can perform specific roles in nuclear compartments. For example, the lncRNAs NEAT2 and TUG1 are components of nuclear speckles (activating compartments) and Polycomb bodies (repressive compartments) respectively. These lncRNAs act as scaffolds, but also help recruit specific genes into these

compartments. This involves Pc2, a subunit of PRC1, which simultaneously interacts with target genes and either TUG1 or NEAT2. Methylation of Pc2 switches its affinity from one lncRNA to the other, and thus acts as a toggle to regulate Pc2, and target gene, association with the two compartments (Yang et al, 2011b).

Telomeres are transcribed into telomeric repeat-containing RNA (TERRA), within which a G-quadruplex interacts with the telomeric protein TRF2 to contribute to telomere structure (Biffi et al, 2012). TERRA also inhibits the telomerase enzyme via binding its RNA component (Redon et al, 2010), and thus illustrates how lncRNAs function both via primary sequence and structural elements.

Paraspeckles are nuclear bodies involved in the retention of hyperedited Alu-containing RNAs, and contain long and short isoforms of the lncRNA NEAT1. These contribute to paraspeckle structure (Clemson et al, 2009; Sasaki et al, 2009; Sunwoo et al, 2009) in a manner requiring ongoing transcription (Mao et al, 2011). NEAT1 recruits paraspeckle proteins such as p54 (Clemson et al, 2009; Murthy et al, 2010), but also acts as a key architectural component by adopting a highly ordered conformation (Souquere et al, 2010).

LncRNAs modulate protein activity

Besides regulating protein localisation, lncRNAs can affect protein function. For example, the *DHFR* upstream transcript binds TFIIB and blocks its interaction with other PIC components (Martianov et al, 2007), whereas the lncRNAs NEAT2 and TUG1 more subtly influence the protein Pc2 by modulating the affinity of its chromodomain for various histone modifications (Yang et al, 2011b). LncRNA binding can also allosterically activate proteins, such as the upstream cyclin D lncRNA that binds and activates the co-repressor TLS by displacing its inhibitory C terminus (Wang et al, 2008).

LncRNAs hybridise with nucleic acids

Many lncRNAs associate with chromatin, and this can involve direct hybridisation with complementary DNA or RNA sequences. Some lncRNAs interact with the DNA major groove via non-canonical (Hoogsteen) hydrogen bonding to form a triplex (triple helix), such as the *DHFR* promoter-associated lncRNA that disrupts PIC assembly (Martianov et al, 2007). LncRNAs within mammalian rDNA promoters also act via triplex formation, displacing the rDNA transcriptional activator TTF1 and specifically recruiting the DNA methyltransferase Dnmt3b (Schmitz et al, 2010). Other lncRNAs hybridise to RNA, including several that bind to mRNAs and block access to processing factors or the ribosome (Nishizawa et al, 2008; Rintala-Maki et al, 2009; Yoon et al, 2012), or TERRA which binds the RNA component of telomerase (Redon et al, 2010). RNA duplexes can also be recognised by factors that bind dsRNA. For example, the RNAi machinery processes an intermolecular Xist:Tsix duplex into siRNAs (Ogawa et al, 2008), and a hairpin in the lncRNA H19 into miRNAs (Keniry et al, 2012). Additionally, binding sites for the dsRNA-binding protein Staufen 1 can be generated via imperfect base pairing between lncRNAs (1/2-sbsRNAs) and Alu elements, resulting in mRNA downregulation (Gong et al, 2011). Finally, lncRNAs could hybridise with a displaced strand of DNA to form R-loops, prevalent structures linked to genomic instability (Mischo et al, 2011).

The act of non-coding transcription is functional

The act of transcribing the lncRNA can also have consequences, independent of the transcript itself. Transcription is a disruptive process, and in many cases lncRNA transcription functions by displacing chromatin-bound factors. For example, non-coding transcription between the rDNA repeats displaces cohesin, leading to hyper-recombination (Kobayashi et al, 2005), and transcription upstream of *FLO11* and *ADH1* ejects regulatory factors (Bird et al, 2006; Bumgarner et al, 2012). This ability to "reset" chromatin might be a

major function of lncRNA transcription (Tuck et al, 2012). Besides simply disrupting protein:DNA interactions, lncRNA transcription also directs the formation of active or repressive chromatin via events that occur during the normal transcription cycle. For example, at the *fbp1* locus a cascade of upstream transcription remodels chromatin progressively downstream to promote expression (Hirota et al, 2008), and an antisense lncRNA at the *PHO5* locus increases chromatin plasticity (Uhler et al, 2007). Additionally, stalled promoter-proximal transcription might maintain promoters in a poised state. Conversely, lncRNA transcription represses transcription by directing nucleosome deposition upstream of *SER3* (Hainer et al, 2012; Hainer et al, 2011; Martens et al, 2005; Pruneski et al, 2011; Thebault et al, 2011), and via H3K36 methylation and Rpd3S-dependent histone deacetylation in the *GAL1-10* locus (Houseley et al, 2008; Pinskaya et al, 2009).

Non-coding transcription can also function indirectly, via sequestering transcription factors and diverting transcriptional output away from productive transcription. For example, a divergent lncRNA is suggested to antagonise *TPI1* transcription by competing for the same components during PIC assembly (Neil et al, 2009). LncRNA transcription can also act downstream of PIC assembly, as PICs assembled at the *IMD2* promoter can initiate at the canonical TSS to generate mRNAs, or at a promoter-proximal non-productive site (Jenks et al, 2008; Kuehner et al, 2008). A similar system exists at the *URA2* locus, but rather than competing with mRNA transcription, upstream non-productive transcription exerts a constant negative effect (Thiebaut et al, 2008).

Most “canonical” transcription is apparently associated with a plethora of bidirectional lncRNA transcription that can span several kilobases. This suggests that lncRNA transcription could also act as a signal to communicate the transcriptional status of a locus to adjacent regions. Indeed, divergent transcription from *GAL80* represses the adjacent *SUR7* locus (Xu et al, 2011), and immediate early gene induction in mouse cells results in a

cascade of intergenic transcription that activates neighbouring genes and spreads histone acetylation (Ebisuya et al, 2008).

In conclusion, the emerging picture is that lncRNAs, and the act of lncRNA transcription, function via diverse mechanisms. These depend upon the ability of lncRNAs to bind proteins, DNA and/or RNA, adopt specific structures, and act as scaffolds, tethers or "sponges", and the ability of non-coding transcription to modify chromatin and displace or sequester transcription factors. LncRNAs function in *cis* (at proximal loci), and in *trans* (at loci throughout the genome) (Guttman et al, 2011), and participate in many biological processes. However, although many studies have characterised the genomic origins of lncRNAs, we know very little about events downstream (processing, transport and turnover), which currently limits out understanding of how lncRNAs function.

1.4 How are lncRNAs distinguished from mRNAs?

Despite performing very distinct functions, lncRNAs and mRNAs have many similarities, and it is therefore not clear how they can be distinguished within cells. For example, mRNAs and lncRNAs are transcribed from similar, often overlapping, promoters, and their expression is highly regulated. Indeed, lncRNAs are regulated by canonical transcription factors, including Gal4, Reb1 and Zap1 in yeast (Bird et al, 2006; Houseley et al, 2008; Pinskaya et al, 2009; Xu et al, 2011), and p53 and pluripotency-associated factors in humans (Guttman et al, 2011; Huarte et al, 2010). The composition of lncRNA and mRNA PICs is similar (Rhee et al, 2012), and their transcription similarly involves H3K4 and H3K36 trimethylation (Guttman et al, 2009; Houseley et al, 2008; Khalil et al, 2009) and Pol II CTD Ser5 and Ser2 phosphorylation (Preker et al, 2011). Furthermore, like mRNAs, many lncRNAs are capped (Carninci et al, 2005; ENCODE, 2007; Miura et al, 2006; Neil et al, 2009) and have poly(A) tails (David et al, 2006; Kapranov et al, 2007; Preker et al, 2008).

Notably, some differences exist between lncRNA and mRNA transcription. H3K79 methylation is present downstream but not upstream of divergent promoters (Pokholok et al, 2005; Preker et al, 2008; Seila et al, 2008), and lncRNA promoters are resistant to experimental sequence disruption, suggesting they are defined by more general chromatin features (Uhler et al, 2007). Furthermore, although some lncRNAs are polyadenylated by the canonical poly(A) polymerase (Houseley et al, 2008; Thiebaut et al, 2006; Wyers et al, 2005), others are subject to "non-canonical" oligoadenylation associated with nuclear surveillance (Davis et al, 2006; Preker et al, 2011) or are not adenylated (Kapranov et al, 2010; Yang et al, 2011a). Finally, whereas some lncRNAs resemble mRNAs in being relatively stable (Tani et al, 2012; Tani et al, 2010; Xu et al, 2009) and detectable in the cytoplasm (Berretta et al, 2008; Geisler et al, 2012; van Dijk et al, 2011), others are rapidly degraded (Neil et al, 2009) and predominantly nuclear (Neil et al, 2009; Preker et al, 2011; Xu et al, 2009). Together, therefore, initial events in lncRNA and mRNA synthesis are relatively similar, but their functions are different, and the localisation and stability of lncRNAs is uncertain.

During their production, maturation and decay, mRNAs interact with a defined series of protein factors. I reasoned that determining which of these also interact with lncRNAs would reveal when and how lncRNAs are distinguished from mRNAs, and how the properties of lncRNAs are defined. I will therefore now consider the various steps in mRNA biogenesis, and the limited evidence for their involvement in lncRNA metabolism.

Pol II CTD modifications coordinate co-transcriptional events

Messenger RNA transcription is coupled to processing and nuclear export via the co-transcriptional recruitment of factors by Pol II CTD PTMs and nascent RNA sequences. Recent studies have provided an updated model of CTD states, which occur at fixed distances from the TSS or poly(A) site, and are similar across all Pol II genes (Figure 1.2). Ser5P levels rise across the promoter, peak at the TSS, remain moderately high across the

gene, and decline just before the poly(A) site. Ser7P is similar, but with a less prominent peak (Bataille et al, 2012; Kim et al, 2010a; Kim et al, 2009a). Conversely, Tyr1P and Ser2P levels rise ~250 nt downstream of the TSS, plateau at 600-1000 nt, then decline at the termination site (Ser2P) or ~180 nt upstream of the pA site (Tyr1P) (Bataille et al, 2012; Kim et al, 2010a; Mayer et al, 2012). Particular combinations of CTD modifications therefore enable proteins to be targeted to precise regions of a gene.

Capping

Shortly after transcription initiation, the enzymes Cet1, Ceg1 and Abd1 are recruited by Ser5P and cap the nascent transcript. This contributes to an early transcription checkpoint by promoting elongation (Kim et al, 2004a; Schroeder et al, 2004). The cap then binds the nuclear cap binding complex (CBC; Sto1 and Cbc2), which influences transcription termination, splicing, RNA stability, and translation.

Cleavage and polyadenylation

The 3' end of an mRNA is generated by endonucleolytic cleavage followed by polyadenylation, which together constitute 3' end processing. This is closely coupled to "poly(A)-dependent termination", whereby Pol II dissociates from the template ~200 nt downstream. The participants in 3' end processing and poly(A)-dependent termination are relatively well characterised (Kuehner et al, 2011; Mandel et al, 2008; Millevoi et al, 2009) (Table 1.3). However, the molecular mechanisms and coupling with other events (transcription, export and surveillance) remain to be fully elucidated.

The cleavage and polyadenylation machinery recognises the Ser2P CTD and sequence elements in the nascent transcript. These include the AU-rich efficiency element (EE), A-rich positioning element (PE), and U-rich upstream and downstream elements (UUE and DUE). The cleavage site itself comprises a pyrimidine residue followed by an adenosine tract, within which cleavage occurs. Many of these elements are not essential, but contribute

Table 1.3: Factors participating in cleavage, polyadenylation and coupled termination, versus Nrd1-dependent termination.

Complex	Component	Function in cleavage/polyadenylation and coupled termination	Function in Nrd1/Nab3/Sen1 termination	References
CFIA (contains two Rna14-Rna15 heterodimers)	Pcf11	Destabilises transcription complex via CID for termination; CID dispensable in some studies; separable cleavage activity; recruits Yra1 then later exchanges it for Clp1; cleavage and polyadenylation; contributes to Rat1 recruitment	Present at snoRNAs, CUTs, 5' ncRNAs; requires CID but not cleavage activity for termination; Nrd1 competes, so requires Nrd1 dissociation before termination can be completed	(Kim et al, 2010a; Kim et al, 2006; Loya et al, 2012; Luo et al, 2006; Steinmetz et al, 2003; Zhang et al, 2005) (Gross et al, 2001a; Luo et al, 2006; Sadowski et al, 2003)
	Rna14	Bridges Rna15 and Hrp1 and stabilises their interaction with PE and EE; cleavage and polyadenylation; termination	Present; required for snoRNA synthesis	(Fatica et al, 2000; Gordon et al, 2011; Gross et al, 2001a; Gross et al, 2001b; Kim et al, 2006; Morlando et al, 2002)
	Rna15	Cleavage and polyadenylation; termination; binds the A-rich positioning element; binds to Hrp1 via direct contacts and through Rna14	Present; required for snoRNA synthesis	(Fatica et al, 2000; Gordon et al, 2011; Gross et al, 2001a; Gross et al, 2001b; Kim et al, 2006; Leeper et al, 2010; Morlando et al, 2002),
	Clp1	Only contacts Pcf11; transmits conformational changes to Pcf11 that affect its interaction with Rna14-Rna15; bridges CFIA and CPF; termination, cleavage and polyadenylation	Present, required for snoRNA termination	(Ghazy et al, 2012; Gross et al, 2001a; Haddad et al, 2012)
CFIB	Hrp1	Binds efficiency element; contributes to selection of poly(A) site; suppresses cryptic poly(A) sites	Required at some Nrd1-dependent sites for termination	(Gross et al, 2001a; Kim et al, 2006; Kuehner et al, 2008; Leeper et al, 2010; Loya et al, 2012; Steinmetz et al, 2006b)
CFII (CPF) Sufficient for cleavage (with CFI), but remainder of CPF increases efficiency	Pta1	Scaffolding protein contacting core-CPF, CFI and APT; cleavage, polyadenylation and termination; N-terminal region inhibits cleavage/polyadenylation, which is relieved by interaction with Ssu72	Required; N-terminal region important for snoRNA processing; bridges core-CPF and the APT (associated with Pta1) complex	(Ghazy et al, 2009; Nedea et al, 2003)
	Ysh1/Brr5 (cleavage)	Cleavage endonuclease; required for termination; binds cleavage site	Conflicting evidence; not generally required, but may be required in some situations	(Garas et al, 2008; Kim et al, 2006; Zhelkovsky et al, 2006)
	Cft1/Yhh1	Binds cleavage site, Pol II CTD and TFIID; required for cleavage		(Dichtl et al, 2002b)
	Cft2/Ydh1	Binds U-rich site; required for cleavage; might act as a scaffolding protein	Not required	(Kyburz et al, 2003)
PFI (CPF)	Pfs2	Scaffold component; required for polyadenylation and cleavage in vivo		(Ohnacker et al, 2000)
	Yth1 (U)	Binds U-rich site; binds Fip1 to modulate switch from cleavage to polyadenylation	Not required for snoRNAs	(Morlando et al, 2002)

Complex	Component	Function in cleavage/polyadenylation and coupled termination	Function in Nrd1/Nab3/Sen1 termination	References
	Fip1 (U)	Tethers Pap1 to the CPF via a flexible central region and enhances Pap1 processivity; binds U-rich site; required for polyadenylation		(Ezeokonkwo et al, 2011)
Additional CPF factors	Pap1 (polyadenylation)	Poly(A) polymerase required for mRNA 3' end polyadenylation		
	Mpe1	Required for cleavage and polyadenylation		(Vo et al, 2001)
APT (CPF) Primary function is in snoRNA termination	Ssu72 (phosphatase)	Promotes recruitment of Pcf11 and Rtt103 to Pol II; also prevents inhibition by binding Pta1; important for termination of some protein-coding genes, and might play a minor role in cleavage/polyadenylation	Required; Ser5P dephosphorylation, enabling Pcf11 to bind for completion of termination	(Ganem et al, 2003; He et al, 2003b; Kim et al, 2006; Kuehner et al, 2008; Loya et al, 2012; Nedea et al, 2003; Steinmetz et al, 2003; Zhang et al, 2012b)
	Ref2	Required for efficient cleavage of weak polyadenylation sites; negative regulator of poly(A) tail synthesis; minor role	Required; keep Glc7 associated with APT; might suppress polyadenylation?	(Dheur et al, 2003; Kim et al, 2006; Mangus et al, 2004b; Nedea et al, 2003; Nedea et al, 2008; Russnak et al, 1995)
	Glc7 (phosphatase)	Required for polyadenylation but not cleavage; suggested to dephosphorylate Pta1 to allow association of Fip1 to APT	Required for snoRNA termination; phosphatase activity on Sen1 might regulate helicase activity?	(He et al, 2005; Nedea et al, 2008)
	Pti1	Not required (but some conflicting evidence); minor role	Required; suppresses polyadenylation activity, so might permit cleavage without polyadenylation?	(Dheur et al, 2003; Kim et al, 2006; Skaar et al, 2002)
	Swd2 (also in COMPASS)	Required for termination of some mRNAs (not cleavage or polyadenylation); ubiquitylation of Swd2 facilitates recruitment of Mex67; required to overcome repressive effect of COMPASS	Required for some snoRNAs; keeps Glc7 associated with APT	(Cheng et al, 2004a; Dichtl et al, 2004; Nedea et al, 2008; Soares et al, 2012)
	Syc1	Negative regulator of cleavage and polyadenylation, but not termination	Potentially suppresses polyadenylation of Nrd1-terminated transcripts?	(Zhelkovsky et al, 2006)
	Ess1	Minor roles; enables Ser5P dephosphorylation by Ssu72, for binding of Pcf11 and Rtt103	Major role; enables Ser5P dephosphorylation by Ssu72, enabling Pcf11 to bind	(Singh et al, 2009; Zhang et al, 2012b)
Poly(A) binding	Nab2	Regulates poly(A) tail length		(Hector et al, 2002)
Rat1-dependent termination	Rat1	Exonucleolysis triggers termination, but not sufficient; enhances Pcf11 and Rna15 recruitment	Not required, but can be detected bound to snoRNA genes	(Kim et al, 2004c; Kim et al, 2006; Luo et al, 2006)
	Rai1	Enhances Rat1 activity and termination; removes unmethylated caps as part of a checkpoint		(Kim et al, 2004c)

Complex	Component	Function in cleavage/polyadenylation and coupled termination	Function in Nrd1/Nab3/Sen1 termination	References
	Rtt103	Binds Ser2P and contributes to Rat1 recruitment; minor role in termination	Recruits Rat1 to Pol II Ser2P	(Kim et al, 2004c)
Sen1-Nrd1-Nab3	Nrd1	Present at 3' end of some mRNAs; present throughout many mRNAs, enriched towards the 5' end	Binds Pol II CTD Ser5P/Ser7P and purine-rich motif; not always required	(Jamonnak et al, 2011; Kim et al, 2010a; Wlotzka et al, 2011)
	Nab3	Present throughout some mRNAs, enriched at 5' end	Binds RNA (UCUUG); facilitates recruitment of Nrd1:Nab3:Sen1 complex	(Jamonnak et al, 2011; Wlotzka et al, 2011)
	Sen1	R loop removal to facilitate Rat1-dependent termination at some mRNAs; interacts with Glc7 subunit of APT	Unwinds RNA-DNA hybrid in Pol II active site	(Jamonnak et al, 2011; Nedea et al, 2003; Nedea et al, 2008; Skourti-Stathaki et al, 2011; Steinmetz et al, 2006b; Ursic et al, 2004)
	Cap binding complex	Prevents polyadenylation at weak termination sites, by impeding recruitment of termination factors Pcf11 and Rna15		(Wong et al, 2007)
Pol II	Rpb1 (CTD)	Not essential, but affects efficiency and site selection; recruits Pcf11 and Rtt103	Recruits Nrd1, Pcf11 and Sen1	(Kim et al, 2004c; Licatalosi et al, 2002; Ursic et al, 2004; Vasiljeva et al, 2008a)
	Rpb3/11	Contain surface required to transmit termination signal	Contain surface required to transmit termination signal	(Steinmetz et al, 2006a)
Miscellaneous	U1 snRNA	Prevents premature cleavage and polyadenylation		(Berg et al, 2012)
	Ras signalling		Blocks Nrd1:Nab3 dependent termination	(Darby et al, 2012)
	Mpk1		Blocks Nrd1:Nab3 recruitment to the Paf1C	(Kim et al, 2011a)
	Npl3	Antagonises Rna15 binding, and thus CFIA-associated Hrp1		(Bucheli et al, 2007; Dermody et al, 2008)
	Paf1C	Required for Ser2P, and thus Pcf11 recruitment; also direct recruitment of Cft1	Contributes to Nrd1 recruitment	(Kim et al, 2011a; Mueller et al, 2004; Sheldon et al, 2005)

to specificity and efficiency. The 3' ends of lncRNAs are poorly defined, and it is uncertain whether they contain similar motifs. In terms of protein components, the cleavage and polyadenylation machinery comprises several subcomplexes (CFIA, CFIB, and CPF). A subset of CPF members constitute a stable subcomplex (APT complex), which interacts with other CPF components via Pta1 (Dichtl et al, 2002a; Gavin et al, 2002; Nedea et al, 2003). Although most CF/CPF factors are nuclear, CFIB (of which Hrp1 is the sole member) shuttles to the cytoplasm (Kessler et al, 1997) where it functions in surveillance (González et al, 2000) and the stress response (Buchan et al, 2011; Henry et al, 2003).

The cleavage and polyadenylation machinery, and poly(A)-dependent termination factor Rtt103, predominantly bind the 3' end of genes (Ahn et al, 2004; Kim et al, 2004b; Kim et al, 2004c; Licatalosi et al, 2002), dependent on the Ser2P modification of the Pol II CTD (Ahn et al, 2004). The CFIA component Pcf11 (Licatalosi et al, 2002) and Rtt103 (Luo et al, 2006) specifically bind Ser2P via their CIDs. This is enhanced towards the 3' end of genes where multiple Ser2P residues exist within a single CTD, enabling Pcf11 or Rtt103 to bind cooperatively (Lunde et al, 2010). Pcf11 distribution does not completely overlap with Ser2P (Mayer et al, 2010), suggesting that other factors contribute to Pcf11 localisation. Indeed, Pcf11 and Rtt103 cannot bind diphosphorylated Tyr1P,Ser2P CTDs in the centre of genes (Mayer et al, 2012). Furthermore, Ser5P and Ser7P impair CID interactions, either directly (e.g. Rtt103 binds only weakly to Ser2P,Ser5P) (Vasiljeva et al, 2008a) or via Ser5P/Ser7P-bound factors competing for binding (Honorine et al, 2011). This inhibition is overcome by the CPF component Ssu72, a Ser5P/Ser7P phosphatase, which promotes termination (but apparently not cleavage/polyadenylation) of protein-coding genes (Bataille et al, 2012; Steinmetz et al, 2003; Zhang et al, 2012b). The peptidyl prolyl isomerase Ess1 also promotes termination by isomerising the CTD Ser5-Pro6 bond, a prerequisite for Ssu72 activity (Ma et al, 2012; Singh et al, 2009). Notably, the Pol II CTD is not essential for cleavage and

polyadenylation (Licatalosi et al, 2002), but contributes to efficiency and helps confine this process to the 3' end of genes.

In addition to the Pol II CTD, the 3' end processing and poly(A)-dependent termination factors can themselves be post-translationally modified, which influences their recruitment and activity (Table 1.4). Furthermore, additional Pol II-associated factors, such as the Paf1C, can assist in CPF recruitment (Nordick et al, 2008). Together, the participation of numerous sequence motifs, proteins and post-translational modifications in 3' end processing allows a great deal of specificity, but also some flexibility.

Following endonucleolytic cleavage by the CPF component Ysh1, mRNAs are polyadenylated by the poly(A) polymerase Pap1. The length of the poly(A) tail requires tight regulation, as short oligo(A) tails can recruit nuclear surveillance factors (Houseley et al, 2009), and if unchecked, Pap1 (Viphakone et al, 2008) and non-canonical poly(A) polymerases (Schmid et al, 2012) can hyperadenylate transcripts. Poly(A) tail length control is dependent on the poly(A) binding proteins Nab2 (Anderson et al, 1993; Brockmann et al, 2012; Kelly et al, 2007) and Pab1, which control the access of nucleases and poly(A) polymerases (Schmid et al, 2012; Viphakone et al, 2008). There is some debate as to which poly(A) binding protein regulates the initial nuclear poly(A) tail length, as although Nab2 is predominantly nuclear (Anderson et al, 1993; Wilson et al, 1994) and Pab1 cytoplasmic, both proteins cross the nuclear envelope (Aitchison et al, 1996; Brune et al, 2005). Hyperadenylation defects in *NAB2* mutants cannot be rescued by Pab1 (Hector et al, 2002), but Pab1 interacts with CFIA and can confer poly(A) tail length regulation *in vitro* (Amrani et al, 1997; Minvielle-Sebastia et al, 1997; Schmid et al, 2012).

Whereas the roles of Nab2 and Pab1 in lncRNA metabolism are uncharacterised, Pap1 is reported to polyadenylate some lncRNAs, including IGS1-R (Houseley et al, 2007), TERRA (Luke et al, 2008), NEL025c$_{long}$ (Wyers et al, 2005), SRG1$_{short}$/SGR1$_{long}$ (Thiebaut et al, 2006), and NGR040w$_{long}$ (Thiebaut et al, 2006). Furthermore, some lncRNAs are stable and

Table 1.4: Post-translational modifications regulating mRNA processing. The protein subject to modification is indicated in the first column, and the enzyme responsible for the modification in the third column.

Protein	Modification	Enzyme	Function	Reference
Npl3	Phosphorylation	Sky1	Promotes disassembly of mRNPs in the cytoplasm	(Gilbert et al, 2001)
	Dephosphorylation	Glc7	Npl3 binds mRNAs in dephosphorylated state; dephosphorylation permits Mex67 binding	(Gilbert et al, 2004)
	Phosphorylation	Ck2	Promotes termination: reduces ability of Npl3 to compete with Rna15, and prevents stimulation of Pol II elongation by Npl3	(Dermody et al, 2008)
	Methylation	Hmt1	Required for nuclear export; disrupts association with Tho2, perhaps licensing the mRNP for export?	(McBride et al, 2005; Shen et al, 1998; Xu et al, 2004; Yu et al, 2004)
Hpr1	Ubiquitylation	Rsp5	Binding site for Mex67-UBA, which when transferred to mRNA reveals the Hpr1 ubiquitin moiety and triggers proteolysis	(Gwizdek et al, 2005; Gwizdek et al, 2006; Neumann et al, 2003)
Yra1	Ubiquitylation	Tom1	Release of Yra1 from complex with Nab2-Mex67-Mtr2	(Duncan et al, 2000; Iglesias et al, 2010)
	Methylation	Hmt1	Unknown	(Yu et al, 2004)
Nab2	Phosphorylation	Mpk1	Coincident with Nab2:Yra1 retention in nuclear foci, and dissociation of Mex67	(Carmody et al, 2010)
	Methylation	Hmt1	Required for Nab2 export	(Green et al, 2002)
Hrp1	Methylation	Hmt1	Enhances Hrp1 interaction with Ccr4:Not1; affects genome-wide binding profile; favours nuclear export of Hrp1 (perhaps indirect, via Npl3 methylation); recognition of TATATA antagonises methylation	(Kerr et al, 2011; Shen et al, 1998; Xu et al, 2004; Yu et al, 2004)
Swd2	Ubiquitylation		Recruitment of Mex67 by Swd2 (APT complex)	(Vitaliano-Prunier et al, 2012)
Pta1	Dephosphorylation	Glc7	Permits association of Fip1 with Pta1, promoting polyadenylation (Fip1 regulates the poly(A) polymerase Pap1)	(Ezeokonkwo et al, 2011; He et al, 2005; Helmling et al, 2001)
Fip1	Ubiquitylation	?	Proteolysis of Fip1 results in depletion, and thus inhibits polyadenylation	(Saguez et al, 2008)

can be exported to the cytoplasm (van Dijk et al, 2011; Xu et al, 2009), resembling mRNAs that undergo cleavage and polyadenylation. However, deletion of the CFIA component Rna14 did not affect the expression of tested lncRNAs (Marquardt et al, 2011). The generality of cleavage and polyadenylation in lncRNA metabolism is therefore uncertain.

Poly(A)-dependent termination is closely coupled to 3' end processing, and both the pA site (Kim et al, 2004b) and CF/CPF factors are required for termination (Garas et al, 2008; Kim et al, 2006; Sadowski et al, 2003). Furthermore, the termination factor Rat1 facilitates recruitment of CFIA components (Rna15 and Pcf11), and Pcf11 contributes to Rat1 recruitment (Luo et al, 2006). Termination is dependent on disruption of the 8 nt DNA:RNA heteroduplex in the Pol II active site (Kireeva et al, 2000), but there are several models as to how this is triggered. In the "torpedo" model (Kim et al, 2004c), the 5'-3' exonuclease Rat1 and its activator Rai1 (Xue et al, 2000) are recruited to the 3' end of genes by Rtt103, and degrade the nascent transcript downstream of the cleavage site. Rat1 is suggested to catch the elongating Pol II to trigger its dissociation from the template. However, Rat1, Rtt103 and Rai1 cannot mediate termination *in vitro* (Dengl et al, 2009), and the exonuclease Xrn1 cannot rescue termination defects in *rat1* mutants despite degrading 3' cleavage products (Luo et al, 2006). This suggests that termination requires factors besides Rat1 exonuclease activity. Pcf11 is a prime candidate, as it colocalises with Rat1 (Kim et al, 2010a) and its CID can disrupt Pol II:DNA interactions *in vitro* (Zhang et al, 2005) and is required for termination (Mariconti et al, 2010; Sadowski et al, 2003). Furthermore, Ess1 and Ssu72, which enhance Pcf11 CID interactions with Pol II by removing inhibitory Ser5P/Ser7P modifications, also promote termination. These observations led to the "allosteric" model, in which Pcf11 transmits conformational changes between the nascent transcript and Pol II CTD to trigger termination. Poly(A)-dependent termination might require a combination of allosteric and exonucleolytic activities, since neither allosteric nor torpedo mechanisms alone are sufficient (Luo et al, 2006).

Nrd1-dependent termination

The 3' ends of short Pol II transcripts, such as snoRNAs, are generated directly by a Rat1-independent termination mechanism, without Ysh1-dependent cleavage or extensive polyadenylation (Kuehner et al, 2011; Lykke-Andersen et al, 2007). Instead, "Nrd1-dependent" termination requires the RNA binding proteins Nrd1 and Nab3, and the superfamily I DNA/RNA helicase Sen1 (Steinmetz et al, 2001).

Nrd1-dependent termination of snoRNAs is coupled to subsequent processing (Houalla et al, 2006; Jamonnak et al, 2011; Steinmetz et al, 2001; Wlotzka et al, 2011), as Nrd1 can directly recruit the exosome (Honorine et al, 2011; Vasiljeva et al, 2006) to trim snoRNA 3' ends to their mature length. SnoRNA-binding proteins, which assemble co-transcriptionally with snoRNAs into snoRNPs, limit the extent of trimming (Ballarino et al, 2005; Morlando et al, 2004). Small nuclear RNA (snRNA) termination is also Nrd1-dependent (Steinmetz et al, 2001), although the pre-snRNA 3' end is generated by the endonuclease Rnt1 (Abou Elela et al, 1998) with which Nrd1 interacts (Vasiljeva et al, 2006).

Nrd1-dependent termination also functions at some protein-coding genes. *HRP1* and *NRD1* are autoregulated by Nrd1-dependent early terminators (Arigo et al, 2006a; Houalla et al, 2006; Kuehner et al, 2008; Steinmetz et al, 2001; Wlotzka et al, 2011), and non-coding transcription upstream of *IMD2* and *URA2* is terminated by Sen1-Nrd1-Nab3 (Jenks et al, 2008; Kuehner et al, 2008; Thiebaut et al, 2008). In these cases, termination is coupled to exosome-dependent turnover. Conversely, Nrd1-dependent termination of *NAB2* and *CTH2* is followed by limited trimming, which generates mature mRNAs (Ciais et al, 2008; Roth et al, 2009; Vasiljeva et al, 2006). Nrd1 and Nab3 also bind Pol III transcripts, perhaps acting post-transcriptionally in surveillance (Jamonnak et al, 2011; Wlotzka et al, 2011).

Nrd1-dependent termination participates in the termination of "short" lncRNAs, including CUTs and divergent transcripts from protein-coding gene promoters (Arigo et al, 2006b; Marquardt et al, 2011; Thiebaut et al, 2006; Wlotzka et al, 2011; Wyers et al, 2005). Nrd1

and Nab3 also participate in the surveillance, and sometimes termination, of longer lncRNAs (Houseley et al, 2007; Kim et al, 2010a; Marquardt et al, 2011; Vasiljeva et al, 2008b; Wlotzka et al, 2011). Nrd1-dependent termination, coupled to surveillance, might therefore be prevalent amongst lncRNAs.

Mechanism of Nrd1-dependent termination

The presence of Nrd1 is not alone sufficient for termination (Gudipati et al, 2008), which also requires the recognition of specific combinations of motifs in nascent RNA (Porrua et al, 2012; Steinmetz et al, 2001), with Nab3 binding UCUUG (Carroll et al, 2004; Hobor et al, 2011; Porrua et al, 2012; Wlotzka et al, 2011) and Nrd1 binding purine-rich motifs including GUA[A/G] (Carroll et al, 2004; Wlotzka et al, 2011). Nrd1 and Nab3 function as a heterodimer, with highest affinity when cooperatively bound to multiple motifs (Carroll et al, 2007; Hobor et al, 2011). Furthermore, Nrd1-dependent terminators are only effective in promoter-proximal positions (Gudipati et al, 2008; Kopcewicz et al, 2007; Porrua et al, 2012), as termination requires interactions between Nrd1 and Pol II CTD Ser5P and/or Ser7P (Gudipati et al, 2008; Kim et al, 2010a; Vasiljeva et al, 2008a). Although Ser5P/7P are present further downstream, here Tyr1P prevents their interaction with Nrd1 (Mayer et al, 2012). Set1-dependent H3K4 trimethylation also promotes Nrd1-dependent termination, perhaps via regulating promoter-proximal Pol II kinetics (Terzi et al, 2011). Amongst these factors, recognition of the Nab3 motif (Arigo et al, 2006b; Kim et al, 2011a; Wlotzka et al, 2011) and interaction of the Nrd1 CID with Pol II (Arigo et al, 2006b) are most important, whereas many terminators lack a Nrd1 motif (Wlotzka et al, 2011) and the Nrd1 RRM is dispensable (Arigo et al, 2006b).

Sen1 also plays a key role in Nrd1-dependent termination (Finkel et al, 2010; Kim et al, 2006; Steinmetz et al, 2001; Steinmetz et al, 2006b), perhaps via its ability to disrupt DNA:RNA hybrids (R-loops) that form behind elongating Pol II and are prevalent over termination sites (Mischo et al, 2011; Skourti-Stathaki et al, 2011). Sen1 might therefore

trigger termination, perhaps assisted by Nrd1 transmitting force between the nascent transcript and Pol II CTD.

Crosstalk between poly(A)-dependent and Nrd1-dependent termination

Poly(A)-dependent and Nrd1-dependent termination were initially thought to be largely distinct, but considerable overlap has now been reported. For example, Nrd1-dependent terminators can direct poly(A)-dependent termination when placed distally from a promoter (Gudipati et al, 2008; Kopcewicz et al, 2007; Porrua et al, 2012; Steinmetz et al, 2006a), and CF/CPF factors bind genes with Nrd1-dependent terminators (Table 1.3). Reciprocally, Nrd1 directs termination at poly(A) sites under some conditions (Honorine et al, 2011), and Nrd1 and Nab3 binds both ends of numerous mRNAs (Jamonnak et al, 2011; Kim et al, 2010a; Wlotzka et al, 2011). Indeed, Nrd1- and poly(A)-dependent termination pathways frequently interact, by sharing factors, competing, or backing each other up.

Some Nrd1-dependent terminators resemble cleavage and polyadenylation sites (Porrua et al, 2012; Steinmetz et al, 2003), and indeed, CFI and the APT subcomplex participate in Nrd1-dependent termination (Table 1.3). The CID of Pcf11 is particularly important (Kim et al, 2006), suggesting that it plays a conserved role in transmitting force between the Pol II CTD and nascent RNA at both types of terminator (Mariconti et al, 2010; Sadowski et al, 2003). In contrast to its role at Nrd1-dependent terminators, the APT subcomplex makes only minor contributions to canonical cleavage, polyadenylation and termination (Table 1.3).

The APT component Ssu72 (Ghazy et al, 2009; Kim et al, 2006) and the Ess1 prolyl isomerase (Singh et al, 2009) are particularly important at Nrd1-dependent terminators, compared to canonical cleavage/polyadenylation sites (Ganem et al, 2003; He et al, 2003b; Steinmetz et al, 2003; Zhang et al, 2012b). Ssu72 and Ess1 cooperate to remove Ser5P marks from the Pol II CTD, favouring Pcf11 binding and excluding Nrd1 (Singh et al, 2009). This suggests that exchange of Nrd1 for Pcf11 is important for completion of Nrd1-dependent termination, and towards the 5' end of genes where Ser2P levels are low, perhaps more

assistance is required from Ssu72 and Ess1 for Pcf11 to overcome the inhibitory effect of Ser5P and/or Nrd1. Additional PTA subunits might adapt the 3' end processing machinery to Nrd1-dependent termination. For example, Glc7 stimulates the activity of Sen1 (Nedea et al, 2008), and Pti1 (Dheur et al, 2003), Syc1 (Zhelkovsky et al, 2006) and Ref2 (Mangus et al, 2004b) suppress poly(A) tail synthesis. Pap1-dependent polyadenylation is, however, evident at some Nrd1-dependent terminators (Grzechnik et al, 2008; Thiebaut et al, 2006; Wyers et al, 2005).

Reciprocally, Nrd1-dependent termination factors can assist at canonical cleavage/polyadenylation sites. For example, Sen1 binds the 3' end of many mRNAs (Creamer et al, 2011; Jamonnak et al, 2011) and contributes to termination at a subset of mRNAs (Kawauchi et al, 2008; Mischo et al, 2011; Steinmetz et al, 2006b).

Factors from one termination pathway can also antagonise the other pathway. For example, an increase in Pcf11 binding to Pol II Ser2P can antagonise Nrd1-dependent termination (Gudipati et al, 2008; Honorine et al, 2011), whereas Pol II Ser5P promotes Nrd1 binding at the expense of Pcf11 (Singh et al, 2009). This suggests that Pol II CTDs either bind Nrd1 and Pcf11, which directly compete for binding. However, as each CTD has multiple heptad repeats, Nrd1 and Pcf11 might bind simultaneously and the relative amount of binding determine the mode of termination.

Finally, the two termination pathways might act as fail-safe mechanisms for each other, as read-through transcripts in *rat1* mutants are terminated by Nrd1 (Rondon et al, 2009), and poly(A)-dependent termination acts on read-through from snoRNA Nrd1-dependent terminators (Grzechnik et al, 2008), which is widespread (Houalla et al, 2006). Defective cleavage/polyadenylation can also be rescued by Rnt1 cleavage coupled to Rat1 termination (Ghazal et al, 2009; Rondon et al, 2009).

In conclusion, there is considerable overlap between various 3' end processing and termination pathways, and the elongating polymerase initially keeps both options open.

LncRNAs could terminate via either pathway, or a combination of the two, and distinct classes of lncRNAs might be defined on this basis.

Nuclear export is coupled to transcription and processing

Whereas snoRNAs and snRNAs are retained in the nucleus, mRNAs are exported to the cytoplasm. Export is coupled to transcription and 3' end processing, and assembly of an export-competent mRNP (messenger ribonucleoprotein particle) occurs co-transcriptionally (Figure 1.3). Firstly, the nascent transcript binds Tho2 within the five subunit THO complex, comprising Tho2, Mft1, Tex1, Thp2 and Hpr1 (Chavez et al, 2000; Pena et al, 2012), which is also tethered to Pol II via the Prp19 complex (Chanarat et al, 2011). Hpr1 can bind the ATP-dependent helicase Sub2 (Straszer et al, 2001), which in turn binds the RNA-binding export adapter Yra1 (Strasser et al, 2002; Zenklusen et al, 2002) (initially recruited by Pcf11 (Johnson et al, 2009)). Sub2 and Yra1 are important export factors, and with the THO complex constitute the transcription/export (TREX) complex. TREX also contains the poorly characterised RNA-binding proteins Gbp2, Hrb1 (Hurt et al, 2004) and Tho1 (Jimeno et al, 2006).

TREX prevents transcription-associated recombination by suppressing R-loop formation (Huertas et al, 2003), facilitates transcription elongation (Rondón et al, 2003), and coordinates the timely recruitment of export factors. Mex67, the primary mRNA export receptor, binds to ubiquitylated Hpr1 (Gwizdek et al, 2006), and to Yra1 in a manner that disrupts the Yra1:Sub2 interaction (Straszer et al, 2000; Zenklusen et al, 2001). Mex67 also interacts with Nab2, which participates in export and poly(A) tail length control (Gallardo et al, 2003; Green et al, 2002; Marfatia et al, 2003), and Yra1 enhances this interaction (Iglesias et al, 2010). Together, therefore, a complex sequence of handovers culminates in Yra1 loading Nab2 and Mex67 onto the mRNA (Figure 1.3). Nab2 and Mex67 then facilitate export, via contacts with nuclear pore complex (NPC) components, including Mlp1 (Fasken et al, 2008; Green et al, 2003) and nucleoporins (Hobeika et al, 2009). The NPC component

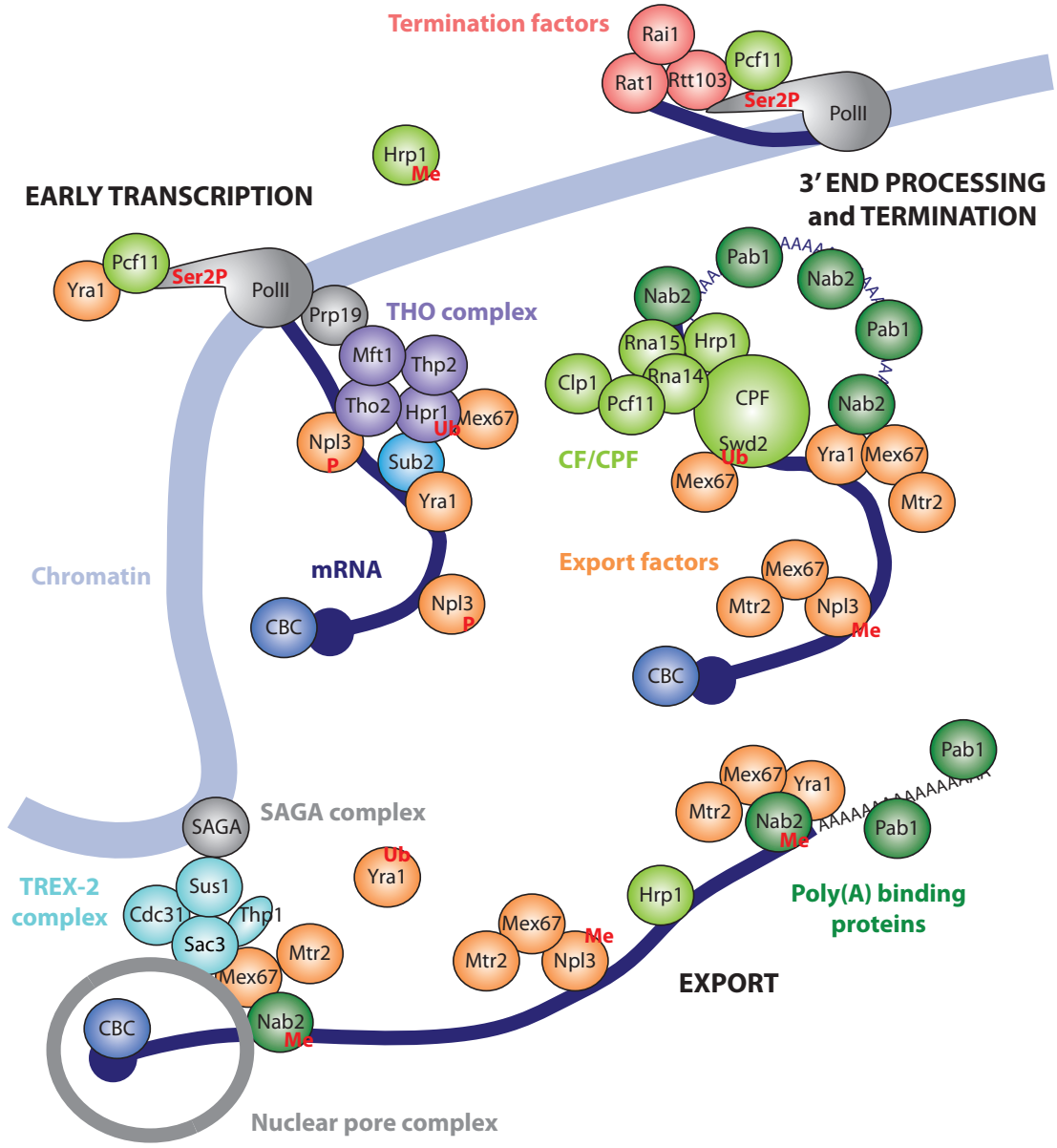


Figure 1.3: mRNA transcription, 3' end processing and export. Early in transcription, CBC binds the mRNA cap, and the THO complex is recruited via Prp19 (binds Pol II) and Tho2 (binds RNA). The THO complex recruits the export cofactor Yra1 via Sub2, and the export receptor Mex67. After co-transcriptional cleavage and polyadenylation by CF and CPF factors, Nab2 and Pab1 bind the poly(A) tail. Transcription is terminated by Rat1 with various accessory factors. Mex67 is then transferred to Nab2, assisted by Yra1 and Swd2 (a CPF component). Npl3 acts as an alternative adapter to recruit Mex67. Export involves contacts between the nuclear pore complex and Mex67, Nab2 and the TREX-2 complex (which assists in the later stages of export and contacts chromatin via the Sus1 subunit of the SAGA complex). Numerous post-translational modifications regulate these events (red text and Table 1.4).

Tom1 ubiquitylates Yra1, triggering its dissociation and retention in the nucleus (Iglesias et al, 2010).

Nab2 associates with most transcripts (Batisse et al, 2009), but interactions with Mex67 or Yra1 are only reported for ~1000 (Hieronymus et al, 2003). This suggests that export pathways exist besides Yra1-Mex67-Nab2. Strong candidate participants include the shuttling factors Npl3 (Flach et al, 1994; Gilbert et al, 2004) and Hrp1 (Henry et al, 1996), which contribute to export and bind many transcripts (Kim Guisbert et al, 2005). Indeed, Npl3 is co-transcriptionally loaded onto nascent transcripts and can recruit Mex67 (Gilbert et al, 2004). Notably, Nab2 can suppress *yra1* but not *npl3* mutants (Iglesias et al, 2008), and Tom1 mutations affect export of Nab2 but not Npl3 (Duncan et al, 2000). This suggests that Nab2-Yra1 and Npl3 are distinct adapters for Mex67 recruitment.

The coupling between transcription and export is reinforced by the TREX-2 complex, comprising Thp1, Sac3, Sem1, Cdc31 and two Sus1 proteins (Faza et al, 2009; Gonzalez-Aguilera et al, 2008) and functioning in export and transcription elongation (Fischer et al, 2002; Rodriguez-Navarro et al, 2004). Sac3 acts as a scaffold (Jani et al, 2009), interacting with Mex67 and nucleoporins (Fischer et al, 2002) to dock mRNPs to the nuclear pore. Conversely, Sus1 is a component of the SAGA elongation complex and therefore contacts chromatin (Pascual-Garcia et al, 2008). TREX-2 also suppresses R loop formation and hyper-recombination (Gonzalez-Aguilera et al, 2008). TREX-2 therefore assists transcription and export, and might physically tether the active locus along with its mRNPs to the nuclear pore.

Export is also coupled to 3' end processing – mutations in CF/CPF factors cause nuclear retention of mRNAs (Brodsky et al, 2000; Hammell et al, 2002; Libri et al, 2002; Milligan et al, 2005), and defective export disrupts 3' processing and polyadenylation (Assenholt et al, 2008; Qu et al, 2009; Saguez et al, 2008). This coupling arises from the participation of 3' end processing factors in export, including (i) Hrp1, (ii) Pcf11, which recruits Yra1 and later

exchanges it for the CFIA component Clp1 (Johnson et al, 2011), (iii) Swd2, which in its ubiquitylated state facilitates recruitment of Mex67 to CPF (Vitaliano-Prunier et al, 2012), and (iv) Glc7, which dephosphorylates Npl3 to permit Npl3:Mex67 interactions (Gilbert et al, 2004). Furthermore, polyadenylation promotes export (Dower et al, 2004), perhaps via recruitment of Pab1 (Brune et al, 2005) and Nab2 (Iglesias et al, 2010).

Remodelling events are important for processing and export

The many changes in mRNP composition during processing and export are achieved via a series of remodelling events. Each of these is a potential opportunity for regulation and quality control.

Defects in TREX components lead to the formation of a stalled export intermediate, containing mRNA, chromatin, nucleoporins and 3' end processing factors (Qu et al, 2009; Rougemaille et al, 2008). This suggests that TREX components facilitate a major remodelling event upon completion of 3' processing that releases the NPC-tethered mRNP from the site of transcription, and perhaps requires Sub2 helicase activity. This remodelling event might incorporate Yra1, Mex67, Nab2 and Npl3 into the mRNP and trigger proteolysis of Hpr1 (Hobeika et al, 2009), and apparently coincides with a quality control checkpoint that retains aberrant mRNPs (Libri et al, 2002).

Further remodelling and quality control occurs at the nuclear pore. Nab2 and methylated Hrp1 associate with Mlp1 and the Ccr4-Not complex, respectively, both of which are NPC-associated surveillance factors (Fasken et al, 2008; Green et al, 2003; Kerr et al, 2011). Additionally, Yra1 is deubiquitylated by Tom1, which is apparently necessary to pass an Mlp1-dependent checkpoint (Iglesias et al, 2010). Together, events at the nuclear pore comprise a major remodelling and surveillance checkpoint, as a study of single mRNPs finds that three quarters of mRNPs probing the NPC are turned away (Siebrasse et al, 2012).

A final remodelling step takes place at the cytosolic face of the NPC, dependent on the DEAD-box ATPase Dbp5. Here, Nab2 interacts with the Dbp5 activator Gle5 and the Dbp5-associated protein Gfd1 (Grant et al, 2008; Suntharalingam et al, 2004; Zheng et al, 2010). This concentrates mRNPs at the cytosolic side of the pore, where Dbp5 facilitates removal of export factors, including Mex67 (Lund et al, 2005) and Nab2 (Tran et al, 2007). This imparts directionality on intrinsically bidirectional diffusion through the NPC.

There is considerable uncertainty as to whether lncRNAs undergo some, all, or none of the canonical steps in mRNA maturation and export. Some lncRNAs are rapidly degraded after termination, whereas others are more stable. Distinct classes of lncRNAs might therefore exist, with variation in the extent to which they resemble mRNAs and assemble with export factors.

1.5 RNA turnover

RNA turnover contributes to surveillance, processing and regulation of gene expression, and plays a significant role in lncRNA metabolism as deletion or mutation of turnover factors results in extensive lncRNA upregulation (Davis et al, 2006; Gudipati, 2012; Neil et al, 2009; van Dijk et al, 2011; Xu et al, 2009). A better understanding of lncRNA turnover might therefore provide insight into when, where and how lncRNAs diverge from mRNAs. Evidence exists for lncRNA turnover in both the nucleus and cytoplasm, but redundancy between decay pathways makes this hard to interpret. There are several major decay pathways in yeast (Houseley et al, 2009; Parker, 2012), and recent studies have provided mechanistic insights into how substrates are targeted to particular pathways.

Nuclear surveillance

The major degradative activity in the nucleus is provided by the exosome (Mitchell et al, 1997), a multiprotein complex which also participates in cytoplasmic decay. The exosome comprises a nine subunit core, with a ring of S1/KH RNA-binding proteins (Csl4, Rrp4 and

Rrp40) atop a hexameric barrel of RNase PH-like proteins (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, Mtr3) (Liu et al, 2006; Wang et al, 2007). The exosome core resembles bacterial PNPase, but is rendered catalytically inert by mutations in the phosphorolytic active sites (Dziembowski et al, 2007). Instead, ribonuclease activity is provided by Rrp44 and Rrp6 (Allmang et al, 1999b). Rrp44 possesses a processive 3'-5' exoribonucleolytic RNB domain and an endoribonucleolytic PIN domain (Lebreton et al, 2008; Lorentzen et al, 2008; Schaeffer et al, 2009; Schneider et al, 2009), and stores the energy from multiple single nucleotide hydrolysis reactions to unwind structured substrates in ~4-nt bursts (Lee et al, 2012). Whereas Rrp44 is present in nuclear and cytoplasmic exosomes, Rrp6, a distributive 3' to 5' exoribonuclease (Liu et al, 2006), is strictly nuclear (Allmang et al, 1999b).

The exosome core regulates the activity of both nucleases, with the current model based on cryo-EM structures (Malet et al, 2010), crosslinking studies (Wasmuth et al, 2012) and RNase protection (Bonneau et al, 2009). RNA threads through the S1/KH-protein ring and the centre of the hexameric barrel to access both nucleolytic domains of Rrp44 (Bonneau et al, 2009; Wang et al, 2007; Wasmuth et al, 2012), and via the S1/KH cap and upper portion of the barrel to access Rrp6 (Wasmuth et al, 2012), which is proposed to dock near the top of the exosome core (Cristodero et al, 2008). In the absence of the exosome core, Rrp6 and Rrp44 are both highly active on unstructured substrates, whereas Rrp6 fares better on poly(A) substrates (which are semi-structured (Seol et al, 2007)) and Rrp44 on structured substrates (Liu et al, 2006). Binding to the exosome core reduces Rrp44 activity, particularly on structured substrates, as a ~33 nt single stranded overhang is required to thread through the channel (Bonneau et al, 2009; Liu et al, 2006; Wasmuth et al, 2012). Conversely, Rrp6 is only mildly affected as its substrates take a more peripheral path (Liu et al, 2006; Wasmuth et al, 2012). Rrp6 and Rrp44 modulate each others' activities when simultaneously present in 11-subunit exosomes, with Rrp6 stimulating Rrp44, but substrates engaged with Rrp44 blocking access to Rrp6 (Wasmuth et al, 2012).

In the nucleus, the exosome acts on diverse substrates (Gudipati, 2012; Schneider et al, 2012) and fulfils numerous functions. These include complete degradation of the substrate, for example (i) surveillance of aberrant transcripts such as misfolded tRNAs (Wlotzka et al, 2011), (ii) turnover of excess pre-mRNAs/pre-tRNAs to regulate expression (Gudipati, 2012; Schneider et al, 2012) and (iii) turnover of RNA processing by-products such as the rRNA 5' ETS (Lebreton et al, 2008; Schaeffer et al, 2009; Schneider et al, 2009). Notably, the exosome degrades many lncRNAs, including CUTs (Neil et al, 2009; Preker et al, 2008; Xu et al, 2009), SUTs (Schneider et al, 2012), MUTs (Lardenois et al, 2011) and PROMPTs (Preker et al, 2008). The exosome also participates in processing of precursor RNAs (Allmang et al, 1999a).

In these various roles, the three exosome catalytic activities make different contributions. For example, within pre-rRNAs, Rrp44 makes the major contribution to 5' ETS turnover, whereas Rrp6 predominantly degrades products of aberrant Pol I stalling in 18S, and Rrp44 and Rrp6 act at sequential steps in 5.8S production (Schneider et al, 2012; Thomson et al, 2010). Overall, however, Rrp6 and Rrp44 targets largely overlap, and both Rrp6 and Rrp44 participate in CUT and SUT turnover (Gudipati, 2012; Schneider et al, 2012). Furthermore, the endonucleolytic and exonucleolytic activities of Rrp44 cooperate at most substrates, although their relative contributions vary (Schneider et al, 2012). For example, CUTs and SUTs depend more on exonuclease activity, whereas both activities cooperate on small, structured substrates such as pre-tRNAs (Gudipati, 2012). This suggests that structured substrates require both Rrp44-dependent exonucleolysis (which powers substrate unwinding) and endonuclease activity (which cleaves stalled substrates). In situations where cooperation between Rrp44 and Rrp6 is not required, these nucleases might function independently of the exosome core (Callahan et al, 2008; Schneider et al, 2012). Furthermore, the nucleases might play non-catalytic roles, as Rrp6 promotes deadenylation and retention of aberrant mRNAs independently of catalytic activity (Hilleren et al, 2001; Milligan et al, 2005).

Exosome cofactors

The exosome has some intrinsic substrate specificity, as Rrp6 and Rrp44 activity is influenced by RNA structure and nucleotide composition (Liu et al, 2006), and the core barrel imposes a requirement for a 31-33 nt single-stranded region (Bonneau et al, 2009. Consistently, genome-wide RNA melting temperature analyses reveal that low Tm substrates are most susceptible to exosome-mediated turnover {Wan, 2012 #2030). Furthermore, isolated Rrp44 can recognise hypomodified tRNAs (Schneider et al, 2007). However, the greatest contribution to exosome targeting arises from cofactors, without which purified exosomes show only weak activity (LaCava et al, 2005).

The Nrd1-Nab3 complex acts as an exosome cofactor, recruiting the exosome to degrade Pol I, II and III transcripts, including CUTs and antisense lncRNAs, and process snoRNAs and snRNAs (Creamer et al, 2011; Grzechnik et al, 2008; Jamonnak et al, 2011; Wlotzka et al, 2011). Notably, the involvement of Nrd1 and Nab3 in the turnover of Pol III transcripts suggests they can function independently of termination (Jamonnak et al, 2011; Wlotzka et al, 2011). Another cofactor, Rrp47, binds preferentially to structured substrates and recruits Rrp6 (Mitchell et al, 2003a; Stead et al, 2007), whereas the cofactor Mpp6 associates with Rrp44- and Rrp6-containing exosomes and binds relatively unstructured substrates (Milligan et al, 2008). Both Mpp6 and Rrp47 contribute to the turnover of CUTs (Milligan et al, 2008).

Perhaps the best studied exosome cofactor is the Trf4/Trf5-Air1/Air2-Mtr4 (TRAMP) complex (LaCava et al, 2005). Various TRAMP compositions exist, each with a poly(A) polymerase (Trf4 or Trf5), an RNA binding protein (Air1 or Air2) and the helicase Mtr4. Together, these activities unwind substrates and append a short oligo(A) tail, both of which contribute to exosome recruitment and progression through structured regions (LaCava et al, 2005; Vaňáčová et al, 2005; Wyers et al, 2005).

The Air proteins contain five zinc knuckles (ZnKs), and these domains typically interact with exposed guanosines in single-stranded regions of RNA, or with RNA duplexes via

linkers between the ZnKs (D'Souza et al, 2005). Biochemical studies, together with an NMR structure of Air2 ZnK1-ZnK5 (Holub et al, 2012), and a crystal structure of Air2 ZnK4-ZnK5 in complex with Trf4 (Hamill et al, 2010), reveal that ZnK5 binds to Trf4, ZnK4 activates Trf4 activity (without directly binding), and ZnK2, ZnK3 and ZnK4 bind RNA. Additionally, the N-terminal region of Air2 binds Mtr4. Trf4 and Trf5, which share 65% sequence identity, were mistakenly reported to have template-dependent DNA polymerase activity (Wang et al, 2000), but actually function as non-canonical poly(A) polymerases (Haracska et al, 2005).

Mtr4 is a helicase that shares 38% sequence identity with its cytoplasmic exosome-associated counterpart Ski2. Two crystal structures of Mtr4 (Jackson et al, 2010; Weir et al, 2010) reveal a conserved helicase core that interacts with Trf4 and Air2, is sufficient for helicase activity, and is suggested to dock to the top of the exosome core. Additionally, Mtr4 and Ski2 have a unique insertion not seen in other Ski2 family helicases, which in Mtr4 adopts a stalk-like structure with a terminal RNA-binding KOW domain. Although dispensable for helicase or polyadenylation activities, the KOW domain is required for exosome-mediated decay of $tRNA_i^{Met}$ (Holub et al, 2012), to which it binds (Weir et al, 2010), and trimming of 5.8S+30 (Jackson et al, 2010). It is therefore suggested to contribute to the recognition or transfer of substrates between TRAMP/exosome subunits.

The activities and affinities of Trf4 and Mtr4 are highly interdependent. Mtr4 regulates the length of the oligo(A) tail added by Trf4, sensing the number of 3'-terminal adenosines and modulating Trf4 activity to promote (for A<4) or suppress (for A>=4) adenylation (Jia et al, 2012). Once the oligo(A) tail reaches five residues, Mtr4 can bind and unwind the substrate. This requires both Mtr4 and Trf4 to frequently sample the 3' end of the transcript, and consistently, Trf4 adds adenylate residues in 1-5 nt bursts (Jia et al, 2012). Besides generating a single-stranded platform for Mtr4 to bind, Trf4 also stimulates Mtr4 unwinding activity (Jia et al, 2012). TRAMP has highest affinity for, and shows fastest unwinding on,

short oligo(A) tails, suggesting that Mtr4 and Trf4 precisely tune each other's activities to ensure that polyadenylation and unwinding is efficiently coupled. The short length of oligo(A) tails (1-5 nt) might ensure that they are not bound by Pab1, which requires at least 12 adenylate residues (Sachs et al, 1987). Mtr4 is more abundant than other TRAMP components, so might also function independently (LaCava et al, 2005).

The TRAMP complex assists the exosome in many of its functions, presumably via generation of an unstructured 3' end (via adenylation/unwinding activities) that can thread through the core exosome barrel. Mtr4 might then continue to sit atop the exosome cap and assist in substrate unwinding. The TRAMP complex also makes non-catalytic contributions to the stimulation of Rrp6 activity (Callahan et al, 2010) and surveillance of defective mRNPs (Rougemaille et al, 2007) and CUTs (Houseley et al, 2007). Indeed, many changes in gene expression in *trf4Δ* or *trf5Δ* yeast can be rescued by a catalytically inactive Trf4 mutant (Paolo et al, 2009). Non-catalytic roles might involve physically recruiting the exosome to targets, or acting as a scaffold for surveillance complexes.

The TRAMP complex participates in the surveillance of diverse transcripts, including pre-rRNA fragments arising from transcriptional pausing (El Hage et al, 2010), aberrant precursor RNAs (Dez et al, 2006; Hilleren et al, 2001; Kadaba et al, 2006; Milligan et al, 2005; Wlotzka et al, 2011), and (iii) many lncRNAs (Davis et al, 2006; Neil et al, 2009; Schneider et al, 2012; Wlotzka et al, 2011; Wyers et al, 2005). Trf4 is reported to make a greater contribution to CUT surveillance than Trf5 (Paolo et al, 2009). The TRAMP complex also contributes to the regulatory turnover of functional transcripts (Schneider et al, 2012), and processing/maturation of structural and some messenger RNAs (Allmang et al, 1999a; Ciais et al, 2008; de la Cruz et al, 1998; Grzechnik et al, 2008).

5' to 3' degradation also contributes to nuclear surveillance

The majority of nuclear surveillance is performed by the exosome, but the 5'-3' exoribonuclease Rat1 also contributes to surveillance and processing in addition to

transcription termination. For example, Rat1 participates in rRNA processing and removal of aberrant tRNAs (Chernyakov et al, 2008), and turnover of some lncRNAs, including TERRA (Luke et al, 2008) and those from the *GAL* locus (Geisler et al, 2012). This is preceded by decapping by Dcp1 and Dcp2, which are predominantly cytoplasmic but can enter the nucleus (Groušl et al, 2009). Rat1 is also involved in co-transcriptional quality control of mRNAs with defective caps (Jiao et al, 2010).

Nuclear surveillance checkpoints

Most, if not all, transcripts are susceptible to nuclear surveillance, with the exosome recruited co-transcriptionally via elongation factors and Hrp59 in *Drosophila* (homologous to yeast Gbp2) (Andrulis et al, 2002; Hessle et al, 2009; Hessle et al, 2012), Nrd1 in yeast (Honorine et al, 2011), and perhaps other factors. Despite the prevalence of the surveillance machinery, however, if a Rnt1 site or self-cleaving ribozyme is inserted into a gene, the non-adenylated upstream product of cleavage is stably exported to the cytoplasm (Dower et al, 2004; Meaux et al, 2011). Any non-adenylated products generated by the canonical CF/CPF machinery would be rapidly degraded. This reveals that the presence of the surveillance machinery is not sufficient for surveillance, but additional factors or events act as triggers, functioning at specific "checkpoints".

Perhaps the best characterised nuclear surveillance checkpoint occurs during 3' end processing, when the mRNP begins to establish contacts with the nuclear pore and is released from the site of transcription (Assenholt et al, 2008; Rougemaille et al, 2008). This checkpoint is triggered by defects in elongation, 3' end processing or export, and retains the transcript at or near the site of transcription, dependent on Rrp6 (Hilleren et al, 2001; Jensen et al, 2001; Libri et al, 2002; Qu et al, 2009; Rougemaille et al, 2008; Rougemaille et al, 2007; Thomson et al, 2003) (in some cases requiring Rrp6 catalytic activity (Assenholt et al, 2008)). The TRAMP complex, Rrp44-containing exosome (Hilleren et al, 2001), Ccr4 deadenylation complex (Assenholt et al, 2011; Azzouz et al, 2009) and Nrd1-Nab3 complex

(Honorine et al, 2011; Vasiljeva et al, 2006) also participate. The intimate coupling of this checkpoint with 3' end processing is reflected by the requirement for a polyadenylation site in many Rrp6-dependent events (Saguez et al, 2008), and the physical interaction between Rrp6 and Pap1 (Burkard et al, 2000).

Following this checkpoint, most aberrant transcripts are degraded by the TRAMP-assisted exosome after Pap1-mediated adenylation is overcome, via downregulation of Fip1 (Saguez et al, 2008) or stimulation of deadenylation (Milligan et al, 2005), and dependent on Rrp6. Indeed, Rrp6 plays a key role in promoting turnover (Libri et al, 2002; Rougemaille et al, 2007; Saguez et al, 2008), either via catalytic (Assenholt et al, 2008) or non-catalytic roles (Milligan et al, 2005)). Notably, some aberrant mRNAs retain their Pap1-dependent poly(A) tails, or are even hyperadenylated, and accumulate in foci (Jensen et al, 2001; Libri et al, 2002; Qu et al, 2009; Rougemaille et al, 2007; Thomson et al, 2003). Here, Pap1 might outcompete Rrp6-dependent deadenylation, nuclear deadenylases might be denied access, or defects might have arisen after the mRNPs acquired a stable poly(A) tail. The hyperadenylated transcripts are translationally incompetent, thus nuclear retention prevents deleterious effects in the cytoplasm (Kallehauge et al, 2012).

Surveillance can also be triggered by nuclear pore-associated factors, such as Mlp1 and Mlp2 (Galy et al, 2004; Vinciguerra et al, 2005). It is currently unclear whether this is a distinct checkpoint, or whether NPC-associated and 3' end processing-associated surveillance constitute a single event.

Together, therefore, nuclear surveillance might play a major role in lncRNA turnover, and it is tempting to speculate that lncRNAs and mRNAs are distinguished during one or more of the key surveillance checkpoints.

Cytoplasmic turnover

In contrast to nuclear turnover, which predominantly functions in surveillance and processing, the primary function of cytoplasmic decay is to regulate the stability, and thus abundance, of mature mRNAs. In the cytoplasm, mRNA poly(A) tails are bound by Pab1, which circularises mRNAs via Pab1-eIF4G-eIF4E-cap interactions and facilitates translation. Bulk mRNA decay is initiated by removal of the poly(A) tail by the Ccr4-Pop2-Not and Pan2-Pan3 deadenylase complexes (catalytic subunits underlined) (Boeck et al, 1996; Tucker et al, 2002). Ccr4 is the major cytoplasmic deadenylase (Goldstrohm et al, 2007; Tucker et al, 2002), although Pan2-Pan3 initially trims the poly(A) tail from ~90 residues to ~65 (Brown et al, 1998).

Deadenylation is stimulated by several factors, including (i) defective translation initiation, (ii) normal translation termination, via the termination factor eRF3 interacting with Pab1 (Funakoshi et al, 2007; Kobayashi et al, 2004), and (iii) 3' UTR sequence elements. Deadenylation is also regulated by poly(A) binding proteins, as Pab1 stimulates Pan2-Pan3 (Boeck et al, 1996) via direct binding (Mangus et al, 2004a) and inhibits Ccr4 (Tucker et al, 2002), whereas Nab2 inhibits Pan2-Pan3 (Schmid et al, 2012; Viphakone et al, 2008). Both Ccr4 and Pan2 are also present in the nucleus (Assenholt et al, 2011; Azzouz et al, 2009), and exchange of poly(A) binding proteins might result in differing mRNP susceptibilities to these deadenylases in the nucleus versus the cytoplasm.

Once the poly(A) tail is reduced to ~10-12 residues, Pab1 is displaced and the 3' end is accessible to the cytoplasmic exosome and Pat1/Lsm1-7 complex, which preferentially binds short oligo(A) tails (Chowdhury et al, 2007). This complex activates decapping, which together with 5'-3' exonucleolysis constitutes the more rapid cytoplasmic decay pathway (Cao et al, 2001). Pab1 eviction might also play a direct role in activating decapping, as without Pab1, decapping and 5'-3' degradation do not require deadenylation. Decapping also requires dissociation of the cytoplasmic cap binding complex (eIF4E-eIF4G), and is

catalysed by Dcp2 with its cofactor Dcp1 (Deshmukh et al, 2008). Together with the oligo(A)-bound Pat1/Lsm1-7 complex and other activators, Dcp1 and Dcp2 contribute to a complex that juxtaposes the mRNA 5' and 3' ends. After decapping, the exoribonuclease Xrn1 binds the resulting 5' monophosphate and processively degrades the transcript, aided by its inherent ability to melt RNA duplexes (Jinek et al, 2011). Notably, Dcp1-Dcp2 and Xrn1 might act co-translationally, behind the elongating ribosome (Hu et al, 2009).

Although decapping and 5' to 3' degradation is considered to be the major decay pathway, *xrn1Δ* or *dcp1Δ* strains have few defects (He et al, 2003a) suggesting that most functions can also be carried out by the cytoplasmic exosome. Consistently, simultaneous ablation of both pathways is lethal (Anderson et al, 1998). Unlike the nuclear exosome, the cytoplasmic exosome associates with just one nuclease, Rrp44, and is regulated by distinct cofactors, including AU-rich element binding proteins, the Ski complex, and the Dom34-Hbs1 complex.

The Ski complex is a heterotetramer of Ski2, Ski3 and two copies of Ski8 (Brown et al, 2000; Synowsky et al, 2008) that assists the exosome in bulk mRNA decay (Anderson et al, 1998; Araki et al, 2001; van Hoof et al, 2000). Ski8 acts as a scaffold (Cheng et al, 2004b), and Ski3 bridges Ski8 and Ski2, a DExH-box RNA helicase related to Mtr4 (Wang et al, 2005). The Ski2 crystal structure reveals an insert similar to that in Mtr4, but lacking a KOW motif and binding RNA with less specificity (Halbach et al, 2012). The Ski complex binds the exosome via the N-terminal domain of an additional factor, Ski7 (Araki et al, 2001), which also contains a C-terminal domain homologous to translation release factors. The Ski complex might assist the exosome via substrate unwinding and/or recruitment.

Pathways for aberrant mRNA removal

In addition to bulk mRNA decay, specialised pathways exist in the cytoplasm to remove aberrant mRNAs. Nonsense-mediated decay (NMD) eliminates mRNAs containing "premature" termination codons, which arise from mutations, long 3' UTRs (Kebaara et al,

2009), endogenous ribosomal frameshift sequences (Belew et al, 2011), short upstream open reading frames (Guan et al, 2006), or alternative out of frame translation initiation sites (Guan et al, 2006). The first step in NMD involves recognition of the premature stop codon by a complex comprising the canonical termination factors eRF1 and eRF3, and the NMD-specific factor Upf1 (Amrani et al, 2004). Essentially, a stop codon is recognised as "premature" if it lacks a canonically configured 3' UTR immediately downstream, but the particular features recognised by the NMD machinery are unclear. Notably, a long 3' UTR favours NMD, and Pab1 protects against NMD when bound proximal to the translation termination site by interacting with eRF3 (Amrani et al, 2004). One model is that a short UTR (typical in yeast) enables Pab1 to compete with Upf1 to bind eRF3 and suppress NMD. However, a recent study refutes this model (Kervestin et al, 2012), and neither Pab1 nor poly(A) tails are essential for NMD substrate specificity (Meaux et al, 2008). Other factors that regulate NMD include Hrp1 (González et al, 2000) and Pub1 (Ruiz-Echevarría et al, 2000). Following Upf1 recruitment, a cascade of events involving additional Upf proteins (Chakrabarti et al, 2011) culminates in deadenylation-independent decapping and Xrn1-dependent decay, or deadenylation and exosome-mediated decay. The latter pathway involves Ski7 recruitment via interactions with Upf1 (Mitchell et al, 2003b; Takahashi et al, 2003).

Another pathway, no-go decay, is triggered by stalled translation elongation, which causes the eRF1 and eRF3 homologues Dom34 and Hbs1 (respectively) to bind the ribosome and trigger ribosomal subunit dissociation (Becker et al, 2011; Shoemaker et al, 2010) and mRNA dissociation. This triggers mRNA cleavage, and subsequent mRNA decay by Xrn1 and the exosome (via Ski7 recruitment) (Meaux et al, 2006)

A third pathway, non-stop decay, targets mRNAs with no stop codon, on which the ribosome continues to elongate. These substrates can arise when alternative poly(A) sites occur within ORFs. Here, the ribosome, stalled at the end of the poly(A) tail, is recognised by the C-

terminal domain of Ski7, perhaps via it mimicking eRF3 and binding in the empty A site (Tsuboi et al, 2012; van Hoof et al, 2002). Exosome-mediated decay ensues, which, unusually, depends mostly on the endonucleolytic, rather than exonucleolytic, activity of Rrp44 (Schaeffer et al, 2011).

Cytoplasmic lncRNA surveillance

The cytoplasmic decay machinery is therefore not only important for bulk mRNA turnover, but also plays a major role in degrading non-translatable transcripts. This suggests that, if lncRNAs are able to reach the cytoplasm, they might be particularly susceptible to turnover. Notably, although nuclear surveillance is more widely reported to participate in lncRNA turnover, some lncRNAs accumulate when cytoplasmic decay factors are deleted or mutated. For example, CUTs such as *SRG1* (Thompson et al, 2007) and those present at metal ion homeostasis (Toesca et al, 2011) and galactose utilisation genes (Geisler et al, 2012), some SUTs (Marquardt et al, 2011) and other lncRNAs (Berretta et al, 2008; van Dijk et al, 2011) are sensitive to Xrn1, Dcp1 and Dcp2. A subset of lncRNAs are also sensitive to NMD (Marquardt et al, 2011; Thompson et al, 2007; Toesca et al, 2011). In many cases, decay is independent of Ski2 and Ccr4 (Thompson et al, 2007) and cytoplasmic decapping activators (Geisler et al, 2012; Marquardt et al, 2011). Together, this suggests that some lncRNA decay occurs in the cytoplasm, via similar, but not identical, pathways to mRNA decay.

1.6 Aims

Overall, the synthesis, transport and turnover of mRNAs and other Pol II transcripts involves numerous factors, which tightly regulate each stage and contribute to a variety of alternative pathways. These factors determine the properties of a transcript, and ultimately its function. It is now apparent that besides the canonical Pol II transcripts, pervasive transcription generates abundant long non-coding RNAs from diverse genomic loci. Numerous functions have been documented for lncRNAs and the act of their transcription. However, information

about lncRNA production, processing and turnover, which would improve our understanding of their functions, is lacking. Notably, despite similarities early in transcription, lncRNAs and mRNAs perform very different functions. I reasoned that lncRNAs, like mRNAs, might interact with a defined series of factors during their biogenesis and decay, and that investigating whether mRNA-binding factors also bind lncRNAs would offer insight into the similarities and differences between lncRNA and mRNA metabolism. I therefore aimed to:

1. Test mRNA binding proteins to see which, if any, interact with lncRNAs, and what lncRNAs they interact with.
2. Establish whether these proteins perform canonical or non-canonical functions (if any) in steady-state lncRNA metabolism.
3. Establish whether different classes of lncRNAs can be defined based on their interactions with different proteins.
4. Investigate whether these factors play dynamic roles in regulating lncRNA expression in response to a nutrient shift.

2: Materials and methods

2.1 Materials

Growth media

Media	Composition
YPD	1% (w/v) yeast extract, 2% (w/v) bacto-peptone, 2% (w/v) glucose
YPD agar	**YPD** + 2% (w/v) agar
YPD-Kan agar	**YPD** agar + 200 μg/ml G418
YPD-Nat agar	**YPD** agar + 100 μg/ml nourseothricin
YPGal	1% (w/v) yeast extract, 2% (w/v) bacto-peptone, 2% (w/v) galactose
YPGal agar	**YPGal** + 2% (w/v) agar
SGlu -U agar	0.69% (w/v) yeast nitrogen base without amino acids, 2% (w/v) glucose, 770 mg/l CSM lacking uracil (Formedium), 2% (w/v) agar
SGlu -W	0.69% (w/v) yeast nitrogen base without amino acids, 2% (w/v) glucose, 740 mg/l CSM lacking tryptophan (Formedium)
SGlu -WLU	0.69% (w/v) yeast nitrogen base without amino acids, 2% (w/v) glucose, 620 mg/l CSM lacking tryptophan, leucine and uracil (Formedium)
SGlu -WLU agar	SGlu -WLU + 2% (w/v) agar
SGlyEtOH -W	0.69% (w/v) yeast nitrogen base without amino acids, 2% (v/v) glycerol, 2% (v/v) ethanol, 740 mg/l CSM lacking tryptophan (Formedium)
SGlyEtOH -WLU	0.69% (w/v) yeast nitrogen base without amino acids, 2% (v/v) glycerol, 2% (v/v) ethanol, 620 mg/l CSM lacking tryptophan, leucine and uracil (Formedium)

Table 2.1: Growth media used in this study.

Plasmids

Plasmid	Purpose and/or description	Source or reference
pBS1539-HTP-URA3	HTP tagging or *URA3* knock in (yAT1)	(Granneman et al, 2009)
pFA6a-natMX6	*RRP6* deletion (for *NAB2::HTP rrpp6Δ*)	(Hentges et al, 2005)
pRS425	Parent of pAT1	(Brachmann et al, 1998)
pAT1	Fui1 overexpression; as pRS425, but with *FUI1* (amplified from genomic DNA) inserted with 597 bp of upstream and 360 bp of downstream sequence	(Swiatkowska et al, 2012)

Plasmid	Purpose and/or description	Source or reference
pFA6a-kanMX6-PGAL1-3HA	*HRP1* depletion (yAT2 construction)	(Longtine et al, 1998)

Table 2.2: Plasmids used in this study.

Strains

Strain	Genotype	Source or reference
BY4741 ("wild-type")	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	(Brachmann et al, 1998)
HTP-tagged strains	As BY4741 but with *gene-HTP::K.I.URA3*	This study
yAT1	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 URA3*	(Swiatkowska et al, 2012)
yAT2	As BY4741 but with *kanMX6::pGAL1::3HA::HRP1*	This study

Table 2.3: Yeast strains used in this study. All strains were derived from BY4741, itself derived from S288C. Strains Hrp1-HTP and Mtr4-HTP were a gift from Wiebke Wlotzka, and Mtr4-HTP a gift from Elizabeth Petfalski.

Oligonucleotides

HTP strain construction (HTP F/R) and checking (C/D) oligos

Cbc2 C	TATAGTTGTCCAGATGAAGCATTGA
Cbc2 D	CTAATTGATAGGCTGGAAGAATTGA
Cbc2 HTP F	TCAGACCAGGTTTCGATGAAGAAAGAGAAGATGATAACTACGTACCTCAGGAGCACCATCACCATCACC
Cbc2 HTP R	TATATATATATCTGTGTGTAGAATCTTTCTCAGATATAAATTGATTGATTTACGACTCACTATAGGG
Cft1 C	ATTCCACCAATAGTTGTATGATGCT
Cft1 D	GGCATATTTAAATTGGACCTTCTTT
Cft1 HTP F	GAGATATCATAAATATTGAATTTTCAATGAGATCTTTATGCCAGGGTAAGGAGCACCATCACCATCACC
Cft1 HTP R	AGATGTATATATCGGGCTAGTTAGCATTAAATATCCTTTATATGTATTGTTACGACTCACTATAGGG
Cft2 C	GAAGAAATTACGGCAAAACTTATCA
Cft2 D	TTCTACCGTGTTCTTTCTCCATAAC
Cft2 HTP F	AACTTTTTGATACTGTCAAAAAATTGGTTACGGATATGTTAGCAAAAATCGAGCACCATCACCATCACC
Cft2 HTP R	CAGTCCACGAACGTAAATTGAACCTTTTATTTGTGCTGTATACAAGAAGGTACGACTCACTATAGGG
Crm1 C	TGGTTTATGATAACAAGATTTCGGT
Crm1 D	ACAGTTTCGTTAAAATTCAAGATGC
Crm1 HTP F	AAAAAGCTGCCAAGATTGGTGGGTTATTAAAACCTTCCGAACTTGATGATGAGCACCATCACCATCACC
Crm1 HTP R	AATAAAAGGGAAAAAATATTGGAAATTTAAAGAATGATACGCCACCGCCTTACGACTCACTATAGGG
Dbp5 C	TGACAATTGGATCTTCCATTATTTT
Dbp5 D	TTCCCGAATTATGTAGAGTTAGCAG
Dbp5 HTP F	CGGATGATTGGGATGAAGTCGAAAAAATAGTTAAGAAAGTGTTAAAGGATGAGCACCATCACCATCACC
Dbp5 HTP R	TGAAATTAGATTAAAGCTTTTACGTATTTTGAGGTATTATGTACTGAATTTACGACTCACTATAGGG
Dcp2 C	AACAGGTCTAGTGCAATAAATGAGC
Dcp2 D	ATTAGGGCAATCTCACACACATTAT
Dcp2 HTP F	CGAATGGAACTTCAGGGTCTAATGAATTATTAAGCATTTTGCATAGGAAGGAGCACCATCACCATCACC
Dcp2 HTP R	CGGCTGCCTTCATTTACAGTGTGTCTATAAAACGTATAACACTTATTCTTTACGACTCACTATAGGG
Fip1 C	CAGATGTATCCAATACCATCACAAA
Fip1 C	CAGATGTATCCAATACCATCACAAA
Fip1 D	AAGTGATTGGTACAATAAGCTCCAG
Fip1 D	AAGTGATTGGTACAATAAGCTCCAG
Fip1 HTP F	TGCCACCCATGAACCAACAGCCTAATCAAAATCAAAATCAAAATTCGAAAGAGCACCATCACCATCACC
Fip1 HTP R	TACGTGAATCATGGACGATATATTTTTATTTCAATTTCACAGATCCCCATTACGACTCACTATAGGG
Gbp2 C	CCTTGAAGATACCAGAGGTACTGAA
Gbp2 D	ATAAAGACAATAGCACAACCCAGAG
Gbp2 HTP F	ATTATAATTATGGTGGTTGTAGTTTACAGATCTCTTATGCTAGACGTGATGAGCACCATCACCATCACC
Gbp2 HTP R	TATACGTTATCATAAAGTACACAGGTCATGGTTCGGTTGGTGCTTAGGAATACGACTCACTATAGGG
Hek2 C	ATCATGATAACAAAGAGGAGCAGTC

Hek2 D	TTCTTGGAGTTTATCATTCATCCAT
Hek2 HTP F	AAGAGAAGGAAGAACCTCAAGAGAATCATGATAACAAAGAGGAGCAGTCGGAGCACCATCACCATCACC
Hek2 HTP R	GATGATAGTTTGTTTTGTCTGTGTGGGACGTGCGCACGCACACGTATATATACGACTCACTATAGGG
Hrb1 C	AACTGGAATAGCTGTCGTTGAATAC
Hrb1 D	AATGAAAACATAGAGCAAAAGCAAC
Hrb1 HTP F	ATTATAACTATGGGGGTTGTGATTTGGATATATCGTACGCTAAACGCCTCGAGCACCATCACCATCACC
Hrb1 HTP R	TACTTGTCGCAGATCCAATAGGTGAGAAAGTATATAGATCGAGAGTAGTTTACGACTCACTATAGGG
Lsm1 C	GCTGCTATTGTAAGCTCAGTAGACC
Lsm1 D	AGTTTTAAAAGTGATGAGCATGGAA
Lsm1 HTP F	AAATGGCCCGCCATGGTATCGTTTACGATTTCCATAAATCTGACATGTACGAGCACCATCACCATCACC
Lsm1 HTP R	TTTGAGAGTTTACTCCAGGATATATGTTGGTAGTATTGTGTTTTTCTTTCTACGACTCACTATAGGG
Lsm8 C	AGCTCCTATAGACGAAAAGAAGGTC
Lsm8 D	CTTTTAAAGATTATCCAGACACATGC
Lsm8 HTP F	TCGAAAATGAGCATGTAATATGGGAAAAAGTGTACGAATCAAAGACAAAAGAGCACCATCACCATCACC
Lsm8 HTP R	GGGTTAATGCTTAAATTTATTGTATGATTTATATACTTCTATACATGGTATACGACTCACTATAGGG
Mex67 C	CATGGTTACAACTCCACATCAAATA
Mex67 D	GGCAAGAGCATTTTGAGAATAAGTA
Mex67 HTP F	CAGAGTAGCATGAATGGCATCCCTAGAGAAGCATTTGTGCAGTTCGAGCACCATCACCATCACC
Mex67 HTP R	TATATTTTTTGTGATACTGTGCGGCTGAAACAGGGAACAATATCATACGACTCACTATAGGG
Mpe1 C	GACTACCAAAGGAAAAGAGAGAACC
Mpe1 D	TGTTAGATATGGAAAGAGAGGATGC
Mpe1 HTP F	CAACGGCTACTATCACAAATCCTCATCAAGCTGACGCAAGCCCTAAGAAAGAGCACCATCACCATCACC
Mpe1 HTP R	ACGTATGTGAAGCCAAGTAGGCAATTATTTAGTACTGTCAGTATTGTTATTACGACTCACTATAGGG
Nab2 C	TTGTCGAAGATCAAGGAAGTAAAAC
Nab2 D	AACCCAGTCTGTCCTTAACTCTTTT
Nab2 HTP F	CTCCTCCGCAAACCAGTTTTACGCACCAAGAACAAGATACGGAAATGAACGAGCACCATCACCATCACC
Nab2 HTP R	ATCAAAAGGGTCACAGGAACATGAATTTCGTTCCGTGATTTTAATAGTAATACGACTCACTATAGGG
Pab1 C	AGCAGCTGGTAAAATTACTGGTATG
Pab1 D	TACCCTCACTTGATTTGTCAATTTT
Pab1 HTP F	CTGCCTATGAGTCTTTCAAAAAGGAGCAAGAACAACAAACTGAGCAAGCTGAGCACCATCACCATCACC
Pab1 HTP R	AAAAAGATGATAAGTTTGTTGAGTAGGGAAGTAGGTGATTACATAGAGCATACGACTCACTATAGGG
Pap1 C	GAAAATGAAAAGAGACCATCAAAGA
Pap1 D	TTTGTTATATTTGTCAATCCTGGGT
Pap1 HTP F	CTGCTTCAGGTGACAACATCAATGGCACAACCGCAGCTGTTGACGTAAACGAGCACCATCACCATCACC
Pap1 HTP R	TGACTGATTAACCTATATTAATAAACTATTCAACTATAAATAGGAATGTCTACGACTCACTATAGGG
Pcf11 C	ACCACAAGGAAAAATATTCAATCAA
Pcf11 D	TCAAGTGAATAAGGAGATAAAACCG
Pcf11 HTP F	CTAATAGTGGCAAGGTCGGTTTGGATGACTTAAAGAAATTGGTCACAAAAGAGCACCATCACCATCACC
Pcf11 HTP R	TAATATAATATATAGTTATTAAATTTAAATGTATATATGCAGTTCTGCTCTACGACTCACTATAGGG
Pti1 C	ATTATCAAACCAAGAAAACATCCAA
Pti1 D	GATCCACGTTATCTAACATTTGAGG
Pti1 HTP F	CAAGTGACCAGCAACTTATGGTGGAAAACTTTAGAAAAGAATATATAATCGAGCACCATCACCATCACC
Pti1 HTP R	AGAACCATCTAATTAGGTGTGGTTTTCTGATACGCTATGGCTCTGATTACTACGACTCACTATAGGG
Ref2 C	TTTGGATGATTCCCAACTAATAAAA
Ref2 D	TGGTTGACTTATACGAACAGATTGA
Ref2 HTP F	AGCATGTCCCCATAGTAAAAAGAAATAAATATCCTCCAAGACCAGTACACGAGCACCATCACCATCACC
Ref2 HTP R	ATATAAAATGAGTATATATACTACATGTTTATGTATCAGCATGTCATAGCTACGACTCACTATAGGG
Rna14 C	CCAGCTGAAGGAAATATACAAGAAA
Rna14 D	TTCTGAAATATTGGCTTTGAATGAT
Rna14 HTP F	TTTTAAATGATCAAGTAGAGATTCCAACAGTTGAGAGCACCAAGTCAGGTGAGCACCATCACCATCACC
Rna14 HTP R	ATAGATGTGTTGGTATAAATATTCATATATACCTATTTATTAACGTAATGTACGACTCACTATAGGG
Rna15 C	ATTGCTAACGAATGGGATATGTAAA
Rna15 D	TGATTAAAACTATAGCCGTCCTCAG
Rna15 HTP F	CTATTTGGGACTTAAAACAAAAAGCATTAAGGGGAGAATTTGGTGCATTTGAGCACCATCACCATCACC
Rna15 HTP R	CTCATCATTGCGGAACCGCATTTTTTTTTTGTATTTTTGCCTCCCTAGTTTACGACTCACTATAGGG
Rnt1 C	GGCTATGCTTCATTACGTCTACATT
Rnt1 D	CCGATTCCATTTTAAGGATTTCTAT
Rnt1 HTP F	TAAAAGATCCCTCACAAAAGAATAAGAAAAGAAAATTCTCAGATACAAGCGAGCACCATCACCATCACC
Rnt1 HTP R	GCTAAAGAAAATCAATGCAAGTTCCATCATGGTTGTGTAAAAGGAACGTTTACGACTCACTATAGGG
Ski2 C	GGTTTATCATTCAAAGAAATCATGG
Ski2 D	GGATGTGGTCTGGTAATAAATTCAG
Ski2 HTP F	CTCAAGAGTTGATTAAGAGAGATATTGTTTTCGCCGCAAGTTTGTATTTAGAGCACCATCACCATCACC
Ski2 HTP R	AAAACATTAACTTTTATAAACATGACTCACATTGAGAATAAATGAGCTCTTACGACTCACTATAGGG
Sto1 C	GGTTTGAAGAATAATGGTTTGATTG
Sto1 D	TCCAAAGCGGATAGTTATAGAAGTG
Sto1 HTP F	AATCCACTAGGAGAACAATTTCGAATTGGATTCAAGAAACAAAGGAAGTTGAGCACCATCACCATCACC
Sto1 HTP R	AAAATTAAAAAGCGGAGTGATAACGAATGTAGTCCATCCTCCGAATCTTTTACGACTCACTATAGGG
Sub2 C	GGTACTAAGGGTTTGGCTATTTCAT
Sub2 D	CAGAAATCTCTTCACTTTTCTTTGC
Sub2 HTP F	GAATTCCCAGAAGAAGGCATTGATCCGTCCACTTATTTGAATAATGAGCACCATCACCATCACC
Sub2 HTP R	AAAATCTTTATATAATCTATATAAAAACGTATCTTTTTTCCTTTATACGACTCACTATAGGG
Swt1 C	TGGATTCATAGATTACAACCCAGAT
Swt1 D	GTTGTTAATGTTGTTCAAAGTGCAG

Swt1 HTP F	TAGAGAAACAGATTCATGAATGGAAAACATCTATCAATGCTATATCAACCGAGCACCATCACCATCACC
Swt1 HTP R	TGTGTGAGTGGCTTCAATAAATCTGATAGATTTACACTAAGCAAAACATTTACGACTCACTATAGGG
Tho1 C	TCAGTAGGAAAAATGAACCTGAAAG
Tho1 D	CCCACCTGCATTACTATGAACTATT
Tho1 HTP F	GCTCCAGAGTAAGTAAAAACAGGAGAGGCAACCGCTCTGGTTACAGAAGAGAGCACCATCACCATCACC
Tho1 HTP R	GGAAAGAACCGAAACTAGAATGAAAAACTCCACCAAAACGGCTTGAGCCTTACGACTCACTATAGGG
Tho2 C	CGAATAAAAGATTCAAGAAGGATGA
Tho2 D	CGATAAAAGAAAGAACGGTTTGTTA
Tho2 HTP F	CGCTTCCGCAAGGTCCCAAGGGTGGGAATTACGTCAGTAGGTACCAGAGGGAGCACCATCACCATCACC
Tho2 HTP R	CTATCAAAGTACACGTTAAAATTCAGCTCGGGTATGTTAAGTACTAGTAATACGACTCACTATAGGG
Thp1 C	TACTGATCGTGCTCCTAGAGAAACT
Thp1 D	CCCTGTTTGAGAGTTGATAATCTTG
Thp1 HTP F	TTAATGAACGAATCACCAAGATGTTTCCTGCCCATTCTCACGTTCTTTGGGAGCACCATCACCATCACC
Thp1 HTP R	AACAGCATAATGTCGCTCCTTCTCTTCTATTTCCTATATATATCTACATATACGACTCACTATAGGG
Tif1 C	TTTCTGCCATCTACTCTGATTTACC
Tif1 D	GCGTTTGATGTACACTTTTTCTTTT
Tif1 HTP F	ACTCCACTCAAATTGAAGAATTGCCATCCGACATCGCTACTTTGTTGAAC**GAGCACCATCACCATCACC**
Tif1 HTP R	GTTATCAAGATAGCCTCACAAGATACTTTTTTAAGAAGTTTTTGTCTCCC**TACGACTCACTATAGGG**
Xrn1 C	AGATCAAGGAAGTCGTATTGTTGTC
Xrn1 D	AATGAGATCAATGAGAAGAAAGTGC
Xrn1 HTP F	AGTCACAAAGCAATGCTGCTGACCGTGATAATAAAAAAGACGAATCTACT**GAGCACCATCACCATCACC**
Xrn1 HTP R	GATATACTATTAAAGTAACCTCGAATATACTTCGTTTTTAGTCGTATGTT**TACGACTCACTATAGGG**
Yra1 C	TGAGACTTAACCTAATCGTTGATCC
Yra1 D	ACACACGTTTAATCAACCTATCCAT
Yra1 HTP F	GTCTTGAAGATCTGGACAAGGAAATGGCGGACTATTTCGAAAAGAAAGAGCACCATCACCATCACC
Yra1 HTP R	AATTAAATTTAATAAAACCAAATTAAATCAAACAAAAAATTGACAATTAATACGACTCACTATAGGG
Yra2 C	CATGTATTTACGAATTTGAAGACCC
Yra2 D	TAGTTTTTGGCATTCTTTCTTCAAC
Yra2 HTP F	TCTGTCGAAGCTCTTGACGCTGAATTAGATGCTTACATGAAAGGTGAGCACCATCACCATCACC
Yra2 HTP R	ACATTGAAAAGACTATAAAGAAATTATAAGTAGATACACAATTGCTACGACTCACTATAGGG
Ysh1 C	GGAGCTAATTTACTGGCACATTTTA
Ysh1 D	TGCAGTTTTGGCTTAGTCTATTTCT
Ysh1 HTP F	GGGTGGAAAGCCTCTTAAATATTGGTGGTAATTTGGTCACACCGCTATGTGAGCACCATCACCATCACC
Ysh1 HTP R	GGTTTGGTATTACTTCTATAAAGTAGTCTACTTAGTATGCGTAACTGTTTTACGACTCACTATAGGG
Yth1 C	ACTGTACACAAAGTCCAGATTGTCA
Yth1 D	AGAACTAATTGATAGCTCCTCCGAT
Yth1 HTP F	TGGATGAAGAAAAGGAAAGGCGTTTAAACGCAATTATAAACGGTGAAGTTGAGCACCATCACCATCACC
Yth1 HTP R	TATGATAATATACATGTCTATGAAATCGAAAACCCCCAGCCATGCATAGATACGACTCACTATAGGG
pBS1539 R2	CTCACCCTGAAAATACAAATTCTC

5’ linkers for CRAC/nascent RNA sequencing

L5Aa	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrUrArArGrC-OH-3'
L5Ab	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrArUrUrArGrC-OH-3'
L5Ac	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrGrCrGrCrArGrC-OH-3'
L5Ad	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrCrGrCrUrUrArGrC-OH-3'
L5Bb	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrGrUrGrArGrC-OH-3'
L5Bc	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrCrArCrUrArGrC-OH-3'
L5Bd	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrUrCrUrCrUrArGrC-OH-3'
L5Cc	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrArCrUrCrArGrC-OH-3'
L5Cd	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrGrArCrUrUrArGrC-OH-3'

miRCat-33 3′ pre-activated, adenylated cloning linker for CRAC

rAppTGGAATTCTCGGGTGCCAAGG/ddC/

Reverse transcription

miRCat-33™ RT	CCTTGGCACCCGAGAAT	CRAC
P3N11HYMM_RT	CTGAACCGCTCTTCCGATCTNNNNNNNNNNNHYMM	Nascent RNA sequencing

PCR amplification of libraries for Solexa sequencing

Solexa F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Solexa R	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGGCCTTGGCACCCGAGAATTCC

pAT1 construction (Fui1 overexpression)

Fui1 F1 XhoI	ATGATCCTCGAGCCCCCACACTTATCTTTAGAGAC	Fui1 amplification
Fui1 R1	TGCGTGAGCTCTCTTGAGAA	Fui1 amplification
pRS42x F2	CGCGTAATACGACTCACTATAGG	Sequencing oligo
Fui1 F2	CTGGACAACACAGGGACATG	Sequencing oligo
Fui1 F3	GCTCAGGATGACGAAAAGGT	Sequencing oligo

Fui1 F4	CTCCATCTGGATTGTCATCAAC	Sequencing oligo
Fui1 F5	CGCTCCTGACTTCACTAGATTC	Sequencing oligo
Fui1 F6	GCTATTGCGGGTGTAATATCAG	Sequencing oligo
Fui1 F7	TTTCCACTGGCTATGAAAAGATAC	Sequencing oligo

yAT1 construction (BY4741 *URA3* knock in)

pBS1539 URA3 DN45	TTTTTTTTTCGTCATTATAGAAATCATTACGACCGAGATTCCCGGCGGGGGATCCACTAGTTCTAG
pBS1539 URA3 UP45	TGACCATCAAAGAAGGTTAATGTGGCTGTGGTTTCAGGGTCCATAGACTCACTATAGGGCGAATTGG

yAT2 construction (*pGAL::3HA::HRP1*)

Hrp1 Gal 3HA F1	CTTGCAGCCCCTCCCTACGAGCAGCGCCGAGATTTCTTGTTGGACAAAGCGAATTCGAGCTCGTTTAAAC	
Hrp1 Gal 3HA R1	GTAGGCTTATCATCGCCGTAGATGTCGTTGAAATCTTCTTCGTCAGAGCTGCACTGAGCAGCGTAATCTG	
Hrp1 A upstream	GGGTCATGCTATCGAAAACC	Checking oligo
HRP1 B	AGTGACGGTACCATACTTACCAAAA	Checking oligo
MX46F	CCTCGACATCATCTGCCCAGAT	Checking oligo

PCR-based template production for riboprobe synthesis

CUT479 F1	GACCTCGAACGATCGTAAGATG
CUT479 R1 T7	GGATCCTAATACGACTCACTATAGGGAGAGGAGAGGTCCTCCTTCTATCATTTCTG
CUT701 F1	CGAGCTTTGATTGTGTATAAGCA
CUT701 R1 T7	GGATCCTAATACGACTCACTATAGGGAGAGGAAGCCGTTAACACTAGCAAGATG

Rrp6 deletion

Rrp6 F1 UP50	TAGACGAAATAGGAACAACAAACAGCTTATAAGCACCCAATAAGTGCGTTCGGATCCCCGGGTTAATTAA
Rrp6 R1 DN50	ATGAAAATTACCATAATTTATAAATAAAAAAATACGCTTGTTTTACATAAGAATTCGAGCTCGTTTAAAC

2.2 Strain generation

Cassettes for recombination-based genomic integration were generated by PCR amplification (Phusion; according to manufacturer's instructions) from the templates indicated in Table 2.2. Yeast were transformed using the lithium acetate method (Schiestl & Gietz, 1989), and transformants selected on the appropriate selective media. Colonies were screened by PCR (Phire; Finnzymes), after lysis in 20 mM NaOH (20 µl, 98 °C, 20 min).

2.3 Cell growth and harvest

Overnight cultures were diluted to A_{600} 0.05 and grown to OD 0.5.

For CRAC, yeast were grown in SGlu -W, and cultures (2.75 l) crosslinked for 100 s as previously described (Granneman et al, 2011). Cells were harvested by centrifugation (3000 xg, 5 min, 4 °C), washed once in ice cold PBS (13 mM NaCl, 2 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4), pelleted (3000 xg, 10 min, 4 °C) and snap frozen in liquid nitrogen.

For 4TU labelling, cultures (1 l) were spiked with 4TU (4-thiouracil (Sigma); 20 µl of 1.0 M suspension in DMSO), and quenched after 3 min by adding to an equal volume of EtOH pre-

chilled to -80 °C. Cells were harvested by centrifugation (3000 xg, 5 min, 4 °C), washed twice in ice cold water, pelleted (3000 xg, 10 min, 4 °C) and snap frozen in liquid nitrogen.

For nutrient shift experiments, cultures were rapidly filtered using a custom built suction-powered filtration device (MF Millipore filter, diameter 147 mm, pore size 0.9 μm), and first washed then resuspended in pre-warmed media.

For checking strains and Hrp1 depletion experiments, yeast were grown in YPD or YPGal, respectively, 1-3 ODs harvested by centrifugation (1 min, 3000 xg), and pellets frozen in liquid nitrogen.

2.4 Northern blotting

Yeast (3 ODs) were lysed by vortexing with zirconia beads (50 μl) in GTC-phenol (Maniatis, 1982) (5 min, 4 °C), then heated (65 °C, 10 min) with additional GTC-phenol (600 μl). Following the addition of 300 μl 24:1 chloroform:IAA and 160 μl 3 M NaOAc, RNA was isolated by vortexing followed by centrifugation (20,000 xg, 5 min), and recovered by ethanol precipitation. After resuspension in water, 5 μg of RNA was glyoxal-denatured and resolved by agarose gel electrophoresis as previously described (Sambrook, 2001). RNA was transferred to Hybond-N+ (GE) by capillary transfer in 6x SSC (450 mM NaCl, 50 mM $Na_3C_6H_5O_7$, pH 7.2), and fixed by UV irradiation. Blots were prehybridised in ULTRAhyb (Ambion) then incubated with a [α-^{32}P]-labelled riboprobe (65 °C, 16 hr). Riboprobes were *in vitro* transcribed using T7 RNA polymerase (Promega) from ~300 bp templates generated by PCR amplification from yeast genomic DNA. Blots were washed once in 6x SSC and twice in 0.2x SSC, 0.1% (w/v) SDS (15 min, 65 °C), and radioactive signal detected using a fluorescent imaging analyser (FLA-5000 scanner, Fujifilm).

2.5 Protein analysis

To verify expression of tagged proteins, yeast were lysed in 25 mM NaOH, 10 mM DTT (10 min, 4 °C) and proteins precipitated with TCA (5 % v/v) and recovered by centrifugation (20,000 xg, 2 min). Pellets were washed in acetone, resuspended in sample loading buffer (25 mM Tris-HCl pH 7.8, 1% (w/v) SDS, 0.1 M DTT, 0.05% bromophenol blue (w/v), 5% (v/v) glycerol) by heating (95 °C, 4 min), and proteins resolved by SDS-PAGE (10 % gel, 150 V) in 1xTG (25 mM Tris, 192 mM glycine). Proteins were transferred to a Hybond C Extra nitrocellulose membrane (GE) in 1xTG with 15 % MeOH (100 V, 1.5 hr), blocked in TBS-T (150 mM NaCl, 50 mM Tris-HCl pH 7.5, 0.1% Tween-20), and probed with peroxidase anti-peroxidase (Sigma) or peroxidase-conjugated anti-HA. Antibodies were detected by enhanced chemiluminescence (Pierce).

HTP-tagged proteins obtained via a two-step purification as part of the CRAC protocol were tested for abundance and purity by SDS-PAGE (NuPAGE 4-12% Bis-Tris gel system, Invitrogen) followed by silver staining (SilverQuest kit, Invitrogen), according to manufacturer's instructions.

2.6 Crosslinking and analyses of cDNAs (CRAC)

The CRAC protocol has been previously described (Granneman et al, 2009; Granneman et al, 2011), but as there are many minor variations possible, I include here the salient details of the version used in this study.

Cell pellets were vortexed with 1 ml TN150 (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 0.1% (v/v) NP-40, 5 mM β-mercaptoethanol, EDTA-free protease inhibitor cocktail (Roche)) and 2.5 ml zirconia beads (Thistle Scientific) for 5x 1 min pulses, cooling on ice in between. Cell lysates were diluted with an additional 3 ml TN150, and debris removed by centrifugation (20 min, 4600 xg; then 20 min, 20,000 xg; 4 °C). Cleared lysates were incubated with 125 µl IgG beads (IgG Sepharose 6 Fast Flow, GE), rotating at 4 °C for 2 hr. Beads were washed

with TN150 (2x 10 ml) then TN1000 (2x 10 ml; as TN150, but with 1 M NaCl), then His-tagged RNA:protein complexes eluted by TEV cleavage in TN150 (1.5 µl homemade GST-TEV (a gift from Simon Lebaron and Claudia Schneider), 2 hr, 18 °C). The eluate was then treated with RNace-IT (Agilent; 0.1 units, 5 min, 37 °C) to fragment protein-bound RNA, and added to 400 mg guanidine-HCl with vortexing to quench RNase activity. The solution was then adjusted for nickel affinity purification by the addition of 27 µl NaCl (5.0 M) and 3 ul imidazole (2.5 M), and added to 50 ul nickel beads (Ni-NTA agarose, Promega). After an overnight incubation (4 °C), the nickel beads were transferred to a spin column (Snap Cap, Pierce) and washed three times with WBI (50 mM Tris-HCl pH 7.8, 300 mM NaCl, 0.1% NP-40, 10 mM imidazole, 5 mM β-mercaptoethanol, 6.0 M guanidine-HCl) then three times with 1xPNK (50 mM Tris-HCl, 10 mM $MgCl_2$, 0.5% NP-40, 5 mM β-mercaptoethanol). Several on-bead reactions (total volume 80 µl in each case) were then performed, washing once with WBI and three times with 1xPNK after each reaction:

i. Tobacco acid pyrophosphatase (Epicentre) treatment (optional) – 2 hr, 37 °C, in 1x TAP buffer (Epicentre).
ii. TSAP (Promega) phosphatase treatment – 30 min, 37 °C, in 1xPNK.
iii. Pre-adenylated 3' linker ligation using T4 RNA ligase – 6 hr, 25 °C, in 1xPNK.
iv. 5' end labelling with [γ^{32}P]-ATP using polynucleotide kinase (Sigma) – 1 hr, 37 °C, in 1xPNK, with addition of 100 nmol ATP after 40 min.
v. 5' linker ligation using T4 RNA ligase – 16 hr, 16 °C, in 1xPNK.

The beads were then washed three times with WBII (50 mM Tris-HCl pH 7.8, 50 mM NaCl, 0.1% (v/v) NP-40, 10 mM imidazole, 5 mM β-mercaptoethanol), then RNA:protein complexes eluted into EB (50 mM Tris-HCl pH 7.8, 50 mM NaCl, 0.1 % NP-40, 150 mM imidazole, 5 mM β-mercaptoethanol) and precipitated with TCA (20 % (v/v) final concentration). For nutrient shift experiments, nickel eluates were pooled. After washing with acetone, pellets were resuspended in NuPAGE 1x LDS sample loading buffer (Invitrogen) and protein:RNA complexes resolved by electrophoresis (4-12 % Bis-Tris NuPAGE gel, Invitrogen; 150 V). After electrophoretic transfer to a Hybond C nitrocellulose

membrane (GE) in 1x NuPAGE transfer buffer (150V; Invitrogen), labelled RNA was detected by autoradiography. The appropriate regions were then excised from the membrane, and treated with Proteinase K (Roche) in WBII containing 1% (w/v) SDS and 5 mM EDTA to elute RNA (55 °C, 2 hr). RNA was isolated by phenol:chloroform extraction followed by ethanol precipitation, resuspending in 11 µl water.

RNA samples were then used for reverse transcription with Superscript III (Invitrogen; 1 hr, 50 °C), using the miRCat-33 RT oligo (IDT). After heat inactivation (15 min, 65 °C), samples were treated with RNase H (NEB; 30 min, 37 °C). The cDNA was then used as a template for PCR amplification, in reactions containing LA Taq (Takara) and Solexa F and R primers (19-24 cycles, 52 °C annealing temperature). PCR products were precipitated using ethanol, resuspended in 1x gel loading dye (NEB) and resolved on a 3% Metaphor agarose (Lonza) gel. A region corresponding to ~120-300 bp was excised from each lane, and DNA extracted using a Qiagen gel purification kit, eluting in 20 µl water.

The libraries were checked by Sanger sequencing. Briefly, 2 µl of the purified PCR product was cloned into a pCR4 TOPO vector and transformed into TOP10 cells (Invitrogen) according to manufacturer's instructions. Colonies were picked, inoculated into LB medium with ampicillin, grown overnight at 30 °C, and plasmid DNA extracted using a Plasmid Mini kit (Qiagen). Sequencing reactions were performed using the Big Dye kit (Applied Biosystems) and the M13 F primer supplied with the pCR4 TOPO vector (Invitrogen).

For Solexa sequencing, libraries were sent to Genepool (University of Edinburgh) or Source Bioscience.

2.7 High-throughput sequencing of newly synthesised RNA

RNA was extracted by resuspending cell pellets in AE (50 mM NaOAc pH 5.3, 10 mM EDTA, 1 % (w/v) SDS) with an equal volume of phenol (pH 7.5) and vortexing (1400 rpm, 45 min, 65 °C). After centrifugation (20,000 xg, 20 min), RNA in the aqueous phase was

further purified by a phenol:chloroform (Ambion; pH 4.5) extraction then a chloroform extraction, and collected by ethanol precipitation. RNA was resuspended in TE2 (50 mM Tris-HCl pH 7.8, 1 mM EDTA) then thio-ketone groups (C=S) reduced to sulphydryl groups (C-SH) by treatment with TCEP (Tris(2-carboxyethyl)-phosphine) immobilised on agarose resin (Pierce). After two hours, the beads were removed by passing the solution through a spin column. The sulphydryl groups were then covalently modified by addition of 3 µl maleimide-PEG11-biotin (Pierce; 125 mM suspension in DMF; 2 hr), and RNA purified by chloroform extraction, collected by ethanol precipitation and resuspended in RBS100 (10 mM Tris-HCl pH 7.8, 100 mM NaCl, 2.5 mM $MgCl_2$, 0.4 % (v/v) Triton X-100). Newly synthesised (biotinylated) RNA was then captured on streptavidin magnetic particles (Roche) pre-blocked with glycogen (200 µg/ml). The particles were washed three times with RBS100 then twice with TEN1000 (10mM Tris-HCl pH 7.5, 1mM EDTA pH8, 1M NaCl), and the immobilised RNA subjected to a series of enzymatic on-bead reactions:

i. Antarctic phosphatase (NEB) treatment – 40 units, 2 hr, 37 °C, in 1xAP buffer (NEB).
ii. Tobacco acid pyrophosphatase (Epicentre) treatment – 2 hr, 37 °C, in 1x TAP buffer (Epicentre).
iii. 5’ linker ligation using T4 RNA ligase – 16 hr, 16 °C, in 1xPNK with 100 nmol ATP.
iv. RQ1 DNase (Promega) treatment – 30 min, 37 °C, in 1x RQ1 buffer (Promega).

RQ1 DNase was inactivated by heating (65 °C, 10 min) after the addition of 8 µl EGTA (20 mM, pH 8.0), then the beads washed with water. The RNA was reverse transcribed (5 min, 42 °C then 1 hr, 50 °C) using Superscript III (Invitrogen) and the P3N11HYMM_RT primer. After inactivating Superscript III (15 min, 65 °C), the sample was RNase treated (RNase H, 37 °C, 30 min). PCR amplification, gel purification, and sequencing were then performed as described for CRAC (Chapter 2.6).

In some cases, newly synthesised RNA was used for Northern analysis. Here, HPDP-biotin was used in place of maleimide-PEG_{11}-biotin, and RNA eluted from streptavidin particles in 100 mM β-mercaptoethanol (Swiatkowska et al, 2012).

2.8 Data analysis

Raw data was pre-processed with the fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html), using fastx_clipper to remove adapters, fastq_quality trimmer (-t 30) to trim low quality nucleotides, and fastq_artifacts filter (-q 23 -p 100) to remove reads without a consistently high quality. Identical reads were collapsed to remove PCR duplicates, then reads sorted by barcode using pySolexaBarcodeFilter (pyCRAC suite; Webb, Kudla, Tollervey and Granneman, in preparation; http://sandergranneman.bio.ed.ac.uk/Granneman_Lab/pyCRAC_software.html), allowing no mismatches. Reads were then mapped to the yeast genome (SGD version 64, http://www.yeastgenome.org/) using novoalign (www.novocraft.com/), and duplicate (same barcode and 5' end coordinate) or low complexity (<8 non-modal nucleotides) reads removed as indicated. Hits were then counted for all genomic features using pyReadCounters, or plotted across individual or multiple genes using pyPileup and gnuplot (http://www.gnuplot.info/). Genomic annotations were downloaded from Ensembl (release EF4.68), and supplemented with lncRNA annotations as described in the text. For pairwise correlation tests, the R package (http://www.r-project.org/) was used. To calculate individual or average hit densities across transcripts (divided into *n* bins), I wrote programmes in Python (http://www.python.org/), which are described in the text and available upon request. Clustering analyses were performed using Cluster 3.0 (de Hoon et al, 2004), and heat maps plotted in gnuplot. Analyses of intron and intron-exon junction abundance were performed as previously described (Schneider et al, 2012).

Analyses of motifs and reads with non-encoded As were performed on a subset of the data for which adapters were detected (and removed) in the initial pre-processing steps. This

retains only reads that are sequenced accurately throughout, and is a more stringent way to remove Solexa artefacts than a simple homopolymer filter. Motif searches were performed using pyMotif, which searches for motifs in a defined group of transcripts (e.g. protein coding) and calculates Z scores, which report on the enrichment of each motif when the sequence bias of each transcript within which it resides is taken into account (Wlotzka et al, 2011). Non-encoded A analyses were performed as previously described (Wlotzka et al, 2011), with non-encoded portions of reads assigned as an "A-tail" if they contained at least two non-encoded As, and fewer than 20 % non-A residues.

To assign transcript 3' ends, the terminal (3' end) coordinate ("3' tag") was extracted for the genome-encoded segment of each A-tailed read in the Pab1 dataset. For each gene, 3' tags were counted for each base, and the position with the greatest number of 3' tags recorded as the 3' end. Any tags mapping to the 5' half of the gene were discarded, and a 3' end was only recorded if there were at least 3 tags at that position. Subsequent analyses of the sequence composition and motif content of regions flanking 3' ends were performed using WebLogo (http://weblogo.berkeley.edu/), DREME (Bailey, 2011) and scripts written by myself in AWK (available upon request).

3: LncRNAs diverge from mRNAs before nuclear export

3.1 Introduction

Although many functions have been identified for lncRNAs, our understanding of their biogenesis, processing and decay is limited. As for mRNAs, the association of lncRNAs with proteins is likely to be important for these steps. However, whereas mRNAs primarily serve as information-carrying intermediates in protein synthesis, lncRNAs perform distinct functions, such as playing structural or targeting roles. Furthermore, in yeast many lncRNAs are apparently byproducts of regulatory circuits dependent on the act of transcription (rather than the transcript). Investigating the interactions of lncRNAs with proteins is important to understand when and how they diverge from mRNAs, whether they function directly or via the act of transcription, and how they perform their various functions. However, whereas the factors involved in mRNA metabolism are extensively characterised (Kelly et al, 2009; Tutucci et al, 2011), few have been tested for any role in lncRNA metabolism. Previous studies have concentrated on lncRNA transcription and degradation, but not addressed other key steps such as processing or subcellular transport. In this chapter, I therefore sought to identify lncRNA interactions with RNA binding proteins from all stages of mRNA metabolism, to determine when in their synthesis lncRNAs and mRNAs diverge, and how they are distinguished.

To identify lncRNA:protein interactions en masse, one can either start with a specific lncRNA or a specific protein as "bait". I opted for the second strategy, in which all lncRNAs interacting with a particular protein are determined, because this permits the identification of interactions with annotated, unannotated or misannotated lncRNAs. This is important, because the annotation of lncRNAs is currently very limited. The published lists include lncRNAs detectable in wild-type yeast (847 SUTs (Xu et al, 2009), 402 antisense lncRNAs (Yassour et al, 2010), and 523 antisense/135 intergenic lncRNAs (Granovskaia et al, 2010)),

925 CUTs (detectable in *rrp6Δ* yeast) (Xu et al, 2009) and XUTs (detectable in *xrn1Δ* yeast) (van Dijk et al, 2011). Each study identified many novel lncRNAs, suggesting that they are far from saturating the noncoding transcriptome. Furthermore, reported 5' and 3' coordinates often differ between studies, so must be viewed as approximate.

To identify protein:RNA interactions, I used the CRAC method (crosslinking and analysis of cDNAs) (Granneman et al, 2009). Here, protein:RNA interactions are fixed in actively growing yeast by UV crosslinking, then a tagged protein of interest is isolated via a two-step denaturing purification, and associated RNAs sequenced. UV crosslinking specifically captures protein-nucleic acid, and not protein-protein or nucleic acid-nucleic acid, interactions (Greenberg, 1979; Hockensmith et al, 1986). The use of a high affinity hexahistidine-TEV-Protein A (HTP) tag permits the purification to be performed under stringent conditions (6 M guanidine-HCl), and tandem affinity purification allows low abundance proteins to be effectively captured. Furthermore, a single set of conditions can be used for CRAC analyses of numerous different proteins, enabling their RNA binding profiles to be compared without bias resulting from variations in purification conditions. The protein-bound RNAs are partially digested by limited RNase treatment during the CRAC procedure so crosslinking sites can be mapped with high precision, revealing not only which transcripts the protein binds, but also the location of the binding sites.

Selection of candidate proteins

Messenger RNAs interact with hundreds of proteins (Castello et al, 2012; Scherrer et al, 2010; Tsvetanova et al, 2010), of which I selected 36 for analysis. I aimed to include key factors from all stages of mRNA metabolism in which lncRNAs might also participate. The available literature suggests that some lncRNAs diverge from mRNAs in the nucleus, by using the alternative Nrd1-dependent termination pathway (Arigo et al, 2006b; Houseley et al, 2007; Kim et al, 2010a; Marquardt et al, 2011; Thiebaut et al, 2006; Vasiljeva et al, 2008b; Wlotzka et al, 2011; Wyers et al, 2005). This mode of termination is often coupled to

nuclear turnover (Arigo et al, 2006b; Honorine et al, 2011; Thiebaut et al, 2006; Vasiljeva et al, 2006), and indeed, many nuclear surveillance factors have been implicated in lncRNA degradation. These include the exosome nucleases Rrp6 and Rrp44 (Lardenois et al, 2011; Neil et al, 2009; Preker et al, 2008; Schneider et al, 2012; Xu et al, 2009), the nuclear exosome cofactors TRAMP (Davis et al, 2006; Neil et al, 2009; Paolo et al, 2009; Schneider et al, 2012; Wlotzka et al, 2011; Wyers et al, 2005), Rrp47 and Mpp6 (Milligan et al, 2008), and the 5' to 3' exonuclease Rat1 (Geisler et al, 2012; Luke et al, 2008). I therefore included nuclear surveillance factors, and "early" mRNA binding proteins (TREX components and the CBC) in my analysis. However, some studies have detected lncRNA accumulation in the absence of cytoplasmic decay factors, primarily Xrn1 (Berretta et al, 2008; Marquardt et al, 2011; Thompson et al, 2007; Toesca et al, 2011; van Dijk et al, 2011), but also decapping activators (Thompson et al, 2007) and nonsense-mediated decay factors (Marquardt et al, 2011; Toesca et al, 2011). This suggests that some lncRNAs resemble mRNAs in terms of being exported to, and degraded in, the cytoplasm, and I therefore included cytoplasmic proteins in my analysis.

Notably, each previous study of lncRNA turnover only tested a limited set of factors, with different studies employing different methods with different biases. Together with the inherent redundancy between decay pathways, this has confounded efforts to estimate the relative contributions of the various decay factors to lncRNA degradation. The ability of CRAC to detect even short-lived protein:RNA interactions in actively growing cells without disruption of cellular decay pathways, and to be applied to many different proteins, partially overcomes these limitations.

The 36 mRNA binding proteins selected for analysis are listed in Table 3.1. For many, the full set of mRNA targets is not known, so this study will offer insight into the canonical roles of these proteins in mRNA metabolism as well as their roles in lncRNA biology. Factors for which high-throughput data are already available were helpful for validating the sensitivity

Process	Protein	Datasets obtained
Nuclear surveillance	Trf4	1
	Mtr4	14
	Rnt1	1
	Swt1	Low crosslinking
Pre-mRNA cap binding	Sto1	11
3' end processing	Hrp1	2
	Nab2	4*
	Rna15	1
	Ref2	1
	Fip1	Low crosslinking
	Mpe1	Low crosslinking
	Ref2	Low crosslinking
	Pap1	Low crosslinking
	Ysh1	Low crosslinking
	Cft1	Low crosslinking
	Pti1	3*
	Pab1	1
	Rna14	1
	Cft2	1
	Pcf11	2*
TREX complex and export	Mex67	2*
	Sub2	Low crosslinking
	Yra1	Low crosslinking
	Yra2	Low crosslinking
	Tho2	1
	Tho1	1
	Crm1	1
	Gbp2	1
	Dbp5	1
Cytoplasmic turnover	Xrn1	2*
	Ski2	2
	Dcp2	Low crosslinking
	Lsm1	Low crosslinking
Translation	Tif1	1
RNA localisation	Hek2	1
Splicing/RNA processing	Lsm8	Low crosslinking

Key
Good quality high-throughput dataset obtained
High-throughput data obtained, but insufficient quality
Low crosslinking, so no high-throughput data obained

* = repeats include independent clones

Table 3.1: Proteins selected for analysis. The number of high-throughput sequencing datasets obtained for each protein is indicated. For some proteins (*) replicates were obtained from two distinct, but identically HTP-tagged, strains. For Mtr4 and Sto1, replicates were obtained under several different growth conditions. Many proteins crosslinked well and good quality sequencing datasets were obtained (green). However, some proteins crosslinked weakly and gave insufficient reads (orange), and others crosslinked very poorly or not at all, so libraries were not sent for sequencing (red).

and specificity of my analyses, and include Hrp1, Nab2 (Batisse et al, 2009; Kim Guisbert et al, 2005), Mex67 (Hieronymus et al, 2003) and Hek2 (Hasegawa et al, 2008). I also included analyses of data from existing CRAC studies: Rat1 (unpublished data, Sander Granneman), Rpo21 (unpublished data, Laura Milligan), Nrd1 and Nab3 (Wlotzka et al, 2011), and Rrp44 and Rrp6 (Schneider et al, 2012).

3.2 Crosslinking and analysis of cDNAs

For each of the proteins selected for analysis, I generated a yeast strain with a HTP cassette (encoding the HTP tag and a selectable *URA3* marker) integrated immediately upstream of the stop codon of the endogenous gene, by homologous recombination. *NAB2* was also tagged in an *rrp6Δ* background, as a recent study suggested that Rrp6 deletion stabilises transient Nab2:RNA interactions (Schmid et al, 2012). I tested the strains (i) for accurate integration of the cassette, by PCR with one gene-specific and one cassette-specific primer, and (ii) for expression of a tagged protein of the expected size, by Western blotting using peroxidase anti-peroxidase, an antibody reactive against protein A (data not shown).

For each protein, I then tested the efficiency of the two-step IgG/nickel purification in the CRAC protocol (Granneman et al, 2009). Yeast were grown to logarithmic phase (A_{600} = 0.5) in SGlu -W (synthetic glucose media lacking tryptophan), then 2.75 litres of culture subjected to UV crosslinking (205 Watt UV-C lamp (van Remmen LL121280); 100 s). I then used crude lysate from one third of the collected yeast for the two-step purification. This comprised an initial IgG-sepharose affinity purification step (binding in buffer containing 150 mM NaCl, and washing in buffer containing 1.0 M NaCl), elution with TEV protease, mild RNase A/T1 treatment to fragment RNA, then a Ni-agarose affinity purification step (binding and washing in buffers containing 6 M guanidine-HCl). I analysed fractions from various stages by SDS-PAGE followed by Western blotting (data not shown). A comparison of the abundance of protein in the crude lysate versus that in the Ni eluate revealed that the yield was typically ~10-50 %. These Western blots also revealed the precise positions at

which the singly tagged proteins migrated following removal of the protein A moiety by TEV protease.

I next assessed the purity of proteins in eight of the Ni eluates by SDS-PAGE separation, followed by silver staining (Figure 3.1A). I observed a specific band in each lane, at positions consistent with the Western blotting analyses (Figure 3.1B/C, red lines). The level of background was low, indicating that the two-step purification is relatively clean. The weak, higher molecular weight bands observed in some lanes (Ski2, Pab1, Tho2) are likely to arise from incomplete denaturation of some proteins in this gel system, since Ski2 and Pab1 gave signals in this region when tested by Western blot (data not shown).

Having verified that the two-step purification was specific and efficient, I next tested whether crosslinking to RNA could be detected for each of the 36 proteins selected for testing. I therefore repeated the two-step purification with the addition of an RNA radiolabelling step, which was performed on the Ni affinity column following replacement of the 6M guanidine-HCl buffer with kinase buffer. Protein:RNA complexes were eluted, resolved by SDS-PAGE, transferred to a nylon membrane, and radiolabelled RNA was detected by autoradiography. A wild-type (untagged; "BY") strain was included as a negative control. Of the 36 proteins tested, 23 gave detectable crosslinking (Table 3.1), and autoradiograms for 18 of these are shown in Figures 3.1B and 3.1C.

Library construction from crosslinked RNAs

I next performed the full CRAC method on each of the 23 strains that gave detectable crosslinking. In addition to the two-step purification and radiolabelling described, this includes (i) a series of on-bead reactions that ligate single-stranded 5' and 3' adapters to the protein-bound RNA fragments, (ii) a third purification step comprising SDS-PAGE, electrophoretic transfer to a nylon membrane, and excision of the protein:RNA band, (iii) Proteinase K digestion to release RNA fragments from the membrane slice, and (iv) reverse transcription and PCR amplification.

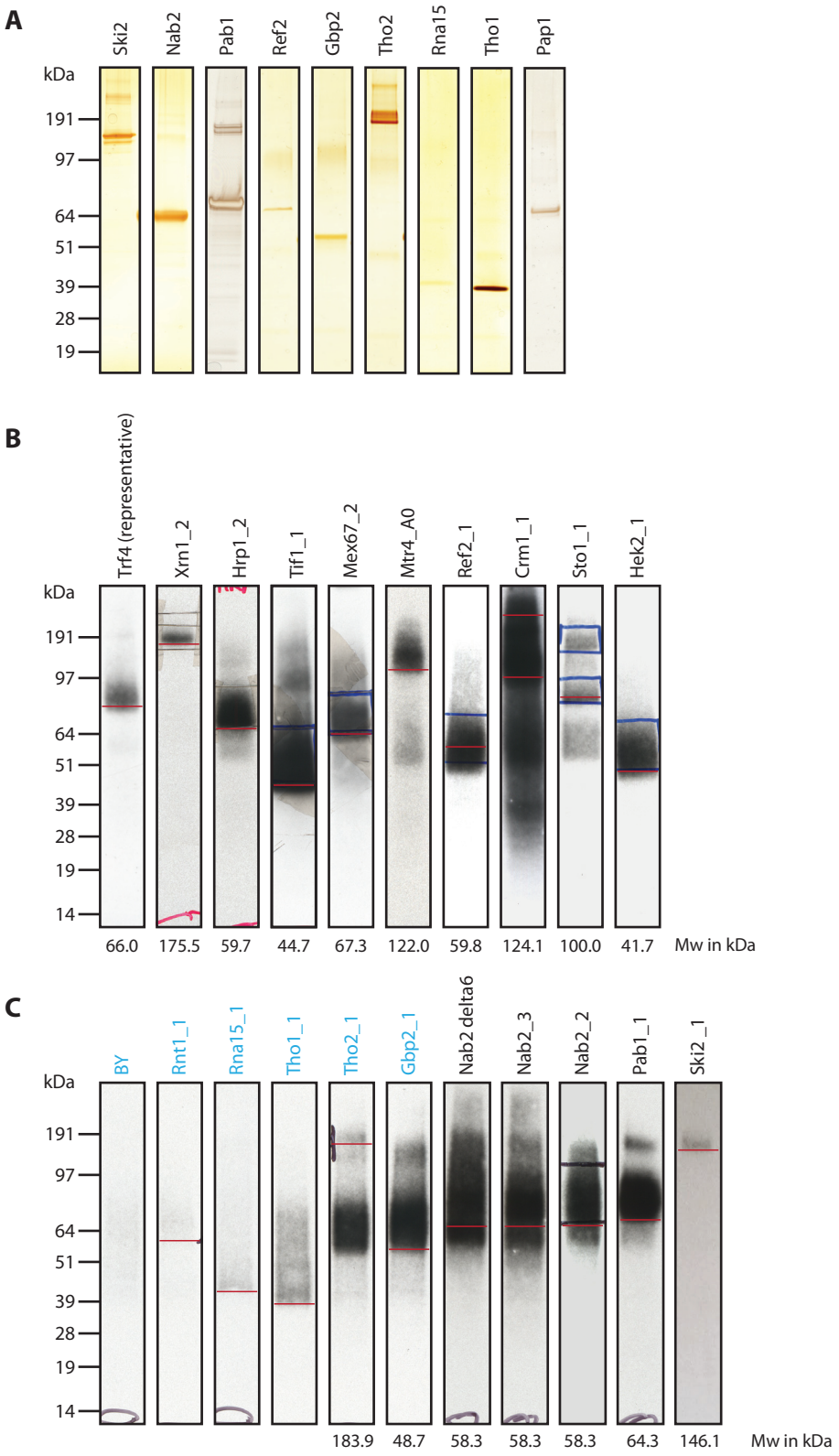


Figure 3.1: Purification and crosslinking efficiencies. A Silver stained SDS-PAGE analysis of eluates from two-step IgG/nickel purifications of HTP-tagged proteins. **B** and **C** Autoradiograms of 5' end-labelled RNAs crosslinked to the indicated HTP-tagged proteins. For each protein, the eluate from a two-step IgG/nickel purification was resolved by SDS-PAGE and transferred to a nylon membrane. Red lines indicate the migration of the protein determined by Western blot analysis using an anti-TAP antibody.

Several steps in this procedure require optimisation for each protein. Firstly, the mild RNase A/T1 digestion performed during the two-step purification must be adjusted so that protein-bound RNA is fragmented to ~50 nt. This optimal length ensures that cDNA sequences can be accurately mapped to the yeast genome, but avoids interference with protein migration during SDS-PAGE that would potentially arise from large residual RNA fragments. The optimal conditions for RNase A/T1 digestion must be determined empirically, but the fragment lengths are only revealed upon analysis of the dsDNA library at the end of the full CRAC procedure. I therefore performed CRAC on Trf4, testing several RNase A/T1 concentrations in parallel. This revealed a negative correlation between RNase concentration and fragment length, with a ~1.7-fold increase in fragment length upon a 5-fold decrease in RNase concentration, and provided an optimal set of digestion conditions (0.1 units RNace-It (Agilent), 5 minutes, 37 °C). For each protein, I therefore performed CRAC using these initial conditions, and if the fragments obtained were too long or too short, I repeated the experiment with adjusted conditions. The majority of proteins did not require any adjustment.

A second step requiring optimisation for each protein is the final PCR amplification of cDNA generated from the isolated RNA fragments. Analyses of CRAC high-throughput sequencing datasets reveal that different RNA fragments are amplified with different efficiencies, but minimising the number of PCR cycles reduces this bias. I therefore performed the minimum number of PCR cycles required to obtain ~20 ng of DNA for each library. For many proteins, I also used a modified 5' linker that contains an internal random 3-mer. This enables PCR duplicates to be filtered out computationally.

I obtained dsDNA libraries for each of the 23 proteins that gave detectable crosslinking, as well as four untagged controls (BY1-4), and these were submitted for Illumina/Solexa sequencing. Some proteins yielded only low amounts of PCR product, but were still sent for sequencing because barcoded 5' linkers enabled them to be mixed with other samples. For

nine proteins, I obtained additional datasets (indicated in Table 3.1). These include (i) technical and/or biological replicates, (ii) analysis of Nab2-bound transcripts in an *rrp6Δ* background, (iii) analysis of Sto1 and Mtr4 targets in yeast subjected to a media shift (Chapter 6), and (iv) a Sto1 replicate where transcripts were enzymatically decapped to facilitate detection of 5'-proximal binding sites ("Sto1_WLUtap"). Altogether, 60 libraries were sent for Solexa sequencing (Table 3.2). The naming convention used for datasets in this study is outlined in Table 3.2.

Data processing

I passed each Solexa dataset through a pre-processing pipeline, which (i) removes 3' linker sequences, (ii) trims low quality bases from the 3' end of reads, (iii) rejects reads with a low overall quality score, (iv) removes homopolymeric reads (common Solexa artefacts), (v) removes PCR duplicates (identical reads), and (vi) splits the dataset by barcode. For some downstream analyses, only reads for which the 3' adapter was identified (and removed) were retained, as these constitute a set of high confidence reads that are correctly sequenced throughout, and free from sequencing artefacts. After pre-processing, I mapped reads to the yeast genome (SGD release 64) using Novoalign, then performed a second round of filtering to remove PCR duplicates. Here, reads with identical random 3-mers in their barcodes and identical 5' end genomic coordinates were collapsed. This removes duplicates that were not detected during pre-processing due to having different quality scores or containing sequencing errors. These steps improve the accuracy of downstream quantitative analyses, and resulted in a 1- to 10-fold reduction in the number of reads. As I expected many transcripts to be adenylated by Pap1 or the TRAMP complex, I used the -s option in Novoalign to permit read trimming from the 3' end and enable adenylated transcripts to be mapped. The numbers of reads remaining after the major filtering and mapping steps are reported in Table 3.2.

	All preprocessed reads			Preprocessed reads with 3' adapter identified		
	Total	Mapped by novoalign		Total	Mapped by blast	
		Total	Duplicates removed		Total	With A-tails
BY_1	132,540	104,828	NA	104,332	87,809	13,344
BY_2	3,279	1,945	NA	2,786	1,722	86
BY_3	177,394	105,339	40,240	147,064	90,799	15,229
BY_4	64,648	49,122	14,706	40,141	29,224	3,840
Cft2_1	121,097	79,017	25,841	78,681	NA	NA
Crm1_1	214,729	163,393	51,635	104,626	NA	NA
Dbp5_1	112,137	71,965	18,143	81,486	NA	NA
Gbp2_1	3,384,410	1,747,127	1,037,113	3,256,107	1,895,962	405,093
Hek2_1	1,203,569	684,274	375,008	1,093,246	660,295	35,210
Hrp1_1	681,331	573,431	NA	588,380	520,449	120,560
Hrp1_2	819,092	577,177	NA	688,510	534,497	153,320
Mex67_1	280,324	231,026	NA	214,190	181,000	13,246
Mex67_2	311,754	247,326	NA	246,480	199,234	14,961
Mtr4_A0	237,030	155,667	38,223	191,856	NA	NA
Mtr4_A16	623,988	365,040	89,647	477,391	NA	NA
Mtr4_A4	545,672	336,275	85,678	418,723	NA	NA
Mtr4_A8	453,619	267,410	69,726	347,657	NA	NA
Mtr4_Amock	553,480	391,895	81,497	444,147	NA	NA
Mtr4_B0	1,005,308	730,591	220,317	768,700	559,544	144,280
Mtr4_B16	550,421	385,962	122,904	409,182	NA	NA
Mtr4_B4	509,397	365,559	128,576	342,353	NA	NA
Mtr4_B8	925,468	614,548	210,963	655,199	NA	NA
Mtr4_Bmock	2,035,764	1,514,999	418,823	1,558,178	1,189,815	331,962
Mtr4_C0	3,783,833	2,591,016	1,044,252	2,945,079	2,085,388	630,954
Mtr4_C16	3,337,634	2,080,960	973,892	2,505,613	NA	NA
Mtr4_C4	4,350,181	2,758,992	1,263,618	3,143,365	NA	NA
Mtr4_C8	987,044	616,341	203,011	703,609	NA	NA
Nab2_1	449,579	372,485	293,766	341,017	289,651	61,440
Nab2_2	6,479,021	4,823,691	2,460,194	4,230,604	3,015,679	596,038
Nab2_3	1,452,862	1,058,178	500,759	1,039,352	766,544	137,137
Nab2_6delta	3,413,860	2,307,904	1,354,443	2,402,276	1,597,320	210,481
Pab1_1	16,702,682	12,855,428	4,242,176	11,112,813	1,884,653	496,561
Pcf11_1	208,068	126,863	34,993	166,334	NA	NA
Pcf11_2	902,976	673,309	151,304	569,097	NA	NA
Pti1_1	152,836	64,509	20,144	130,912	NA	NA
Pti1_2	98,733	71,540	15,173	68,722	NA	NA
Pti1_3	114,554	77,856	19,332	80,013	NA	NA
Ref2_1	273,175	187,995	62,704	213,236	NA	NA
Rna14_1	589,367	423,689	103,653	380,989	NA	NA
Rna15_1	84,938	38,339	9,073	78,816	NA	NA
Rnt1_1	404,222	227,655	64,098	348,818	NA	NA
Ski2_1	1,016,310	819,307	233,012	655,741	524,448	55,600
Ski2_2	1,351,395	1,044,648	702,299	1,010,362	801,907	89,406
Sto1_1	1,436,824	646,108	326,492	1,240,659	580,101	112,902
Sto1_W0	1,056,734	618,448	505,566	907,453	565,123	92,579
Sto1_W16	527,908	316,369	261,903	446,838	NA	NA
Sto1_W4	285,078	176,388	148,655	240,170	NA	NA
Sto1_W8	547,602	322,176	263,793	457,433	NA	NA
Sto1_WLU0	364,865	286,451	202,867	247,787	NA	NA
Sto1_WLU16	643,684	519,964	337,222	421,869	NA	NA
Sto1_WLU4	593,067	487,670	319,171	352,872	NA	NA
Sto1_WLU8	580,092	477,339	307,776	319,524	NA	NA
Sto1_WLUtap	1,288,888	882,165	713,529	NA	NA	NA
Tho1_1	282,649	148,526	32,380	237,432	115,048	1,574
Tho2_1	744,315	543,570	158,320	458,954	293,286	37,829
Tif1_1	201,342	116,724	NA	183,559	134,160	16,757
Trf4_1	313,153	247,673	NA	263,600	224,583	81,692
Xrn1_1	518,797	406,833	NA	369,764	301,196	46,341
Xrn1_2	343,514	249,924	NA	279,853	216,963	25,812

Table 3.2: Total and processed read counts for Solexa datasets obtained in this study. As standard, datasets were obtained from yeast grown in glucose media lacking tryptophan, with repeats indicated by "_1", "_2" etc. However, Sto1 datasets were obtained over a time course (0, 4, 8, 16 minutes) during a glucose to glycerol/ethanol shift, in media lacking tryptophan ("Sto1_Wx") or tryptophan, leucine and uracil ("Sto1_WLUx"), where x indicates the time point. Three similar time courses were performed for Mtr4 ("Mtr4_Ax", "Mtr4_Bx" and "Mtr4_Cx"), in media lacking tryptophan, leucine and uracil. Finally, the "Sto1_WLUtap" CRAC experiment included an enzymatic decapping step.

The datasets generated from proteins that crosslinked very weakly contained a large proportion of reads commonly detected as background in negative controls (BY strain), and were thus excluded from subsequent analyses (indicated in Table 3.1). This left 43 datasets, corresponding to 14 proteins (39 % of the set selected for analysis). The lowest success rate was obtained with canonical mRNA 3' end processing factors, of which only 20 % yielded usable datasets, suggesting that this complex may be generally refractory to analysis by CRAC. This was apparently due to low crosslinking efficiency rather than poor protein expression or purification, as some of these 3' end processing factors were abundantly detected in Ni eluates (e.g. Pap1, Figure 3.1A).

3.3 Evaluation of CRAC datasets

I next performed several analyses to evaluate the quality and reliability of the datasets, using the pyCRAC software suite (Webb, Kudla, Tollervey and Granneman, in preparation) and my own AWK and Python scripts.

Comparison of transcriptome-wide binding profiles

For each protein, the number of reads overlapping with each genomic feature (hits) was counted to generate a hit table. I downloaded genomic annotations from Ensembl (release EF4.68), and added published UTRs (Xu et al, 2009) and antisense and intergenic lncRNAs, CUTs, SUTs, and XUTs (Granovskaia et al, 2010; van Dijk et al, 2011; Xu et al, 2009; Yassour et al, 2010). To ensure that hits in mRNAs, CUTs or SUTs were not missed due to inaccuracies in annotated transcript boundaries, 5' flanks of up to 50 nt and 3' flanks of up to 300 nt were added to these features when counting hits. Importantly, flanks were not extended into neighbouring features, so were often shorter than these maximum values. Furthermore, I classified regions not annotated as canonical genes, or encoding CUTs or SUTs, as "antisense" or "intergenic" features (with or without an annotated feature on the opposite strand, respectively).

For each protein, I summed the hits overlapping with each major transcript class, and expressed these values as a percentage of total hits (Figure 3.2). The transcript profiles are remarkably similar for replicate experiments (Mex67, Hrp1, Nab2, Mtr4, Ski2, Xrn1 and Sto1 repeats), demonstrating that experiments are reproducible. Furthermore, there are marked differences between the profiles for different proteins, and individual profiles are consistent with the available literature. For example, Pab1 binds predominantly to mRNAs, whereas Mtr4 binds to many tRNAs, consistent with recent reports of high tRNA turnover in the nucleus (Gudipati, 2012). There are also striking differences in binding to lncRNAs; Gbp2, Hrp1, Mtr4, Nab2, Pab1, Sto1 and Trf4 bind numerous lncRNAs, whereas Mex67, Hek2, Crm1, Ski2 and Tif1 bind very few.

To further examine the specificity and reproducibility of datasets, I evaluated the similarity between hit tables for pairwise combinations of datasets, by calculating Spearman rank correlation coefficients (displayed as a matrix in Figure 3.3). I limited the analysis to CUTs, SUTs and mRNAs (Figure 3.3A). This reveals that the binding profiles (hits for each gene, normalised to total Pol II hits) are most highly correlated between replica datasets for Mtr4 ($\rho = 0.77$), Sto1 ($\rho = 0.89$), Nab2 ($\rho = 0.91$), Ski2 ($\rho = 0.90$) and Xrn1 ($\rho = 0.88$), but replicates for Mex67 ($\rho = 0.63$) and Hrp1 ($\rho = 0.53$) are less similar. However, published (and validated) datasets for Rrp44 have a correlation coefficient of 0.53, suggesting that this is an acceptable value for replicates. Furthermore, proteins involved in distinct biological processes, such as Crm1 and Mex67 ($\rho = 0.37$) (distinct export pathways), or the decay factors Mtr4 and Xrn1 ($\rho \leq 0.21$) (nuclear versus cytoplasmic turnover), have lower correlation coefficients. Conversely, proteins participating in closely related biological functions have high correlation coefficients, such as the cytoplasmic decay factors Xrn1 and Ski2 ($\rho = 0.87$), or the Pol II subunit Rpo21 and associated RNA-packaging factors, TREX complex members Gbp2 and Tho2 ($\rho \geq 0.73$).

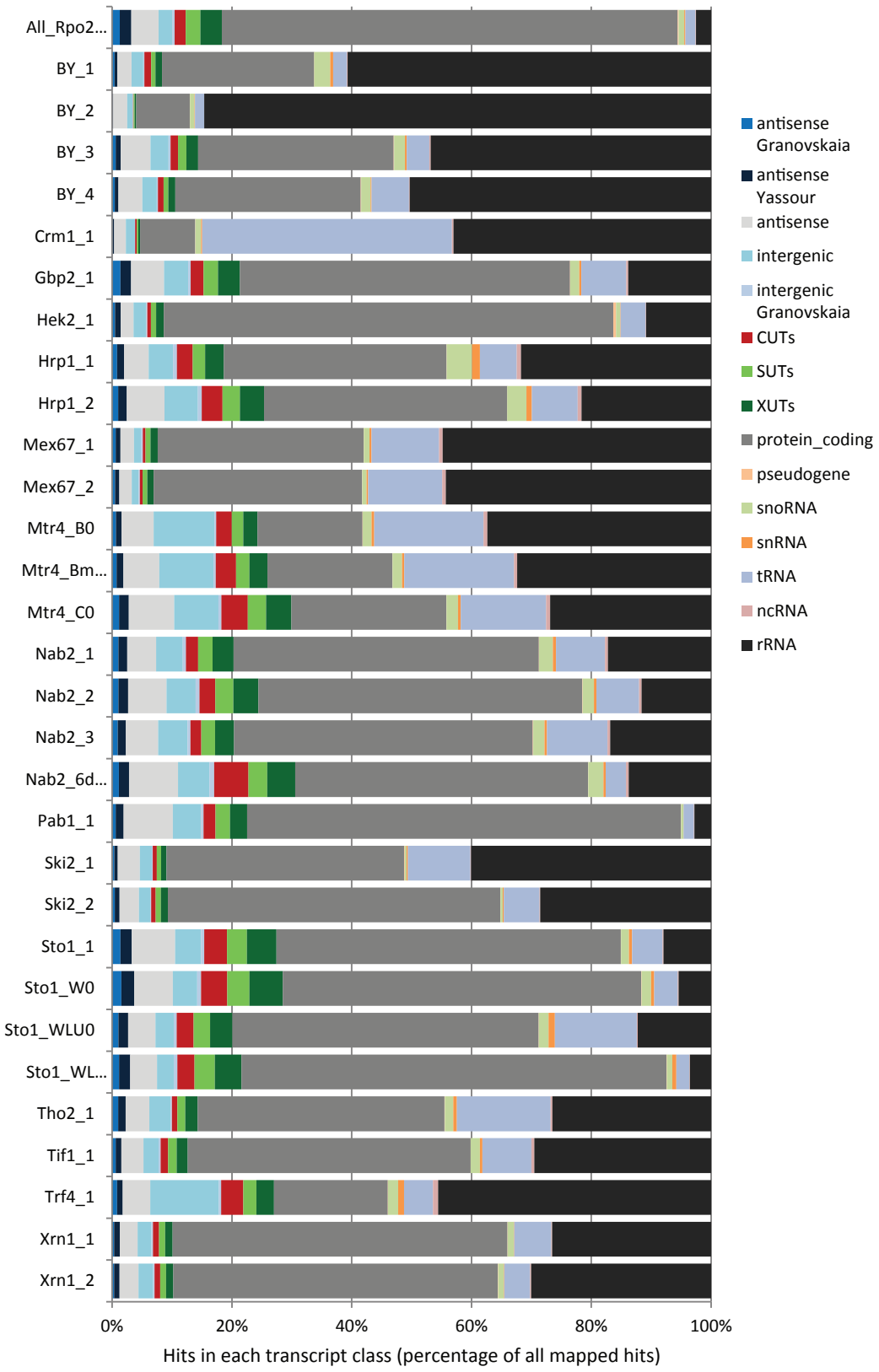


Figure 3.2: Transcriptome-wide hit distributions. For each protein, hits for each transcript class are expressed as a percentage of total hits. This representation provides the clearest overview of the datasets, but as hits are counted more than once if they map to a region with overlapping annotations, later analyses (normalised to total Pol II hits) are used for quantitative comparisons between specific transcript classes.

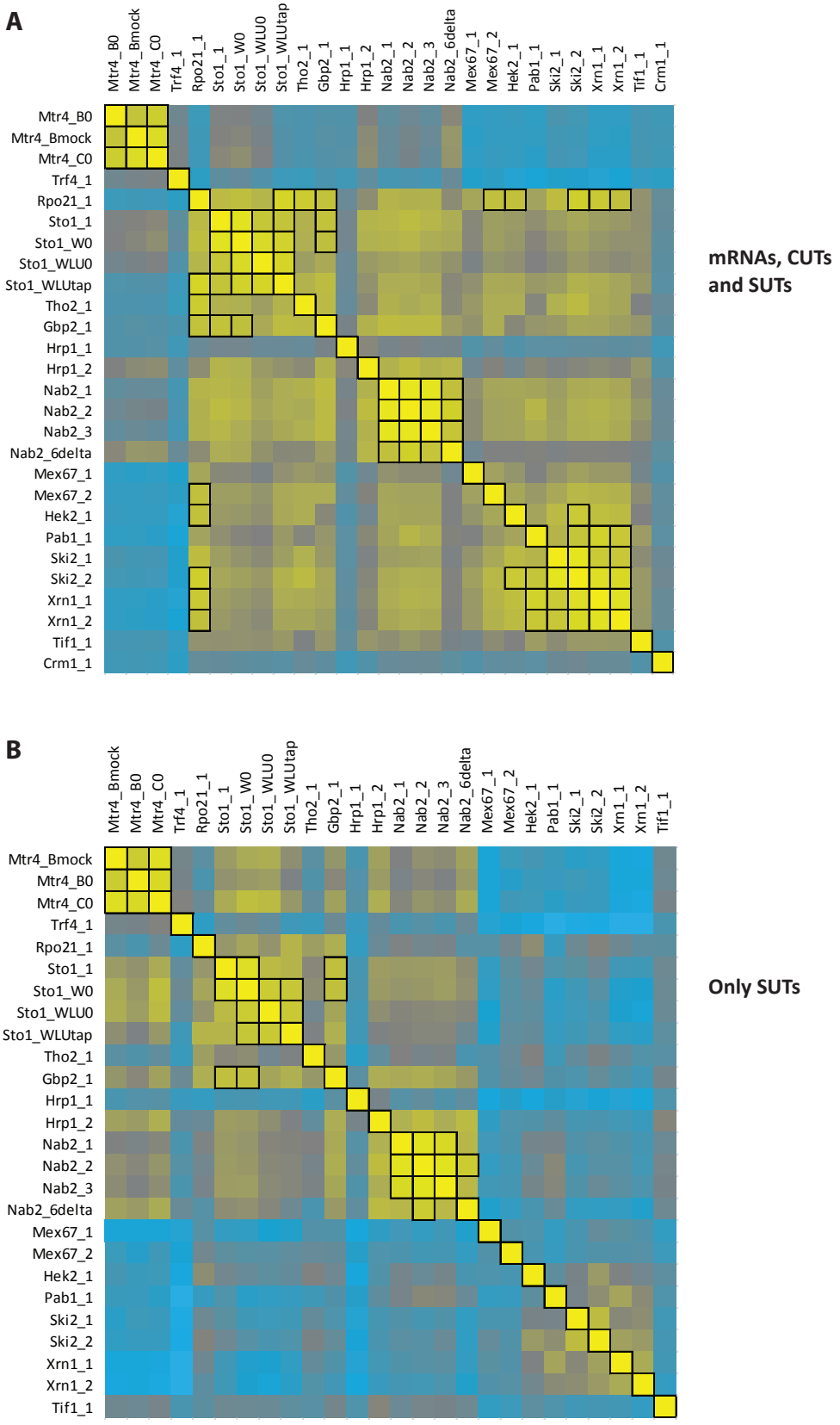


Figure 3.3: Spearman rank correlation scores. A For pairs of proteins, the Spearman rank correlation coefficient was calculated based on the number of hits in SUTs, CUTs and mRNAs. Yellow, $\rho = 1$; blue, $\rho = 0$; boxed, $\rho > 0.75$. **B** The analysis was repeated, considering only SUTs.

I repeated this analysis, but considered only hits in SUTs when calculating correlation coefficients (Figure 3.3B). This reveals that SUT binding profiles are well correlated for Mtr4, Sto1 and Nab2 replicates, suggesting that these proteins bind reproducibly to SUTs. Notably, Mtr4 is more highly correlated with early nuclear mRNA biogenesis factors (Sto1, Gbp2, Hrp1 and Nab2) in this analysis (compared to Figure 3.3A), suggesting that early steps in SUT biogenesis (Figure 3.3B) are commonly coupled to nuclear turnover, whereas bulk Pol II transcript biogenesis (Figure 3.3A) is linked predominantly to cytoplasmic turnover.

For some proteins, RNA targets have been identified in previous high-throughput studies. To assess the correlation between my datasets and published datasets for Hek2 (Hasegawa et al, 2008), Hrp1 (Kim Guisbert et al, 2005) and Nab2 (Batisse et al, 2009), I calculated the degree of overlap between the top 10 % of mRNA targets detected in both analyses. For Hek2, 1168 mRNA targets were detected in both my dataset and that published (Hasegawa et al, 2008), and of the top 10 % (117) in each dataset, 42 % overlapped ($\chi^2 = 1.0 \times 10^{-33}$; n = 1168). I repeated this analysis for Nab2 (dataset Nab2_2; 25 % overlap; $\chi^2 = 2.2 \times 10^{-35}$; n = 2595) (Batisse et al, 2009) and Hrp1 (dataset Hrp1_2; 15 % overlap; $\chi^2 = 0.001$; n = 3079) (Kim Guisbert et al, 2005). I also compared replicate pairs within the current study, including Nab2_1 vs Nab2_2 (85 % overlap; $\chi^2 < 10^{-300}$; n = 6605) and Mex67_1 vs Mex67_2 (57 % overlap; $\chi^2 < 10^{-300}$; n = 5516). Together, therefore, the data obtained in this study are in good agreement with previously published datasets, but there is better agreement between replicates in this study. This is not surprising, because culture conditions and strain backgrounds vary between studies.

Ribosomal RNA hits

Although a detailed analysis of the possible roles of the tested factors in ribosome biogenesis is beyond the scope of this study, an examination of hits mapping across the rDNA locus,

which encodes 25S, 5.8S and 18S rRNA, provides further evidence of the specificity and reproducibility of the datasets. The high abundance of rRNA hits ensures a detailed, high resolution binding profile (Figure 3.4). This reveals specific peaks for Mex67 and Crm1 (boxed regions), consistent with their documented roles in rRNA export. Mex67 is reported to bind both 60S (Yao et al, 2007) and 40S (Faza et al, 2012) preribosomal particles, which is in good agreement with the peaks observed by CRAC (Figure 3.4, "Mex67_1" and "Mex67_2"). The peak of Xrn1 in ITS1 is consistent its role in cytoplasmic degradation of the excised spacer region. However, the function of the apparently specific association of the Ski2 helicase remains unclear.

Additionally, the nuclear surveillance factors Mtr4 and Trf4 bind abundantly to the 5' ETS consistent with their documented role in degradation of this excised spacer, while Mtr4 also bound ITS2, consistent with its role in 3' processing of the 5.8S rRNA. Hrp1, Tho2, Gbp2, Nab2 and Sto1 also give specific binding patterns across the rDNA transcripts, which tended to be more distributed, but still with distinct peaks. Rpo21 exhibited binding across the rDNA locus, suggesting that low level Pol II transcription might explain the rRNA binding patterns observed for some mRNA biogenesis factors. Two peaks commonly detected as background are indicated with asterisks (Figure 3.4).

For various proteins, I also analysed the distribution of hits across specific (non-ribosomal) RNAs (e.g. Figures 3.8-3.10), and the average binding profile across the length of SUTs, CUTs and mRNAs (Chapters 4-6). These analyses, like those for the rDNA locus, reveal specific and reproducible binding patterns.

Sequence motifs

Hrp1 and Hek2 are reported to bind specific sequence motifs (UA and CNN repeats, respectively). To validate my Hrp1 and Hek2 datasets, and test whether other proteins in this study exhibit sequence-specific binding, I searched for sequence motifs in a number of datasets. I used the pyMotif module of the pyCRAC suite, which calculates statistical

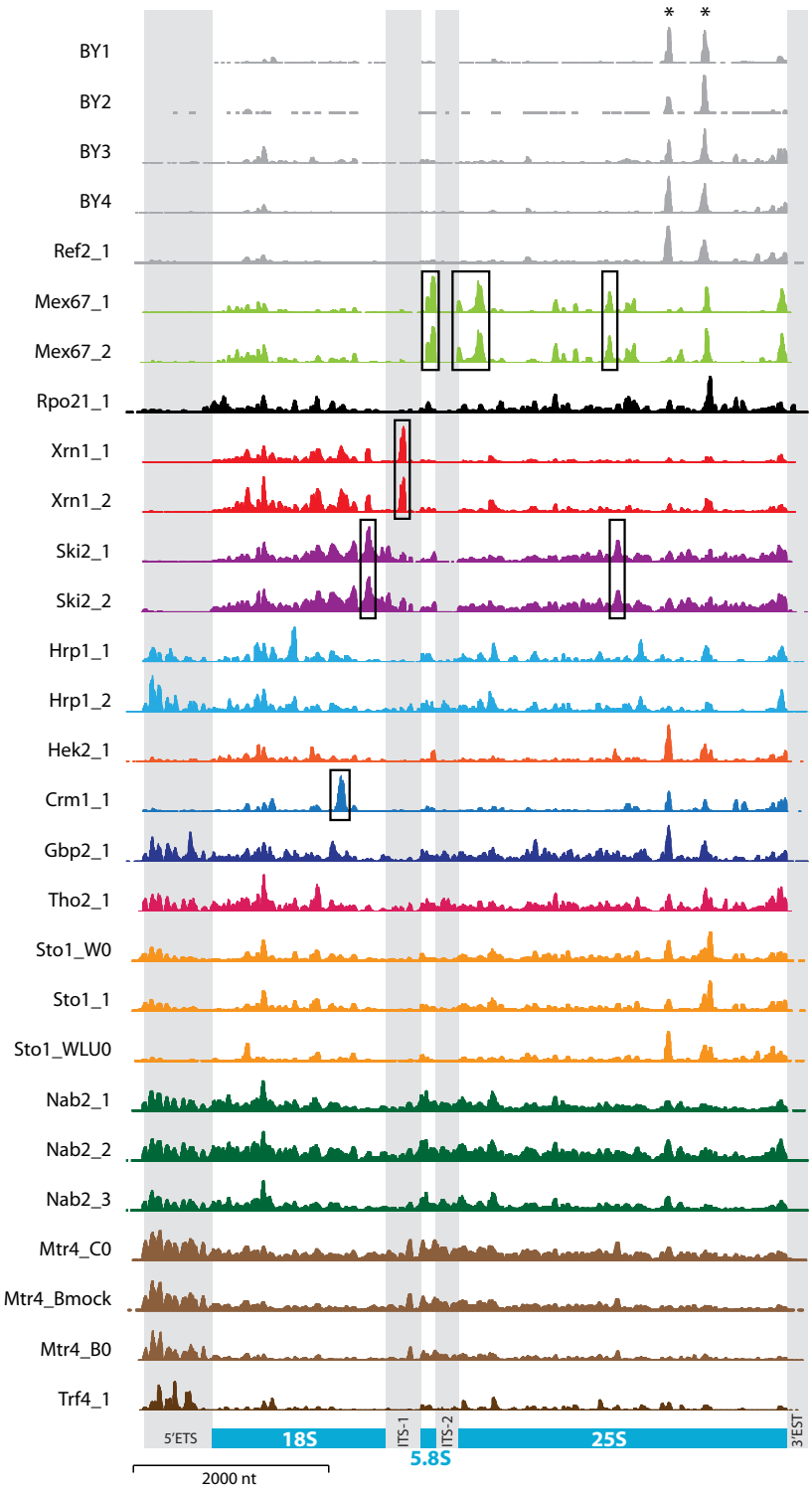
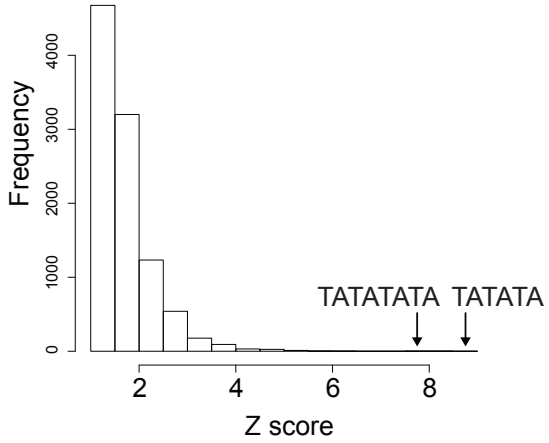


Figure 3.4: Binding profiles across ribosomal RNAs. ITS, internal transcribed spacer; ETS, external transcribed spacer.

overrepresentation scores for each possible k-mer (Wlotzka et al, 2011). High Z scores indicate that a motif is significantly more abundant within mRNA hits than would be expected by chance, taking into account the sequence composition of each mRNA to which the protein binds. To avoid detecting spurious motifs arising from sequencing artefacts, I used only reads for which the 3' adapter was detected. Furthermore, the datasets for some proteins, such as Trf4, Mtr4 and Pab1, had a high proportion of non-encoded A-tails (Table 3.2), consistent with these proteins binding oligo(A) or poly(A) tails. To restrict the present analysis to encoded motifs, I used the genomic sequence corresponding to each mapped read, and excluded low complexity reads (with fewer than 7 non-modal nucleotides, e.g. "GTCCGAAAAAAAAA"). This latter step removes reads with non-encoded As that are otherwise (incorrectly) mapped to A-rich regions of the genome.

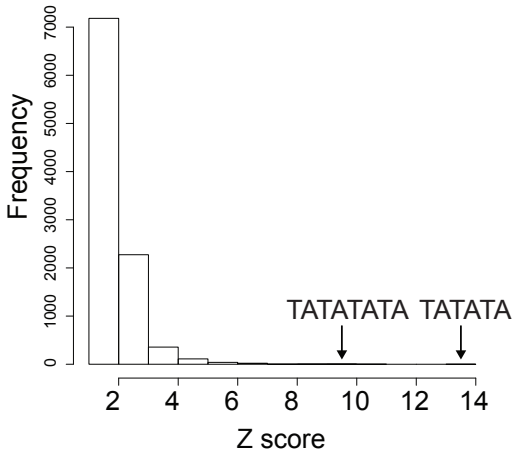
This analysis detected the sequence UAUAUAUA as highly enriched in Hrp1-bound RNA fragments (Figures 3.5A and 3.5B), and CNN repeats in Hek2-bound fragments (Figure 3.5C). Indeed, 21.1 % of all Hek2 reads (after removal of low complexity sequences) contained the motif CNNCNNC. In contrast, no significantly enriched motifs were reproducibly detected for Mex67, Gbp2, Mtr4, Ski2, Tho2 or Trf4, consistent with the requirement for these factors to bind a broad range of substrates in their roles as transcription, export and/or surveillance factors. For Pab1 hits, the most highly enriched motif was UAUAUA. Mutations in cDNAs generally represent errors in reverse transcription, and their frequency is often increased as the enzyme traverses the crosslinked nucleotide during library preparation. A high mutation frequency is therefore often observed at the precise binding site of the protein. Mutations were detected in 37-42 % of all occurrences of the UAUAUA motif in the Hrp1 dataset, indicating direct binding at this site. The mutation frequency was lower for Pab1 (29 %), suggesting that it binds in the vicinity of, but not always precisely at, this motif. I present a more in-depth analysis of Pab1 binding sites in Figures 4.7 and 4.8.

A **Hrp1_1**



	Z score	Mutation frequency (%)
TATATA	8.99	32.32
TATATAT	8.49	41.86
ATATATA	8.43	32.09
ATATATAT	8.03	44.34
TATATATA	7.92	45.1
GAGA	7.66	11.69
GTTCGA	7.59	27.78
ATATAT	7.47	29.24

B **Hrp1_2**



	Z score	Mutation frequency (%)
TATATA	13.34	37.04
ATATA	10.92	28.64
ATATATA	10.87	36.54
TATATATA	9.96	35.19
TATATAT	9.88	39.77
TATAT	9.68	28.86
ATATAT	9.24	34.63
TATAG	9.08	31.8

C **Hek2_1**

	Z score	Mutation frequency (%)
CAGCAGCAA	6.47	10.59
CAGCAGCA	6.46	10.48
ACAGCAGC	6.38	9.68
AACAGCAGC	6.26	13.16
AGCAGCAA	6.24	8.57
CAGCAGCAACA	6.22	14.75
CAGCAGCAAC	6.18	14.49
CAACAGCAGC	6.18	13.04

Percentage of reads containing CNNCNNC	
Hek2	21.1
Mex67_1	12.9
Hrp1_1	9.4
Mtr4_B0	7.7
Mex67_2	12.4

Figure 3.5: Motif analyses. A Statistical overrepresentation scores (Z scores) for motifs within mRNA hits from the Hrp1_1 dataset (top), and examples of the top motifs and the frequency with which these motifs are mutated (bottom). **B** The same analysis was repeated for the Hrp1_2 dataset. **C** Top motifs detected in the Hek2_1 dataset, with mutation frequencies and Z scores for each motif (left); abundance of CNNCNNC motifs within various datasets (right, top); sequence logo representation of the top Hek2 motifs (right, bottom).

Overall, the analyses of global binding patterns (Figures 3.2 and 3.3), the location of hits in specific transcripts (Figure 3.4), and the enrichment of sequence motifs (Figure 3.5) support the conclusion that the CRAC datasets are reliable, reproducible and specific.

3.4 LncRNAs bind a subset of mRNA biogenesis and turnover factors

Having established the quality of the CRAC datasets, I next investigated how lncRNAs and mRNAs differ in which proteins they bind. For each dataset, I normalised the total hits for various coding and non-coding transcript classes to total Pol II hits. The Pol II total is calculated by summing the hits in those CUTs, SUTs, mRNAs, snRNAs and snoRNAs for which there is minimal overlap between annotations, allowing hits to be unambiguously assigned. The normalised values are then expressed as % of Pol II hits (Figure 3.6). The inclusion of additional classes of lncRNAs in Figure 3.6 (XUTs, antisense transcripts (Granovskaia et al, 2010; Yassour et al, 2010), and intergenic transcripts (Granovskaia et al, 2010)) can lead to a total greater than 100 %. This method of data normalisation and plotting permits a quantitative comparison of transcript classes both between and within datasets, and does not over- or under-represent transcript classes that overlap. Replicate datasets, where available, were merged by calculating the mean of the two (or more) values for each class. The normalised values for SUTs, CUTs and XUTs have been replotted in Figure 3.7 in a manner that facilitates comparison of transcript classes within individual datasets.

The nuclear surveillance factors Trf4 and Mtr4 bound extensively to lncRNAs, which account for ~20 % of Pol II hits (Figure 3.6). This is comparable to the early termination factors Nrd1 and Nab3, which are reported to participate in the termination and turnover of CUTs and some longer lncRNAs (Creamer et al, 2011; Marquardt et al, 2011; Wlotzka et al, 2011). Notably, lncRNAs are more prevalent in the Trf4 and Mtr4 datasets than datasets from the major nuclear exonucleases Rrp44, Rrr6 and Rat1 (Figure 3.6), suggesting that the

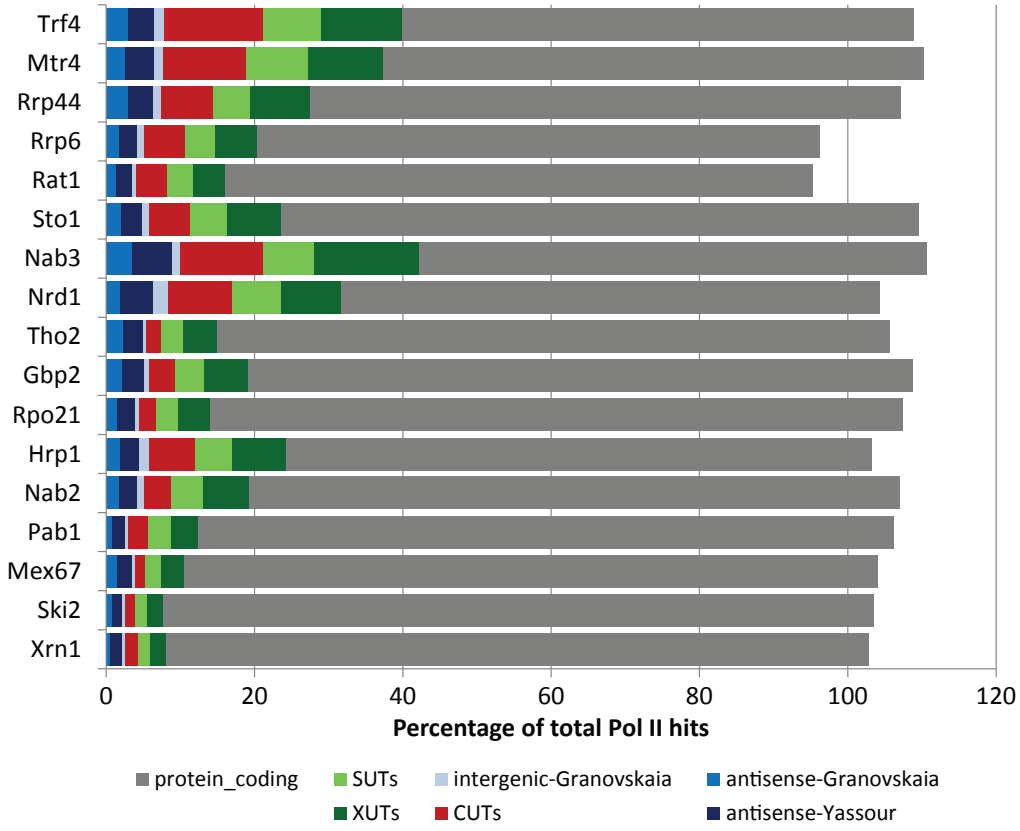


Figure 3.6: Abundance of lncRNA and mRNA hits. Hit counts for each class are normalised to total Pol II hits for each protein, and expressed as a percentage. Values can sum to greater than 100 % because some hits map to, and are counted for, multiple classes of lncRNA.

surveillance cofactors TRAMP, Nrd1 and Nab3 are particularly important for lncRNA turnover, and less so for nuclear (pre-)mRNA turnover.

LncRNAs were also moderately abundant amongst Tho2 and Gbp2 targets, with CUTs and SUTs together accounting for 4.9-7.3 % of their Pol II hits (Figure 3.6). Rpo21, the largest Pol II subunit, binds CUTs and SUTs to the same degree as Gbp2 and Tho2. This suggests that lncRNAs and mRNAs are bound by the TREX complex, of which Tho2 and Gbp2 are members, and binding is in proportion to their level of transcription. In comparison to Rpo21, Gbp2 or Tho2, lncRNAs are bound more abundantly by the nuclear cap binding complex large subunit Sto1 (CBC80), the 3' end processing factor Hrp1, and the nuclear poly(A) binding protein Nab2 (CUTs + SUTs = 8.1-11.2 % of Pol II hits). For Sto1, this might be due to its binding exclusively to the 5' ends of all transcripts, whereas Rpo21, Tho2 and Gbp2 binding occurs across the whole transcript (Chapters 4 and 5), and is therefore is lower for shorter transcripts such as CUTs. However, the enrichment of lncRNAs amongst Hrp1 and Nab2 targets suggests these proteins might play specific roles in lncRNA metabolism.

The most striking result, however, is the scarcity of lncRNAs bound to the export receptor Mex67 and to the cytoplasmic surveillance factors Ski2 and Xrn1 (CUTs + SUTs = 3.1-3.4 % of Pol II hits) (Figure 3.6). LncRNAs are 3- to 6-fold more abundantly bound to nuclear surveillance factors (Trf4, Mtr4, Nrd1, Nab3, Rrp44 and Rrp6), strongly suggesting that lncRNAs are predominantly retained and degraded in the nucleus. The poly(A) binding protein, Pab1, which is present in both the nucleus and cytoplasm, also binds more lncRNAs (CUTs + SUTs = 5.7 % of Pol II hits) than do the strictly cytoplasmic factors. I conclude that the level of binding to lncRNAs is highest for nuclear surveillance factors, moderate for transcription and nuclear processing factors, and lowest for export and cytoplasmic surveillance factors.

Generally, the various annotated classes of lncRNAs exhibit the same profiles when binding is compared across the tested proteins (Figures 3.6 and 3.7). However, SUTs are underrepresented relative to CUTs in Trf4, Mtr4, Rrp44, Rrp6, Nrd1 and Nab3 datasets, but equally or overrepresented in Rpo21, Tho2, Gbp2, Mex67, Ski2 and Xrn1 datasets (Figure 3.7). This suggests that SUTs are less actively retained and degraded in the nucleus than are CUTs, which is consistent with reports that SUTs are more susceptible to the cytoplasmic surveillance machinery (Marquardt et al, 2011). However, these differences between SUTs and CUTs are minor, suggesting that SUTs more closely resemble CUTs than mRNAs.

The genome-wide differences revealed in Figures 3.6 and 3.7 are also borne out at an individual gene level, which is apparent from plots of hits across representative mRNAs (Figure 3.8) and lncRNAs (Figures 3.9 and 3.10). For example, *TEF1* mRNAs were bound abundantly by Xrn1, Ski2 and Mex67, and moderately by Mtr4 and Trf4 (Figure 3.8A). *RPL9B* mRNAs also exhibited Mex67, Ski2 and Xrn1 binding, and a high level of Trf4 and Mtr4 binding. This is consistent with a recent report that, besides terminating at either of two canonical cleavage and polyadenylation sites to produce stable mRNAs, *RPL9B* can also terminate in a Nrd1-dependent manner ~120 nt downstream, with the resultant "CUT-like" transcript rapidly degraded by Rrp6 (Gudipati et al, 2012). Indeed, I detected multiple Pab1 binding sites around the canonical *RPL9B* 3' end, and downstream signal from Mtr4 and (in the *rrp6Δ* background) Nab2 (Figure 3.8B). Inspection of CRAC hits for the *URA2* upstream CUT (Figure 3.9A) (Thiebaut et al, 2008), the *SRG1* CUT at the *SER3* locus (Figure 3.9B) (Martens et al, 2005), and the archetypal CUT *NEL025c* (Figure 3.10A) (Thiebaut et al, 2006; Wyers et al, 2005), reveal abundant Mtr4, Trf4, Sto1 and Hrp1 binding, but little or no Mex67, Ski2 or Xrn1 binding. Finally, I analysed hits across the loci encoding *SUT477* (Figure 3.10B) and *XUT_15F-14* (Figure 3.10C), both reported to be susceptible to cytoplasmic surveillance factors (Marquardt et al, 2011; van Dijk et al, 2011). For *SUT477*, binding is again dominated by the nuclear surveillance factors, whereas *XUT_15F-14* and the

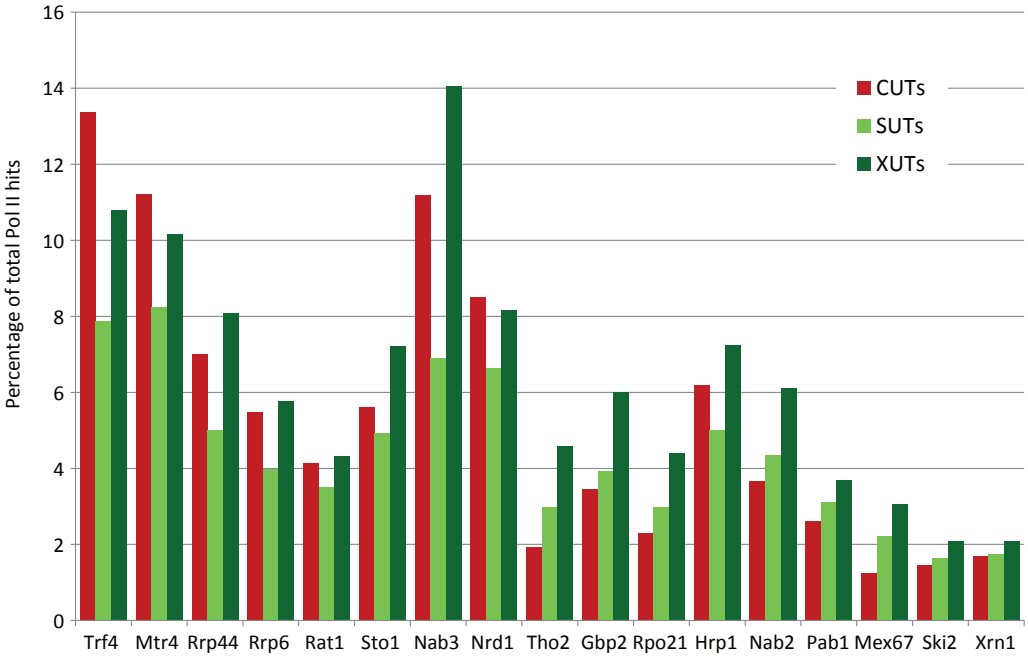


Figure 3.7: Abundance of lncRNA hits. Hit counts for each class are normalised to total Pol II hits for each protein, and expressed as a percentage.

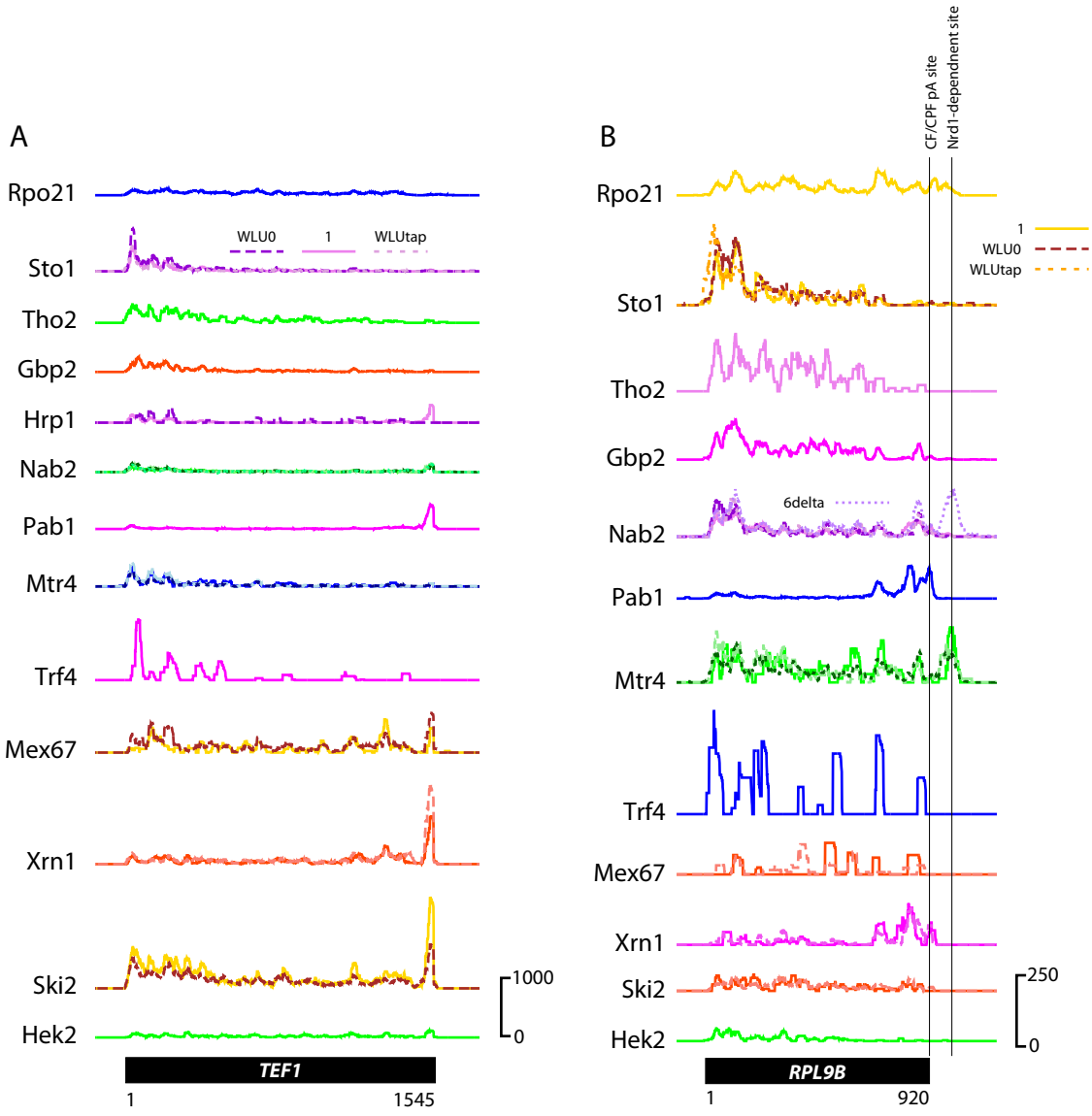


Figure 3.8: Binding profiles across mRNAs. The distribution of CRAC hits mapping to two mRNAs, **A** *TEF1* and **B** *RPL9B*, is plotted for each of the 13 proteins tested (hits per million Pol II hits). *RPL9B* is reported to have a long and a short form arising from termination at alternative sites, which are indicated.

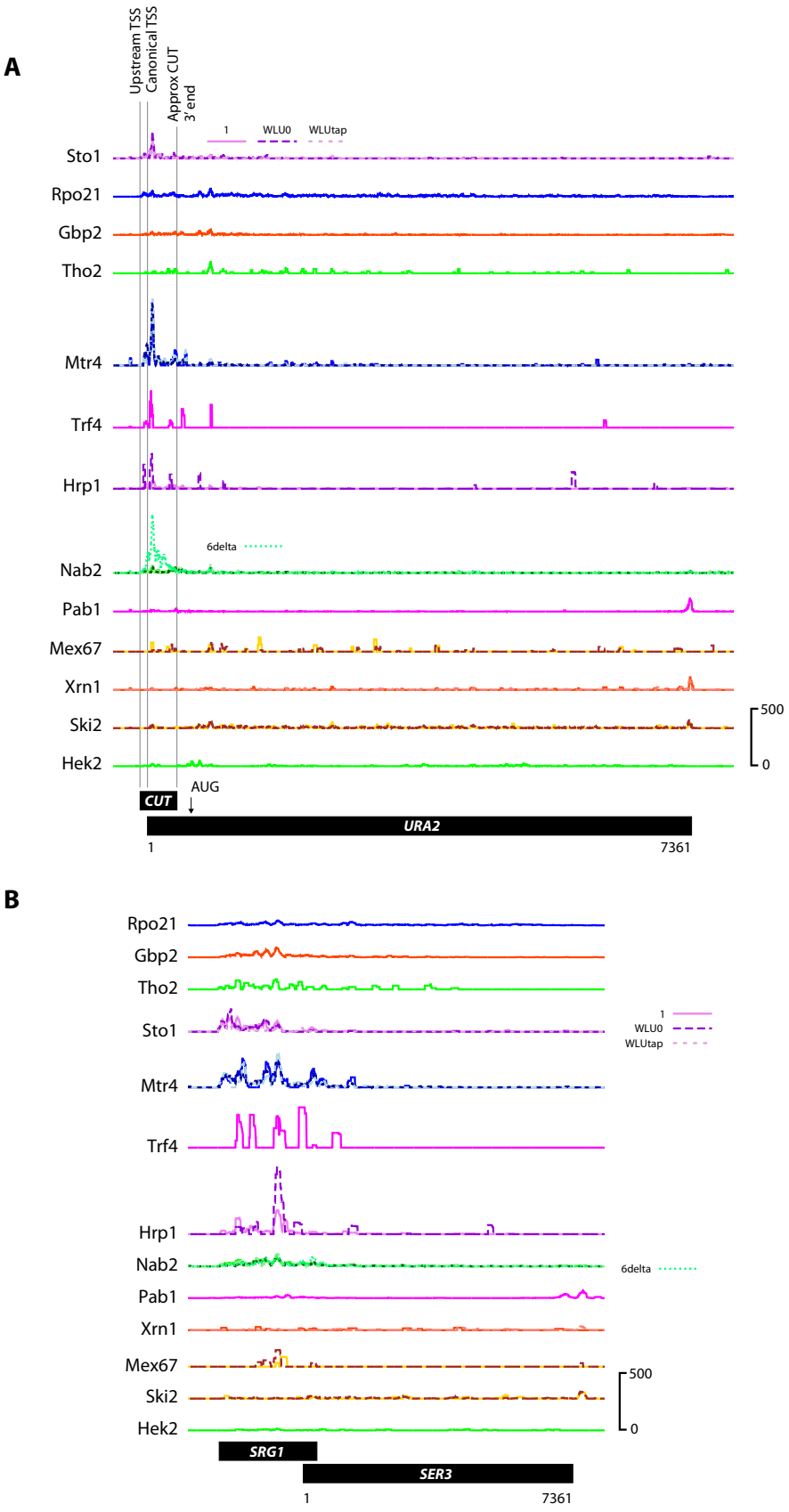


Figure 3.9: Binding profiles across upstream CUTs. The distribution of CRAC hits mapping to CUTs (cryptic unstable transcripts) transcribed from sites upstream of two protein coding genes, **A** *URA2* and **B** *SER3*, is plotted for each of the 13 proteins tested (hits per million Pol II hits). The *SER3* upstream CUT is also known as SRG1 (*SER3* regulatory gene 1).

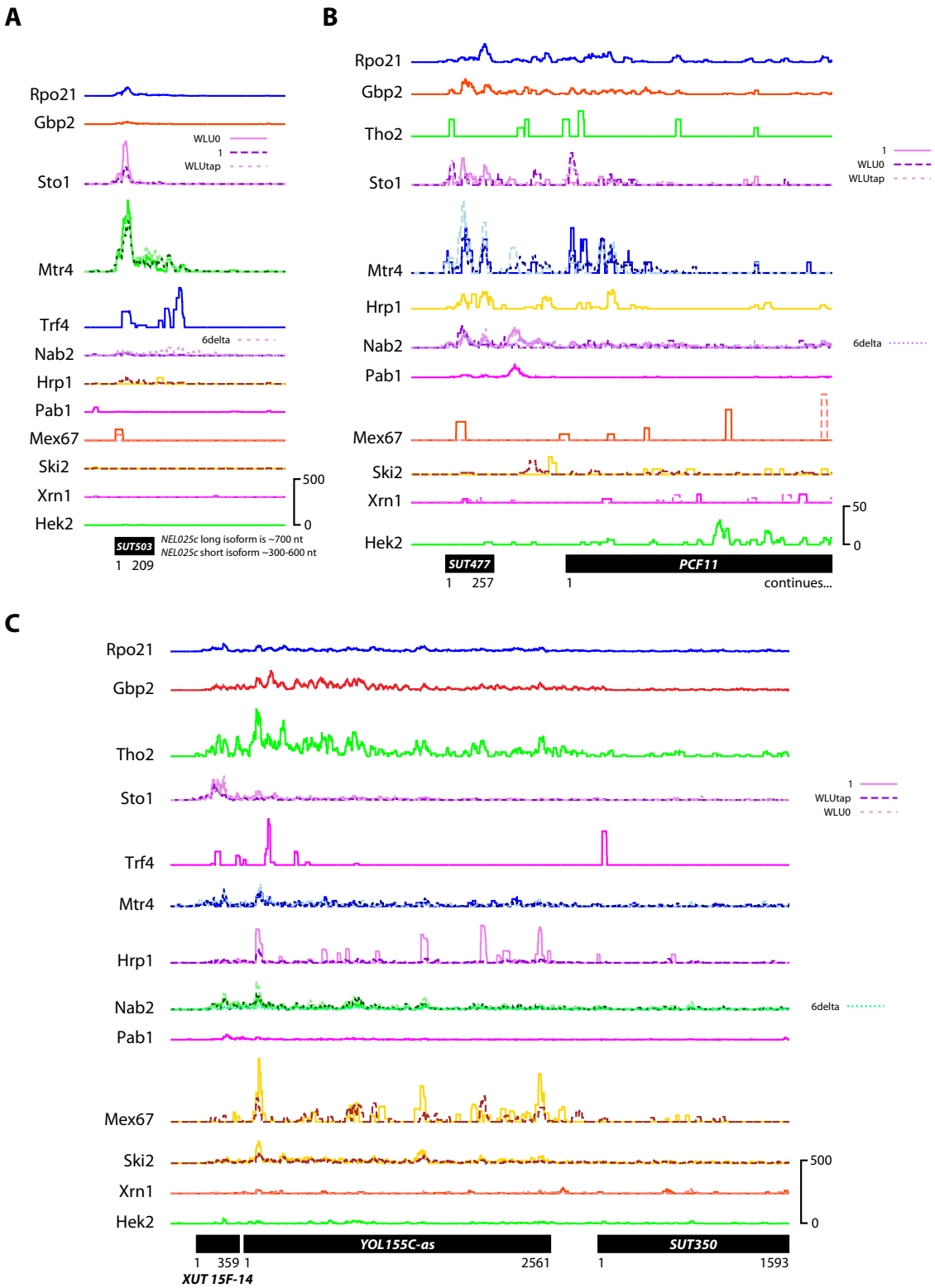


Figure 3.10: Binding profiles within stable or "cytoplasmic" lncRNAs. The distribution of CRAC hits mapping to three lncRNAs, **A** SUT503 (NEL025c), **B** SUT477 and **C** XUT_15F-14, is plotted for each of the 13 proteins tested (hits per million Pol II hits). Notably, SUT503 and SUT477 are detectable in wild-type cells, and XUT_15F-14 is detectable when the cytoplasmic factor Xrn1 is deleted, suggesting it reaches the cytoplasm. SUT477 is encoded upstream of *PCF11*, whereas the *XUT_15F-14* locus also encodes an antisense transcript (YOL155C-as) and a SUT (SUT350).

adjacent *YOL155C-as* antisense transcript identified by (Granovskaia et al, 2010) are also bound by Mex67 and Ski2. However, the extent of Ski2 binding does not exceed that of Mtr4 (Figure 3.10C), suggesting that even when lncRNAs can be detected bound to cytoplasmic surveillance factors, nuclear surveillance still plays a major role in their turnover. Furthermore, abundant binding of Mex67, Ski2 and/or Xrn1 to lncRNAs is rare, as lncRNAs are scarce amongst Mex67, Ski2 and Xrn1 hits (Figure 3.6).

To further investigate the prevalence of lncRNAs with evidence for cytoplasmic localisation (binding to Ski2 and Xrn1), I compared Mtr4, Trf4, Ski2 and Xrn1 binding for individual CUTs and SUTs. I selected CUTs and SUTs with at least 100 hits per million (relative to total Pol II hits) in 2 of the Mtr4, Trf4, Ski2 and Xrn1 datasets. The 708 transcripts that passed this filter were then clustered (hierarchical clustering by Spearman rank), and the results represented as a heat map (Figure 3.11). I normalised each row to enable binding profiles of both high and low abundance transcripts to be easily visualised. Strikingly, the majority of CUTs and SUTs were bound more abundantly by Trf4 and/or Mtr4 than Xrn1 or Ski2, with very few displaying a preference for Ski2 and Xrn1. Together, this suggests that the majority of SUTs and CUTs are predominantly nuclear.

3.5 Discussion

In this chapter, I have presented a comprehensive analysis of the RNA targets of key Pol II transcription, termination, 3' end processing, export and decay factors. The high sensitivity and specificity of the CRAC method enabled me to detect mRNAs, but also much less abundant lncRNAs, and thus compare the behaviour of mRNAs and lncRNAs. This revealed, for all classes of lncRNA tested, a gradual decline in association with RNA binding proteins along the standard mRNA biogenesis pathway (Figures 3.6 and 3.7). Analyses of individual members of each lncRNA class (Figures 3.8-3.11) revealed that this behaviour is typical for most members of each class, with few exceptions.

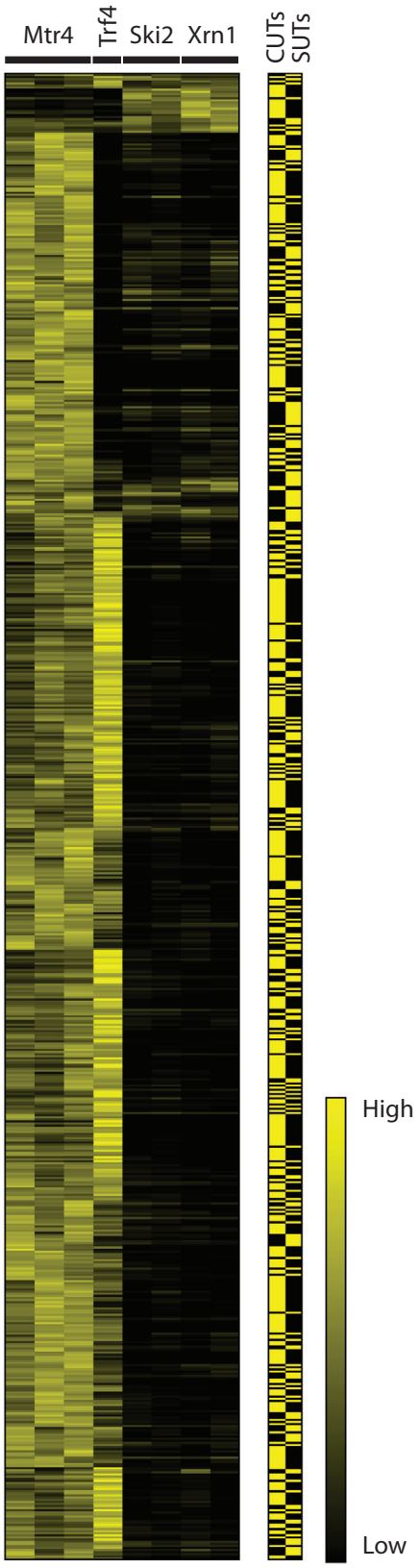


Figure 3.11: CUT and SUT binding to nuclear and cytoplasmic surveillance factors. The abundance of hits in Mtr4, Trf4, Xrn1 and Ski2 datasets is plotted for all CUTs and SUTs with at least 100 hpm (hits per million Pol II hits) in two of these datasets. Each row represents an individual CUT or SUT, and rows were ordered by hierarchical cluster analysis based on Spearman rank correlations. The identity of each row (CUT or SUT) is indicated by yellow bars in the right hand panel.

Early events in transcription are the same for lncRNAs and mRNAs

The previously reported detection of lncRNAs in libraries of capped transcripts (Miura et al, 2006; Neil et al, 2009) is consistent with my observation that lncRNAs are abundantly bound to Sto1, constituting ~10 % of its targets (Figure 3.6). Thus although some lncRNAs are reported to resemble sn/snoRNAS in their mode of transcription termination, their cap structure is apparently more similar to the CBC-bound m_7Gppp cap of mRNAs than the trimethylated cap of mature sn/snoRNAs, which does not usually bind CBC (Schwer et al, 2011). For Sto1, the inclusion of an enzymatic decapping step during library preparation, which enhances the detection of *bona fide* interactions with capped substrates, led to an enrichment of mRNAs and lncRNAs at the expense of snRNAs, snoRNAs, tRNAs and rRNAs (Figure 3.2, "Sto1_WLU0" vs "Sto1_WLUtap"). This confirms that Sto1 is bound to mRNA and lncRNA caps, whereas the low level of binding observed to sn/snoRNAs might reflect a second, partly cap-independent, mode of interaction by the CBC. Indeed, a low level of CBC has been detected at snoRNA genes by ChIP (Kim et al, 2006).

During transcription, nascent mRNAs are bound and packaged by the TREX complex, which suppresses R-loop formation to facilitate transcription elongation and prevent hyper-recombination (Gomez-Gonzalez et al, 2011; Huertas et al, 2003; Rondón et al, 2003). The TREX complex is recruited to active protein coding genes via a number of interactions. Notably, Tho2, a member of the five subunit THO subcomplex (Chavez et al, 2000; Strasser et al, 2002), was recently shown to bind RNA (Pena et al, 2012). The transcriptome-wide results presented in this study confirm this direct interaction of Tho2 with mRNAs, and reveal that Tho2 also interacts with lncRNAs (Figures 3.6, 3.9 and 3.10). THO can also be recruited to chromatin via the Prp19 complex, which bridges Pol II and THO (Chanarat et al, 2011), and it will be interesting to test whether this complex is also associated with lncRNA genes.

In addition to Tho2, the TREX complex contains four additional RNA-binding proteins - Yra1, Sub2 (Strasser et al, 2002), Gbp2 and Hrb1 (Hurt et al, 2004). The RNA crosslinking efficiencies of Sub2 and Yra1 were too low to generate useable CRAC datasets (data not shown), but Gbp2 crosslinked well and the high complexity dataset obtained revealed binding to both mRNAs and lncRNAs (Figure 3.2). The abundance of Gbp2 and Tho2 hits was well correlated for mRNAs, CUTs and SUTs (Figure 3.3). The analyses of Tho2 and Gbp2 reveal that lncRNAs, like mRNAs, are bound by the TREX complex. The pervasive, interleaved nature of non-coding transcription suggests that a failure to package ncRNAs could potentially interfere with transcription of canonical genes and lead to instability at many genomic loci. The association of the TREX complex with lncRNAs might therefore minimise these deleterious effects, by ensuring that lncRNA transcription proceeds efficiently, is rapidly cleared, and does not lead to R-loop formation. I also detect Tho2 and Gbp2 bound to snoRNAs, snRNAs, and Pol I (Figure 3.4) and Pol III transcripts (Figure 3.2), suggesting that the TREX complex might be a ubiquitous component that interacts with each of the transcription systems.

The TREX complex also participates in RNA export pathways (Strasser et al, 2002; Zenklusen et al, 2002) (Figure 1.3) via the interaction of Hpr1 with Sub2 (Zenklusen et al, 2002). Sub2 in turn binds Yra1, which is handed over from Pcf11 towards the 3' end of genes (Johnson et al, 2011; Straszer et al, 2001). Finally, Yra1 recruits the export receptor Mex67 (Strasser et al, 2002; Straszer et al, 2001; Zenklusen et al, 2001), and enhances the interaction of Mex67 with Nab2 (Iglesias et al, 2010). However, the abundance of lncRNAs bound to Mex67 (Figure 3.6) was much lower than that bound to Gbp2 and Tho2. This suggests that for lncRNAs, TREX binding is not coupled to Mex67 recruitment. This could reflect the absence of one or more TREX components (e.g. Yra1 or Sub2), or the failure of lncRNPs to undergo the required remodelling and post-translational modification steps that occur during maturation of mRNPs towards export competency.

Overall, the early events in lncRNA transcription appear to closely resemble those in mRNA transcription. This is consistent with observations that many lncRNA promoters resemble the promoters of protein coding genes, in terms of nucleosome depletion and the composition of the Pol II pre-initiation complex (Murray et al, 2012; Rhee et al, 2012). Additionally, lncRNA transcription, like that of mRNAs, is associated with H3K4 di- and trimethyation and H3K36 trimethylation (Guttman et al, 2009; Houseley et al, 2008; Khalil et al, 2009; Kim et al, 2012; Kirmizis et al, 2007; Pinskaya et al, 2009; van Dijk et al, 2011; van Werven et al, 2012), and Pol II CTD Ser5 and Ser2 phosphorylation (Kim et al, 2010a; Preker et al, 2011).

LncRNAs are predominantly nuclear

The low association of lncRNAs with the mRNA export receptor Mex67 suggests that lncRNAs and mRNAs diverge prior to export from the nucleus. However, a variety of export mechanisms exist in yeast, and it is therefore possible that lncRNAs exit the nucleus via an alternative pathway. Indeed, several groups have detected lncRNA accumulation following the deletion of cytoplasmic surveillance factors, and have interpreted this as evidence for a role of lncRNAs in the cytoplasm (Berretta et al, 2008; Geisler et al, 2012; Marquardt et al, 2011; Matsuda et al, 2009; Thompson et al, 2007; Toesca et al, 2011; van Dijk et al, 2011). Although high-throughput studies of fractionated cell extracts have detected cytoplasmic lncRNAs in mammalian cells (Kapranov et al, 2007), the difficulties in obtaining a cytoplasmic extract free from nuclear contaminants in yeast have so far precluded such analyses. An understanding of where in the cell lncRNAs are localised would help us to address key functional questions about lncRNAs.

A comparison of the CRAC data for the cytoplasmic surveillance factors Ski2 and Xrn1, versus that for the nuclear surveillance factors Mtr4 and Trf4 (and Rrp6, Rat1 and Rrp44), revealed that lncRNAs are far more abundantly bound to nuclear factors than cytoplasmic factors (Figures 3.6, 3.7 and 3.11). The lncRNAs constitute ~20 % of transcripts bound to

Trf4 and Mtr4, and 11 % of those bound by Rrp6. In comparison, less than 4 % of Ski2 or Xrn1 hits map to lncRNAs. This reveals that in comparison to mRNAs, a far greater proportion of lncRNA turnover occurs in the nucleus and, together with the scarcity of lncRNAs among Mex67 hits, suggests that lncRNAs are predominantly nuclear.

This is consistent with the numerous roles reported for lncRNAs in the nucleus compared to the small number of reported cytoplasmic functions. Notable examples of cytoplasmic functions include translational regulation by lincRNA-p21 or *KCS1* as-ncRNA (Huarte et al, 2010; Nishizawa et al, 2008; Yoon et al, 2012). However, even lncRNAs that accumulate upon disruption of the cytoplasmic surveillance machinery predominantly have nuclear functions. For example, *SRG1* transcription directs nucleosome deposition (Thebault et al, 2011), the Ty1 *RTL* lncRNA modulates retrotransposition (Berretta et al, 2008), "CD-CUTs" interfere with transcription factor binding (Toesca et al, 2011), XUTs exert repression via H3K4 dimethylation (van Dijk et al, 2011), and *GAL10*-as ncRNA transcription elicits histone deacetylation (Houseley et al, 2008; Pinskaya et al, 2009). Together with the RNA binding data reported here this suggests that, although some lncRNAs reside in the cytoplasm this is largely non-functional and simply represents a low level of leakage from the nucleus. Studies in which the accumulation of lncRNAs is compared in cytoplasmic versus nuclear surveillance mutants can be very misleading, as nuclear surveillance mutants can be rescued by "fail-safe" surveillance in the cytoplasm, whereas there are no alternative turnover pathways available to cytoplasmic surveillance mutants. The CRAC approach reported here is more likely to reflect the endogenous situation, as it captures a snapshot of RNA binding in wild-type cells.

The accumulation of lncRNAs in *dcp1Δ* or *dcp2Δ* cells (Berretta et al, 2008; Geisler et al, 2012; Marquardt et al, 2011), which has been cited as evidence for the cytoplasmic presence of lncRNAs, might also be explained by these factors participating in nuclear surveillance. Indeed, although predominantly cytoplasmic, Dcp2 can shuttle under some conditions

(Groušl et al, 2009), and participates with Rat1 in lncRNA turnover (Geisler et al, 2012) (consistent with Rat1 CRAC hits in lncRNAs; Figure 3.6). Consistently, for some lncRNAs the Dcp1- and Dcp2-dependent pathway is apparently distinct from bulk mRNA turnover as it is variously independent of canonical decapping activators, deadenylases and/or Ski2 (Geisler et al, 2012; Marquardt et al, 2011; Thompson et al, 2007). It is unlikely, however, that Xrn1 participates in nuclear surveillance, as it is reported to localise exclusively to the cytoplasm (Johnson, 1997), and this is supported to some extent by the low abundance of intronic hits in CRAC datasets for Xrn1 (Chapter 5). Overall, therefore, I suggest that lncRNA turnover is primarily carried out by the canonical nuclear surveillance machinery (TRAMP, Rrp44, Rrp6 and Rat1), perhaps assisted by Dcp1 and Dcp2, but with cytoplasmic surveillance mainly providing a fail-safe pathway.

LncRNAs recruit export adapters but not the export receptor Mex67

The data presented here reveal that although lncRNPs initially resemble mRNPs, they are not exported from the nucleus, and do not bind the export receptor Mex67. The pertinent question then becomes why?

In addition to interacting with ubiquitylated Hpr1 (Gwizdek et al, 2006), Mex67 is recruited by a variety of adapter proteins. These include Yra1 (Zenklusen et al, 2001), Npl3 (Gilbert et al, 2004; Kim Guisbert et al, 2005), Nab2 (Iglesias et al, 2010) and perhaps Hrp1 (Henry et al, 1996). These adapters are suggested to contribute to partially distinct pathways for Mex67 recruitment, based on differences in their genetic interactions, requirements for ubiquitylation (Duncan et al, 2000; Iglesias et al, 2008) and RNA binding profiles (Kim Guisbert et al, 2005). However, a more recent high-throughput study detects Nab2 associated with the majority of mRNAs (Batisse et al, 2009), suggesting that earlier observations suffered from a lack of sensitivity. Indeed, our lab detects Mex67, Nab2 and Npl3 (Rebecca Holmes, unpublished observations, and this study) bound to 75 %, 95 % and 87 % of all mRNAs respectively. The low level of background for Mex67 and Nab2, evident from

inspection of rRNA hits (Figure 3.4), suggests that most of these mRNA hits are genuine. Notably, I detect Hrp1 and Nab2 bound abundantly to lncRNAs (Figure 3.6), and the same is true for Npl3 (Rebecca Holmes, unpublished observations). Thus irrespective of whether Hrp1, Nab2 and Npl3 recruit Mex67 via a common pathway or parallel pathways, these data suggest that a lack of Mex67 recruitment to lncRNAs is not due to the absence of these adapters. Recruitment of Mex67 to mRNPs is also dependent on post-translational modifications, such as the ubiquitylation of the APT and Set1 complex subunit Swd2 (Vitaliano-Prunier et al, 2012) and the THO subunit Hpr1 (Gwizdek et al, 2005; Neumann et al, 2003), and dephosphorylation of Npl3 (Gilbert et al, 2004). Perhaps lncRNPs are deficient for one or more of these steps, or other regulatory events that remain to be discovered.

Hrp1 and Nab2 might play non-canonical roles in lncRNA metabolism

In addition to export, Hrp1 and Nab2 perform various roles in mRNA metabolism, including 3' end processing and poly(A) tail length control, respectively. Thus when bound to lncRNAs, they might still perform these functions despite not recruiting Mex67. Somewhat surprisingly, whereas the binding of Tho2 and Gbp2 to mRNAs and lncRNAs is in proportion to their level of transcription (judged by Rpo21 binding, Figure 3.6), lncRNAs are apparently overrepresented in Nab2, Hrp1 and Sto1 datasets (Figure 3.6). This suggests that Nab2 and Hrp1, and perhaps also Sto1, might perform additional, lncRNA-specific functions. Indeed, Hrp1 can participate in Nrd1-dependent transcription termination (Kim et al, 2006; Kuehner et al, 2008), the CBC complex contributes to nuclear turnover of aberrant and/or retained pre-mRNAs (Das et al, 2006; Das et al, 2000; Kuai et al, 2005), and Nab2 is implicated in the surveillance of aberrant pre-mRNAs (e.g. those with retained introns) at the nuclear pore via interaction with Mlp1 (Galy et al, 2004; Schmid et al, 2012; Vinciguerra et al, 2005). Perhaps these factors therefore contribute to lncRNA termination or decay? In the

following chapter, I test for lncRNA-specific roles for Hrp1, Sto1 and Nab2 using a combined bioinformatic and biochemical approach.

Alternatively, the enrichment of lncRNA versus mRNA reads in Hrp1/Sto1/Nab2 but not Tho2/Gbp2 datasets might reflect a high abundance of lncRNAs in the nucleoplasm. In this model, Tho2 and Gbp2 bind predominantly co-transcriptionally, so their CRAC hits reflect the abundance of nascent transcripts. Following transcription termination, Tho2 and Gbp2 would be recycled via dissociation from the mRNP, perhaps by Hmt1 methylation of mRNA binding proteins which is reported to disrupt interactions with Tho2 and release the mRNP from the site of transcription (McBride et al, 2005; Yu et al, 2004). Conversely, Nab2, Hrp1 and Sto1 bind co-transcriptionally and remain bound until Nab2 is removed at the cytoplasmic face of the NPC (Tran et al, 2007) or site of translation (van den Bogaart et al, 2009), and Sto1 and Hrp1 are displaced by translation factors (Fortes et al, 2000; Gao et al, 2005; González et al, 2000; van den Bogaart et al, 2009). Sto1, Hrp1 and Nab2 CRAC hits therefore reflect RNP composition in the nucleoplasm. Together this suggests that lncRNAs accumulate more in the nucleus than mRNAs, given their respective rates of transcription, and that lncRNA turnover is slower than mRNA export. Notably, although lncRNA turnover is suggested to be extremely rapid, experimental analysis of the half-lives of two CUTs, *NEL025c* and *NMR026w*, finds they are surprisingly stable ($t_{½}$ = 20-30 minutes) (Thompson et al, 2007). Additional kinetic analyses are required to determine whether this result applies only to a minor pool of lncRNAs, or is more generally applicable. One possibility, however, is that lncRNA turnover in the nucleus is a passive consequence of their not being exported, rather than a more rapid, active mechanism.

LncRNAs are particularly abundant amongst TRAMP targets

Trf4 and Mtr4 also bind lncRNAs in excess of the transcription rate inferred from Rpo21 binding (Figure 3.6). This suggests that Trf4 and Mtr4 specifically target lncRNAs, although the source of this specificity is unclear. Indeed, motif analyses (Figure 3.5 and data not

shown) failed to detect any sequence specificity for Mtr4 and Trf4, or indeed for Tho2, Gbp2, Ski2 or Mex67. This likely reflects the requirement for these factors to participate in the metabolism of many different transcripts, which is difficult to reconcile with sequence-specific binding. Substrate targeting is more likely to depend on structural features, and indeed, Trf4 requires a 1 nt ssRNA overhang for adenylation (Jia et al, 2011), the Mtr4 helicase is most active on substrates with a short ssRNA overhang (Jia et al, 2012), and reconstituted TRAMP specifically adenylates incorrectly folded $tRNA_iMet$ (Vaňáčová et al, 2005). Recent advances in high-throughput analyses of RNA secondary structures and melting temperature might soon reveal the defining characteristics of TRAMP and exosome substrates (Kertesz et al, 2010; Wan et al, 2012).

Trf4 and Mtr4 binding to lncRNAs exceeds that of the downstream exonucleases, Rrp6 and Rrp44, suggesting that even amongst exosome targets, TRAMP preferentially binds to lncRNAs. This suggests that TRAMP is particularly important for lncRNA surveillance. TRAMP can affect the expression of numerous genes in an adenylation- and helicase-independent manner (Callahan et al, 2010; Paolo et al, 2009), possibly by directly recruiting the degradation machinery to substrates including lncRNAs. The lower frequency of lncRNAs amongst Rrp44/Rrp6 substrates versus those of Trf4/Mtr4 might reflect very rapid turnover of lncRNAs (compared to mRNAs) upon commitment to decay, in which case TRAMP would be bound for longer than Rrp44/Rrp6.

Towards a complete lncRNP proteome

In summary, many of the mRNA-binding proteins tested here also bound lncRNAs. This revealed that lncRNAs are predominantly retained and degraded in the nucleus, and interact with early mRNA biogenesis factors. However, this approach cannot identify novel lncRNP components that do not bind mRNAs. Moreover, proteome-wide analyses of poly(A)+ RNA binding proteins identify hundreds of factors in both yeast and human cells (Castello et al, 2012; Scherrer et al, 2010; Tsvetanova et al, 2010). Although these are assumed to represent

the mRNA interactome, many lncRNAs are also polyadenylated (evident from their interaction with Pab1, Figure 3.6). Thus amongst these RNA binding proteins, there may also be novel lncRNA-specific factors. Analyses of the transcriptome-wide targets of these newly identified RNA binding proteins, together with mass spectrometric analyses of proteins bound to specific bait lncRNAs, might allow the determination of the complete lncRNP proteome.

4: Non-canonical roles of mRNP proteins in lncRNA metabolism

4.1 Introduction

The data presented in Chapter 3 suggest that the mRNA binding proteins Nab2, Hrp1 and Sto1 play both canonical and non-canonical roles in lncRNA metabolism, as their binding to lncRNAs exceeds their share of the Pol II transcriptional output (Rpo21 hits) and TREX binding. Distinct functions can potentially be distinguished by an examination of where within transcripts these proteins bind, and such analyses are the main focus of this chapter. For each of these proteins, several functions have been documented in addition to their canonical role. I sought to determine whether these, or other, non-canonical functions contribute to lncRNA metabolism.

Canonical and non-canonical functions of Sto1

The nuclear cap binding complex (CBC) comprises a heterodimer of Sto1/Cbc1 (yeast homologue of human CBP80) and Cbc2 (yeast homologue of human CBP20), and is thought to primarily protect the 5' cap structure from nuclear decay. Nuclear 5' to 3' turnover is therefore limited to transcripts that have been decapped by the pyrophosphatase Rai1 (Jiao et al, 2010) and perhaps Dcp2 (Geisler et al, 2012). The CBC accompanies mRNAs to the cytoplasm (Shen et al, 2000) where it is replaced by eIF4E (Fortes et al, 2000), perhaps after participating in a pioneer round of translation (Gao et al, 2005). The CBC is also reported to contribute to translation during osmotic stress (Garre et al, 2012), stimulate transcription initiation (Lahudkar et al, 2011), participate in nuclear turnover of aberrant mRNAs (Das et al, 2006; Das et al, 2000; Kuai et al, 2005), and interact with the U1 snRNA to promote splicing at 5' proximal sites (Lewis et al, 1996). Finally, the CBC contributes to suppression of termination at weak poly(A) sites (Das et al, 2000) by functioning with Npl3 to impede

Pcf11 and Rna15 recruitment (Shen et al, 2000; Wong et al, 2007). However, the synthetic lethality of Npl3 and CBC (McBride et al, 2005) suggests that a second, Npl3-indepedent pathway might exist, perhaps involving the U1 snRNA, which plays a role in suppressing early termination in metazoa (Berg et al, 2012).

Canonical and non-canonical functions of Hrp1

In contrast to Sto1, which functions early in transcription, the best documented role of Hrp1 is in cleavage and polyadenylation where it promotes the use of "major" poly(A) sites and suppresses cryptic poly(A) sites (Kessler et al, 1997; Kim Guisbert et al, 2006; Minvielle-Sebastia et al, 1998). The two Hrp1 RBDs bind in tandem to the polyadenylation efficiency element (EE) UAUAUA (Kim Guisbert et al, 2006), and also contact Rna15 which binds the A-rich polyadenylation element (PE) AAUAAA (Leeper et al, 2010). Notably, although Hrp1 binds specifically to UAUAUA (Chen & Hyman, 1998; Kim Guisbert et al, 2005; Kim Guisbert et al, 2006; Valentini et al, 1999), this interaction can be outcompeted by transcripts lacking this sequence (Chen & Hyman, 1998) and Hrp1 can bind indirectly via Rna15 (Bucheli et al, 2007). Hrp1 is also required for Nrd1-dependent termination at the *NRD1* and *HRP1* attenuators (Kuehner & Brow, 2008; Steinmetz et al, 2006b) and perhaps on snoRNA genes, which crosslink to Hrp1 in ChIP experiments (Kim et al, 2006) and contain TA repeats in their terminator elements (Steinmetz et al, 2006a). However, Hrp1 was not required for termination of an artificial CUT, despite the presence of an AU-rich motif (Porrua et al, 2012). Additionally, Hrp1 can shuttle to the cytoplasm, where it accumulates upon stress (Buchan et al, 2011; Henry et al, 2003) and can target transcripts for NMD via binding within the coding region (González et al, 2000).

Canonical and non-canonical functions of Nab2

Analogous to Hrp1, Nab2 is recruited to mRNAs during 3' end processing and accompanies them to the cytoplasm. Nab2 is a poly(A) binding protein with a tandem array of seven zinc

fingers, three of which constitute a high affinity poly(A) binding domain (Brockmann et al, 2012; Kelly et al, 2010; Kelly et al, 2007; Marfatia et al, 2003). Nab2 is predominantly nuclear (Anderson et al, 1993; Wilson et al, 1994), and is displaced from mRNPs in the cytoplasm by Dbp5 (Tran et al, 2007) and the importin Kap104 (Dheur et al, 2005). Nab2 is therefore suggested to represent the nuclear counterpart of Pab1, a cytosolic poly(A) binding protein that protects mRNA 3' ends from degradation and stimulates translation. In this capacity, Nab2 is thought to primarily act in poly(A) tail length control, as depletion or mutation of Nab2 results in hyperadenylation (Amrani et al, 1997; Hector et al, 2002; Minvielle-Sebastia et al, 1997). Pab1 can shuttle into the nucleus (Brune et al, 2005) and interact with CFIA (Amrani et al, 1997; Minvielle-Sebastia et al, 1997), so there is some debate as to whether it also participates in nuclear poly(A) tail length control. *In vitro*, either Pab1 or Nab2 can regulate the poly(A) tail length (Dheur et al, 2005), but Nab2 is more efficient, and can also protect against subsequent rounds of polyadenylation by Pap1 (Viphakone et al, 2008) or trimming by the PAN complex (Schmid et al, 2012; Viphakone et al, 2008). Furthermore, the *in vivo* hyperadenylation observed upon Nab2 depletion cannot be rescued by Pab1 overexpression (Hector et al, 2002). A recent study, however, finds that Nab2 cannot regulate poly(A) tail length *in vitro*, and that only Pab1 binds short poly(A) tails suggested to represent early intermediates in poly(A) tail synthesis (Schmid et al, 2012). Such *in vitro* assays must be interpreted with caution, however, as *in vivo*, polyadenylation occurs in the nucleus where Nab2 is abundant and Pab1 scarce. *In vivo* depletion studies also have their limitations. For example, Nab2 depletion leads to upregulation of genes involved in 3' end processing, making it hard to distinguish direct effects (Gonzalez-Aguilera et al, 2011).

In addition to poly(A) tail length control, Nab2 is required for mRNA export (Green et al, 2002) and surveillance. This is dependent on its ability to interact with the export receptor Mex67 (Iglesias et al, 2010), as well as the NPC-associated pre-mRNA retention factor Mlp1

(Fasken et al, 2008; Grant et al, 2008; Green et al, 2003) and the mRNP disassembly factor Gfd1 (Grant et al, 2008; Suntharalingam et al, 2004; Zheng et al, 2010). Nab2 initially tethers the mRNP at the nuclear side of the NPC for surveillance, and then at the cytosolic face for disassembly. The ability of Nab2 to facilitate correct mRNP assembly and promote export is illustrated by *nab2* mutations that lead to nuclear retention (Vinciguerra et al, 2005). However, Nab2 can also oppose export by participating in the detection and retention of aberrant mRNPs, evident from *nab2* mutants that constitutively export improperly folded mRNPs (Brockmann et al, 2012; Tran et al, 2007). Indeed, Nab2 depletion results in misexpression of many genes (Gonzalez-Aguilera et al, 2011), but particularly the upregulation of intron-containing pre-mRNAs (Schmid et al, 2012), which are well documented targets of Mlp1-mediated quality control (Galy et al, 2004). Furthermore, Nab2 interacts with Trf4 and Rrp6, promoting Trf4-dependent oligoadenylation (Schmid et al, 2012) as well as exosome-mediated turnover of its own mRNA in an autoregulatory loop (Roth et al, 2009; Roth et al, 2005).

Hrp1, Nab2 and Sto1 therefore play important roles in transcription termination, mRNP assembly and surveillance, and their binding to lncRNAs merits further investigation. By analysing the distribution of these and other RNA binding proteins across two classes of lncRNAs, CUTs and SUTs, I sought to determine what roles these proteins might be playing, and whether they behave differently when binding CUTs versus SUTs.

4.2 Early mRNA biogenesis factors bind CUTs and SUTs in a canonical manner

I used two related bioinformatic approaches to examine the distribution of Sto1, Gbp2 and Tho2 CRAC hits across CUTs and SUTs, with both analyses performed on the 500 most abundantly bound CUTs or SUTs for each protein (Figure 4.1).

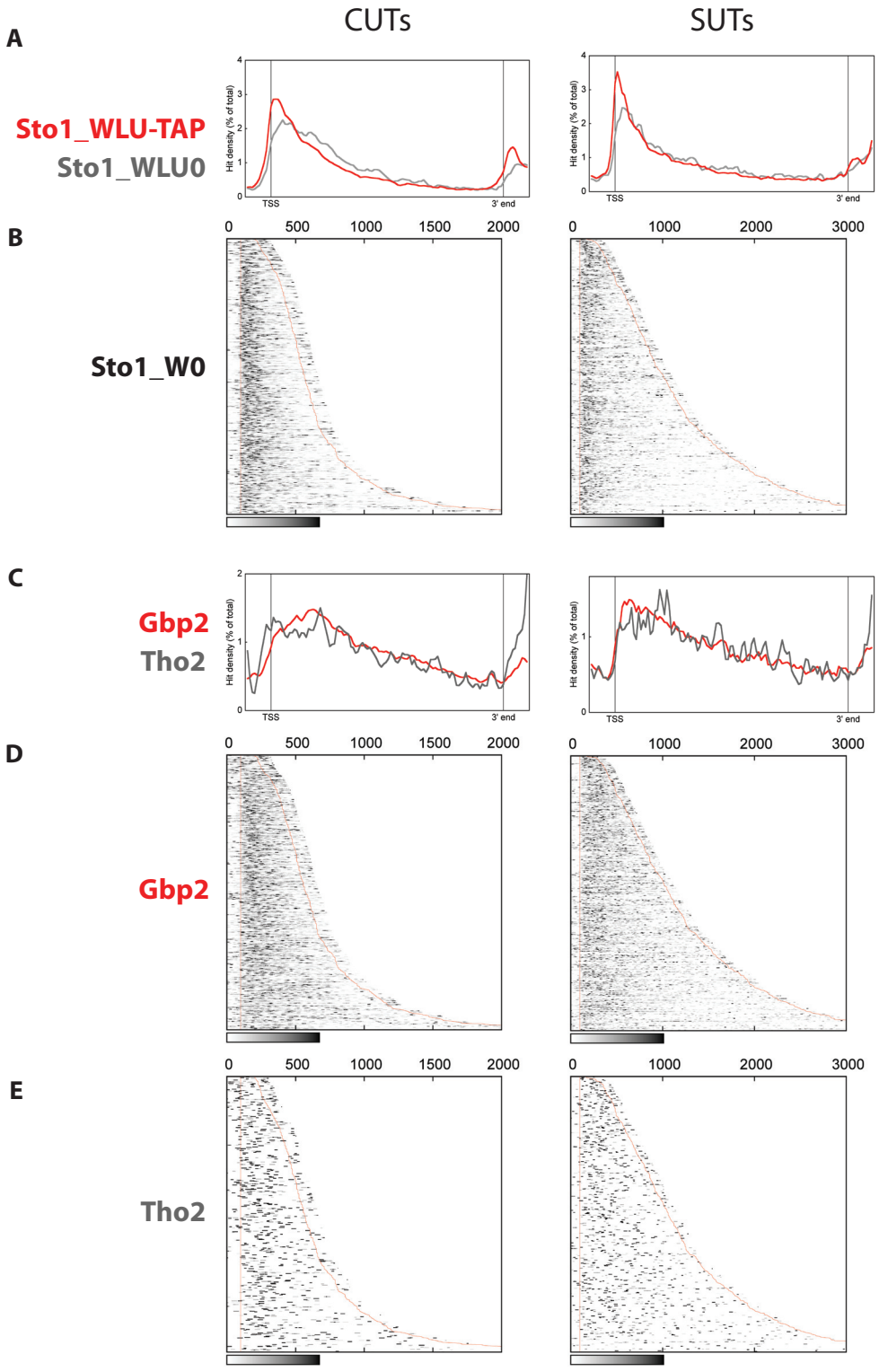


Figure 4.1: Gbp2, Tho2 and Sto1 binding distributions across CUTs and SUTs. A Average Sto1 hit densities across the 500 most abundantly bound CUTs and SUTs, including 100 nt upstream and downstream flanking regions. The "Sto1_WLU-TAP" sample was subjected to an enzymatic decapping step to improve detection of 5' ends. **B** Sto1 hits across the 500 most abundantly bound CUTs and SUTs, plotted individually (one transcript per row). Annotated 5' and 3' ends are indicated by red lines. **C** Average Gbp2 and Tho2 hit densities across CUTs and SUTs. **D** Gbp2 hits across the 500 most abundantly bound CUTs and SUTs. **E** Tho2 hits across the 500 most abundantly bound CUTs and SUTs.

In the first approach, which provides an average binding profile across all transcripts, each transcript was divided into 100 bins of equal length, and 100 nt 5' and 3' flanking regions divided into 10 bins (120 bins in total). Considering the first transcript, hits were counted for each nucleotide, then a mean value calculated for each bin to give the hit density (hits nt^{-1}). I calculated hit densities rather than totals for each bin to account for the different lengths of bins in the flanking versus transcribed regions. This was repeated for all 500 transcripts, then each transcript normalised by linear scaling so that the densities for that transcript summed to 100. This expresses the hit density for each bin as a percentage of the total of all 120 hit densities. Finally, I averaged the 500 individual profiles, and the resulting plot (e.g. Figure 4.1A) reflects the typical hit distribution across all 500 transcripts, with the normalisation step ensuring that each transcript contributes equally.

I complemented this approach with a simpler analysis, in which transcripts were sorted by length, hits counted at each position and scaled to the maximum value for each transcript, and the data plotted as a two dimensional heat map (e.g. Figure 4.1B). Here, each row represents a transcript, and each column the absolute position from the aligned TSSs. This enables the individual hit distributions of all 500 transcripts to be displayed on one plot, without scaling by length.

The plots of Sto1 average binding distribution (Figure 4.1A) and detailed (per transcript) binding distribution (Figure 4.1B) for CUTs (left) and SUTs (right) reveal a strong bias for Sto1 binding toward lncRNA 5' ends. The CRAC procedure biases against the detection of binding at the extreme 5' end, as transcripts require cleavage upstream of the crosslinking site during the RNase fragmentation step to ligate to the 5' adapter. The closer the crosslinking site is to the 5' end of the transcript, the lower the probability of an upstream cleavage event. The inclusion of an enzymatic decapping step (TAP treatment) overcomes this bias, and so is likely to better reflect the physiological binding distribution of Sto1. Indeed, the binding distribution analyses for the TAP-treated Sto1 dataset revealed an

increased level of Sto1 binding at the extreme 5' ends of CUTs and SUTs (Figure 4.1A, red vs grey). This confirms that Sto1 largely interacts with capped lncRNAs. However, even with TAP treatment, the binding sites detected for Sto1 on both CUTs and SUTs extend ~200 nt downstream of the transcript 5' end (Figure 4.1A). This is surprising, as Sto1 is thought to bind exclusively to the cap structure. Inspection of the distribution of Sto1 binding across individual lncRNAs (Figure 3.10B) and mRNAs (Figure 3.8B) reveals that this diffuse pattern is apparent for individual transcripts in both lncRNA and mRNA classes, and is not simply an artefact of misannotated 5' end coordinates leading to a "fuzzy" average plot. Furthermore, the broad peaks observed for Sto1 contrast to the much sharper binding peaks of Pab1 (for example, Figures 3.8 and 3.10), suggesting that they are not due to a limited resolution of the CRAC method. This suggests that Sto1 might make contacts across the 5' region of transcripts, in addition to the cap. Notably, the *in vivo* UV crosslinking performed during CRAC analyses captures even transient interactions (in contrast to other approaches that only detect stable interactions, with the highest affinity or slowest dissociation kinetics). Overall, the Sto1 binding profiles reveal that binding to CUTs and SUTs closely resembles binding to mRNAs, and importantly, confirm that the reference set of annotated lncRNA 5' end coordinates used in this study is relatively accurate.

In contrast to Sto1, the Gbp2 and Tho2 binding profiles across CUTs and SUTs (Figure 4.1C) reveal a more gradual decrease in binding from the 5' to 3' end. This is consistent with their proposed role in coating nascent transcripts, which all start at the same nucleotide but have been elongated for various lengths downstream. Within the pool of nascent transcripts, bases are therefore decreasingly represented with increasing distance from the TSS. Furthermore, the Tho2 and Gbp2 binding profiles appear to be scaled to transcript length, decreasingly linearly over the full length of both CUTs and SUTs (Figures 4.1D-E). This contrasts to Sto1 binding, which is restricted to the 5' ~200 nts of transcripts, regardless of their length (Figure 4.1B). Further evidence for Tho2 and Gbp2 binding predominantly co-

transcriptionally is provided by their abundant binding to mRNA introns (Chapter 5), which are removed during or shortly after transcription. Together, these results indicate that Sto1 binding is limited to lncRNA 5' ends, whereas Tho2 and Gbp2 apparently bind throughout nascent transcripts with a profile defined by the composition of the chromatin-associated transcriptome. This is largely consistent with their reported canonical roles.

4.3 Hrp1 antagonises the production of CUTs

I next examined the binding of Hrp1 across CUTs, SUTs, and mRNAs, with analyses of all Hrp1 hits (Figure 4.2, left) or the subset containing the UAUAUA motif (Figure 4.2, right) that was significantly enriched in Hrp1 datasets (Figure 3.5A). Consistent with the documented role of Hrp1 in mRNA 3' end processing, inspection of the binding profiles across mRNAs revealed a peak at the 3' end, which was particularly prominent when considering only UAUAUA-containing reads (Figure 4.2A). Surprisingly, however, the majority of total Hrp1 hits were located towards the 5' ends of mRNAs. For SUTs (Figure 4.2B), no peak of Hrp1 binding at the 3' end was visible in plots generated from all hits, although a weak 3' peak was discernible when only UAUAUA-containing hits were plotted. For CUTs (Figure 4.2C), no peak of 3' Hrp1 binding was detectable in either type of analysis. These differences were particularly apparent when the average binding profiles were plotted (Figure 4.2D).

The lack of Hrp1 binding to the 3' end of CUTs suggested that its role, if any, in CUT metabolism might be different to its canonical function in mRNA 3' end processing. To test the function of Hrp1 bound to CUTs, I therefore generated a conditional depletion strain (*pGAL::3HA::HRP1*) and tested the effect of a reduced level of Hrp1 on the abundance of several CUTs detected amongst the Hrp1 CRAC hits. Upon shifting the strain to glucose containing media, the level of Hrp1 protein was reduced to below detectable levels within 1 hour (Figure 4.3C). Total protein expression remained constant (judged by Ponceau S staining), and the yeast did not show a growth defect until ~90 minutes. After two hours of

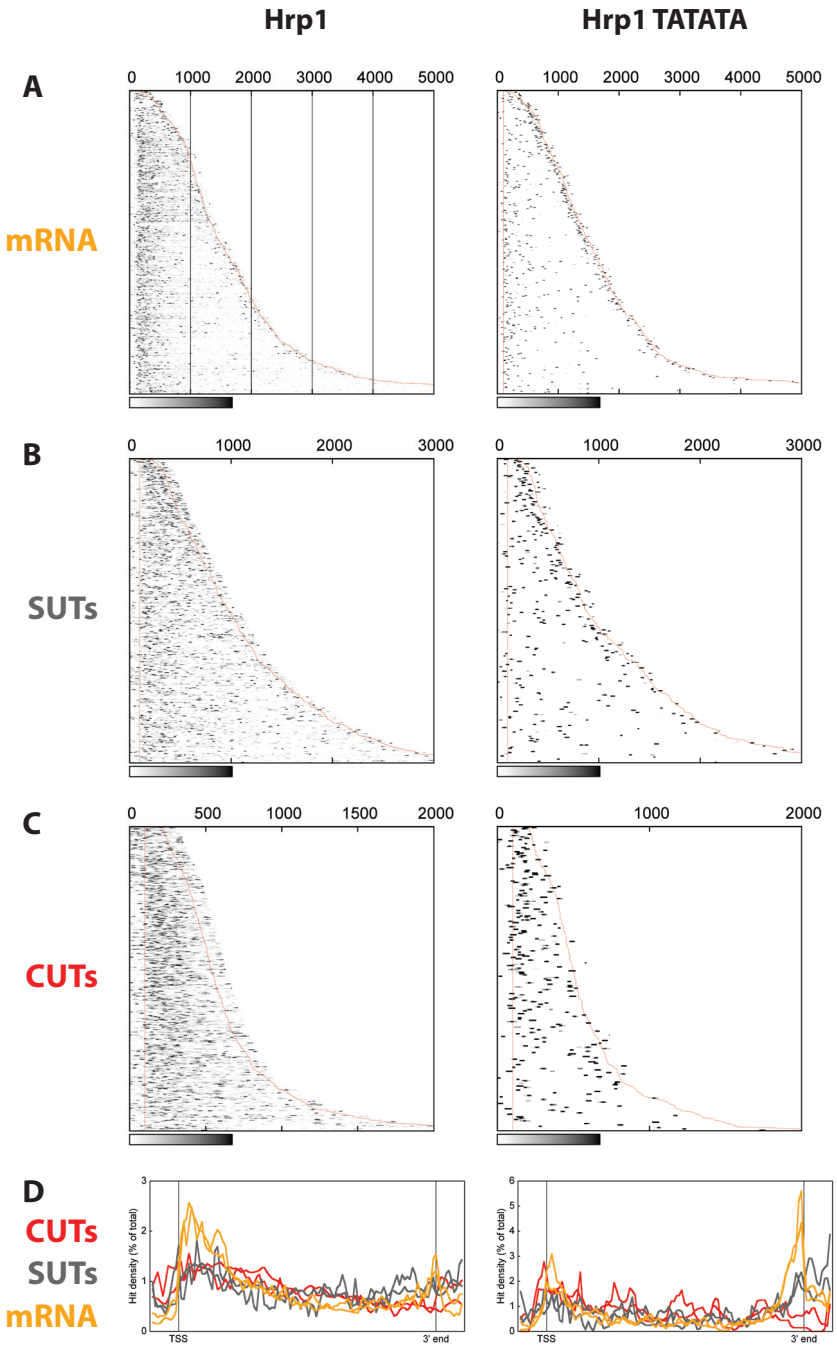


Figure 4.2: Hrp1 binding distributions across mRNAs, CUTs and SUTs. Hrp1 hits across the 500 most abundantly bound **A** mRNAs, **B** SUTs and **C** CUTs, including 100 nt flanking regions, plotted individually (one transcript per row). The analysis was performed for all Hrp1 hits in the Hrp1_2 dataset (left), or just those containing the UAUAUA motif (right). Annotated 5' and 3' ends are indicated by red lines. D Average binding profiles across mRNAs (orange), SUTs (grey) and CUTs (red), from Hrp1_1 and Hrp1_2 datasets.

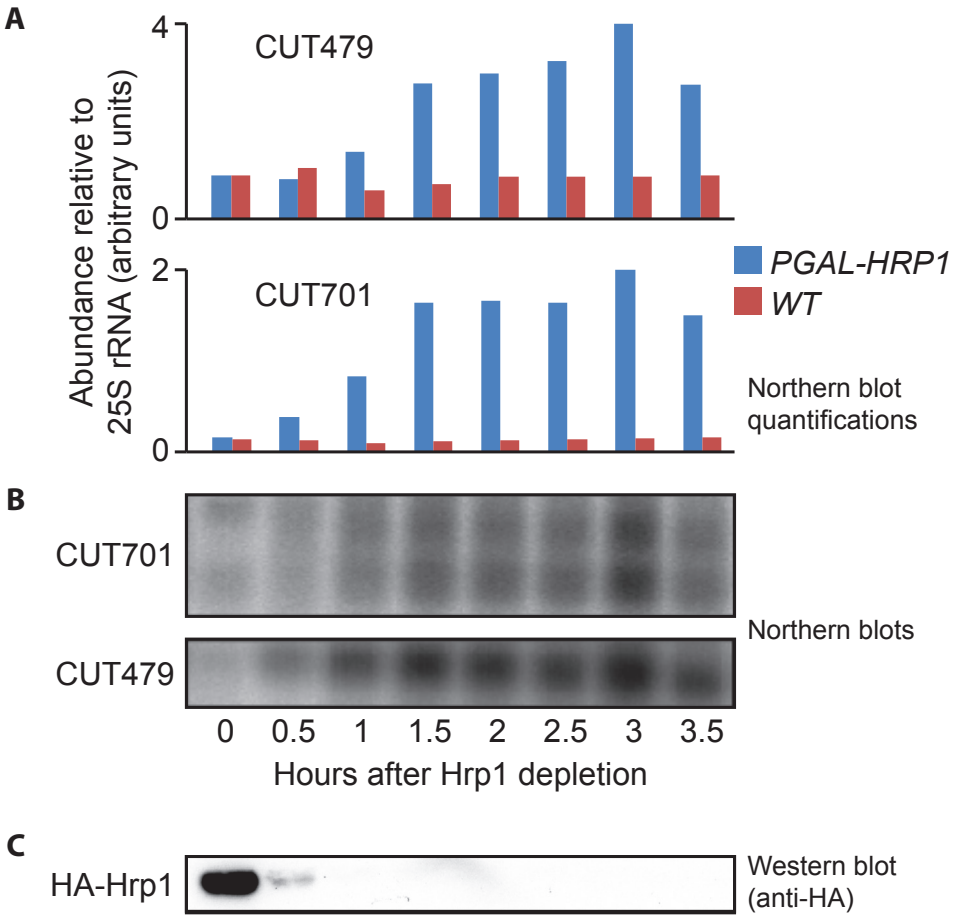


Figure 4.3: Hrp1 depletion upregulates CUT expression. A Quantitative northern analysis of CUT479 (top) and CUT701 (bottom) abundance in *pGAL::3HA:HRP1* or wild-type cells after shifting to glucose, normalised to 25S rRNA. **B** Autoradiogram used for the quantitative analysis in (A). **C** HA-Hrp1 abundance assayed by Western blot (anti-HA).

Hrp1 depletion, the proportion of total RNA with poly(A) tails (binding to oligo(dT) beads) was drastically reduced, confirming that the depletion ablates the canonical function of Hrp1 in 3' end cleavage coupled to polyadenylation (data not shown). Northern analysis of three CUTs revealed that their abundance in cellular RNA extracts increased 4- to 10-fold within 90 minutes of depletion. CUT701 was detectable as two diffuse bands in wild-type cells, and Hrp1 depletion resulted in a 4-fold increase in total CUT signal (Figure 4.3A). This was coincident with a slight increase in migration of the upper band. CUT479 was also detectable as two bands, and upon depletion the lower molecular weight band increased ~10-fold in abundance (Figure 4.3B), whereas the slower migrating band disappeared. Finally, CUT200 migrated as two bands, but upon Hrp1 depletion both increased in abundance (data not shown). The results for CUT200 were consistent between two biological replicates, and analyses of additional CUTs are ongoing. Together, the binding profiles and Northern analyses suggest that Hrp1 helps select the termination site for CUTs, but also acts post-transcriptionally (it is detected bound to the transcripts) to downregulate CUT abundance. This would be consistent with it participating in a Nrd1-dependent termination mechanism.

The Hrp1 binding profile for SUTs (Figure 4.2B), which is intermediate between mRNA and CUT profiles, suggests that SUT 3' end processing might share some but not all features with mRNA 3' end processing. To further investigate the differences between mRNA, CUT and SUT 3' end processing, I next examined the binding profiles for the poly(A) binding proteins Nab2 and Pab1.

4.4 Pab1 specifically binds transcript 3' ends, whereas Nab2 is more promiscuous

Analyses of Nab2 binding (three replicate datasets) across mRNAs revealed a tripartite profile, with (i) a broad peak immediately downstream of the TSS, (ii) moderate binding throughout the rest of the transcribed region, and (iii) a sharp peak at the 3' end (Figure

4.4A, “mRNA”). The 3’ peak of Nab2 binding, and the detection of non-encoded A tails for 18-21 % of Nab2 hits (Table 3.1), is consistent with its proposed role in binding 3’ poly(A) tails and regulating their length. However, the abundant binding across the rest of the transcript, particularly towards the 5’ end, suggests that Nab2 might perform additional functions involving more delocalised binding. In comparison, the Nab2 binding profiles across SUTs (Figure 4.4A, “SUTs”) retain the broad 5’ peak and moderate binding throughout the transcribed region, but largely lack the 3’ peak. More strikingly, the binding profiles across CUTs (Figure 4.4A, “CUTs”) completely lack the 3’ peak, comprising only the broad 5’-proximal peak.

In yeast lacking the surveillance factor Rrp6 (Figure 4.4B), which is reported to antagonise Nab2 binding and degrade CUTs, the profiles for Nab2 binding across mRNAs and SUTs resemble those for wild-type yeast. However, Nab2 binding is now present throughout CUTs, rather than exclusively towards their 5’ ends (Figure 4.4B, “CUTs”). This suggests that the lack of Nab2 binding towards CUT 3’ ends in wild-type cells is caused by the surveillance machinery, which degrades CUTs in a 3’ to 5’ direction. Notably, even in the absence of Rrp6, a peak of Nab2 binding is not detected at the extreme 3’ end of CUTs or SUTs (Figure 4.4B, “CUTs” and “SUTs”), suggesting that in contrast to mRNAs, they might never possess Nab2-bound 3’ poly(A) tails.

In contrast to Nab2, Pab1 was bound exclusively at the 3’ end of mRNAs (Figure 4.4C). Furthermore, 27 % of Pab1 reads contained non-encoded A tails (Table 3.1). This is likely an underestimate, as reads with a high proportion of As cannot be mapped to the yeast genome and so are excluded from the analysis. If these restrictions are lifted, 72 % of Pab1 reads have 10 or more consecutive adenosine residues at the 3’ end, compared to just 5 % for Nab2. This suggests that the majority of Pab1-bound fragments are A-tailed. Pab1 therefore exhibits a much greater specificity towards poly(A) sequences than does Nab2. Furthermore, whereas Nab2 does not bind SUT 3’ ends, Pab1 binds them moderately abundantly (Figure

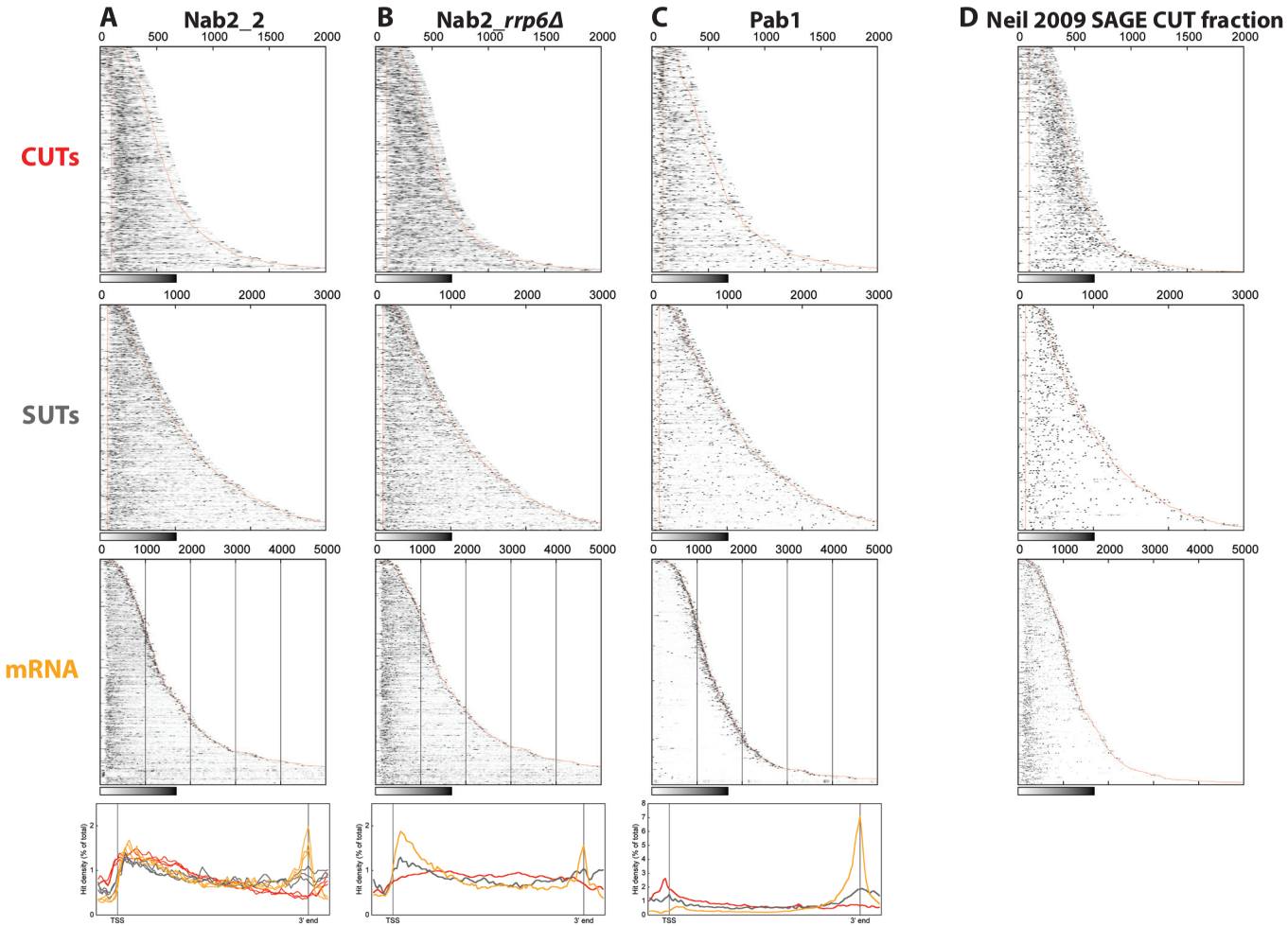


Figure 4.4: Nab2 and Pab1 binding distributions across mRNAs, CUTs and SUTs. The distribution of hits across the 500 most abundant CUTs, SUTs and mRNAs in the **A** Nab2_2, **B** Nab2_deltaRrp6 and **C** Pab1 CRAC datasets, as well as **D** the 3' SAGE dataset described in (Neil et al, 2009), are plotted. For the CRAC datasets, the average distributions are also plotted (bottom).

4.4C, “SUTs”). In contrast, Pab1 does not detectably bind to the 3’ end of CUTs (Figure 4.4C, “CUTs”). To test whether this is a technical artefact arising from inaccurate CUT 3’ end annotations, I examined SAGE tags from a CUT-enriched fraction which reveal the precise 3’ ends of numerous CUTs (Neil et al, 2009). This analysis (Figure 4.4D) revealed that both CUT and SUT 3’ end SAGE tags overlap with annotated 3’ ends, and agree similarly well with the lncRNA reference coordinates used here.

In conclusion, the CRAC data reveal that (i) Pab1 binds the 3’ ends of mRNAs and SUTs, but not CUTs, and (ii) although Nab2 binds along mRNAs, CUTs and SUTs, a prominent 3’ peak of binding is seen only for mRNAs.

4.5 Nab2 binding profiles identify distinct groups of transcripts

Nab2 and Pab1 both participate in poly(A) tail metabolism, but it was not clear whether they bind the same targets or distinct groups of transcripts. The complex average binding profile of Nab2 across the top 500 mRNAs (Figure 4.4A) suggested that there might be distinct groups of mRNAs with different modes of Nab2 binding. I therefore selected all mRNAs, CUTs and SUTs detected in the Nab2_1, Nab2_2, Nab2_3 or Pab1_1 datasets above a threshold level (100 hits per million Pol II hits), and for each transcript, calculated the hit density in 30 bins covering the transcribed region and 100 nt either side (using the same algorithm reported for Figure 4.1). I then plotted the individual profiles for all transcripts in four heat maps (one for each dataset) (Figure 4.5). The gene order is the same for each heat map, and is based on a k-means clustering analysis of the profiles obtained from the Nab2_1 dataset, with each cluster then being subclustered by total Nab2 and Pab1 hits. Strikingly, this analysis reveals that almost all mRNAs have a 3’-proximal Pab1 peak, irrespective of their Nab2 binding profile. The most conspicuous exceptions are SUTs and CUTs, for which Pab1 binding at the 3’ end was less well defined (SUTs) or absent (CUTs).

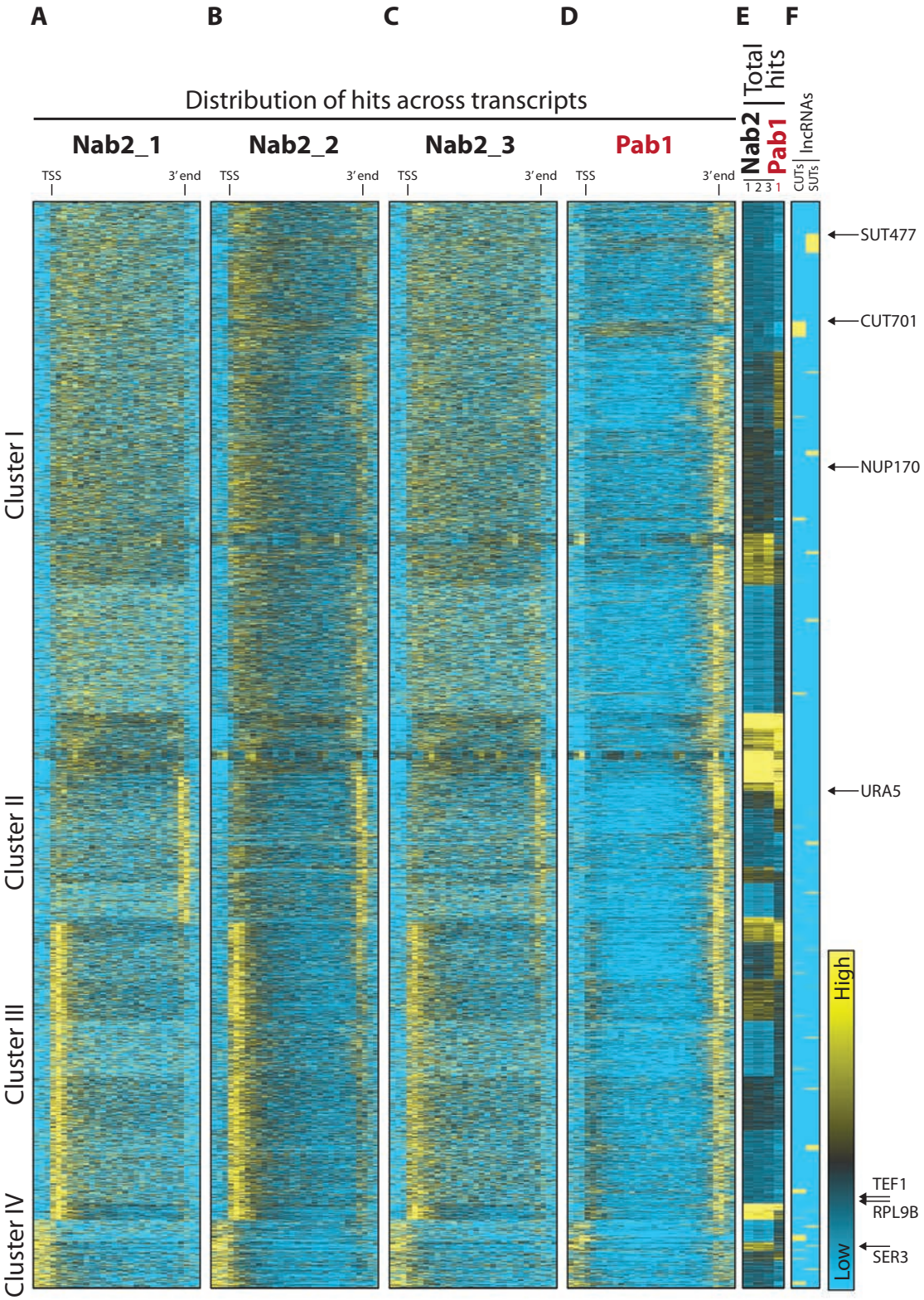


Figure 4.5: Nab2 binding distributions define distinct groups of transcripts. The Nab2 (three datasets) and Pab1 hit densities across 4267 mRNAs, CUTs and SUTs (with 100 nt flanking regions) are plotted in panels **A-D**, with each row representing one transcript, and the order of rows the same for each panel. Transcripts were clustered (**Clusters I-IV**) by k-medians analysis (k = 4) based on the hit distributions in the Nab2_2 dataset. Clusters I-IV were further subclustered by total Nab2 and Pab1 binding (panel **E**). Panel **F** indicates rows corresponding to CUTs and SUTs. Transcripts that feature in other figures are indicated (arrows).

When ranked by Nab2 binding distribution, transcripts fall into four highly reproducible groups (clusters I to IV), from each of which I have selected representative transcripts and plotted the binding distribution for all proteins tested (Figures 3.8, 3.9, 3.10 and 4.6). Cluster I is the largest, containing ~50 % of the transcripts included in this analysis, and is characterised by a distributed pattern of Nab2 binding (e.g. NUP170, Figure 4.6C; SUT477, Figure 3.10B; CUT701, Figure 4.6B). Notably, most CUTs and SUTs belong to this cluster. In contrast, cluster II contains transcripts with a prominent Nab2 peak at the 3' end (e.g. URA5, Figure 4.6A), and cluster III contains transcript with a prominent peak at the 5' end (e.g. TEF1 and RPL9B, Figure 3.8). Cluster III transcripts also exhibit a weak Pab1 peak a short distance downstream of the TSS, perhaps indicative of early transcription termination (discussed in Chapter 5). Finally, transcripts in cluster IV exhibit Nab2 binding upstream of the annotated TSS, which appears to predominantly arise from the presence of upstream, regulatory CUTs (e.g. *SRG1* at the *SER3* locus; Figure 3.9B). The Nab2 binding distribution thus reveals unusual transcription units and upstream CUTs, and suggests that the major role of Nab2 does not involve binding to canonical poly(A) tails. It appears likely that, like Hrp1, Nab2 performs distinct functions when bound at different positions on transcripts.

4.6 SUTs and mRNAs contain similar 3' end processing signals

Whereas the abundant, but distributed, binding of Nab2 can be used to identify non-canonical transcripts, the specific binding of Pab1 to poly(A) tails facilitates a high resolution analysis of 3' ends. Previous genome-wide analyses of transcript 3' ends (Ozsolak et al, 2010) have used oligo(dT) priming to select for A-tailed transcripts, but potentially contain false positives arising from priming at internal A tracts. Using the Pab1 CRAC dataset, I attempted to define a transcriptome-wide set of 3' ends by analysing Pab1-bound RNA fragments with non-encoded A tails (two or more adenosines) (Figure 4.7A). The requirement for Pab1 binding effectively acts as a second (biological) filter to reduce false

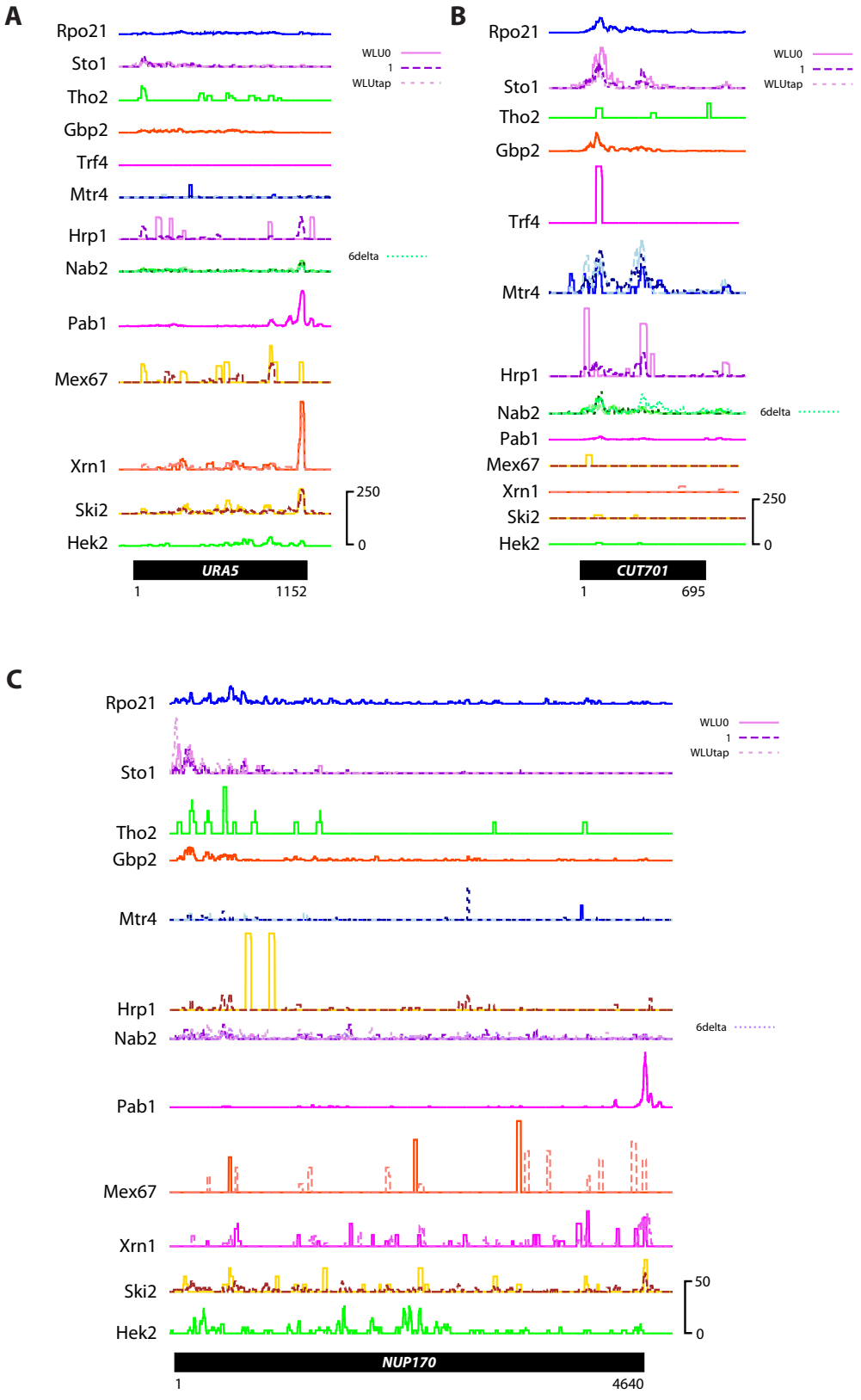


Figure 4.6: Binding profiles for representative transcripts from Figure 4.5. The distribution of CRAC hits mapping to **A** *URA5*, **B** *CUT701* and **C** *NUP170* is plotted for each of the 13 proteins tested (hits per million Pol II hits).

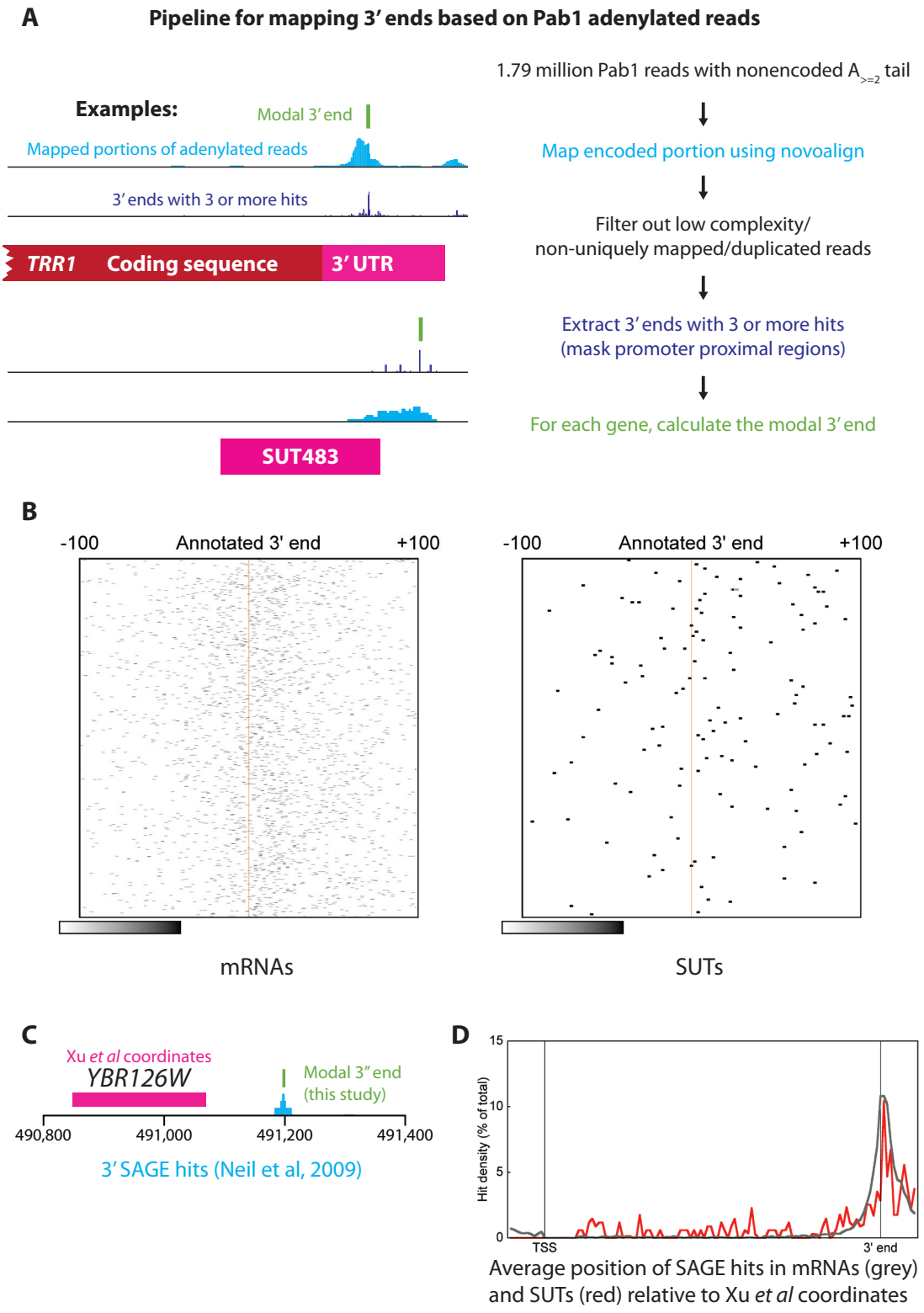


Figure 4.7: Identification of 3' ends based on non-encoded A-tails and Pab1 binding. A Pipeline for defining 3' ends, with two examples (TRR1 and SUT483). **B** Distribution of 3' ends defined in this study, aligned to previously annotated 3' end coordinates (Xu et al, 2009). **C** Comparison of 3' ends defined in this study, (Neil et al, 2009) and (Xu et al, 2009), for a representative mRNA. **D** Average distribution of 3' SAGE tags (Neil et al, 2009), aligned to previously annotated 3' end coordinates (Xu et al, 2009). This reveals that mRNA and SUT coordinates listed in (Xu et al, 2009) are similarly accurate.

positives, as previous studies have used only the presence of a non-encoded A tail to identify 3' ends.

Briefly, I used blast to identify the genome-encoded and non-encoded portions of each read, retained reads with two or more non-encoded As and mapped the encoded regions using novoalign. The analysis excluded (i) low complexity A-tailed reads that might artefactually map to genome-encoded A tracts, (ii) reads mapping to more than one position, (iii) PCR duplicates with identical start positions and identical, random 3-mer barcodes, and (iv) reads mapping near to annotated transcription start sites (which likely reflect the 3' ends of upstream, flanking transcripts). I then extracted the 3' end position for the remaining mapped reads, and for each gene identified the modal 3' end position (excluding positions with <3 hits) (Figure 4.7A). This identified 3' ends for 186 SUTs and 4932 mRNAs. The 3' ends of many additional SUTs could be obtained if a lower threshold is used, which is still likely to yield accurate results due to the extremely low background in this dataset.

To validate these 3' end positions, I compared their location with 3' ends annotated in (Xu et al, 2009) (Figure 4.7B) and (Neil et al, 2009) (Figure 4.7C), since both datasets contain 3' end coordinates for rare transcripts. Although there was a good general agreement between the 3' end coordinates obtained from Pab1 CRAC data and (Xu et al, 2009), the positions differed by up to 100 nt (Figure 4.7B), with similar results for SUTs and mRNAs. This is likely a result of the lower resolution of tiling arrays versus high-throughput sequencing. Indeed, I observed a better agreement of my data with the 3' SAGE data in (Neil et al, 2009), with many 3' ends agreeing to within a few nucleotides (for example, Figure 4.7C). Plotting the average 3' SAGE tag distribution (Neil et al, 2009) across transcription units listed in (Xu et al, 2009) (Figure 4.7D) revealed a similar match to that between my data and (Xu et al, 2009) (Figure 4.7B). Together, this suggests that the Pab1-derived 3' end positions offer a significant improvement in accuracy over those listed in (Xu et al, 2009), and can therefore provide greater insight into mRNA and SUT 3' end formation. However, the annotations in

(Xu et al, 2009) are very comprehensive, including ~1900 lncRNAs, and so remain the best choice for many other transcriptome-wide analyses.

The high-resolution Pab1-defined 3' end coordinates enabled me to address the question of whether SUT 3' ends resemble mRNA 3' ends in terms of nucleotide composition and the presence of specific motifs. The sequence elements surrounding mRNA polyadenylation sites are well characterised (Dichtl & Keller, 2001; Graber et al, 1999; Graber et al, 2002), and primarily include (i) the UAUAUA efficiency element (~50 nt upstream of the cleavage site), (ii) an A-rich positioning element (PE; AAUAAA or AAAAAA) (10-30 nt upstream of the cleavage site), (iii) two degenerate U-rich sequences (flanking the cleavage site), and (iv) the cleavage site itself, typically a U or C residue followed by one or more A residues.

I aligned the genomic sequence flanking the Pab1-determined mRNA and SUT 3' ends, and generated a sequence logo to reveal the information content and nucleotide bias at each position (Figure 4.8A, left). This reveals an AU-rich region upstream of the mRNA 3' ends, consistent with the literature. Strikingly, the pattern for SUTs was almost identical. Additionally, this revealed that the last encoded residue is almost always an A (~90-95 %), and is typically in a short A tract with 2-3 As (data not shown). I then searched for motifs flanking the 3' ends (100 nt either side) using DREME (Bailey, 2011), and this identified the UAUAUA sequence as most highly enriched for both mRNAs and SUTs (Figure 4.8A, right). I plotted the frequency of the UAUAUA motif around the 3' end positions, as well as that of the most frequent A-rich positioning element (AAUAAA), and obtained remarkably similar patterns for mRNAs and SUTs (Figure 4.8B). The position of these elements was in excellent agreement with the literature. I also plotted the individual nucleotide frequencies around these 3' end positions, and again, both mRNAs and SUTs were highly similar, with the well documented U-rich elements flanking the poly(A) site (Figure 4.8C) Together, therefore, Pab1-bound mRNA and SUT 3' ends are highly similar in terms of nucleotide bias and the presence of motifs.

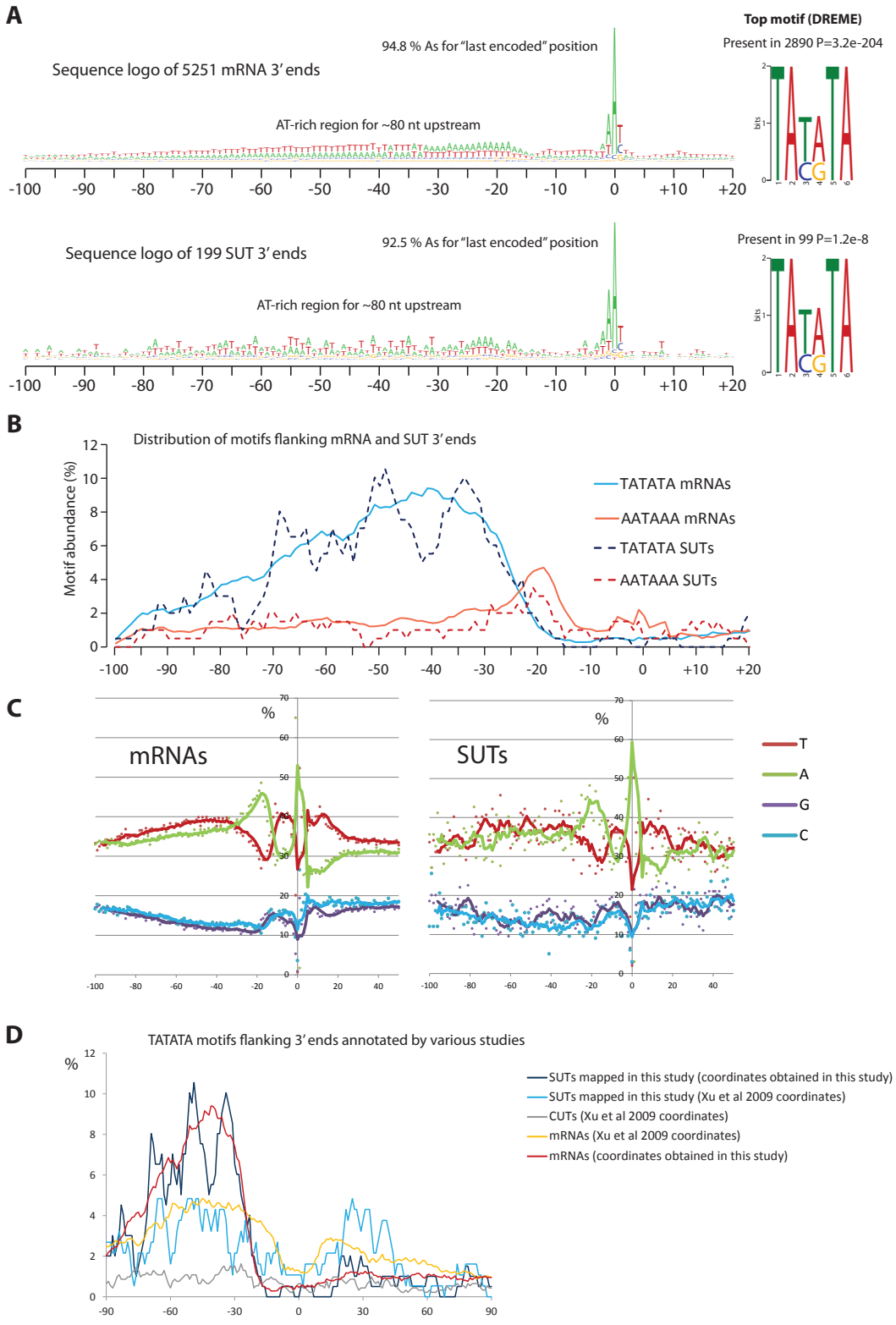


Figure 4.8: Nucleotide bias and motif enrichment flanking mRNA and SUT 3' end positions. A Sequence logos indicating the nucleotide bias in the region flanking mRNA (top) and SUT (bottom) 3' ends defined in this study. Distances, in nts, are relative to the 3' end position (at 0). Sequence logos were also generated for the most highly enriched motifs (right). **B** Frequency of TATATA and AATAAA motifs in the regions flanking mRNA and SUT 3' ends. **C** Nucleotide frequencies in the regions flanking mRNA and SUT 3' ends defined in this study. **D** Distribution of TATATA motifs around 3' ends defined by various studies, for mRNAs, CUTs and SUTs (as indicated).

However, I was only able to define 3' ends for 186 SUTs using the Pab1 CRAC data, and I could not obtain 3' end coordinates for CUTs as Pab1 did not bind their 3' ends. I therefore sought to determine whether the 3' end coordinates defined by (Xu et al, 2009) were sufficiently accurate to be used in an analysis including CUTs. I plotted the frequencies of UAUAUA motifs around mRNA, CUT and SUT 3' ends, using either the Pab1-derived coordinates from this study, or coordinates from (Xu et al, 2009). This revealed that although the UAUAUA enrichment upstream of the 3' end of mRNAs and SUTs was less obvious when using the (Xu et al, 2009) coordinates versus the Pab1-derived coordinates, it was still clearly visible. This indicates that the annotations in (Xu et al, 2009) are sufficiently accurate for motifs to be detected upstream of 3' ends. However, no UAUAUA enrichment was visible for CUTs (Figure 4.8D, grey), strongly suggesting that their 3' ends are not defined by the same motifs as mRNAs and SUTs. Finally, to investigate how the genome-wide patterns in Figure 4.8 translate into nucleotide and motif patterns at the level of individual genes, I randomly selected 50 mRNAs and SUTs and searched for UAUAUA and AAUAAA motifs in the sequence flanking their 3' ends (Pab1-determined coordinates) (Figure 4.9). This reveals that although these motifs are prevalent, many genes contain only one, or neither, and there is significant gene-to-gene variation in their arrangement.

4.7 Discussion

In this chapter I have investigated the potential canonical and non-canonical roles of proteins that were found to bind lncRNAs in Chapter 3, using a combination of biochemical and bioinformatic analyses. Comparison of the binding profiles of these proteins across mRNAs, CUTs and SUTs revealed both similarities and potentially important differences between these transcript classes.

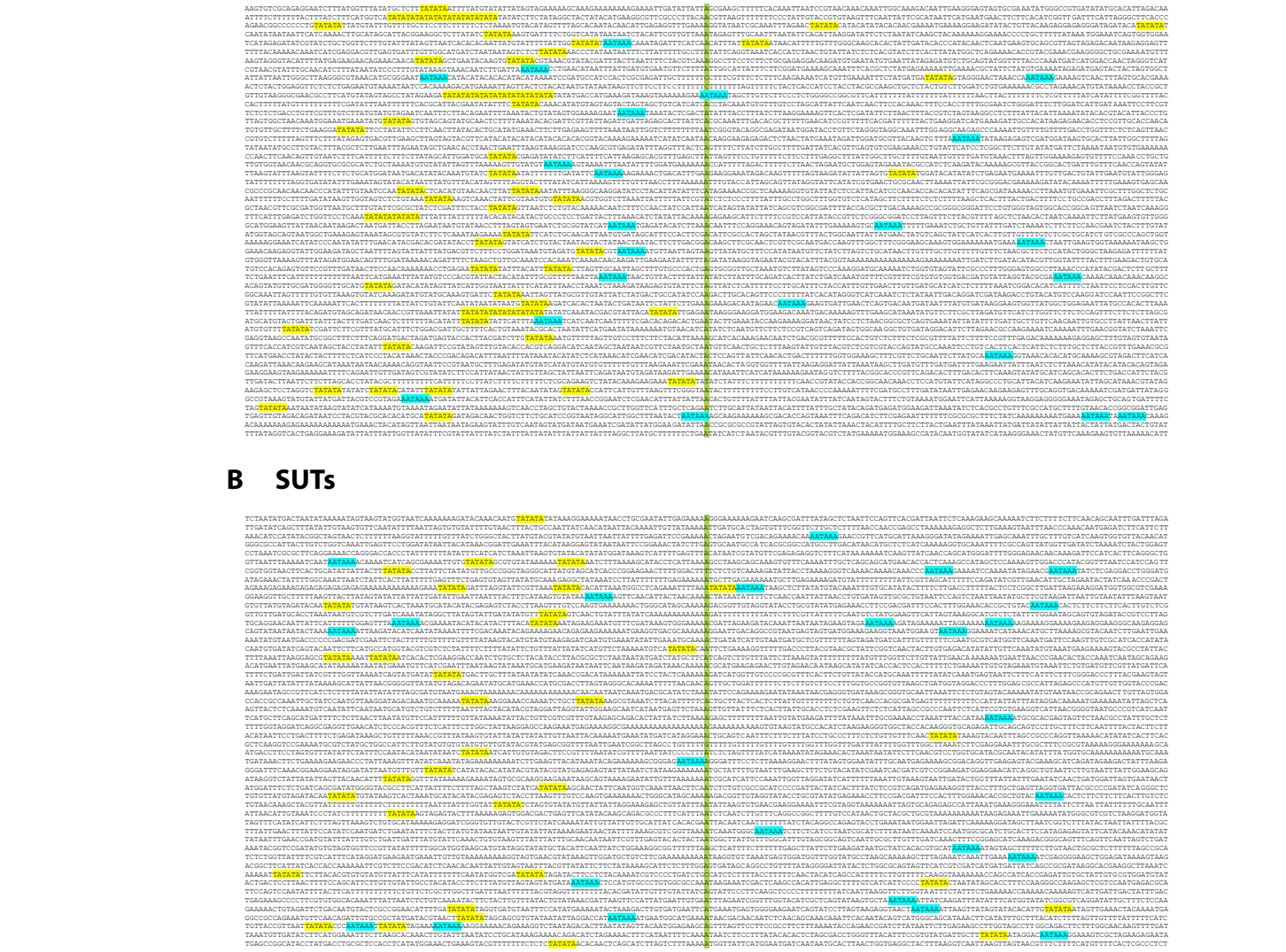


Figure 4.9: Examples of 3' motifs in 50 randomly selected mRNAs and SUTs. TATATA and AATAAA motifs in regions flanking the 3' ends of 50 randomly selected **A** mRNAs and **B** SUTs are highlighted in yellow (TATATA) and blue (AATAAA). The sequences are centred on the 3' end positions determined in this study (aligned, and highlighted in green).

Non-canonical roles of the nuclear cap binding complex

The nuclear cap binding complex component Sto1 bound to a broad region at the 5' end of mRNAs, CUTs and SUTs (Figures 3.8 and 4.1). CRAC is sensitive to even transient interactions, due to the high power UV crosslinking step performed *in vivo*, so the broad Sto1 peak might reflect weak or transient contacts made by Sto1 in addition to stronger binding to the cap. The sharp peaks detected for Pab1 suggest that the broader distribution of Sto1 is not a technical artefact. One possible explanation is that mRNPs and lncRNPs adopt a more globular fold at the 5' versus the 3' end. Indeed, mRNA 5' ends on average have a higher melting temperature than 3' ends, consistent with the 5' end being the more structured (Wan et al, 2012). Furthermore, Sto1 binding to the body of genes is reported to contribute to transcription initiation (Lahudkar et al, 2011) and to antagonise recruitment of Pcf11 and Rna15 to weak poly(A) sites (Das et al, 2000; Wong et al, 2007). Both of these functions are consistent with the observed Sto1 binding pattern, perhaps helping to mask binding sites on the nascent RNA.

Sto1 bound particularly abundantly to CUTs and SUTs, suggesting that it might perform some lncRNA-specific roles. Notably, many CUTs are terminated by the Nrd1-dependent pathway. Perhaps, via its ability to exclude canonical 3' end processing factors (Das et al, 2000; Wong et al, 2007) and interact with Nrd1 (Vasiljeva & Buratowski, 2006), Sto1 promotes Nrd1-dependent termination for CUTs. Sto1 also participates in nuclear RNA decay (Das et al, 2006; Das et al, 2000; Kuai et al, 2005), so binding to CUTs and SUTs might contribute to their turnover.

Nab2 is a component of diverse RNPs

Pab1 and Nab2 are both poly(A) binding proteins. However, whereas Pab1 bound specifically to poly(A) tails at the 3' end of mRNAs and SUTs, only a small number of mRNAs bound Nab2 at their 3' ends (Figure 4.5). Instead, Nab2 bound abundantly across

the full length of mRNAs, CUTs and SUTs, with a slight bias towards the 5' end (Figures 4.4 and 4.5). This is consistent with the reported binding of Nab2 across the entire transcribed region of Pol II genes (Gonzalez-Aguilera et al, 2011), and the formation of weak, non-specific interactions with RNA (Kelly et al, 2010; Viphakone et al, 2008). Nab2 binds most, if not all, Pol II transcripts (Batisse et al, 2009) (Figures 3.2 and 4.5), is expressed at high levels (Figure 3.1A), and crosslinks efficiently to RNA (Figure 3.1C). EM analyses of Nab2-containing mRNPs revealed an elongated ribbon structure, with each mRNA binding ~12 Nab2 molecules (Batisse et al, 2009), whereas Nab2 assembles into a compact structure on poly(A) tails (rather than simple linear deposition) (Viphakone et al, 2008). Together, these observations suggest that Nab2 is a core architectural component of mRNA-containing and lncRNA-containing RNPs, perhaps forming distinct structures on the body and tail regions. This model could explain the complex effects of various Nab2 mutations.

Nab2 is required for mRNA export (Hector et al, 2002) and this is dependent on the N-terminal domain, which interacts with Mlp1 and Gfd1 (Fasken et al, 2008; Grant et al, 2008; Green et al, 2003; Suntharalingam et al, 2004; Zheng et al, 2010), and the RGG domain, which interacts with Mex67 (Gonzalez-Aguilera et al, 2011; Green et al, 2002; Iglesias et al, 2010; Vinciguerra et al, 2005). In contrast, mutations in zinc fingers 5-7 of Nab2 that abolish poly(A) binding (Brockmann et al, 2012; Kelly et al, 2010; Kelly et al, 2007) have a negligible effect on export (Kelly et al, 2010; Tran et al, 2007). Together, this is consistent with a model whereby Nab2 participates in export by making non-specific contacts along the RNA, and acts as an adapter for export factors.

Although dispensable for export, poly(A) binding by the Nab2 zinc finger domain is required for poly(A) tail length control (Brockmann et al, 2012; Hector et al, 2002; Kelly et al, 2010; Viphakone et al, 2008). However, poly(A) tail length regulation does not correlate with the poly(A) binding affinity of various *nab2* Zn5-7 mutants (Brockmann et al, 2012). Aberrant

poly(A) tail length regulation in these mutants is better correlated with their ability to suppress the growth defects of *dbp5 (rat8-2)* and *yra1-8* strains, which suffer from defective mRNP remodelling (Brockmann et al, 2012; Qu et al, 2009; Tran et al, 2007; Vinciguerra et al, 2005). This suggests that besides poly(A) binding, Zn5-7 performs an additional role that is required for both poly(A) tail regulation and tight mRNP assembly. Perhaps Zn5-7 contributes to Nab2-dependent poly(A) packaging, acting together with other factors to restrict access to the 3' end and block PAN-dependent trimming or additional rounds of Pap1-dependent polyadenylation (Schmid et al, 2012; Viphakone et al, 2008).

Together, EM analyses, studies of mutants and the CRAC analyses reported here suggest that the predominant role of Nab2 is to bind throughout transcripts and assemble them into a compact fold, for export and poly(A) tail length control.

Nab2 functions in lncRNA metabolism

Although Nab2 bound abundantly to CUTs and SUTs, it failed to recruit Mex67 to these lncRNAs (Figure 3.6). Furthermore, Nab2 does not appear to contribute directly to the regulation of lncRNA poly(A) tail length, as it does not bind at lncRNA 3' ends (Figure 4.5). Nab2 might only function as an architectural component of CUT and SUT lncRNPs. Compact folding is likely to be important for all transcripts, which might otherwise make spurious interactions with other transcripts and/or DNA. This could in turn lead to defects including genome instability and, indeed, Nab2 mutations lead to hyper-recombination (Gallardo et al, 2003). Nab2 compaction of lncRNAs might be particularly important, as they are nuclear restricted and may have a reduced propensity to form secondary structures. This role in compaction might also extend to nascent Pol I and Pol III transcripts, to which Nab2 was also localised by CRAC (Figures 3.2 and 3.4) and ChIP (Gonzalez-Aguilera et al, 2011). Notably, Nab2 apparently plays a sufficiently important role to drive the evolution of a specialised import system (Aitchison et al, 1996; Lange et al, 2008; Truant et al, 1998).

In addition to acting as a general architectural component for CUT and SUT lncRNPs, Nab2 might contribute to their surveillance, via interactions with Rrp6, Trf4 and Mlps (Green et al, 2003; Roth et al, 2009; Roth et al, 2005; Schmid et al, 2012). Although a recent study did not find any changes in lncRNA expression following a seven hour tetracycline-mediated Nab2 depletion, this might reflect redundancy in lncRNA surveillance pathways (Schmid et al, 2012). Another study found more widespread changes in gene expression when Nab2 was depleted using a temperature-sensitive degron strain, but did not examine the levels of SUTs (Gonzalez-Aguilera et al, 2011). I have generated a glucose-repressible *pGAL::NAB2* strain in which Nab2 is depleted by > 90 % within ~1.5 hours of shifting to glucose, and am currently testing for any effects on the steady state abundance of SUTs. This should reveal whether Nab2 plays a specific role in lncRNA surveillance, in addition to its general contribution to RNP architecture.

SUTs diverge from mRNAs during 3' end remodelling

The analyses in this chapter have also provided insight into the timing of the divergence between lncRNAs and mRNAs during RNP maturation. SUTs resembled mRNAs in several analyses, with (i) Sto1 bound to their 5' ends (Figure 4.1A), (ii) TREX components bound throughout, consistent with co-transcriptional binding (Figure 4.1C), (iii) Pab1 bound to their 3' ends (Figure 4.4), (iv) Hrp1 bound, albeit less abundantly, to AUAUAU motifs towards their 3' ends (Figure 4.2B), and (v) a canonical configuration of sequence elements at their 3' ends, including upstream UAUAUA and AAUAAA motifs and U-rich regions flanking the cleavage site (Figure 4.8). This is consistent with a previous study that detected similar motifs in the regions flanking the 3' ends of mRNAs and poly(A)+ antisense and intergenic lncRNAs (Ozsolak et al, 2010). Together, this suggests that mRNAs and SUTs undergo many of the same events up to, and perhaps including, 3' end processing. However, whereas Nab2 binding was clearly enriched at the 3' ends of mRNAs (Figure 4.4), binding to the 3' ends of SUTs was barely discernible (Figure 4.4). This suggests that, despite the apparent

similarities, polyadenylation of SUTs is functionally distinct from that of mRNAs. Furthermore, unlike mRNAs, SUTs do not recruit the export receptor Mex67, or interact with cytoplasmic surveillance factors (Figure 3.6). Together, this suggests that SUTs and mRNAs diverge during 3' end processing. Perhaps the distinction between mRNAs and SUTs is made during the Sub2/Yra1-dependent remodelling events coincident with the completion of 3' end cleavage and polyadenylation, after which Mex67 is normally recruited and Nab2 might interact with the poly(A) tail to prevent hyperadenylation.

The molecular events that lead to these differences in processing of SUTs and mRNAs are not immediately apparent. The lower level of Hrp1 detected at SUT 3' ends might indicate that SUTs lack the full complement of 3' processing factors, or they are not canonically bound. Another possibility is that proteins bound to SUTs lack particular post-translational modifications, many of which are associated with 3' end processing (e.g. Nab2 phosphorylation or methylation (Carmody et al, 2010; Green et al, 2002)). This could be tested by examining kinase and/or methyltransferase mutants. Furthermore, by combining the Pab1-determined 3' ends with other datasets (e.g. (Ozsolak et al, 2010) and (Neil et al, 2009)), I hope to obtain a more comprehensive list of accurate 3' ends that can be used to investigate the differences between mRNAs, CUTs and SUTs (in this study, 3' ends were only defined for ~20 % of SUTs).

CUTs diverge from mRNAs during early transcription

Like SUTs, CUTs were bound by Sto1, Gbp2, Tho2 (Figure 4.1) and Nab2, with Nab2 binding across CUTs but not specifically at their 3' ends (particularly apparent in the Nab2 *rrp6Δ* dataset) (Figure 4.4). However, unlike SUTs, CUTs did not bind Pab1 (Figure 4.4C) or Hrp1 (Figure 4.2) at their 3' ends, and were not associated with any of the tested cleavage and polyadenylation motifs (Figure 4.8D). This suggests that CUTs diverge from mRNAs at an earlier stage than SUTs, and that although CUTs and SUTs share early events in transcription (TREX, Hrp1 and Nab2 binding) they are terminated by distinct mechanisms.

Numerous CUTs are reported to terminate by the Nrd1-dependent pathway (Arigo et al, 2006b; Creamer et al, 2011; Kim et al, 2010a; Marquardt et al, 2011; Thiebaut et al, 2006; Vasiljeva et al, 2008b; Wlotzka et al, 2011; Wyers et al, 2005), and my results would be consistent with this being a general property of most, if not all, CUTs.

The stark difference between mRNAs and CUTs is further illustrated by the different roles that Hrp1 plays in their biogenesis and turnover. When bound to mRNAs, Hrp1 enhances cleavage and polyadenylation or diverts it from cryptic to major sites (Bucheli et al, 2007; Kim Guisbert et al, 2006; Minvielle-Sebastia et al, 1998). However, depletion of Hrp1 (Figure 4.3) increased the abundance of CUTs with little, if any, decrease in alternative poly(A) site usage, suggesting that Hrp1 acts to reduce the total transcriptional output of CUT loci. Notably, Hrp1 is implicated in Nrd1-dependent termination (Kim et al, 2006; Kuehner & Brow, 2008; Steinmetz et al, 2006b), suggesting that these factors might function in the same pathway. However, Hrp1 depletion apparently stabilised CUTs, rather than resulting in transcription read-through (Figure 4.3), and a tested artificial CUT was not dependent on Hrp1 for termination (Porrua et al, 2012). This suggests that the role of Hrp1, when bound to CUTs, might be to recruit the surveillance machinery following Nrd1-dependent termination (Wlotzka et al, 2011). Hrp1 thus plays a very different role in CUT metabolism than in mRNA 3' end processing.

In conclusion, the binding profiles of Nab2, Hrp1 and Pab1 across mRNAs, CUTs and SUTs reveal that CUTs and mRNAs diverge early in transcription, whereas SUTs and mRNAs share many features of 3' end processing and apparently diverge during a remodelling event in which mRNPs become export-competent, but SUT lncRNPs remain nuclear restricted and are committed to degradation.

5: Promoter-proximal transcription termination within protein-coding genes

5.1 Introduction

In the previous two chapters, I have presented evidence suggesting that the 3' end processing of SUTs bears some similarity to that of mRNAs, but in contract to mRNAs, CUTs and SUTs are largely retained and degraded in the nucleus. These conclusions are based on analyses of the overall level of binding of cytoplasmic versus nuclear factors to CUTs, SUTs and mRNAs, described in Chapter 3. However, inspection of the distribution of hits across the length of transcripts is also informative, and in Chapter 4 I discussed how this offers insights into non-canonical roles of mRNA binding proteins in lncRNA (and general RNA) metabolism. For these analyses, I used a published set of non-overlapping mRNA, CUT and SUT annotations, obtained by dividing the genome into segments based upon transcriptome tiling array data (Huber et al, 2006; Xu et al, 2009). However, this segmented view of the genome is an oversimplification as transcripts often overlap, and interleaved transcription units are prevalent throughout eukaryotic genomes. Notably, the high resolution of CRAC analyses can help to reveal which particular transcript, within a group of overlapping transcripts, a protein binds. Indeed, plotting the distribution of Nab2 hits across mRNAs, CUTs and SUTs (Figure 4.5) successfully identified known examples of sense-oriented CUTs that overlap mRNA 5' ends (Figure 4.5, group IV). In this chapter, I extended this approach to other RNA binding proteins, to investigate more fully the prevalence of lncRNAs overlapping mRNAs. This provided insight into the roles of these proteins in mRNA metabolism, but also identified an abundant class of promoter-proximal mRNA fragments with distinct properties from full-length mRNAs, which I argue are produced by early termination. This is consistent with a growing body of evidence for sense-oriented lncRNAs overlapping protein coding genes.

Previously characterised sense-oriented lncRNAs overlapping mRNAs

Many studies have focused on lncRNAs transcribed from intergenic regions or antisense to protein coding genes, as their lack of overlap with mRNAs makes them easy to distinguish and suggests they perform distinct functions. However, there are also reports of lncRNAs that overlap mRNAs in the sense direction, with the number of characterised examples rapidly increasing. These can be divided into three categories, based on their mode of transcription initiation (for a full list of examples with references, see Table 5.1). Firstly, a small number of genes have been characterised where premature transcription termination generates a transcript with the same 5' end as the full length mRNA, but a truncated 3' end. The production of these short isoforms typically downregulates expression, as these transcripts do not encode a functional protein (Mayer & Dieckmann, 1989; Sparks et al, 1997) and are rapidly degraded (Steinmetz et al, 2001). Notably, this is used as an autoregulatory mechanism for some termination factors (Arigo et al, 2006a; Creamer et al, 2011; Kuehner & Brow, 2008; Steinmetz et al, 2001). This early termination is typically mediated by the Sen1-Nrd1-Nab3 complex (Arigo et al, 2006a; Creamer et al, 2011; Darby et al, 2012; Kim & Levin, 2011; Kuehner & Brow, 2008; Steinmetz et al, 2001; Wlotzka et al, 2011), and is therefore distinct from the phenomenon of alternative poly(A) site selection by the canonical 3' end processing machinery, which is prevalent in yeast and typically occurs further downstream (Ozsolak et al, 2010).

A second class of sense-oriented lncRNAs arises from genes with a single promoter but multiple TSSs, and is epitomised by *IMD2* (Davis & Ares, 2006; Jenks et al, 2008; Kuehner & Brow, 2008) and *URA2* (Thiebaut et al, 2008). Here, PICs assembled at the promoter can initiate transcription from either a promoter-proximal upstream TSS, to generate a short transcript that is rapidly degraded, or from a promoter-distal TSS, to generate a full-length mRNA. The non-productive promoter-proximal transcription is terminated by the Nrd1 pathway in close proximity to the downstream TSS, and contributes to repression (Jenks et

al, 2008; Kuehner & Brow, 2008; Thiebaut et al, 2008). Inspection of transcriptome tiling array and Nrd1 binding data revealed a number of other genes where a similar mechanism is likely to occur (Creamer et al, 2011; Davis & Ares, 2006; Kim et al, 2010a; Thiebaut et al, 2008).

Thirdly, lncRNAs can initiate from independent promoters upstream of protein-coding genes, either producing short upstream CUTs that terminate in close proximity to the mRNA TSS, or longer chimeric CUTs that terminate at the canonical poly(A) site of the downstream gene. Notable examples include the *SRG1* CUT, upstream of *SER3* (Martens et al, 2004), and CUTs produced at subtelomeric metal ion homeostasis genes (Toesca et al, 2011). Transcription from an independent upstream promoter can regulate the downstream mRNA promoter, via mechanisms including ejection of transcription factors (Bird et al, 2006; Bumgarner et al, 2009; Bumgarner et al, 2012; Toesca et al, 2011), deposition of nucleosomes (Hainer et al, 2011), and histone deacetylation (Houseley et al, 2008; Kim et al, 2012; Pinskaya et al, 2009; van Werven et al, 2012). Thus lncRNAs produced from independent promoters might have the greatest capacity for regulation.

In addition to these three classes of lncRNA that initiate at, or upstream of, mRNA transcription start sites, some sense-oriented lncRNAs initiate from positions within the CDS (despite the many mechanisms that suppress cryptic initiation). For example, a sense-oriented intragenic lncRNA transcribed from the *ASP3* locus promotes H3K4 trimethylation that favours expression of the mRNA (Huang et al, 2010).

LncRNAs overlapping mRNAs might be prevalent

Overall, approximately 30 sense-oriented lncRNAs overlapping protein coding genes in yeast have been characterised in detail (Table 5.1), and they are suggested to perform a variety of regulatory roles. However, there is evidence to suggest that these 30 lncRNAs are part of a more widespread phenomenon of promoter-proximal non-coding transcription, which reflects a post-initiation regulatory step in the general Pol II transcription cycle.

The genome-wide distribution of Pol II has been mapped by a variety of methods, with many studies detecting more Pol II in the promoter-proximal region of genes than further downstream. Several of these studies employed methods based on transcriptional run on, which only detects Pol II that has initiated and is able to actively elongate (McKinlay et al, 2011; Pelechano et al, 2009; Rodriguez-Gil et al, 2010). The authors therefore propose that the 5' enrichment of Pol II represents temporary transcriptional pausing, with the polymerases typically resuming transcription. The nascent transcripts associated with these stalled polymerases are susceptible to cleavage by the anti-termination factor TFIIS (Dst1), which might trigger release from the pause (Churchman & Weissman, 2011). Further evidence for a 5' enrichment of Pol II in some genes comes from Pol II ChIP (Kim et al, 2011; Pelechano et al, 2009) and sequencing of chromatin-associated transcripts (Carrillo Oesterreich et al, 2010; Churchman & Weissman, 2011). The proportion of genes subject to pausing, and the positions of the pauses detected, varies between studies, but highly-transcribed intron-containing genes are consistently detected as having a high 5' bias (Brannan et al, 2012; Creamer et al, 2011; Kim et al, 2011).

Although these analyses reveal that paused Pol II can resume transcription under run on conditions, this does not necessarily imply that this is always the case *in vivo*. It is also possible that promoter-proximal Pol II pausing could result in termination. Indeed, pausing is thought to contribute to termination coupled to cleavage and polyadenylation at the 3' end of genes. In yeast, high-throughput sequencing of polyadenylated transcript 3' ends detected ~15 % within coding regions (Ozsolak et al, 2010; Yoon & Brem, 2010), and a 3' SAGE analysis of unstable transcripts in *rrp6Δ trf4-depl* mutants detected many promoter-proximal 3' ends (Neil et al, 2009). In humans, abundant sense-oriented promoter-associated non-coding RNAs have been detected, up to ~1 kb long (Flynn et al, 2011; Kanhere et al, 2010; Kapranov et al, 2007; Seila et al, 2008; Taft et al, 2011). Furthermore, the termination factors Nrd1 (Creamer et al, 2011; Wlotzka et al, 2011) and Xrn2 (Brannan et al, 2012) (the

human homologue of Rat1) bind towards the 5' end of mRNAs, and in human cells, depletion of Xrn2 relieves the 5' Pol II bias at many genes (Brannan et al, 2012). Together, these observations suggest that some polymerase pausing can be resolved by termination.

In this chapter, I sought to identify lncRNAs overlapping with protein coding genes, and test the hypothesis that early termination is prevalent in yeast. I present evidence that truncated 5' mRNA fragments are abundantly bound to early mRNA binding proteins, coincide with the localisation of termination factors, are retained in the nucleus, and are degraded by the nuclear surveillance machinery.

5.2 Canonical functions of mRNA binding proteins

I first examined the average binding profile across mRNAs for the 13 factors tested in this study and Rpo21 (Figure 5.1), using the algorithm described for Figure 4.1. For factors that bind co-transcriptionally, such as Rpo21, Tho2 and Gbp2, one must bear in mind that the binding profiles are influenced by the overrepresentation of 5' ends within the pool of nascent transcripts. Notably, where replicate datasets were analysed, binding profiles were highly reproducible, and comparing different proteins revealed clear differences (Figure 5.1).

Considering the nuclear factors, Rpo21 binding was enriched towards the 5' end of genes (Figure 5.1), consistent with numerous previous analyses of Pol II distribution (Carrillo Oesterreich et al, 2010; Churchman & Weissman, 2011; Kim et al, 2011; McKinlay et al, 2011; Pelechano et al, 2009; Rodriguez-Gil et al, 2010). Tho2 has a similar distribution, whereas Gbp2 is more enriched at the 5' end. The greater degree of 5' enrichment for Gbp2 versus Tho2 might reflect different stoichiometries of binding to nascent mRNAs, with more molecules of Gbp2 bound per transcript than Tho2. Sto1 bound almost exclusively towards mRNA 5' ends, and the inclusion of an enzymatic decapping step facilitated the detection of interactions at the extreme 5' end (Figure 5.1). Conversely, Pab1 bound exclusively at mRNA 3' ends, and Nab2 gave a minor peak at 3' ends. The surveillance factors Trf4 and

Table 5.1: Documented examples of non-coding RNAs overlapping mRNAs in the sense orientation. Non-coding transcripts associated with the indicated protein coding genes (left column) are organised into groups with shared transcription architectures.

Early termination: single TSS, but different termination sites		
NRD1	Nrd1-dependent early termination	(Arigo et al, 2006a; Kim et al, 2010a; Kuehner & Brow, 2008; Steinmetz et al, 2001)
HRP1	Early termination dependent on Hrp1, Nrd1, Nab3, Ssu72	(Kim et al, 2010a; Kuehner & Brow, 2008)
CBP1	Produces a 2.2 kb and truncated 1.3 kb isoform	(Mayer & Dieckmann, 1989; Sparks et al, 1997)
CLN3	Premature termination in response to nutrient depletion; Nrd1-dependent	(Darby et al, 2012)
PCF11	Nrd1-dependent early termination	(Creamer et al, 2011)
RPB10	Nrd1-dependent early termination	(Creamer et al, 2011)
FKS2	Sen1-dependent termination inhibited by Mpk1-Paf1C interaction upon stress	(Kim & Levin, 2011)
Alternative TSS selection: single promoter (TATA box and PIC), but transcription initiates at different TSSs		
IMD2	LncRNA/mRNA expression anticorrelated, with competition between TSSs; Nrd1-dependent termination	(Davis & Ares, 2006; Jenks et al, 2008; Kuehner & Brow, 2008; Loya et al, 2012)
IMD3		(Thiebaut et al, 2008)
URA2	Constant negative repression; CUT ends downstream of mRNA TSS	(Thiebaut et al, 2008)
ADE12	Nrd1-dependent termination	(Kim et al, 2010a; Thiebaut et al, 2008)
URA8	Nrd1-dependent termination	(Kim et al, 2010a; Thiebaut et al, 2008)
LEU4		(Davis & Ares, 2006)
HPT1, GUA1, ADE17	Presumed to involve alternative TSS selection because the alternative TSSs flank Nrd1-binding sites	(Creamer et al, 2011)
Distinct upstream promoter gives rise to CUT, downstream promoter to mRNA		
ZRR1	Upstream CUT displaces Rap1 from the *ADH1* promoter; short CUT terminates upstream of TSS, whereas a longer CUT terminates at the mRNA 3' end	(Bird et al, 2006)
ZRT1	CD-CUT displaces Rap1 from the *ZRT1* promoter and overlaps entire ORF; TSS at -2kb	(Toesca et al, 2011)
FIT3	CD-CUT displaces Aft1 from the *FIT3* promoter and overlaps entire ORF	(Toesca et al, 2011)
IME1	CUT terminates within promoter, upstream of *IME1* TSS	(van Werven et al, 2012)
DCI1	CUT extends across whole ORF	(Kim et al, 2012)
FLO11	*ICR1* CUT ejects activators in the *FLO11* promoter; originates ~3.3 kb upstream, and terminates at various sites flanking the mRNA start codon	(Bumgarner et al, 2009; Bumgarner et al, 2012)
TPI1	CUT overlaps the mRNA TSS (starts ~150 nt upstream and ends ~150 nt downstream)	(Neil et al, 2009)
SER3	SRG1L/SRG1S and SRG1-SER3 chimera; *SRG1* 3' ends are ~50 nt either side of *SER3* TSS	(Martens et al, 2005; Thiebaut et al, 2006; Thompson & Parker, 2007)
GAL1	Antisense *GAL10* lncRNA transcribed across, and represses, *GAL1* promoter	(Houseley et al, 2008; Pinskaya et al, 2009)
Miscellaneous/mechanism unknown		
GPM1	TS-CUT – unsure of TSS arrangement	(Neil et al, 2009)
FBA1	TS-CUT – unsure of TSS arrangement	(Neil et al, 2009)
ASP3	Completely intragenic, in sense direction	(Huang et al, 2010)

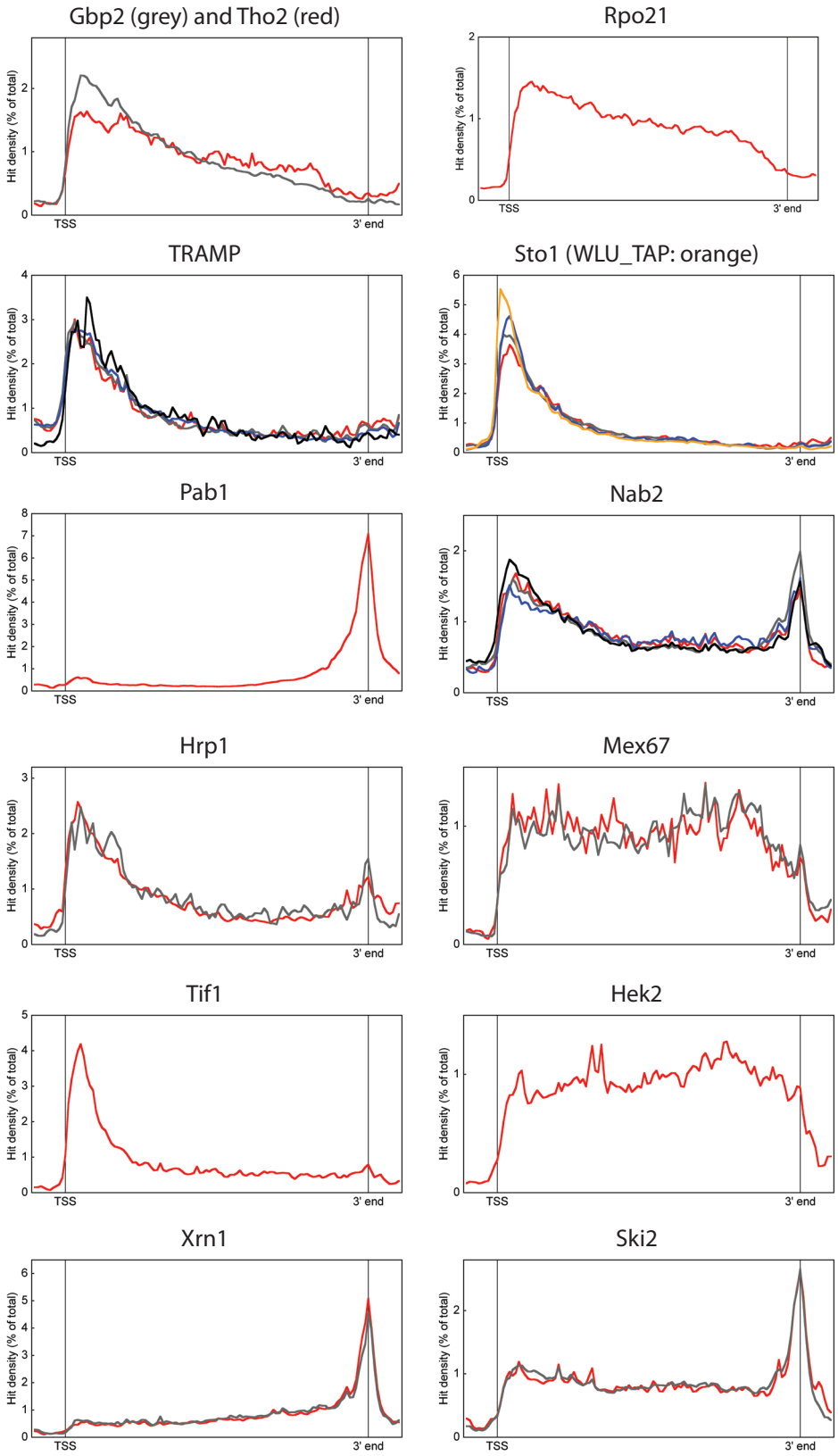


Figure 5.1: Average binding profiles across mRNAs. For the indicated proteins, the average hit density across the 500 most abundantly bound mRNAs are plotted. Replicate experiments are represented by different coloured lines. The "TRAMP" plot contains data from TRAMP complex members Trf4 (one dataset) and Mtr4 (three replicates).

Mtr4 were highly enriched towards the 5' ends of mRNAs, whereas Hrp1 and Nab2 exhibited less prominent 5' peaks. The 5' binding bias of Trf4, Mtr4, Hrp1 and Nab2 might reflect their binding to nascent transcripts, degradation intermediates and/or products of premature termination.

I next examined factors with a more cytoplasmic localisation. Notably, the export receptor Mex67 and shuttling protein Hek2 were not biased towards either end of mRNAs. Furthermore, although their average binding profiles were similar, on many individual transcripts Mex67 bound throughout whereas Hek2 bound at specific positions (data not shown). Tif1 (eIF4A), a translation initiation factor that participates in scanning for the start codon, bound towards the 5' end of mRNAs. The cytoplasmic decay factors Xrn1 and Ski2 both exhibited prominent 3' peaks on mRNAs, consistent with the recruitment of both being regulated by the poly(A) tail. However, whereas Ski2 hits otherwise mapped evenly across protein coding genes, Xrn1 hits were depleted towards the 5' end. This might indicate that Xrn1 activity is slowed towards the 3' end of transcripts, perhaps by structural elements or ribosomes, whereas Ski2, which is a cofactor for the cytoplasmic exosome, apparently progresses at a consistent rate.

Together, these analyses provide a detailed picture of a "typical" mRNP, in which different regions of the transcript are bound by a characteristic set of proteins. Notably, the majority of individual transcripts included in these average plots each exhibit a binding profile similar to the overall profile. This suggests that mRNPs largely have a common, "signature" organisation, and this arrangement of factors is likely to be critical for correct mRNA biogenesis and function, and the recognition of properly assembled mRNPs.

Binding to spliced versus unspliced transcripts reveals when proteins bind mRNAs

In addition to examining the spatial distribution of mRNA binding proteins across transcripts, I also sought to determine when in the mRNA lifecycle they bind. During or shortly after transcription, splicing removes introns from mRNAs. The relative abundance of spliced versus unspliced (intron-containing) mRNAs within CRAC datasets can therefore reveal when, in relation to splicing, proteins bind. I performed two complementary analyses (Schneider et al, 2012), both of which were restricted to hits in intron-containing genes. In the first (Figure 5.2A), I calculated the proportion of hits for each protein that mapped to introns, relative to the total mapped reads. In the second (Figure 5.2B), I calculated the ratio between hits across intron-exon (IE) boundaries (present only in unspliced mRNAs) and hits across exon-exon (EE) boundaries (only present in spliced mRNAs). Each analysis has biases, so the two should be considered together. For example, in the first analysis, a low proportion of hits in introns does not necessarily mean that a protein binds only after splicing, as it could instead reflect binding exclusively to exons within unspliced pre-mRNAs. In the second analysis, by considering only the subset of hits that map to IE or EE junctions, this bias is overcome. However, this might be biased against proteins that bind the central region of introns, and not IE junctions.

Despite these caveats, both analyses gave similar results (Figure 5.2). Considering the IE versus EE junction analysis (Figure 5.2B), the RNA polymerase subunit Rpo21 serves as a good control, as nascent transcripts interact with the polymerase before any other factors. Indeed, for Rpo21, IE junctions are highly enriched (14.3-fold) over EE junctions. In comparison, IE junctions in Mtr4, Trf4, Sto1, Hrp1 and Gbp2 datasets are 1.8- to 3.6-fold enriched over EE junctions. This suggests that these factors bind before splicing, and are displaced shortly after, consistent with their roles in nuclear surveillance and pre-mRNA processing. Conversely, Ski2 and Xrn1 are enriched for EE junctions (10- and 5-fold

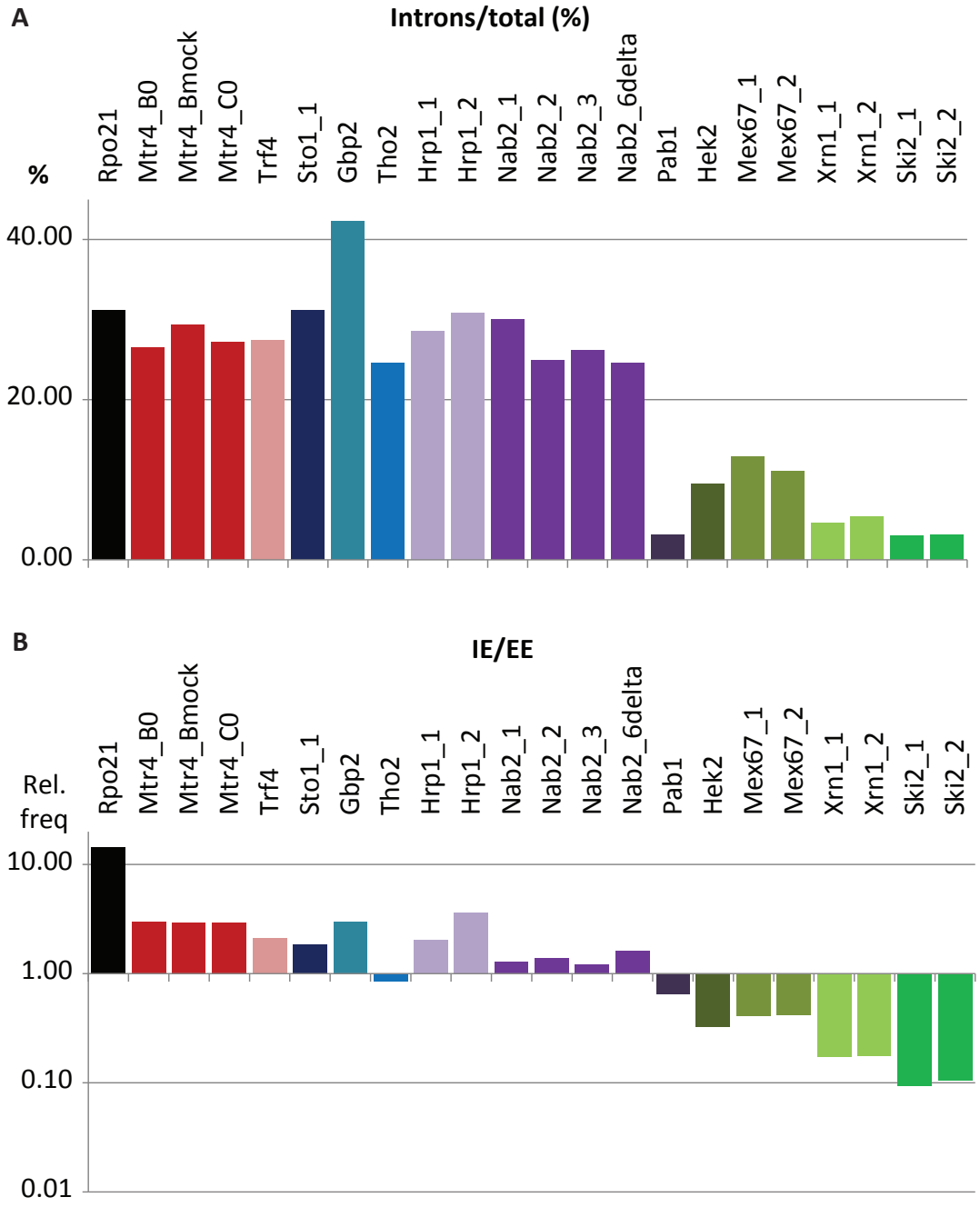


Figure 5.2: Analyses of hits in unspliced versus spliced mRNAs. A Propotion of hits in introns, considering all hits that map to intron-containing genes. **B** Relative frequency of hits that map to intron-exon junctions versus exon-exon junctions.

respectively), consistent with their cytoplasmic roles. This bias is reproducibly less strong for Xrn1 than for Ski2, consistent with 5' degradation playing a more important role than 3' degradation in the cytoplasmic turnover of aberrantly exported unspliced or partially spliced pre-mRNAs (Harigaya & Parker, 2012; Hilleren & Parker, 2003; Sayani et al, 2008). The export receptor Mex67 is also enriched for EE junctions, suggesting that it predominantly associates with spliced mRNAs, perhaps indicative of coupling between the completion of splicing and export competency. Pab1 is slightly enriched for binding to EE junctions, whereas Nab2 slightly favours IE junctions, suggesting that Nab2 generally functions at an earlier stage in mRNA biogenesis than Pab1. The bias of Nab2, but not Mex67, towards unspliced IE junctions indicates that Nab2 is recruited to transcripts before Mex67, since both Nab2 and Mex67 dissociate from the mRNP at a similar time (Lund & Guthrie, 2005; Tran et al, 2007). This is consistent with a model in which Nab2 binding alone is not sufficient for Mex67 recruitment, but requires a subsequent remodelling step (Chapter 4).

Tho2 appears slightly anomalous, with a modest (1.17-fold) bias towards EE junctions, which is perhaps inconsistent with its early role in mRNA biogenesis. However, when total hits in introns are considered (Figure 5.2A), Tho2 appears similar to other early mRNA biogenesis factors such as Gbp2 and Sto1, binding abundantly to introns. Inspection of the distribution of Tho2 and Gbp2 hits across individual intron-containing mRNAs suggest that Tho2, but not Gbp2, is specifically excluded from some intron-exon junctions, perhaps via competition from *bona fide* splicing factors, but otherwise binds abundantly across introns (data not shown). Notably, Gbp2 contains SR domains present in many splicing factors, and Hrp1 functionally interacts with the splicing-associated factor Npl3 (Bucheli et al, 2007). Together, this suggests that Hrp1, Gbp2 and Tho2 all bind mRNAs at a similar time relative to splicing, but Gbp2 and Hrp1 might play more specific roles in regulating splicing and thus be enriched at splice sites (intron-exon junctions), whereas Tho2 binds central regions of introns that may be less important for splicing regulation. For all other factors tested, the

proportion of hits in introns (Figure 5.2A) is consistent with the IE versus EE junction analysis (Figure 5.2B).

In conclusion, the analyses of hits in spliced versus unspliced mRNAs reveal the temporal order of mRNA:protein interactions, with early pre-mRNA interactions by TRAMP, Sto1, Gbp2, Tho2, Hrp1 and Nab2, whereas and Pab1, Mex67, Ski2 and Xrn1 largely act at later steps.

5.3 Non-encoded A-tails distinguish stable versus unstable 3' ends

The observation that TRAMP, Hrp1, Nab2, Gbp2 and (to a lesser extent) Tho2 show a 5' bias in mRNA binding, suggests that promoter-proximal regions are more actively transcribed than downstream regions. This is consistent with the documented enrichment of Pol II at the 5' end of genes, and suggests that either (i) there are abundant stalled transcription elongation complexes, or (ii) promoter-proximal transcription termination occurs at many genes, releasing 5' mRNA fragments. The high enrichment of TRAMP towards the 5' end of mRNAs (Figure 5.1) indicates that the nuclear surveillance machinery is active on these promoter-proximal transcripts. This supports a model in which the 5' mRNA fragments have undergone termination and possess 3' ends accessible to the TRAMP and exosome complexes.

Terminated versus paused transcripts can potentially be distinguished via analysis of their 3' ends. In the yeast nucleus, three poly(A) polymerases can adenylate transcript 3' ends. Pap1, the canonical poly(A) polymerase, generates A-tails that are ~60-80 nt long, whereas the non-canonical poly(A) polymerases Trf4 and Trf5 preferentially generate shorter (~1-5 nt) A-tails (Jia et al, 2011; LaCava et al, 2005; Vaňáčová et al, 2005; Wlotzka et al, 2011; Wyers et al, 2005). Generally, long A-tails are associated with stable, export competent transcripts, whereas short oligo(A) tails are hallmarks of nuclear surveillance targets. Neither

type of A-tail is genome-encoded, so they can be identified in CRAC datasets as unmapped tracts of As at the 3' end of otherwise confidently mapped reads. The RNA fragmentation step in the CRAC protocol preserves the length of A-tails, as RNase A and T1 only rarely cut after A residues. In my analyses < 4 % of cuts occurred after an A, probably reflecting *in vivo* generated 3' ends. Moreover, deadenylation of poly(A) tails usefully proceeds to around A_{10-12} followed by rapid degradation, so poly(A)+ transcripts are not expected to make a major contribution to the short A-tail (A_{1-5}) population. I therefore examined the CRAC datasets for reads with non-encoded A-tails (2 or more non-encoded As), reasoning that A-tailed promoter-proximal reads would indicate early termination, whereas a lack of A-tails amongst promoter-proximal hits would support the pausing model. Furthermore, short oligo(A) tails would be indicative of surveillance, whereas longer A-tails would suggest a more stable fate.

Proportion of hits with A-tails

After extracting A-tailed reads from each CRAC dataset, I used novoalign to map them and the pyCRAC suite to count hits in each transcript. Plotting the abundance of A-tailed reads in each dataset as a proportion of all mapped reads (Figure 5.3) revealed that A-tailed reads are abundant in some datasets (e.g. Mtr4, Trf4, Nab2 and Pab1), and relatively scarce in others (e.g. Mex67 and Hek2). This is consistent with the documented roles of TRAMP, Nab2 and Pab1 in binding and/or generating A-tails. Furthermore, it suggests that the Trf4- and Mtr4-bound promoter-proximal transcripts might be A-tailed, and thus represent termination products. Indeed, Hrp1, Sto1 and Gbp2, all of which display a bias towards binding near the 5' end of mRNAs (Figure 5.1), are also bound to A-tailed transcripts (Figure 5.3). I therefore sought to characterise the A-tailed reads for each protein in more detail, by examining (i) whether A-tailed hits are enriched amongst particular transcript classes, (ii) whether long and short A-tails can be distinguished, and (iii) where within genes the A-tailed reads map.

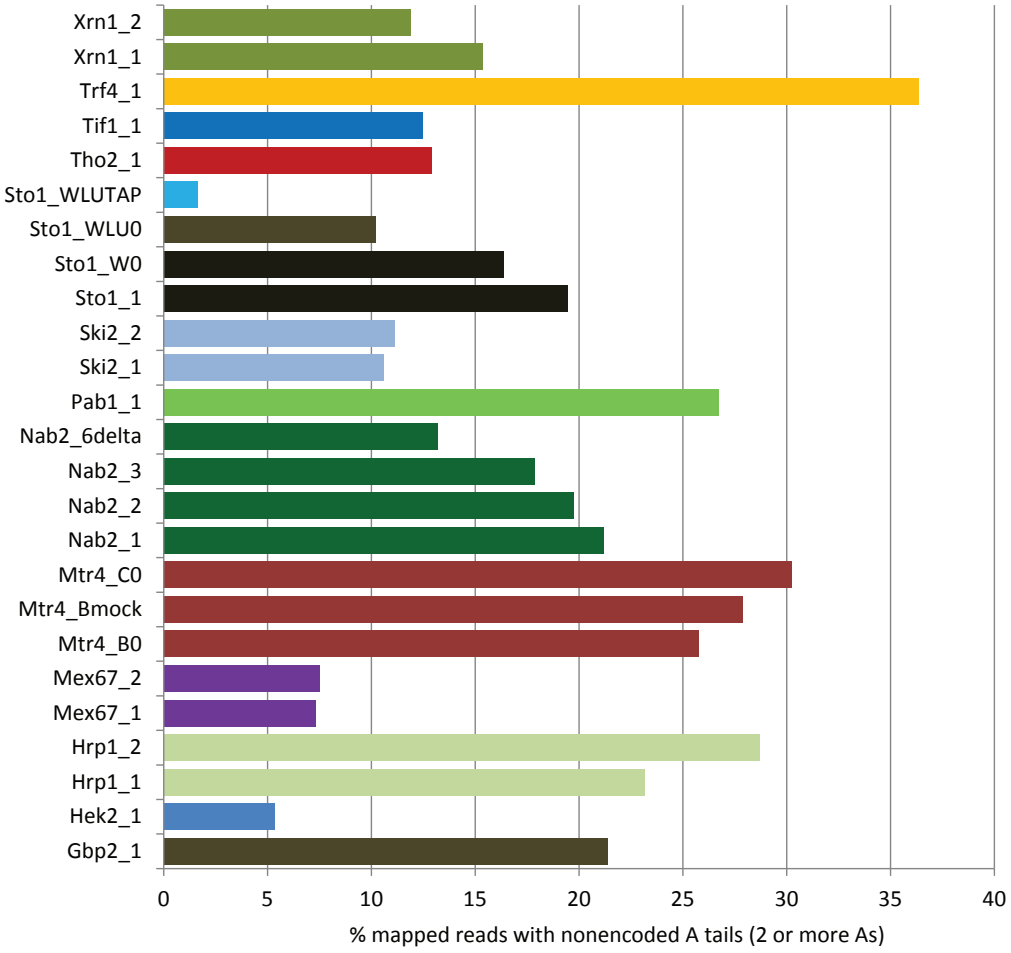


Figure 5.3: Abundance of non-encoded A-tails in CRAC datasets. For each protein, the proportion of hits containing non-encoded A-tails (two or more As) was calculated.

A-tails counted for each transcript class

For each protein, I calculated the proportion of A-tailed reads mapping to each transcript class (Figure 5.4). For nuclear proteins, this produced a similar profile as seen for all reads (Figure 3.2). However, rRNA reads were rare amongst cytoplasmic decay factor A-tailed reads (3 % of Xrn1 hits; Figure 5.4) in comparison to their total reads (30 % for Xrn1; Figure 3.2), and a similar trend was observed for A-tailed tRNAs. This suggests that A-tails are prevalent amongst most transcript classes in the nucleus, but A-tailed rRNA and tRNA fragments are absent from the cytoplasm. CUTs were generally enriched among A-tailed reads for many proteins, relative to SUTs, which is consistent with the more abundant binding of the poly(A) polymerase Trf4 to CUTs versus SUTs (Figure 3.2). Furthermore, there is a particularly high (~4-fold) enrichment of CUTs (but no other lncRNA classes) amongst Xrn1 and Ski2 A-tailed reads (~ 4 %) versus all Xrn1 and Ski2 reads (~ 1 %). This suggests that cytoplasmic CUTs have frequently been targeted by the nuclear TRAMP complex, but have evaded exosome degradation sufficiently long to escape to the cytoplasm. Together, this analysis is consistent with A-tails being prevalent in the nucleus due to their widespread role in surveillance, whereas in the cytoplasm they are predominantly restricted to mRNAs where they function in translation and regulated turnover.

A-tail length distributions

I next analysed the length distribution of non-encoded A-tails in several CRAC datasets (Figure 5.5). This revealed that A-tails up to ~12 nt are abundantly detected for Xrn1, Ski2 and Pab1, which are predicted to bind poly(A) tails on mature mRNAs. The distribution for Mtr4 is narrower, with few A-tails longer than 5 nt, which is consistent with the documented preference of TRAMP to generate short oligo(A) tails. For Mtr4, Ski2 and Pab1, these results were consistent between replicate datasets (data not shown). Notably, A-tails longer than ~15 nts cannot be detected by this analysis due to the limited length of sequence reads (43 nt,

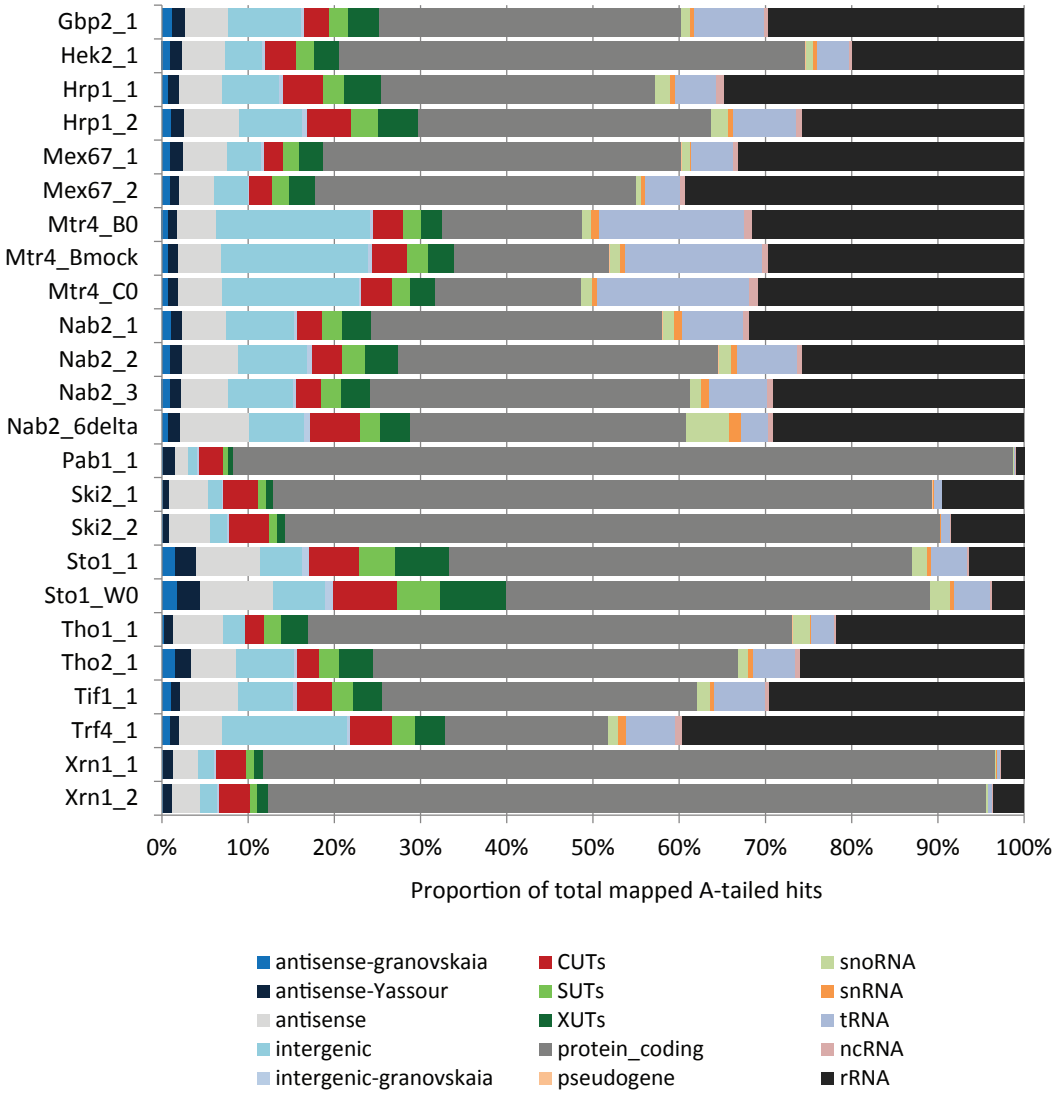


Figure 5.4: Breakdown of A-tailed reads by transcript class. For each protein, reads with non-encoded A-tails (2 or more As) were mapped, and totals counted for each transcript class.

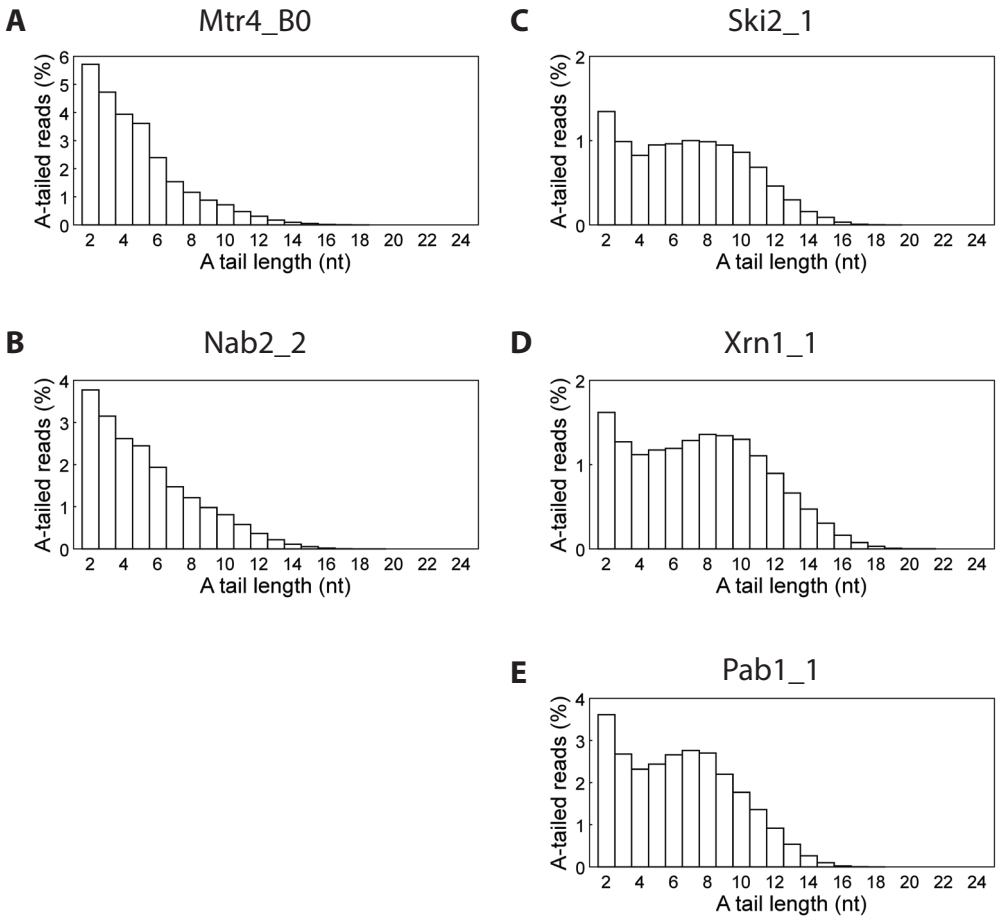


Figure 5.5: Length distribution of non-encoded A-tails detected in reads from various CRAC datasets.

once barcodes are removed), and the requirement for reads to contain a mappable region at the 5' end (>~15-20 nt) and the linker at the 3' end (~10 nt) in order to confidently detect a non-encoded A-tail. These constraints might lead to underestimations in the length of Xrn1-, Ski2- and Pab1-bound A-tails. Indeed, in a crude analysis considering all reads, 19 % of Pab1 reads end in $\geq A_{20}$, suggesting that the difference between Xrn1/Ski2/Pab1 and Mtr4 is even greater than suggested by Figure 5.5. Nonetheless, the different distributions are clearly distinguishable in Figure 5.5, confirming that A-tail analyses can not only distinguish interactions with A-tailed versus non-A-tailed transcripts, but they can also specifically and reliably distinguish binding to different lengths of A-tails.

The short length of A-tailed reads for Mtr4 (Figure 5.5) suggests that they predominantly represent degradation intermediates, and thus the promoter-proximal transcripts bound by TRAMP (Figure 5.1) are targeted for active degradation. The distribution of A-tail lengths amongst Nab2-bound fragments (Figure 5.5) is intermediate between that of Pab1 and Mtr4. Together with the Nab2 binding profile across mRNAs (Figure 5.1), this indicates that Nab2 binds both long A-tails at the 3' end of mature mRNAs, and short A-tails present on promoter-proximal unstable fragments.

Distribution of A-tailed hits within transcripts

In the above discussion, I have assumed that the A-tailed reads for Xrn1, Ski2 and Pab1 map to the 3' end of transcripts, and the A-tailed reads for Mtr4 closely track the distribution of all Mtr4 hits. To test whether these assumptions are correct, I plotted the average distribution of A-tailed reads across mRNAs for several proteins, and compared it to the total hit distribution (Figures 5.6-5.8). For Xrn1 and Ski2, A-tailed hits exclusively mapped to the 3' end of mRNAs (Figure 5.6), despite these proteins being abundantly bound throughout mRNAs. However, for Nab2, A-tailed hits coincided with both 5' and 3' peaks of total binding, but were largely absent from mid-CDS regions (Figures 5.7A-C). To test whether the lengths of A-tails in 5' versus 3' positions were different, I plotted the location of hits

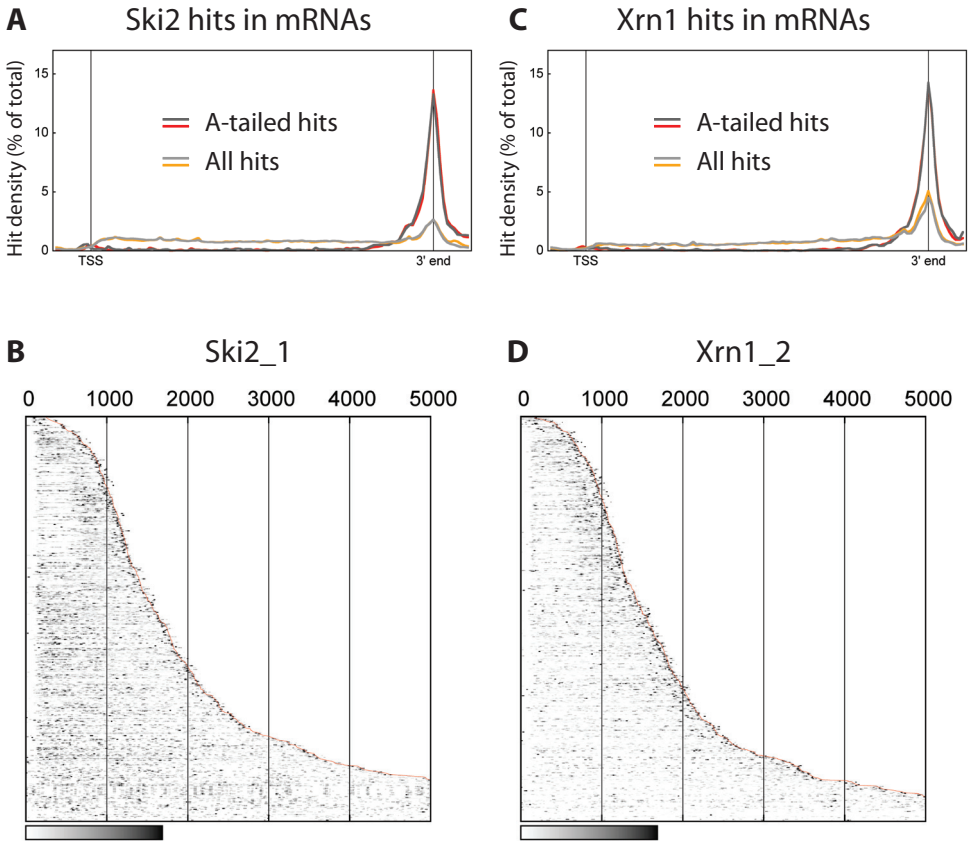


Figure 5.6: Distribution of Ski2 and Xrn1 A-tailed hits across mRNAs. A Average hit density across the 500 mRNAs most abundantly bound by Ski2, considering all reads or just those with non-encoded A-tails (as indicated). Two replicates are shown (different coloured lines). **B** The distribution of Ski2 hits across each of the mRNAs used for the average plot in (A) (each row represents one gene; Ski2_1 dataset). **C** Average hit density across the 500 mRNAs most abundantly bound by Xrn1 D The distribution of Xrn1 hits across each of the mRNAs used for the average plot in (C) (Xrn1_2 dataset).

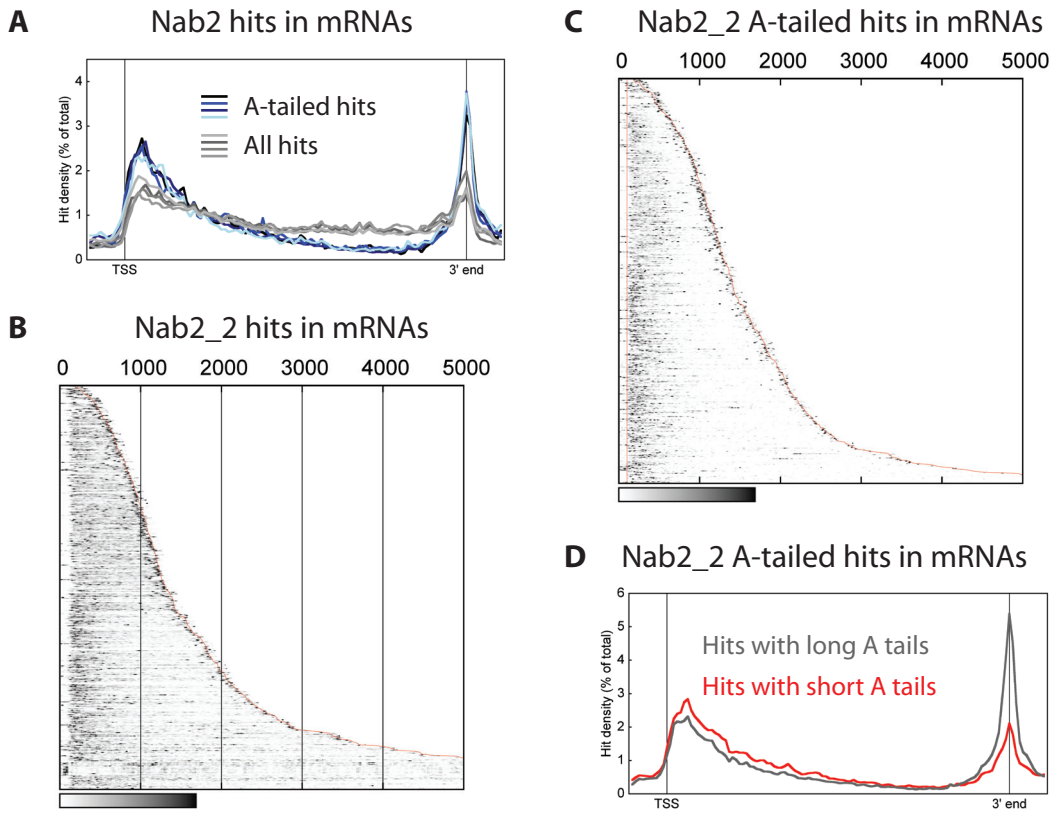


Figure 5.7: Distribution of Nab2 A-tailed hits across mRNAs. A Average hit density across the 500 mRNAs most abundantly bound by Nab2, considering all reads or just those with non-encoded A-tails (as indicated). Four replicates are shown (different coloured lines). **B** The distribution of Nab2 total hits across each of the mRNAs used for the average plot in (A) (each row represents one gene; Nab2_2 dataset). **C** The distribution of Nab2 A-tailed hits across each of the mRNAs used for the average plot in (A) (each row represents one gene; Nab2_2 dataset). **D** Average hit density across the 500 mRNAs most abundantly bound by Nab2, considering only reads with short (≤5 nt) or long (>5 nt) A-tails (Nab2_2 dataset).

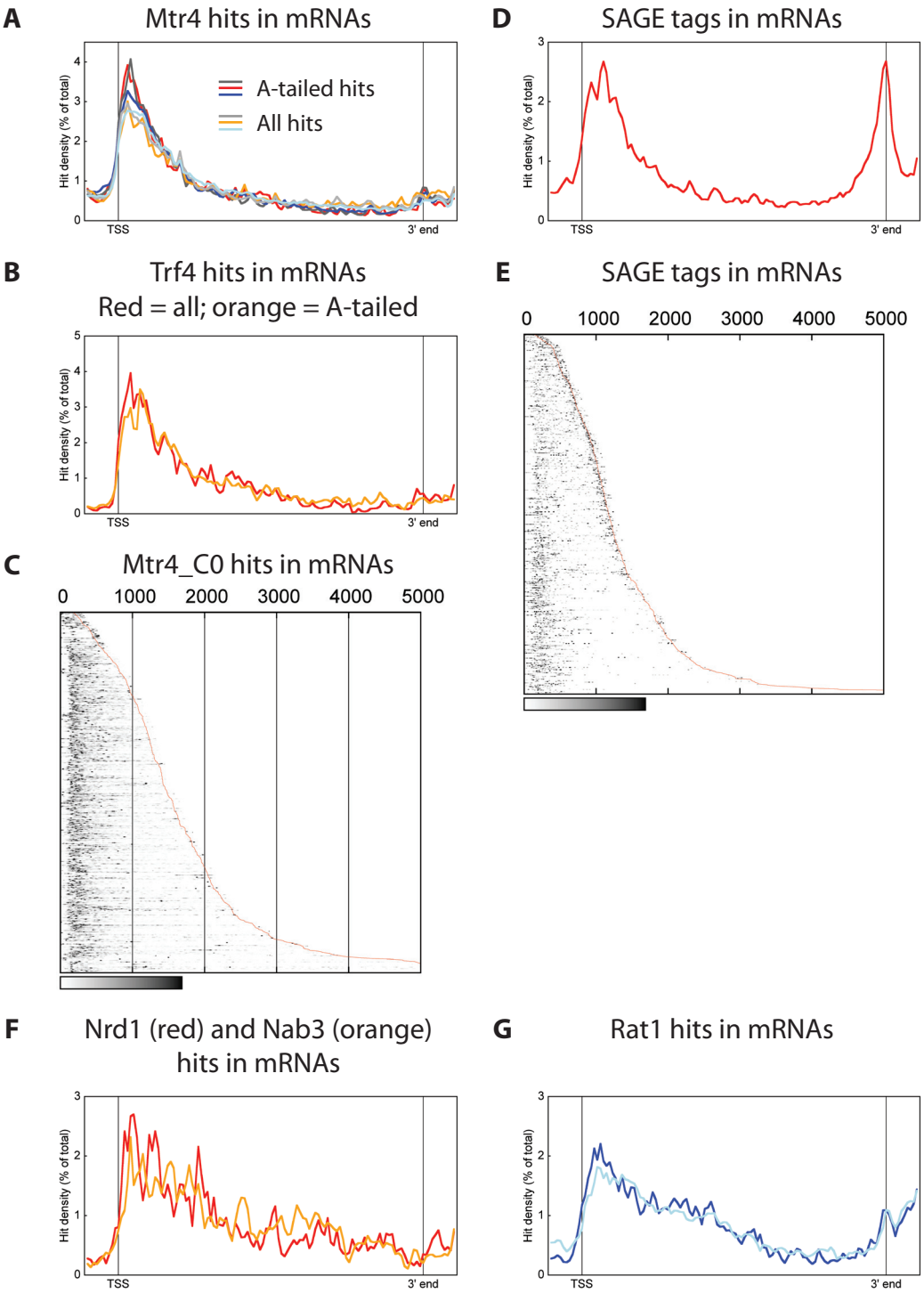


Figure 5.8: Surveillance and termination factors bind promoter-proximal transcripts. A Average hit density across the 500 mRNAs most abundantly bound by Mtr4, considering all reads (dark lines) or just those with non-encoded A-tails (light lines). Three replicates are shown (different coloured lines). **B** Average hit density across the 500 mRNAs most abundantly bound by Trf4, considering all reads or just those with non-encoded A-tails, as indicated. **C** The distribution of Mtr4 hits across each of the mRNAs used for the average plot in (A) (each row represents one gene; Mtr4_C0 dataset). **D** Average distribution of 3' SAGE tags of a CUT fraction described in (Neil et al, 2009) within mRNAs. **E** Individual SAGE tag positions for the mRNAs used in the average plot in (D). **F** Average hit density across the 500 mRNAs most abundantly bound by Nrd1 and Nab3. **G** Average hit density across the 500 mRNAs most abundantly bound by Rat1 (two replicates shown).

with short A-tails ($\leq A_5$) versus hits with long A-tails ($>A_5$) (Figure 5.7D). This revealed that short A-tails are enriched in Nab2-bound promoter-proximal transcripts, whereas long A-tails are prevalent in Nab2 hits at mRNA 3' ends. Mex67 also bound A-tailed fragments from both 5' and 3' ends of mRNAs, but their absolute abundance was ~3-fold less than in Nab2 datasets (data not shown). Finally, I compared the distribution of Mtr4- and Trf4-bound A-tailed and total transcripts (Figures 5.8A-C), and obtained almost identical profiles in each case. Mtr4 forms separate TRAMP complexes with Trf4 and Trf5, but this result suggests that the Trf4 complex may play the major role in nuclear surveillance of Pol II transcripts.

Together, the A-tail analyses identify a population of promoter-proximal 5' mRNA fragments, that are bound by TRAMP (Figure 5.8), appended with short A-tails (Figures 5.5 and 5.7), and are not detectably associated with cytoplasmic factors (Figure 5.6). To test whether these fragments coincide with the unstable transcripts whose 3' ends were mapped by (Neil et al, 2009), and the enriched population of Pol II towards the 5' end of genes, I plotted the average profiles of 3' SAGE hits (Figure 5.8D) and Rpo21 CRAC hits (Figure 5.1) across mRNAs. I also plotted the hits across 759 individual mRNAs that were bound abundantly to Mtr4, and for which Rpo21 and 3' SAGE data was also available (Figure 5.9). In Figure 5.9, genes are ordered by their Mtr4 binding profile, and the order is the same for all three panels. These analyses reveal a high degree of similarity between the location of Mtr4 promoter-proximal binding, and the 3' CUT ends mapped by (Neil et al, 2009), within individual transcripts and across the whole mRNA class. The 5' end enrichment observed for Rpo21 (Figures 5.1 and 5.9) was less strong than that of Mtr4 or the CUT 3' SAGE tags, so is perhaps consistent with a model whereby ~50 % of polymerases stall and terminate in the 5' promoter-proximal region, while the remainder transcribe full-length mRNA.

Finally, I analysed the average binding profiles for Nrd1, Nab3 (Wlotzka et al, 2011) and Rat1 (Sander Granneman, unpublished data) across mRNAs (Figure 5.8F and 5.8G). This

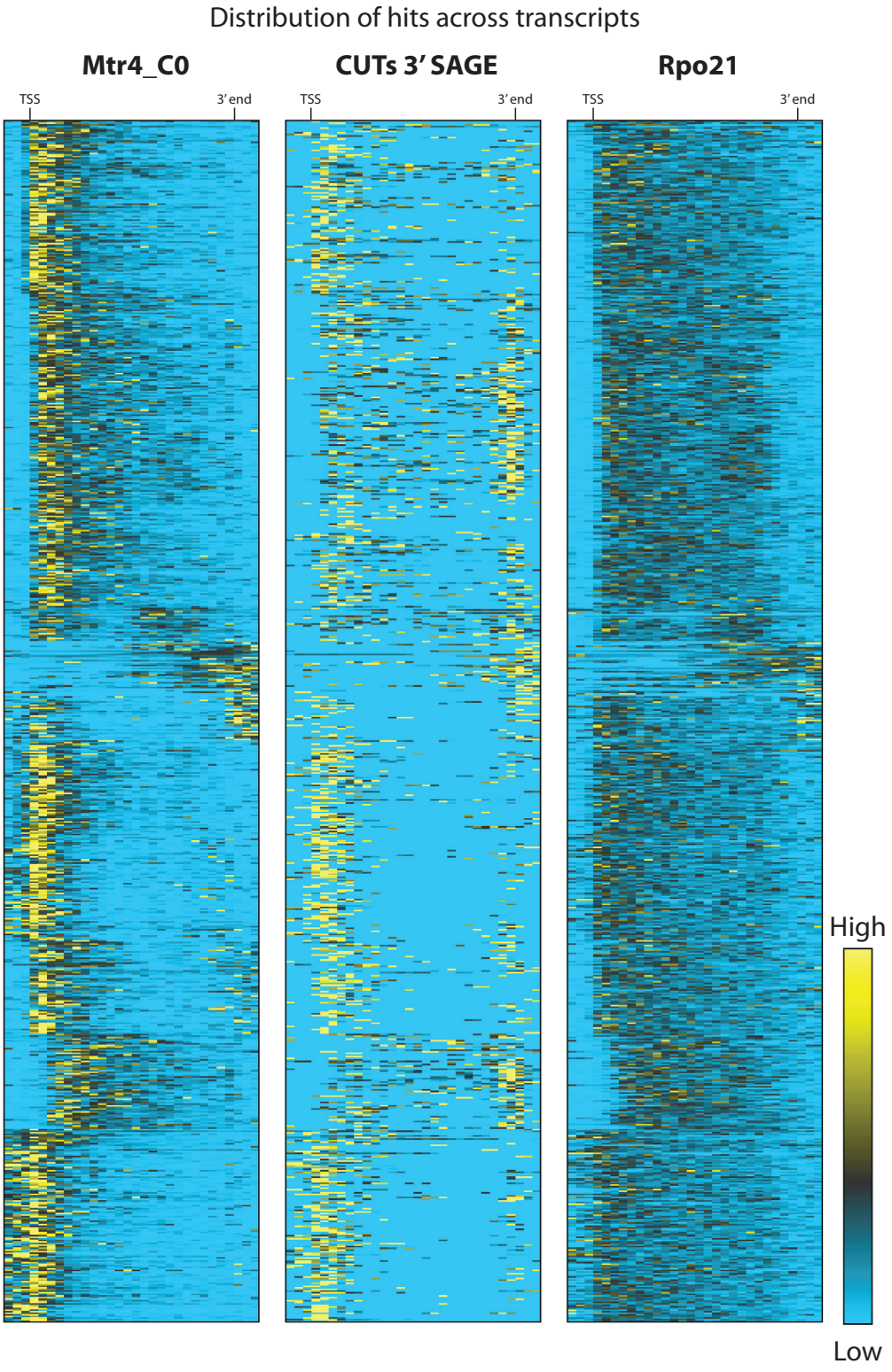


Figure 5.9: Mtr4 and Rpo21 CRAC hits, and 3' SAGE tags, mapping to protein coding genes. The distribution of CRAC hits/SAGE tags is plotted across 759 individual mRNAs, including 100 nt flanking regions. Each row represents a single gene, and rows are ordered by a k-medians clustering (k=10; Spearman rank) of the Mtr4_C0 profiles. Within each panel, each row is normalised (total sum of squares = 1), and the order of genes is the same.

revealed strong 5' enrichment for all three factors, suggesting that either, or both, Nrd1- and Rat1-dependent termination could be prevalent in promoter-proximal regions.

5.4 Discussion

In this chapter, analyses of the mRNA hits in CRAC datasets have provided insight into where, when and how various mRNA biogenesis and turnover factors function. I obtained a coherent picture from several complementary approaches, which examined total mRNA hits, hits in spliced versus unspliced mRNAs, and hits derived from A-tailed transcripts. I detected Pab1, Xrn1 and Ski2 abundantly bound to long A-tails at the 3' end of mature mRNAs (Figures 5.5 and 5.6), suggesting that the poly(A) tail acts as a hub from which various factors orchestrate the regulated turnover of mRNAs. This is consistent with deadenylation preceding both exosome- and Xrn1-mediated cytoplasmic mRNA turnover, the latter of which involves the assembly of a decapping complex contacting both the 5' end and the 3' $A_{10\text{-}12}$ stub. Nab2 also bound poly(A) tails, and an analysis of Nab2 hits in spliced versus unspliced mRNAs (Figure 5.2) suggests that Nab2 binding precedes that of Pab1. However, Nab2 hits in some regions of mRNAs (Figure 5.7A, central region) were not adenylated, consistent with the model proposed in Chapter 4 in which Nab2 binds both poly(A) tails and non-sequence-specific sites throughout transcripts.

Perhaps the most striking observation was the abundant binding of Sto1, Hrp1, Gbp2, Tho2, Nab2, Trf4 and Mtr4 to transcripts from promoter-proximal regions of protein coding genes (Figure 5.1). These transcripts possessed short (1-5 nt) oligo(A) tails (Figures 5.5, 5.7 and 5.8A-B), and were not detectable in CRAC datasets derived from cytoplasmic factors. These observations suggest that Pol II can terminate transcription in promoter-proximal regions to generate abundant mRNA 5' fragments. Many of the Mtr4 binding sites overlap with the 3' ends of unstable transcripts detected by a previous 3' SAGE analysis (Neil et al, 2009), confirming that these 5' mRNA fragments are rapidly turned over. A-tailed transcripts bound by Nab2 (Figure 5.7C) and TRAMP (Figure 5.8C) almost exclusively mapped to the first

~500 bp of the transcribed region, and the extreme 3' end, but not the region in between. This rules out an alternative hypothesis, whereby the 5' mRNA fragments are intermediates arising from degradation of mature mRNAs, as this would result in TRAMP hits across the entire length of transcripts (as seen for the cytoplasmic surveillance factors Ski2 and Xrn1, which degrade full-length mRNAs (Figure 5.6)). Indeed, RNA binding analyses of the termination factors Nrd1 and Nab3 (Creamer et al, 2011; Wlotzka et al, 2011) and Rat1 (Sander Granneman, unpublished data) also reveal a prominent 5' bias in mRNA hits (Figures 5.8F and 5.8G), further supporting a model whereby early termination generates truncated 5' mRNA fragments. This pathway is likely to be conserved in humans, as Xrn2, the human homologue of Rat1, is also preferentially bound towards the 5' end of genes (Brannan et al, 2012).

The well documented Pol II enrichment towards the 5' end of protein coding genes (Churchman & Weissman, 2011; Kim et al, 2011; McKinlay et al, 2011; Pelechano et al, 2009; Rodriguez-Gil et al, 2010) has previously been interpreted as reflecting promoter-proximal pausing. Considering a single gene, this would result in a high Pol II density in the promoter-proximal region, with a greater spacing between polymerases downstream. However, this is at odds with a recent study of Pol II spacing along individual genes. The authors performed a sequential ChIP protocol to isolate DNA bound to two Pol II molecules. The probability of recovery was constant along the tested gene, strongly indicating that Pol II spacing is even (an uneven spacing would give a higher recovery of DNA segments with higher Pol II density) (Peil et al, 2011). This supports the early termination model, in which the promoter-proximal region is more frequently transcribed than the downstream region. However, the two models might be reconciled if promoter-proximal pausing can be resolved either by termination or the resumption of transcription. Intriguingly, the terminal nucleotide of paused nascent transcripts is most commonly adenosine (Churchman & Weissman, 2011), which is the preferred substrate for adenylation by Trf4 (Haracska et al, 2005), suggesting

that pausing and turnover could be closely coupled. In humans, promoter-proximal pausing is associated with the generation of very short (~18-90 nt) RNAs (Preker et al, 2008; Taft et al, 2011), and many longer promoter-associated ncRNAs are also produced (Kanhere et al, 2010; Kapranov et al, 2007), indicating that early termination is prevalent in metazoa.

What is the mechanism of early termination?

The detection of Nrd1, Nab3 and Rat1 bound to transcripts from promoter-proximal regions (Figures 5.8F and 5.8G) suggests that either, or both, Nrd1- and Rat1-dependent mechanisms might contribute to early termination.

Nrd1-dependent termination

The most obvious candidate for eliciting widespread promoter-proximal termination is the Nrd1, Nab3, Sen1 complex. Nrd1-dependent termination has been reported for many of the documented upstream CUTs, including those upstream of *IMD2*, *URA* and *SER3* (Jenks et al, 2008; Kuehner & Brow, 2008; Thiebaut et al, 2006), and for early termination of *HRP1*, *NRD1*, *FKS2* and *PCF11* (Arigo et al, 2006a; Creamer et al, 2011; Houalla et al, 2006; Kim & Levin, 2011; Kuehner & Brow, 2008; Steinmetz et al, 2001). Furthermore, Nrd1 and Nab3 bind the 5' regions of many mRNAs (Creamer et al, 2011; Wlotzka et al, 2011). Nrd1-dependent termination does not require prior cleavage, but involves destabilisation of the elongating polymerase to liberate the nascent transcript with a free 3' end. Nrd1 and Nab3 provide RNA-binding activities, whereas Sen1 is thought to play a key role in the termination mechanism (Finkel et al, 2010; Kim et al, 2006; Steinmetz et al, 2001; Steinmetz et al, 2006b), perhaps by disrupting contacts between the nascent RNA and DNA template (Mischo et al, 2011; Skourti-Stathaki et al, 2011).

Nrd1-dependent termination is predominantly restricted to promoter-proximal regions (Gudipati et al, 2008; Kopcewicz et al, 2007; Porrua et al, 2012; Steinmetz et al, 2006a), due to the preference for the Nrd1 CID to bind the Pol II CTD with Ser5P modifications

(Gudipati et al, 2008; Vasiljeva et al, 2008a) and possibly Ser7P (Kim et al, 2010a), and the exclusion of CTD-CID interactions from the middle of genes by Y1P modification (Mayer et al, 2012). This agrees well with the distribution I see for promoter-proximal Mtr4-bound fragments (Figure 5.8C). Additionally, the ability of Nrd1 to recruit the exosome (Arigo et al, 2006b; Vasiljeva & Buratowski, 2006) is consistent with the instability of 5' mRNA fragments (Neil et al, 2009) and their binding to TRAMP (Figure 5.1). Finally, Nrd1 interacts with the transcription factor Spt5 (Vasiljeva & Buratowski, 2006), which acts together with Spt4 to promote the escape of Pol II from promoter-proximal regions (Rodriguez-Gil et al, 2010). Nrd1 is therefore well placed to function in a promoter-proximal decision between pausing, elongation and termination. Nrd1-dependent termination is, however, also dependent on the presence and organisation of binding sites for Nrd1 (GUA[A/G]), Nab3 (UCUUG) and AU-rich motifs in the nascent transcript, which are bound by single or perhaps multiple Nrd1:Nab3 heterodimers (Carroll et al, 2007; Porrua et al, 2012). This potentially limits the use of the Nrd1-dependent pathway in promoter-proximal termination.

Rat1-dependent termination

Another major termination pathway in yeast involves Rat1, which is proposed to function by degrading nascent transcripts from the 5' end, to catch and destabilise the elongating polymerase (Kim et al, 2004c). Rat1 can only act on transcripts with a free 5' monophosphate, so this mode of termination requires an entry site in the nascent RNA. At mRNA 3' regions, this is classically provided by the mRNA cleavage and polyadenylation machinery, but entry sites can also be generated by the endonuclease Rnt1 (Ghazal et al, 2009; Rondon et al, 2009) or decapping proteins. Rai1 and the homologous Dxo1 are pyrophosphatases that remove unmethylated caps and trigger Rat1-dependent termination. This is prevalent in cells with capping defects (Jiao et al, 2010; Jimeno-González et al, 2010), but also occurs in wild-type yeast (Chang et al, 2012). Another pyrophosphatase,

Dcp2 has recently been shown to trigger Rat1-dependent degradation of lncRNAs in yeast (Geisler et al, 2012), and Xrn2-dependent early termination of lncRNAs and many mRNAs in humans (Brannan et al, 2012; Davidson et al, 2012). Together with the promoter-proximal binding observed for Xrn2 (Brannan et al, 2012) and Rat1 (Figure 5.8G), this suggests that Rat1 might function in widespread early termination at mRNAs in yeast, triggered by decapping. The free 5' end required for Rat1-dependent termination might alternatively be provided by endonuclease cleavage. The human endoribonucleolytic Microprocessor complex was recently found to trigger early termination by Xrn2, and apparently acts at hundreds of protein coding genes (Wagschal et al, 2012). In yeast, similar roles could be played by Rnt1, the PIN domain endonuclease activities of the exosome component Rrp44 or Swt1, or an as yet uncharacterised endonuclease.

However, RNAs detected in CRAC analyses are presumably not themselves released by Rat1-mediated termination, since the "torpedo" mechanism degrades the nascent transcript. The binding of Mtr4 to adenylated promoter-proximal fragments shows that the termination mechanism liberates 3' ends that can be accessed by the TRAMP complex (Figures 5.8A-C). This would be consistent with endonucleolytic cleavage preceding TRAMP- and Rat1-dependent turnover of upstream and downstream fragments respectively. It is also possible that some combination of Nrd1-dependent and Rat1-dependent termination occurs. Human Sentaxin (the Sen1 homologue) cooperates with Xrn2 in Microprocessor-dependent early termination (Wagschal et al, 2012), and the transcriptome-wide binding distribution of Sen1 suggests that it often functions in collaboration with Rat1 in yeast (Creamer et al, 2011; Jamonnak et al, 2011).

What is the function of early termination?

Early termination of mRNA transcription could serve numerous roles. Non-productive transcription might allow gene expression to be rapidly altered in response to stress, if the pool of Pol II complexes associated with a gene can be switched between non-productive

and productive transcription (Kim & Levin, 2011; Kim et al, 2011). However, another study suggests that changes in RNA stability play a greater role (Garcia-Martinez et al, 2012). Promoter-proximal transcription could also contribute to regulation by directing activating or repressive chromatin modifications, as has been documented for several upstream CUTs (Jenks et al, 2008; Kuehner & Brow, 2008; Thebault et al, 2011; Thiebaut et al, 2008). Regardless of their specific functions, promoter-proximal transcripts apparently sequester a large proportion of nuclear surveillance and RNA binding factors. They might therefore have indirect consequences by reducing the availability of these factors at other genes/transcripts. The results in this study suggest that when considering Pol II transcripts, the nuclear surveillance machinery is predominately occupied with promoter-proximal transcripts and lncRNAs, in stark contrast to the cytoplasmic decay factors, which act on mature mRNAs and have very few lncRNAs as substrates. Such a high rate of production and turnover of promoter-proximal transcripts in yeast could explain previous calculations suggesting that only ~10 % of Pol II engaged on chromatin produces detectable, stable transcripts (Pelechano et al, 2010; Struhl, 2007).

6: The dynamic interplay between coding and non-coding transcription

In the previous chapters, I have focused on mRNA and lncRNA metabolism in yeast during growth in synthetic glucose media. However, in their natural environment, yeast are subject to a complex, and continually changing, mixture of nutrients. Consequently, much of their regulatory circuitry is dedicated to sensing and responding to external conditions. Recent studies have revealed that some lncRNAs play a major role during periods of change, when there is extensive reprogramming of gene expression, rather than steady-state metabolism (Geisler et al, 2012; Kim et al, 2012). Indeed, there might be a vast number of lncRNAs that only exist fleetingly, and it is now apparent that even apparently simple molecular events can involve complex cascades of non-coding transcription (Hirota et al, 2008). Furthermore, changes in mRNA expression might be achieved not only by changes in the transcriptional output from a locus, but also changes in mRNA stability (Garcia-Martinez et al, 2012) and redistribution of Pol II from non-productive upstream or antisense transcription to transcription of full-length mRNAs (Darby et al, 2012; Kim & Levin, 2011; Kim et al, 2011; Yoon & Brem, 2010). To fully understand how gene expression is regulated in response to a change in growth conditions, for each locus we must therefore consider changes in the transcription and decay rates of coding and non-coding transcripts, and how these transcripts interact.

I therefore sought to investigate the effects of a nutrient shift on (i) how mRNAs and lncRNAs are regulated via changes in transcription and decay rates, and (ii) how changes in lncRNA expression relate to changes in mRNA expression. In particular, I reasoned that given the predominantly nuclear and unstable nature of lncRNAs, the nuclear surveillance machinery would play a major role in regulating their expression. I therefore performed three analyses: (i) metabolic labelling of nascent transcripts to obtain genome-wide transcription

rates, (ii) CRAC on the exosome cofactor Mtr4, to analyse transcriptome-wide nuclear surveillance activity, and (iii) CRAC on Sto1, to obtain an estimate of nuclear transcript abundance. I opted for a nutrient shift from synthetic glucose media, in which yeast grow rapidly via fermentative growth, to synthetic glycerol and ethanol media, in which yeast grow slowly via aerobic growth. There are large differences in mRNA and lncRNA expression during steady-state growth under these two conditions (Xu et al, 2009), suggesting that significant changes during a rapid (4-16 minute) shift would be anticipated.

6.1 Genome-wide transcription rate measurements via metabolic labelling

Changes in RNA abundance following a shift in growth conditions are largely due to changes in global transcription rates, although changes in RNA decay rates can also contribute (Castells-Roca et al, 2011; Garcia-Martinez et al, 2012; Miller et al, 2012; Munchel et al, 2011). I therefore sought to establish a reliable, non-invasive method to determine transcription rates. Recently, this has been achieved for mammalian and insect cells by following the incorporation into nascent RNA of the non-toxic sulphur-containing nucleoside analogue 4-thiouridine or modified base 4-thiouracil (Cleary et al, 2005; Dolken et al, 2008; Miller et al, 2009). I therefore tested the efficacy of this approach in yeast, using a modified version of a protocol developed by the Begg's laboratory (David Barrass, personal communication). During the course of this study, we and others have published variants of this method (Miller et al, 2012; Munchel et al, 2011; Swiatkowska et al, 2012).

I first tested whether 4-thiouracil (4TU) could be taken up by yeast and efficiently incorporated into nascent RNA. Overexpression of the uridine permease Fui1 is reported to facilitate 4TU uptake (Swiatkowska et al, 2012), and so I generated a yeast strain overexpressing Fui1 from a multicopy plasmid (strain yAT1). After the addition of 4TU to an exponentially growing yAT1 culture, I harvested yeast at 2 minute intervals, fixing in -80

°C ethanol, and extracted RNA. I treated the RNA extracts with HPDP-biotin, which forms a non-covalent adduct with thiouracil, then purified the biotinylated, 4TU-containing nascent RNA using streptavidin beads. I examined the eluate (nascent RNA fraction) by Northern blotting, probing for the 18S rRNA and its precursor, 20S pre-rRNA. This revealed rapid accumulation of 20S pre-rRNA within 1-2 minutes, followed by accumulation of 18S (data not shown), and is consistent with published results from ^{3}H-uracil pulse-chase experiments (Kos & Tollervey, 2010).

I next adapted this approach to enable the nascent RNA to be analysed by deep sequencing (Figure 6.1A). I used maleimide-PEG_{11}-biotin in place of HPDP-biotin, to form a covalent adduct with thiouracil-containing RNA that enables it to be permanently captured on streptavidin beads. I then performed a series of on-bead enzymatic reactions, comprising dephosphorylation, decapping, 5' linker ligation, and random-primed reverse transcription (which appended a 3' adapter). The resultant cDNA was used as a template for PCR amplification, the products size selected on an agarose gel, and ~50-200 bp amplicons obtained for Solexa sequencing. Extensive optimisation revealed that 3 minutes of labelling with 20 μM 4TU gave usable yields. I generated libraries from yeast labelled either in synthetic glucose (SGlu) or synthetic glycerol and ethanol (SGlyEtOH) media, and submitted these for sequencing. As a negative control, I also submitted libraries generated from yeast that had not been treated with 4TU. By using a cap-dependent method for 5' linker ligation, I obtained reads exclusively from the 5' end of transcripts, enabling overlapping transcripts to be distinguished and maximising read depth. For cells harvested after a short pulse of 4TU labelling, levels of degradation are expected to be low and 4TU incorporation should therefore largely reflect the transcription rate.

Comparison of the SGlu +4TU dataset to the negative control (SGlu -4TU) revealed an enrichment for short-lived transcripts such as CUTs (Figure 6.1B), consistent with this representing a nascent RNA fraction. Furthermore, reads almost exclusively mapped to the

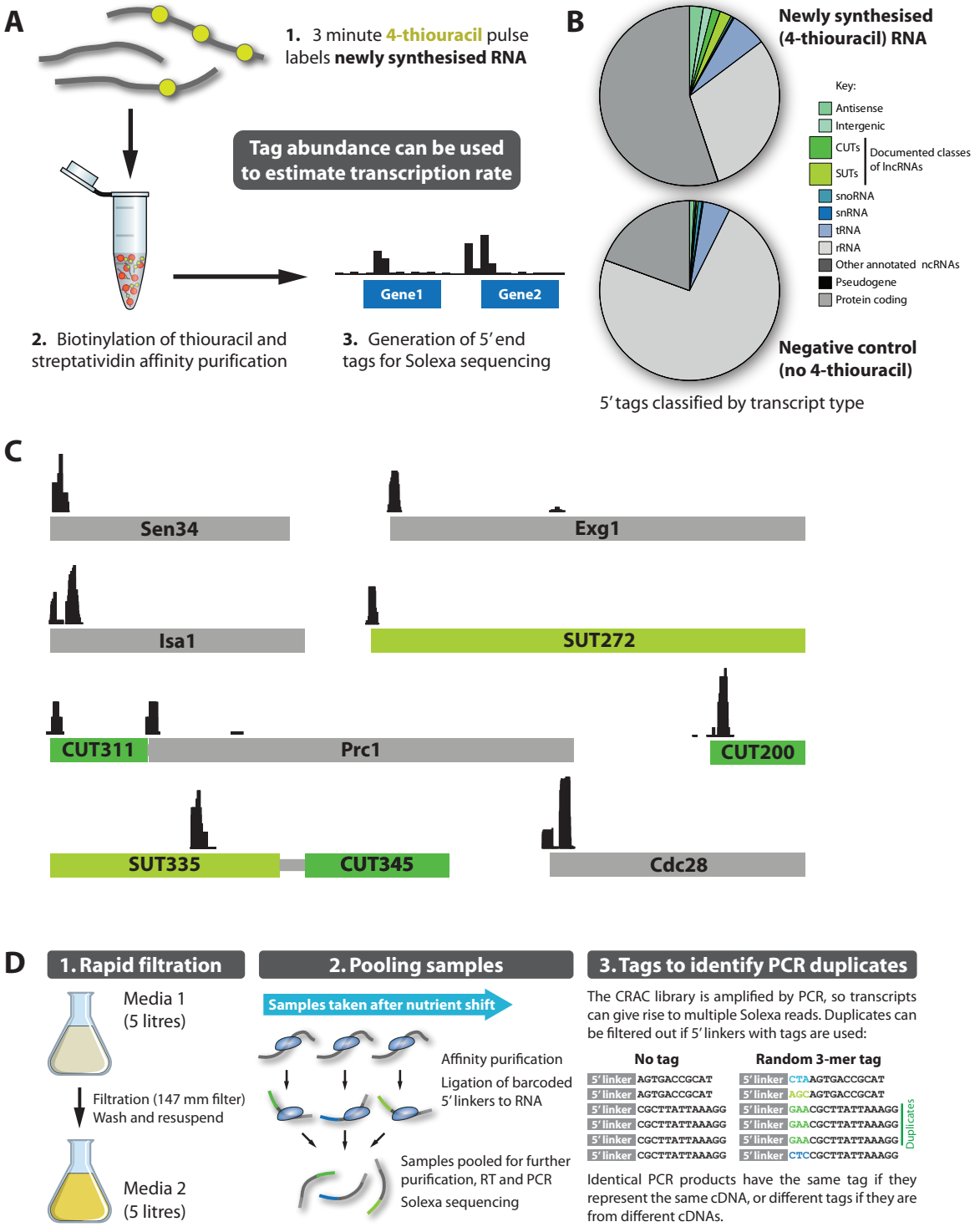


Figure 6.1: Metabolic labelling and CRAC analyses to measure RNA synthesis and decay rates in yeast during a nutrient shift. A To measure global transcription rates, newly synthesised transcripts were labelled by exposing yeast to a short 4-thiouracil pulse, and analysed by high-throughput sequencing of 5' tags. **B** Distribution of 5' tags by transcript type, for a newly synthesised RNA fraction (+4TU) and negative control (no 4TU). **C** Examples of 5' tags that map to mRNAs and lncRNAs. **D** Modifications to the standard CRAC protocol that facilitated the analysis of Mtr4- and Sto1-bound transcripts in yeast collected at various time points during a nutrient shift.

5’ end of transcripts (examples in Figure 6.1C), indicating that the cap-dependent cloning was effective. In cases where reads did not align with annotated transcription start sites (e.g. SUT335, Figure 6.1C), visual inspection of transcriptome tiling array data (Xu et al, 2009) revealed a step in probe intensity, suggesting that most 5’ tags in the 4TU dataset do indeed represent *bona fide* TSSs.

Unfortunately, however, the read depth was very low for many genes, especially in the SGlyEtOH dataset (data not shown). This suggests that global transcription is downregulated during growth in SGlyEtOH, consistent with the prolonged doubling time (~7.5 hrs). Despite numerous attempts, I was unable to increase the overall yield from this method in the time available. In future, the yield might be increased by using a different labelling chemistry or a different method of library construction, and background could be reduced by using a proteinase K digestion to specifically elute nascent RNA, via disruption of the streptavidin:biotin interaction. At present, however, this method is not suitable for quantifying the transcription rate of low abundance transcripts, such as lncRNAs, although it has provided a high quality set of 5’ end coordinates for many transcripts. In the absence of direct experimental determination, transcription rates can be inferred from RNA abundance and decay rate, and I therefore focused my efforts on experiments to determine these parameters.

6.2 Analyses of changes in surveillance activity and transcript abundance

I attempted to identify changes in nuclear surveillance activity and transcript abundance during a shift from SGlu to SGlyEtOH using CRAC. This requires a large volume of culture (750 ml per time point). I therefore used a custom-built, dead-end filtration device (Darwin workshop, University of Edinburgh) fitted with a 0.9 μm filter (147 mm diameter; MF-Millipore). With this, the yeast could be rapidly collected from ~5 litres of culture, washed,

and resuspended in fresh medium (Figure 6.1D). In a typical experiment, I grew Sto1-HTP or Mtr4-HTP yeast to logarithmic phase in SGlu, then shifted 2.5 litres of culture to SGlyEtOH, and 800 ml to SGlu (mock shift). I collected five samples for UV irradiation, including one pre-shift sample (grown in SGlu), one mock shift sample (16 minutes growth in SGlu after filtration and resuspension), and three SGlyEtOH samples (4, 8 and 16 minutes after resuspension in SGlyEtOH). I collected three sets of samples for the Mtr4-HTP strain, and one for the Sto1-HTP strain. For each sample the Mtr4- or Sto1-bound transcripts were identified by CRAC. Notably, I grew the yeast in conditions amenable to 4TU labelling, so that future 4TU datasets could be analysed alongside these CRAC datasets ("Sto1_WLU" and "Mtr4_A/B/C", Table 3.1). Specifically, the HTP-tagged strains contained the Fui1 plasmid and were grown in media lacking tryptophan, uracil and leucine. As a control to test whether the CRAC results obtained under these conditions resembled those obtained under standard conditions, I also repeated the shift for Sto1-HTP yeast, with no Fui1 plasmid, in media lacking tryptophan ("Sto1_W" datasets, Table 3.1).

To reduce experimental variation between time points, for each time course I pooled the crosslinked RNA:protein complex samples after 5' linker ligation, and all subsequent experimental steps were performed on this single sample. The time points could be separated computationally after sequencing, as they were prepared with barcoded 5' linkers. I then processed the Solexa datasets as described in Chapter 3.2. The inclusion of random 3-mer tags in the 5' linkers enabled PCR duplicates to be removed, improving the quantitative ability of this method (Figure 6.1D). One experiment, "Mtr4_A", produced a low number of reads, so I focused on the higher complexity Mtr4 datasets ("Mtr4_B" and "Mtr4_C"), together with the Sto1 datasets, for further analyses.

I first assessed the correlation between Mtr4 datasets, by calculating pairwise Spearman rank coefficients based on the number of hits in each dataset for each Pol II gene (normalised to total Pol II hits) (Figure 6.2A). This revealed generally stronger correlations between two t =

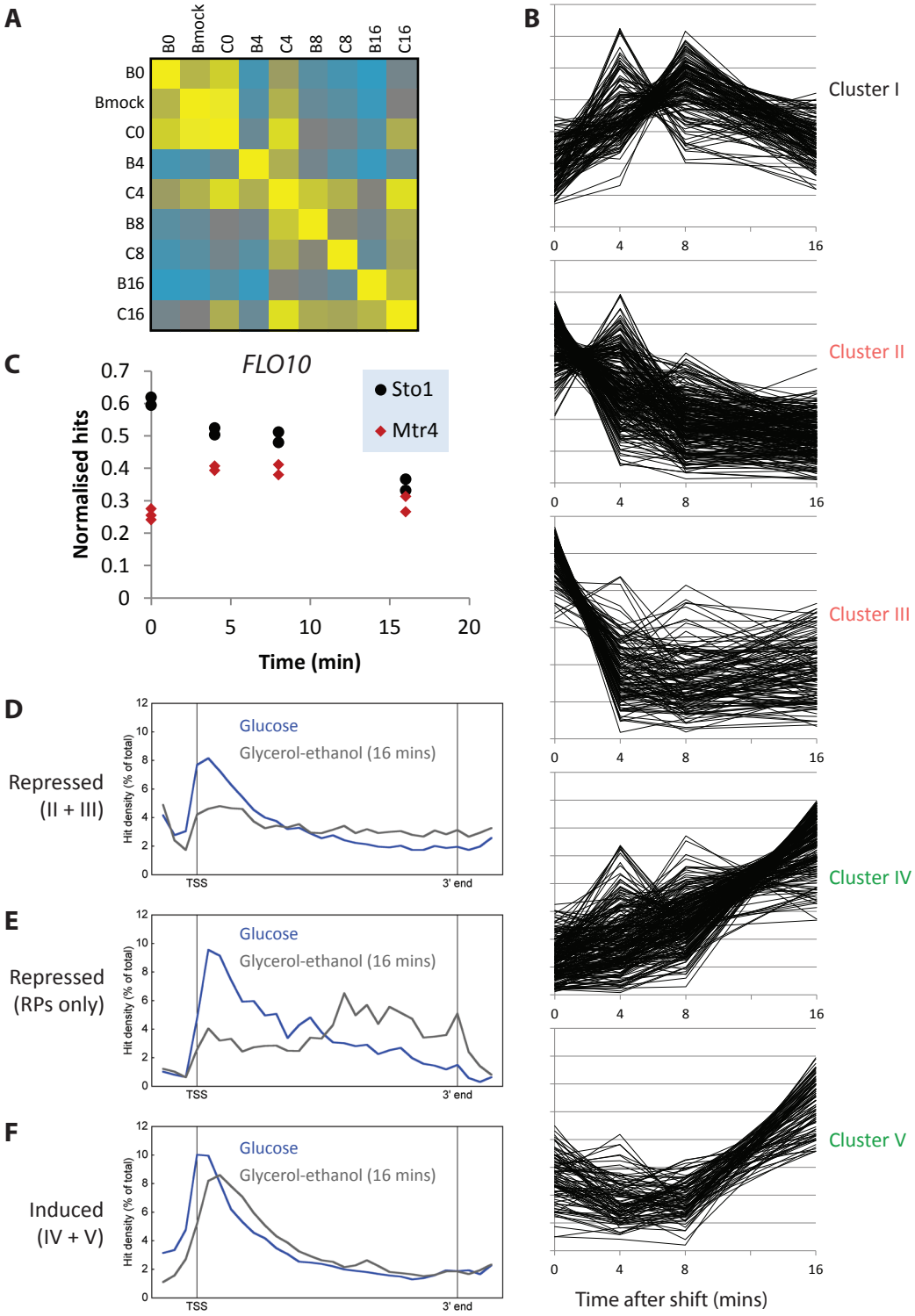


Figure 6.2: Transcriptome-wide changes in nuclear surveillance activity in yeast subjected to a media shift. A Correlation matrix summarising the similarity between CRAC analyses of Mtr4 targets in yeast collected at various time points during a media shift (SGlu to SGlyEtOH). For pairs of samples, the Spearman rank correlation coefficients were calculated based on the number of Mtr4 hits in Pol II transcripts. Yellow, $\rho = 1$; blue, $\rho = 0$. **B** Transcripts were arranged into five clusters based on the level of Mtr4 binding (CRAC hits) across the time course. The graphs show the binding profiles (hits versus time) for all members of each cluster. **C** Example of a transcript for which changes in surveillance activity (Mtr4 hits) and abundance (Sto1 hits) were not correlated. **D**, **E** and **F** Average distribution of Mtr4 CRAC hits across Pol II transcripts before and after the nutrient shift, for the indicated clusters.

0 replicates (ρ = 0.79-0.88), or two t = 16 replicates (ρ = 0.79), than between t = 0 and t =16 datasets (ρ = 0.58-0.77). To define a high quality dataset with reproducible changes over the time course, I selected Pol II genes with >200 hits per million in at least two Mtr4 datasets (1763 genes), and with a Pearsson correlation coefficient of at least 0.8 between Mtr4_B and Mtr4_C time series (849 genes). To gain an overview of the changes in Mtr4 binding during the nutrient shift, I clustered these 849 genes by their hit profile across the Mtr4 datasets (k-medians clustering; k = 5; Spearman rank) (Figure 6.3A). This yielded two clusters in which Mtr4 binding decreased during the time course (clusters II and III), two in which Mtr4 binding increased (clusters IV and V), and one in which Mtr4 binding transiently increased (cluster I). These general trends were particularly clear when the profiles of individual genes in a cluster were overlaid (Figure 6.2B).

The Mtr4 CRAC data reflect changes in surveillance activity for each transcript. To determine how these changes relate to changes in the nuclear abundance of each transcript, I plotted the Sto1 (nuclear CBC) data alongside (Figure 6.3B). I subclustered each major Mtr4 cluster (I-V), based on the Sto1_WLU binding profile across the time course, to facilitate the identification of groups of transcripts with different Sto1 binding profiles but similar Mtr4 binding profiles. This revealed that, generally, changes in Mtr4 binding are closely correlated to changes in transcript abundance (Sto1 binding). Therefore, despite the absolute abundance of many transcripts changing, the proportion bound by Mtr4 (and thus the surveillance rate) remains relatively constant, and the major influence on transcript abundance (including that of lncRNAs) is apparently, therefore, a change in transcription rate. This is consistent with conclusions from recent transcription run on experiments (Garcia-Martinez et al, 2012). Furthermore, these data are consistent with the TRAMP complex playing a ubiquitous role in surveillance. However, there were exceptions for which the proportion of transcripts bound by Mtr4 varied. For example, in cluster I (Figure 6.3) some transcripts underwent a transient increase in binding to Mtr4 coincident with a decrease in binding to Sto1 (e.g. *FLO10*,

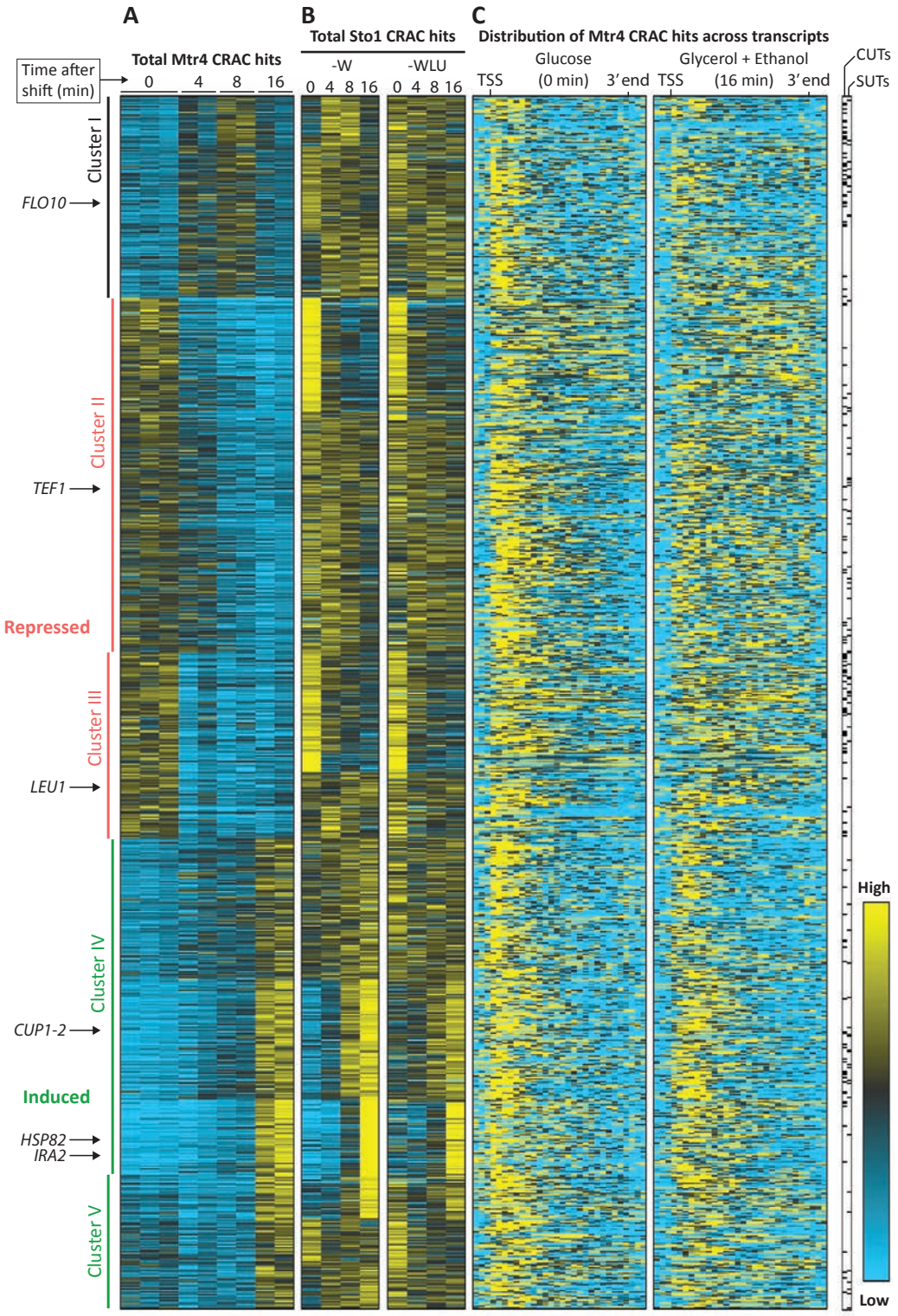


Figure 6.3: CRAC analyses of Mtr4 and Sto1 binding to Pol II transcripts during a nutrient shift. A The heat map shows the level of Mtr4 binding to Pol II transcripts during a shift from SGlu to SGlyEtOH. Transcripts were arranged by k-medians clustering (k = 5; Spearman rank). For each time point, replicate experiments are shown. **B** Sto1 binding to Pol II transcripts during an identical nutrient shift was also analysed by CRAC, and the abundance of hits in each transcript is shown. This experiment was repeated for yeast grown in –W (left panel) and –WLU (right panel) drop out media. **C** The position of Mtr4 hits across each transcript before (SGlu, t=0) and after (SGlyEtOH, t=16) the shift. For each panel in **A**, **B** and **C**, each row is normalised (sum of squares = 1). Transcripts that feature in other figures are indicated (arrows on left).

Figure 6.2C). For these transcripts, Mtr4-dependent turnover might be upregulated to elicit rapid downregulation. CUTs and SUTs underwent a very rapid reduction in abundance (Figure 6.3, clusters II and III), which suggests they are particularly unstable amongst nuclear transcripts. GO term analyses reveal that upregulated genes were enriched for protein folding activity (e.g. HSC82 and HSP82) and functions at the cell periphery (e.g. HXT2, a high affinity glucose transporter). Conversely, downregulated genes were enriched for ribosome biogenesis factors, snoRNP proteins, snoRNAs, and ribosomal proteins. This is largely consistent with the shift to glycerol and ethanol media eliciting an acute stress response.

Recent studies have suggested that changes in non-productive transcription, which produces upstream CUTs and early termination products instead of full length mRNAs, contribute to stress responses in yeast (Darby et al, 2012; Kim & Levin, 2011; Kim et al, 2011; Yoon & Brem, 2010). For each gene, I therefore plotted the distribution of Mtr4 hits across transcripts, including 100 nt up- and downstream flanking sequences, for the 0 and 16 minute time points (Figure 6.3C). This revealed an enrichment of Mtr4 hits towards the 5' end of most transcripts in the SGlu dataset. The 5' bias was strikingly reduced for clusters II and III in the 16 minute SGlyEtOH dataset (e.g. *TEF1* and *LEU1*, Figures 6.4A and 6.4B) but was maintained for clusters IV and V. These changes are particularly apparent when the average profiles for clusters II+III, and IV+V, are plotted (Figures 6.2D and 6.2F). Furthermore, the reduction in 5' bias is particularly prominent for Mtr4 binding across ribosomal protein genes (Figure 6.2E). This suggests that for repressed transcripts (clusters II and III), Mtr4-dependent surveillance shifts from promoter-proximal transcripts to full-length mRNAs, perhaps helping to accelerate repression. Alternatively, the rapid degradation of promoter-proximal transcripts might leave only full-length mRNAs for Mtr4 to bind, resulting in an apparent decrease when the results are tabulated as hit per million.

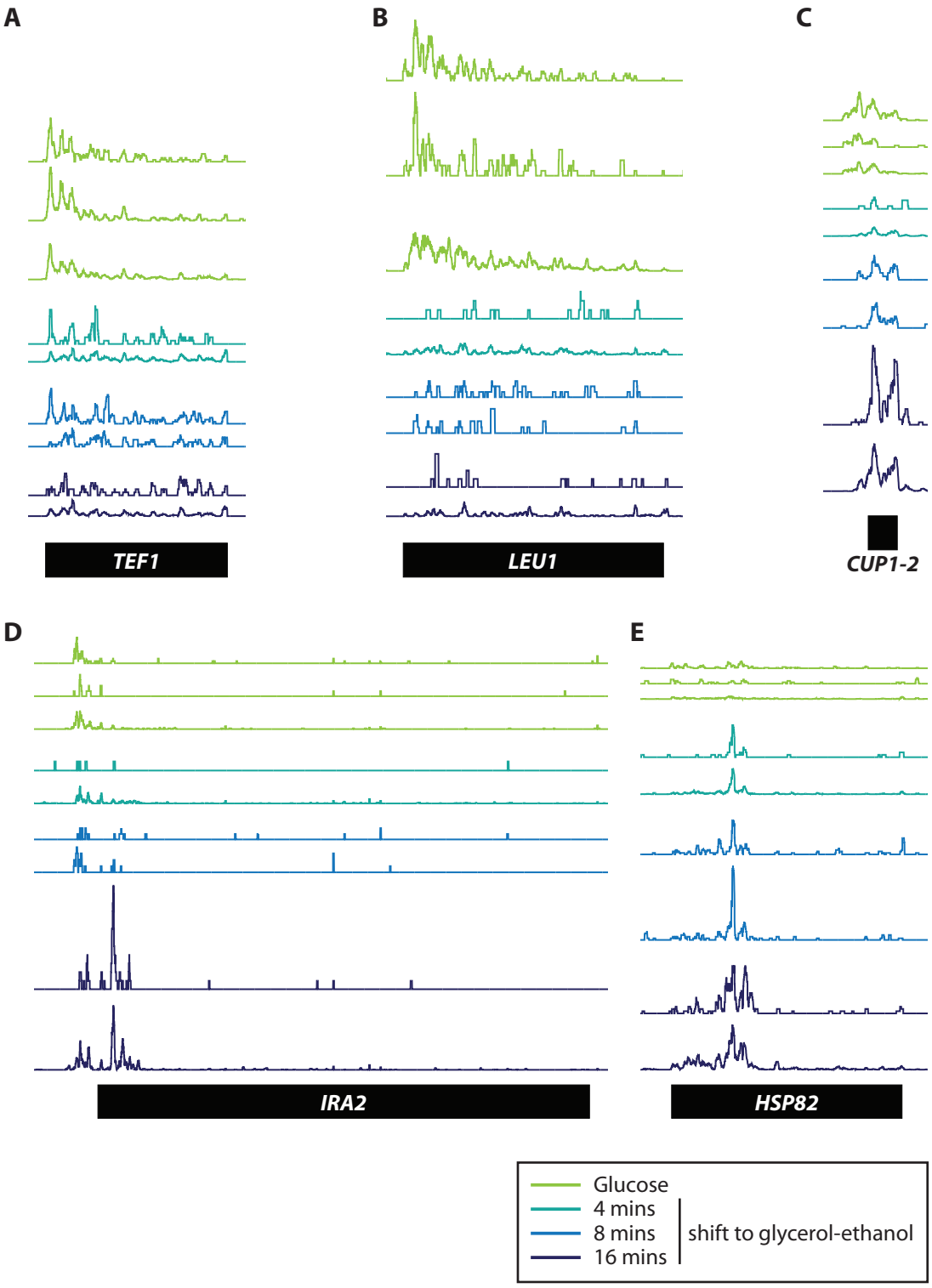


Figure 6.4: Mtr4 binding across representative transcripts during a 16 minute SGlu to SGlyEtOH shift.

Inspection of the average binding profile for clusters IV+V (Figure 6.2F) revealed that induction of genes in these clusters is accompanied by a shift in Mtr4 binding from the region over the TSS to slightly more downstream sequences. The Mtr4 binding profiles across individual transcripts, such as *CUP1-2* and *IRA2* (Figures 6.4C and 6.4D), suggest that this reflects the production and turnover of upstream CUTs when the gene is repressed. Indeed, the region immediately upstream of *IRA2* is annotated as a CUT (Xu et al, 2009). This suggests that the induction of some genes in clusters IV+V is accompanied by a shift from upstream, non-coding transcription to productive, canonical transcription. Approximately one third of genes in clusters IV+V have high signal (> 10 % of total) in the 100 nt upstream of their TSS, suggesting that this mechanism is common. The other two thirds of genes in clusters IV+V (e.g. *HSP82*, Figure 6.4E) are presumably subject to more classical regulatory mechanisms.

6.3 Discussion

A transcriptome-wide analysis of nuclear RNA abundance (Sto1 binding) and surveillance (Mtr4 binding) revealed many alterations in response to nutritional down shift (SGlu to SGlyEtOH) (Figure 6.1D). Changes in transcript abundance were predominantly accompanied by a correlated change in surveillance activity (i.e. upregulated transcripts were also bound more abundantly by Mtr4) (Figure 6.3). This suggests that the primary function of the nuclear surveillance machinery is the ubiquitous monitoring of gene expression. However, some transcripts (e.g. *FLO10*, Figure 6.2C) exhibited changes in the proportion bound to Mtr4, consistent with the nuclear surveillance machinery playing specific roles in modulating the expression of a subset of individual genes. LncRNAs such as CUTs and SUTs (Figure 6.3) were predominantly rapidly downregulated, indicating that instability is a general property of lncRNAs in yeast. This is consistent with the high enrichment of lncRNAs among TRAMP targets (Figure 3.6). This rapid clearance of lncRNAs might make

them particularly suited to function in regulatory pathways during rapid or transient responses.

An analysis of the binding profile of Mtr4 across the length of individual transcripts before and after the nutrient shift (Figure 6.3C) was particularly informative. Firstly, many upregulated genes displayed an upstream peak of Mtr4 binding in their repressed state, suggesting that they might be regulated via alternative TSS selection, as has been documented for a number of nucleotide biosynthetic genes (Jenks et al, 2008; Kuehner & Brow, 2008; Thiebaut et al, 2008). These regulatory circuits typically involve the generation of upstream CUTs that are terminated by Nrd1. Notably, I observed upstream Mtr4 binding to *IRA1*, a positive regulator of the Nrd1-dependent termination pathway (Darby et al, 2012), prior to nutrient down shift. Following transfer to SGlyEtOH, Mtr4 binding shifted downstream (Figure 6.4A) and a peak of Sto1 binding appeared that mapped to the annotated *IRA1* TSS (data not shown). This suggests that Nrd1-dependent termination of the *IRA1* upstream CUT might act in a feedback loop to reinforce or temper Nrd1 activity via up- or downregulation of *IRA1* mRNA expression.

I observed a different pattern of Mtr4 binding for transcripts that were downregulated following the shift to SGlyEtOH, whereby Mtr4 binding at the 5' end was reduced, and instead binding was distributed across the full length of mRNA coding transcripts (Figures 6.3C and 6.2D). RP genes were particularly strongly affected. This suggests that the nuclear surveillance machinery is primarily occupied with the turnover of promoter-proximal transcripts during active expression (SGlu), but after shifting to SGlyEtOH, is mostly bound to full-length mRNAs. When yeast are starved of glucose, as is the case following a rapid shift to SGlyEtOH, transcription is drastically reduced (as shown by decreased 4TU incorporation, data not shown) (Jona et al, 2000). The reduction in Mtr4 binding to promoter-proximal transcripts may therefore arise simply because transcription is globally downregulated, and these transcripts decay more rapidly than full-length mRNAs. However,

during stress or growth in nutrient poor media, Pol II becomes more evenly distributed across genes, particularly RP genes (Kim et al, 2011; Rodriguez-Gil et al, 2010). This suggests that reduced Mtr4 binding to promoter-proximal fragments reflects a shift in transcription from promoter-proximal to downstream regions. Furthermore, RP transcripts are destabilised during glucose starvation, suggesting that in the absence of promoter-proximal transcripts, surveillance activity is diverted to full-length mRNAs (Munchel et al, 2011). Thus downregulation of transcription and upregulation of nuclear surveillance might cooperate to rapidly downregulate genes in response to stress. I speculate that promoter-proximal transcription usually sequesters surveillance factors, whereas the clearance of promoter-proximal transcripts during stress results in surveillance factor release, promoting degradation of the full-length transcripts.

The CRAC results are apparently inconsistent with previous studies of stress responses in which downregulated genes undergo increased early termination (Darby et al, 2012), and upregulated genes less early termination (Kim & Levin, 2011; Yoon & Brem, 2010). However, the Sto1 CRAC data do not offer any insight into the abundance of short versus long transcripts, only the total abundance of transcripts from each TSS. It is therefore possible that, despite the continued presence of a 5' peak of Mtr4 binding in induced transcripts (Figure 6.3C), the abundance of full length versus truncated transcripts increases. Future work should address the association of HTP-tagged Pab1, to determine the ratio between Sto1 and Pab1 binding and thus the proportion of full length versus truncated mRNAs. It is also conceivable that under stress conditions, full-length mRNAs are very rapidly exported from the nucleus, potentially reducing their association with to the predominantly nuclear Sto1. However, this seems less likely.

The transcriptome-wide analyses of Mtr4 and Sto1 binding suggest that non-coding transcription plays a significant role in the response to changing environmental conditions. A major function of the nuclear surveillance machinery appears to be the degradation of the

various non-canonical transcripts associated with this regulation. The nuclear surveillance machinery also appears to be instrumental in ensuring rapid downregulation of highly expressed transcripts during the stress response, consistent with recent reports that highly expressed intron-containing genes are particularly prone to nuclear turnover (Gudipati, 2012). Furthermore, regulated nuclear turnover appears to contribute to modulating the expression of some genes (e.g. *FLO10*). These results indicate that RNA metabolism in the nucleus is substantially different from cytoplasmic turnover, and appears to perform very different roles.

7: Discussion

In addition to mRNAs, pervasive Pol II transcription generates abundant lncRNAs. Although lncRNAs and mRNAs resemble each other in some respects, their fates and functions are very different. I analysed the transcriptome-wide targets of a number of key factors in mRNA metabolism, to determine whether they also interact with lncRNAs, and to gain insight into how and when lncRNAs and mRNAs are distinguished. For this, I used the CRAC method, which is able to identify weak or transient *in vivo* RNA:protein interactions in actively growing cells due to the inclusion of a UV crosslinking step. This revealed that distinct classes of lncRNAs diverge from mRNAs at various stages, and that mRNA binding proteins perform both canonical and non-canonical roles in lncRNA biogenesis and turnover.

7.1 Distinct lncRNA classes are defined during 3' end processing

The early events in lncRNA and mRNA transcription have much in common, including assembly of a typical Pol II PIC (Rhee & Pugh, 2012), regulation by the same transcription factors (Bird et al, 2006; Houseley et al, 2008; Pinskaya et al, 2009; Xu et al, 2011), capping of the nascent transcript (Neil et al, 2009), and similar post-translational modifications of histones (Kim et al, 2012) and the Pol II CTD (Kim et al, 2010a). Consistent with this, I found the TREX components Tho2 and Gbp2 and the cap binding protein Sto1 bound similarly to lncRNAs and mRNAs (Figures 3.6, 4.1 and 5.1). Furthermore, the 3' end processing factors Nab2 and Hrp1 bound abundantly throughout lncRNAs and mRNAs, suggesting that they perform non-canonical roles early in transcription that are common to both transcript classes. These data are in agreement with ChIP analyses of Hrp1 and Nab2 binding (Gonzalez-Aguilera et al, 2011; Kim et al, 2004b). I conclude that lncRNAs and mRNAs bind a common set of proteins early in transcription, supporting a model in which similar events define the initial stages of lncRNA and mRNA biogenesis.

Despite these similarities in early transcription, lncRNAs and mRNAs differed greatly in the extent to which they bound decay factors (Figures 3.6 and 3.11). In contrast to mRNAs, lncRNAs bound ~9-fold more abundantly to TRAMP components than to Xrn1 and Ski2, suggesting that they diverge from mRNAs prior to export and are predominantly degraded in the nucleus. Analyses of lncRNAs bound to intermediate factors in mRNA metabolism revealed that different classes of lncRNA diverge from mRNAs at different stages. CUTs were least similar to mRNAs, lacking canonical polyadenylation signals (Figure 4.8) or poly(A) tails bound by Pab1 (Figure 4.4). Furthermore, CUTs were subject to negative regulation by Hrp1 (Figure 4.3), which was previously shown to promote mRNA poly(A) site selection. CUTs also bound ~5-fold less abundantly than mRNAs to the export receptor Mex67, when considering their nuclear abundance inferred from Sto1 binding (Figure 3.6). Together, this suggests that CUTs terminate via a mechanism that is distinct from mRNAs and bears little resemblance to canonical cleavage and polyadenylation. Nrd1/Nab3-dependent termination is the most likely candidate.

In contrast to CUTs, SUTs more closely resembled mRNAs. Many were bound at the 3' end by Pab1 (Figure 4.4) and possessed cleavage and polyadenylation signals similar to those of mRNAs (Figure 4.8). In other analyses, SUTs appeared intermediate between mRNAs and CUTs. For example, mRNAs bound abundantly to Mex67 (Figure 3.2) and possessed 3' UAUAUA motifs bound by Hrp1 (Figure 4.2) and poly(A) tails by Nab2 (Figure 5.7). For CUTs, these characteristics were not apparent, but for SUTs, I detected a low level of binding to Mex67 and (at 3' ends) to Hrp1. Together, this suggests that although SUTs are predominantly retained and degraded in the nucleus, their 3' end processing resembles that of mRNAs and thus they diverge at a later stage than CUTs. Many events in transcription are dependent on distance from the promoter, and the similar lengths of mRNAs and SUTs, in contrast to the generally shorter CUTs, might explain why their biogenesis pathways are more similar. It is not apparent what causes SUTs and mRNAs to diverge during or shortly

after 3' end processing, although aberrant RNP assembly and/or different post-translational modifications of RNP components might contribute.

In addition to the independently transcribed CUTs and SUTs, analyses of CRAC hits mapping to protein coding genes revealed that some proteins were bound abundantly to promoter-proximal transcripts, which apparently constitute an additional class of lncRNAs. In many respects, the promoter-proximal transcripts resemble independently transcribed CUTs. For example, they possessed short oligo(A) tails and were bound to TRAMP components, and to Hrp1 and Nab2, but not Xrn1, Ski2 or Mex67 (Figures 5.1 and 5.7), suggesting they are predominantly nuclear and highly unstable. Indeed, shortly after yeast cultures were transferred to media lacking glucose, many of the promoter-proximal fragments were no longer detectable, indicative of rapid turnover (Figure 6.3).

Despite these similarities, closer inspection of the data suggested that promoter-proximal transcripts differ from CUTs in some respects; terminating closer to the TSS and/or being degraded by a different mechanism. Firstly, 5' peaks of Nab2, Mtr4 and Trf4 binding extend only ~100-200 nt downstream from annotated mRNA TSSs (Figures 5.7 and 5.8). Promoter-proximal transcripts are apparently undergoing active degradation, so from these data alone it is not clear whether they terminate further downstream. However, in *rrp6Δ trf4*Δ strains, independently transcribed CUTs and promoter-proximal transcripts are stabilised, and their modal 3' end positions map to ~450 nt and ~120 nt downstream of the CUT and mRNA TSSs, respectively (Figure 4.4D) (Neil et al, 2009). Furthermore, in *rrp6Δ* strains (Figure 4.4), Nab2 hits mostly remained within ~200 nt of the TSS for mRNAs, but extended throughout CUTs. These data indicate that full length CUTs are stabilised by loss of Rrp6, whereas promoter-proximal transcripts accumulate but are not extended. This suggests that promoter-proximal transcripts either (i) terminate within ~100-200 bp of the mRNA TSS, or (ii) terminate further downstream, and are then degraded in a complex manner, distinct from CUTs, perhaps involving handover to Rrp6 for the final ~150 nt. Besides offering insight

into where promoter-proximal transcripts terminate, the CRAC analyses in this study provide evidence that they initiate at the same TSSs as mRNAs. Most notably, the upstream boundary of Mtr4, Trf4, Nab2 and Sto1 binding is predominantly within ~20 nt of the associated mRNA TSS, even in an *rrp6Δ* strain (for Nab2) or when an enzymatic decapping step is included in the CRAC protocol (for Sto1) (Figure 5.1). This apparently conflicts with 5' RACE data for 16 sense-oriented promoter-proximal transcripts, which reportedly initiate ~80 bp upstream of mRNA TSSs (Neil et al, 2009). Inspection of the CRAC data for these 16 loci (data not shown) reveals a low level of Mtr4, Nab2, Trf4 and Sto1 binding in this upstream region, but much greater binding at the mRNA TSS. To reconcile these data, I suggest that promoter-proximal transcription initiates mostly from mRNA TSSs, but also at a lower level from heterogeneous TSSs up to ~100 bp upstream. This is reminiscent of heterogeneous, promoter-associated transcription in higher eukaryotes (Kanhere et al, 2010; Kapranov et al, 2007; Seila et al, 2008; Taft et al, 2009; Taft et al, 2011). In summary, promoter-associated lncRNAs in yeast initiate at or near to mRNA TSSs, terminate ~100-150 bp downstream, and are ~150-200 nt long. CUTs are longer (~450 nt) and subject to a partially distinct turnover mechanism, and I therefore speculate that CUTs and promoter-proximal lncRNAs constitute distinct classes.

The apparent termination of promoter-proximal transcripts ~150 bp downstream of many Pol II TSSs might reflect a universal checkpoint or regulatory step in the yeast Pol II transcription cycle. In metazoa, a variety of sense-oriented, promoter-proximal ncRNAs are produced, ranging from 18 nt to ~1 kb and initiating at or just upstream of the mRNA TSS (Figure 1.1, and references therein). Although some of these reflect Pol II pausing at +30 to +60 directed by DSIF and NELF, the latter of which is absent from yeast, other promoter-proximal ncRNAs apparently arise from obstacles to elongation further downstream, such as nucleosomes. It is therefore possible that events in the yeast Pol II transcription cycle also lead to stalling and termination at ~+150 bp. Indeed, transitions in chromatin structure and

Pol II properties reportedly occur in this region. Firstly, Pol II is prone to pause just before the dyad axis of the +2 nucleosome (Churchman et al, 2011), which is typically ~200 bp downstream of the TSS (Rhee et al, 2012). Secondly, many elongation factors are only fully recruited to Pol II ~150 bp downstream of the TSS. These include the histone chaperones Spt6 and Spt16, which facilitate elongation past nucleosomes (Mayer et al, 2010). Thirdly, marks of active chromatin such as H3K4me3 and histone acetylation are high across and just downstream of the promoter, but less abundant further downstream (Kirmizis et al, 2007; Pokholok et al, 2005). Together, this suggests that Pol II is relatively unimpeded for the first ~100-200 bps, but might require a full complement of elongation factors to successfully negotiate stable nucleosomes further downstream. In this model, a Pol II pause around +150 nt would coincide with a remodelling event in which initiation factors are exchanged for elongation factors. If this remodelling was aberrant or incomplete, the elongating Pol II complex would be destabilised and terminate, generating a promoter-proximal lncRNA. This putative checkpoint might prevent transcription blockages further downstream, which would be more wasteful and perhaps harder to resolve. As discussed in Chapter 5, promoter-proximal termination might proceed via the Nrd1-dependent pathway or an alternative mechanism, perhaps involving Rat1 and an endonuclease.

In conclusion, I propose the model shown in Figure 7.1. LncRNA transcription initiates at or near the mRNA TSS, and early events in mRNA and lncRNA transcription are similar (TREX, Hrp1 and Nab2 recruitment). The Pol II holoenzyme might undergo a remodelling step after traversing the first ~150 bp that, if successful, facilitates elongation past downstream nucleosomes. However, some polymerases are destabilised at this point and undergo promoter-proximal termination, perhaps dependent on Nrd1 or endonucleolytic cleavage and Rat1 recruitment, which generates ~100-200 nt promoter-proximal lncRNAs. CUTs are produced by Nrd1-dependent termination within ~500 bp of the promoter. Both

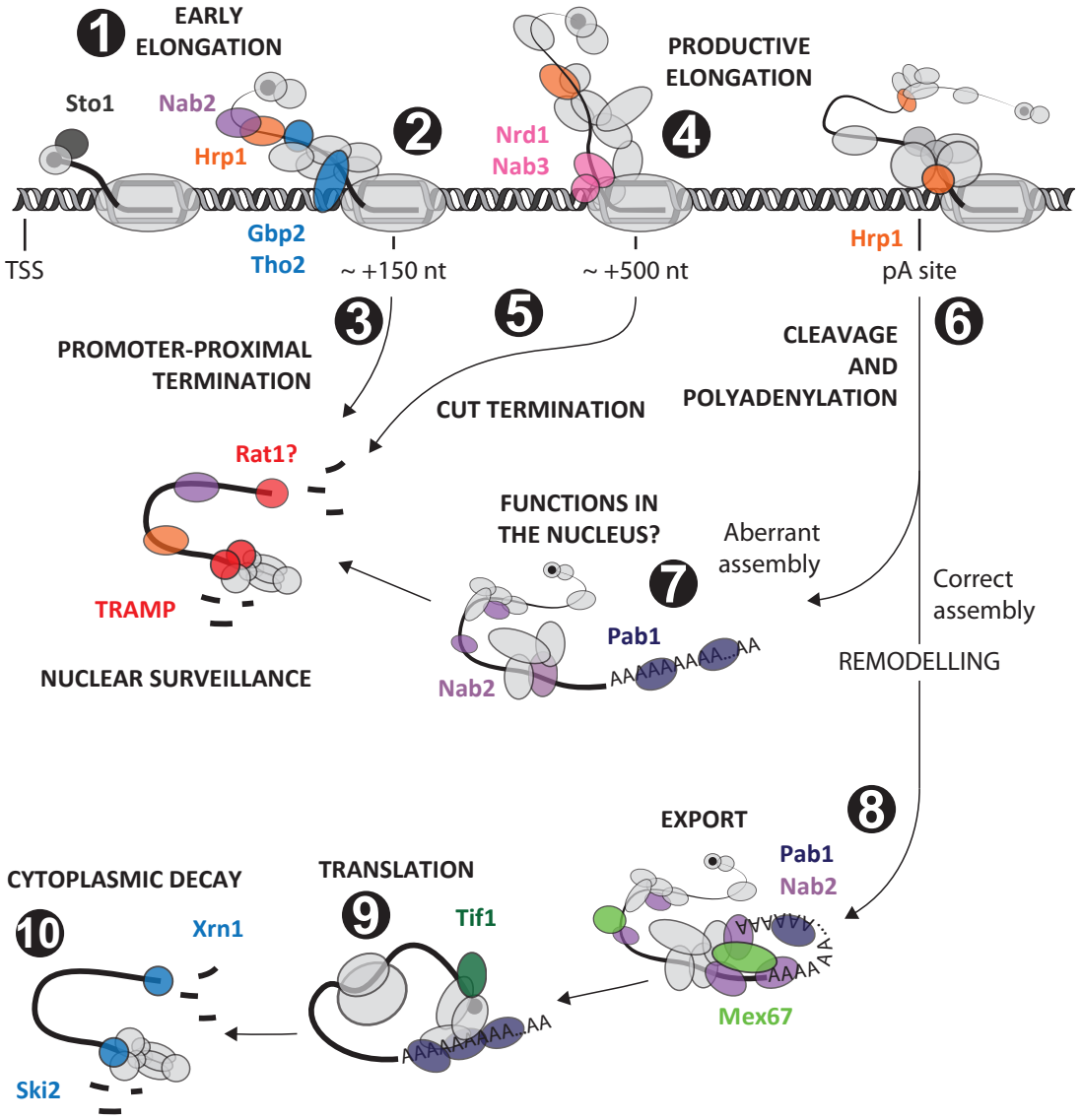


Figure 7.1: Model for lncRNA and mRNA biogenesis and turnover. 1 LncRNA transcription initiates at or near an mRNA TSS or from a dedicated promoter. Early events in mRNA and lncRNA transcription are similar. **2** Pol II pauses ~150 bp downstream of the TSS, and remodelling of the holoenzyme is important for penetration of chromatin downstream. **3** Any polymerase failing to attain full elongation competency is ejected from the template, via an unknown mechanism, producing a promoter-proximal lncRNA. **4** Elongation continues, and **5** for some genes (CUTs) is terminated ~500 bp downstream of the TSS by Nrd1-dependent termination. For mRNAs and SUTs, transcription continues further downstream, and **6** 3' end processing is performed by the CF/CPF machinery upon recognition of cleavage and polyadenylation signals. **7** During or shortly after 3' end processing, SUT lncRNPs are recognised as aberrant, perhaps because in comparison to mRNPs they lack the full complement of protein factors or the appropriate PTMs. SUTs are thus retained in the nucleus, where they might function, and are eventually degraded. **8** Conversely, mRNAs "correctly" assemble into mRNPs and undergo remodelling, during which Nab2-assisted folding might regulate poly(A) tail length, and Mex67 is recruited. **9** mRNAs are thus exported and translated, and 10 eventually turned over by the cytoplasmic decay machinery.

CUTs and promoter-proximal transcripts are rapidly turned over by the nuclear exosome assisted by TRAMP, but CUTs are more dependent on Rrp6. Both mRNAs and SUTs terminate further downstream, via cleavage and polyadenylation. For mRNAs, this is accompanied by a remodelling step that renders the mRNP competent for export and involves Nab2, which recruits Mex67 and interacts with the poly(A) tail to regulate its length. The mRNA is then translated in the cytoplasm and eventually degraded by Xrn1 and/or the cytoplasmic exosome. Conversely, SUTs lack either the requisite post-translational modifications or a full complement of 3' end processing factors, and thus assemble into RNPs that are detected as aberrant and retained in the nucleus. At least some SUTs presumably function within the nucleus, before eventually being degraded by TRAMP and the exosome.

In future work it will be important to more fully characterise promoter-proximal transcripts and independently transcribed CUTs, to establish whether these are indeed distinct classes of lncRNA. A lncRNA-enriched fraction from yeast lysates can be obtained via affinity purification of CBC components with bound RNAs (Neil et al, 2009) (this study). In combination with surveillance mutants, this might enable the lengths, genomic coordinates, and turnover pathways to be distinguished for CUTs and promoter-proximal transcripts. During the work reported here, termination and 3' end processing have emerged as key steps during which transcripts diverge and assume different fates. CRAC analyses of additional 3' end processing factors might help resolve precisely when and how transcript classes diverge. To this end, I have obtained preliminary data for Rna14 and Cft2, and these analyses are ongoing. A more comprehensive analysis of the sequence elements associated with lncRNA 3' ends is also required, and I am now compiling 3' end coordinates from a variety of sources to facilitate this analysis. Finally, a reciprocal study in which lncRNAs are used as bait to identify novel protein partners would help define the full repertoire of proteins participating in lncRNA metabolism.

7.2 Non-coding regulators

The results in this study highlight the complexity of transcripts that can be associated with a single genomic locus, including full-length mRNAs, stable or unstable lncRNAs, and shorter promoter-proximal transcripts. Furthermore, changes in gene expression often coincide with changes in the relative abundance of these various species (Figure 6.3), suggesting that the "inner workings" of even simple genetic switches are surprisingly complex. Several results in this study suggest that lncRNAs are well suited to play non-canonical roles in gene regulatory circuits. Firstly, they are predominantly nuclear (Figure 3.11), and so can function via mechanisms such as hybridising to complementary nucleic acid targets without directly affecting translation in the cytoplasm. Secondly, they can be rapidly degraded (e.g. Figure 6.3) so are ideal for performing short-lived roles, perhaps participating in complex regulatory cascades where each step must be completed quickly. Rapid turnover also enables lncRNAs to function as rapid switches during the response to stress or changing environmental conditions, and ensures that any lncRNAs generated as byproducts of transcription-dependent regulatory mechanisms are quickly cleared. Thirdly, non-coding transcription can be used to alter the transcriptional output of a locus after Pol II has bound, either via selection of alternative TSSs or non-productive promoter-proximal termination. Promoter-proximal termination might also provide a way to eject aberrantly assembled elongation complexes, perhaps acting during a kinetic proofreading step in which a long pause in elongation is interpreted as aberrant and triggers termination. Inspection of transcriptome-wide Nab2 and Mtr4 binding suggests that ~1000 genes are subject to early termination, and in each case, ~50 % of elongating Pol II complexes are affected (Figures 4.5 and 5.9). Finally, the abundant interactions between mRNA biogenesis factors and lncRNAs suggests that lncRNA levels might globally regulate the availability of these factors for canonical pre-mRNA transcription and packaging. This potential role is highlighted by the behaviour of Mtr4 upon nutritional down shift (Figure 6.3); Mtr4 is predominantly bound to promoter-

proximal transcripts before down shift, but becomes active on full-length mRNAs following glucose withdrawal. LncRNAs might therefore act as a "sponge" to sequester processing and turnover factors, and release them either in a regulated way to nearby loci, or perhaps as part of a global response to stress.

7.3 Non-canonical surveillance targets

In addition to providing insight into the biogenesis, localisation and turnover of lncRNAs, the CRAC results in this study reveal that the nuclear and cytoplasmic decay machineries perform strikingly different roles. Cytoplasmic decay factors are predominantly occupied with the turnover of mature, full-length transcripts (Figures 5.1 and 5.2), whereas the nuclear surveillance machinery is occupied with abundant lncRNAs, including CUTs, SUTs and promoter-proximal transcripts (Figures 3.6 and 5.8). This suggests that there is a very high turnover of non-canonical transcripts in the nucleus. The rapid progress currently being made in understanding the biogenesis, function and turnover of these transcripts is likely to reveal many surprising feature of nuclear RNA turnover.

An important and largely unanswered question is, how are transcripts targeted for decay? The results in this study provide some valuable insights. Firstly, the high proportion of intron-containing transcripts amongst TRAMP targets (Figure 5.2) supports a model in which nuclear surveillance occurs co-transcriptionally or shortly after transcription. Furthermore, Gbp2, Nab2 and Hrp1 similarly show early binding (Figure 5.2) to most, if not all, Pol II transcripts (Figure 3.2). All three of these factors are implicated in nuclear surveillance: Hrp1 depletion results in upregulation of lncRNAs (Figure 4.3), the *Drosophila* homologue of Gbp2 recruits the exosome to nascent transcripts (Hessle et al, 2009), and Nab2 participates in surveillance of intron-containing pre-mRNAs (Schmid et al, 2012). This suggests that nuclear surveillance factors are recruited to transcripts co-transcriptionally, perhaps via interactions with pervasive RNA binding proteins. However, as discussed above, transcript fate is predominantly determined by the mode of termination. Perhaps the

surveillance machinery, once recruited, waits until 3' end processing and termination, at which point a combination of protein factors, RNA folding and post-translational modifications are interrogated to determine whether the transcript should be exported, immediately degraded or transiently retained in the nucleus.

Finally, the analyses of non-encoded A-tails highlight their importance as "hubs" that regulate and coordinate surveillance. Approximately one third of TRAMP substrates were found to possess non-encoded A-tails (Figure 5.3), and this estimate is conservative as RNA cleavage prior to linker ligation is likely to remove many A-tails. Furthermore, A-tails were prevalent amongst all tested TRAMP substrates (Figures 5.4 and 5.8), suggesting that A-tails are important for the decay of most, if not all, nuclear surveillance substrates. The prevalent generation of transcripts by non-canonical termination (e.g. promoter-proximal termination in protein coding genes) also explains the requirement for non-canonical poly(A) polymerases (Trf4 and Trf5) in the nucleus, as the majority of nuclear surveillance substrates are apparently not accessible to Pap1-dependent adenylation. The cytoplasmic decay factors Ski2 and Xrn1 abundantly bound to oligo(A) tails at the 3' end of mRNAs (Figure 5.6). This is consistent with the documented role of deadenylation in displacing poly(A) binding proteins to expose an oligo(A) tail to Ski2 and decapping activators along with Xrn1. The length distribution of Ski2-associated A-tails is in very good agreement with previous reports that mRNA deadenylation proceeds to A_{12}-A_{10} prior to binding of the 5' and 3' degradation machinery (Decker et al, 1993). However, the high level of Ski2 and Xrn1 binding to 3' oligo(A) tails suggests that much of their time is spent there, and that the poly(A) tail is the site of a rate-limiting checkpoint that presumably ensures decay is tightly controlled. Together, therefore, A-tails are key regulators of both nuclear and cytoplasmic decay.

In summary, I have presented a transcriptome-wide analysis of the targets of mRNA binding proteins, which has provided insights into when and how lncRNAs and mRNAs are

distinguished, how coding and non-coding transcription interact, and how the timely degradation of lncRNAs and mRNAs is achieved.

Bibliography

Abou Elela S, Ares M (1998) Depletion of yeast RNase III blocks correct U2 3' end formation and results in polyadenylated but functional U2 snRNA. *EMBO J* **17:** 3738-3746

Ahn SH, Kim M, Buratowski S (2004) Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3′ End Processing. *Mol Cell* **13:** 67-76

Aitchison JD, Blobel G, Rout MP (1996) Kap104p: a karyopherin involved in the nuclear transport of messenger RNA binding proteins. *Science* **274:** 624-627

Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ (2009) TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* **34:** 387-393

Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D (1999a) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J* **18:** 5399-5410

Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D, Mitchell P (1999b) The yeast exosome and human PM–Scl are related complexes of 3′ → 5′ exonucleases. *Genes & Development* **13:** 2148-2158

Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A (2004) A faux 3'-UTR promotes aberrant termination and triggers nonsense- mediated mRNA decay. *Nature* **432:** 112-118

Amrani N, Minet M, Le Gouar M, Lacroute F, Wyers F (1997) Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. *Mol Cell Biol* **17:** 3694-3701

Anderson JS, Parker RP (1998) The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J* **17:** 1497-1506

Anderson JT, Wilson SM, Datar KV, Swanson MS (1993) NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. *Mol Cell Biol* **13:** 2730-2741

Andrulis ED, Werner J, Nazarian A, Erdjument-Bromage H, Tempst P, Lis JT (2002) The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. *Nature* **420:** 837-841

Araki Y, Takahashi S, Kobayashi T, Kajiho H, Hoshino S-i, Katada T (2001) Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast. *EMBO J* **20:** 4684-4693

Arigo JT, Carroll KL, Ames JM, Corden JL (2006a) Regulation of Yeast NRD1 Expression by Premature Transcription Termination. *Mol Cell* **21:** 641-651

Arigo JT, Eyler DE, Carroll KL, Corden JL (2006b) Termination of Cryptic Unstable Transcripts Is Directed by Yeast RNA-Binding Proteins Nrd1 and Nab3. *Mol Cell* **23:** 841-851

Assenholt J, Mouaikel J, Andersen KR, Brodersen DE, Libri D, Jensen TH (2008) Exonucleolysis is required for nuclear mRNA quality control in yeast THO mutants. *RNA* **14:** 2305-2313

Assenholt J, Mouaikel J, Saguez C, Rougemaille M, Libri D, Jensen TH (2011) Implication of Ccr4-Not complex function in mRNA quality control in Saccharomyces cerevisiae. *RNA* **17:** 1788-1794

Audas Timothy E, Jacob Mathieu D, Lee S (2012) Immobilization of Proteins in the Nucleolus by Ribosomal Intergenic Spacer Noncoding RNA. *Mol Cell* **45:** 147-157

Azzouz N, Panasenko OO, Colau G, Collart MA (2009) The CCR4-NOT complex physically and functionally interacts with TRAMP and the nuclear exosome. *PLoS One* **4:** e6760

Bailey TL (2011) DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27:** 1653-1659

Ballarino M, Morlando M, Pagano F, Fatica A, Bozzoni I (2005) The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in Saccharomyces cerevisiae. *Mol Cell Biol* **25:** 5396-5403

Banfai B, Jia H, Khatun J, Wood E, Risk B, Gundling WE, Jr., Kundaje A, Gunawardena HP, Yu Y, Xie L, Krajewski K, Strahl BD, Chen X, Bickel P, Giddings MC, Brown JB, Lipovich L (2012) Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res* **22:** 1646-1657

Bantignies F, Roure V, Comet I, Leblanc B, Schuettengruber B, Bonnet J, Tixier V, Mas A, Cavalli G (2011) Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell* **144:** 214-226

Bataille Alain R, Jeronimo C, Jacques P-É, Laramée L, Fortin M-È, Forest A, Bergeron M, Hanes Steven D, Robert F (2012) A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. *Mol Cell* **45:** 158-170

Batisse J, Batisse C, Budd A, Böttcher B, Hurt E (2009) Purification of Nuclear Poly(A)-binding Protein Nab2 Reveals Association with the Yeast Transcriptome and a Messenger Ribonucleoprotein Core Structure. *J Biol Chem* **284:** 34911-34917

Becker T, Armache J-P, Jarasch A, Anger AM, Villa E, Sieber H, Motaal BA, Mielke T, Berninghausen O, Beckmann R (2011) Structure of the no-go mRNA decay complex Dom34–Hbs1 bound to a stalled 80S ribosome. *Nat Struct Mol Biol* **18:** 715-720

Belew AT, Advani VM, Dinman JD (2011) Endogenous ribosomal frameshift signals operate as mRNA destabilizing elements through at least two molecular pathways in yeast. *Nucleic Acids Res* **39:** 2799-2808

Beniaminov A, Westhof E, Krol A (2008) Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA* **14:** 1270-1275

Berg Michael G, Singh Larry N, Younis I, Liu Q, Pinto Anna M, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L, Dreyfuss G (2012) U1 snRNP Determines mRNA Length and Regulates Isoform Expression. *Cell* **150:** 53-64

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

Biffi G, Tannahill D, Balasubramanian S (2012) An Intramolecular G-Quadruplex Structure Is Required for Binding of Telomeric Repeat-Containing RNA to the Telomeric Protein TRF2. *J Am Chem Soc* **134:** 11974-11976

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

Boeck R, Tarun SJ, Rieger M, Deardorff JA, Müller-Auer S, Sachs AB (1996) The Yeast Pan2 Protein Is Required for Poly(A)-binding Protein-stimulated Poly(A)-nuclease Activity. *J Biol Chem* **271:** 432-438

Bonneau F, Basquin J, Ebert J, Lorentzen E, Conti E (2009) The Yeast Exosome Functions as a Macromolecular Cage to Channel RNA Substrates for Degradation. *Cell* **139:** 547-559

Brachmann C, Davies A, Cost G, Caputo E, Li J, Hieter P, Boeke J (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14:** 115-132

Brannan K, Kim H, Erickson B, Glover-Cutter K, Kim S, Fong N, Kiemele L, Hansen K, Davis R, Lykke-Andersen J, Bentley D (2012) mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Mol Cell* **46:** 311-324

Brock HW, Hodgson JW, Petruk S, Mazo A (2009) Regulatory noncoding RNAs at Hox loci. *Biochem Cell Biol* **87:** 27-34

Brockmann C, Soucek S, Kuhlmann SI, Mills-Lujan K, Kelly SM, Yang JC, Iglesias N, Stutz F, Corbett AH, Neuhaus D, Stewart M (2012) Structural basis for polyadenosine-RNA binding by Nab2 Zn fingers and its function in mRNA nuclear export. *Structure* **20:** 1007-1018

Brodsky AS, Silver PA (2000) Pre-mRNA processing factors are required for nuclear export. *RNA* **6:** 1737-1749

Brown CE, Sachs AB (1998) Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. *Mol Cell Biol* **18:** 6548-6559

Brown JT, Bai X, Johnson AW (2000) The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. *RNA* **6:** 449-457

Brune C, Munchel SE, Fischer N, Podtelejnikov AV, Weis K (2005) Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export. *RNA* **11:** 517-531

Buchan JR, Yoon JH, Parker R (2011) Stress-specific composition, assembly and kinetics of stress granules in Saccharomyces cerevisiae. *J Cell Sci* **124:** 228-239

Bucheli ME, He X, Kaplan CD, Moore CL, Buratowski S (2007) Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. *RNA* **13:** 1756-1764

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

Bumgarner SL, Neuert G, Voight FR, Symbor-Nagrabska A, Grisafi P, van Oudenaarden A, Fink GR (2012) Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment. *Mol Cell* **45:** 1-13

Burkard KTD, Butler JS (2000) A Nuclear 3′-5′ Exonuclease Involved in mRNA Degradation Interacts with Poly(A) Polymerase and the hnRNA Protein Npl3p. *Mol Cell Biol* **20:** 604-616

Cairns BR (2009) The logic of chromatin architecture and remodelling at promoters. *Nature* **461:** 193-198

Callahan KP, Butler JS (2008) Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p. *Nucl Acids Res* **36:** 6645-6655

Callahan KP, Butler JS (2010) TRAMP Complex Enhances RNA Degradation by the Nuclear Exosome Component Rrp6. *J Biol Chem* **285:** 3540-3547

Camblong J, Beyrouthy N, Guffanti E, Schlaepfer G, Steinmetz LM, Stutz F (2009) Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. *Genes Dev* **23:** 1534-1545

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

Cao D, Parker R (2001) Computational modeling of eukaryotic mRNA turnover. *RNA* **7:** 1192-1212

Carmody SR, Tran EJ, Apponi LH, Corbett AH, Wente SR (2010) The mitogen-activated protein kinase Slt2 regulates nuclear retention of non-heat shock mRNAs during heat shock-induced stress. *Mol Cell Biol* **30:** 5168-5179

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest AR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, Ambesi-Impiombato A, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, Della Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T, Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminiecki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasawa Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SP, Kruger A, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Madan Babu M, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rost B, Ruan Y, Salzberg SL, Sandelin A, Schneider C, Schonbach C, Sekiguchi K, Semple CA, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammoja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegen E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusic V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M, Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J, Hayashizaki Y (2005) The transcriptional landscape of the mammalian genome. *Science* **309:** 1559-1563

Carrillo Oesterreich F, Preibisch S, Neugebauer KM (2010) Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* **40:** 571-581

Carroll KL, Ghirlando R, Ames JM, Corden JL (2007) Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA* **13:** 361-373

Carroll KL, Pradhan DA, Granek JA, Clarke ND, Corden JL (2004) Identification of cis Elements Directing Termination of Yeast Nonpolyadenylated snoRNA Transcripts. *Mol Cell Biol* **24:** 6241-6252

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann Benedikt M, Strein C, Davey Norman E, Humphreys David T, Preiss T, Steinmetz Lars M, Krijgsveld J, Hentze Matthias W (2012) Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149:** 1393-1406

Castells-Roca L, Garcia-Martinez J, Moreno J, Herrero E, Belli G, Perez-Ortin J (2011) Heat Shock Response in Yeast Involves Changes in Both Transcription Rates and mRNA Stabilities. *PLoS One* **6:** e17272

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I (2011) A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA. *Cell* **147:** 358-369

Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H, Conti E (2011) Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2. *Mol Cell* **41:** 693-703

Chanarat S, Seizl M, Strasser K (2011) The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes Dev* **25:** 1147-1158

Chang JH, Jiao X, Chiba K, Oh C, Martin CE, Kiledjian M, Tong L (2012) Dxo1 is a new type of eukaryotic enzyme with both decapping and 5′-3′ exoribonuclease activity. *Nat Struct Mol Biol* **19:** 1011-1017

Chavez S, Beilharz T, Rondon AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, Aguilera A (2000) A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. *EMBO J* **19:** 5824-5834

Chen LL, Carmichael GG (2009) Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell* **35:** 467-478

Chen S, Hyman LE (1998) A specific RNA-protein interaction at yeast polyadenylation efficiency elements. *Nucleic Acids Res* **26:** 4965-4974

Cheng H, He X, Moore C (2004a) The Essential WD Repeat Protein Swd2 Has Dual Functions in RNA Polymerase II Transcription Termination and Lysine 4 Methylation of Histone H3. *Mol Cell Biol* **24:** 2932-2943

Cheng Z, Liu Y, Wang C, Parker R, Song H (2004b) Crystal structure of Ski8p, a WD-repeat protein with dual roles in mRNA metabolism and meiotic recombination. *Protein Sci* **13:** 2673-2684

Chernyakov I, Whipple JM, Kotelawala L, Grayhack EJ, Phizicky EM (2008) Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. *Genes Dev* **22:** 1369-1380

Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, Hughes TR, Winston F (2008) Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the *Saccharomyces cerevisiae* Genome. *PLoS Biol* **6:** e277

Chowdhury A, Mukhopadhyay J, Tharun S (2007) The decapping activator Lsm1p-7p–Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* **13:** 998-1016

Chu C, Qu K, Zhong Franklin L, Artandi Steven E, Chang Howard Y (2011) Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol Cell* **44:** 667-678

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

Ciais D, Bohnsack MT, Tollervey D (2008) The mRNA encoding the yeast ARE-binding protein Cth2 is generated by a novel 3' processing pathway. *Nucleic Acids Res* **36:** 3075-3084

Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC (2005) Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol* **23:** 232-237

Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* **33:** 717-726

Conrad T, Akhtar A (2012a) Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet* **13:** 123-134

Conrad T, Cavalli FMG, Vaquerizas JM, Luscombe NM, Akhtar A (2012b) Drosophila Dosage Compensation Involves Enhanced Pol II Recruitment to Male X-Linked Promoters. *Science* **337:** 742-746

Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* **322:** 1845-1848

Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, Wheelan SJ, Corden JL (2011) Transcriptome-Wide Binding Sites for Components of the Saccharomyces cerevisiae Non-Poly(A) Termination Pathway: Nrd1, Nab3, and Sen1. *PLoS Genet* **7:** e1002329

Cristodero M, Bottcher B, Diepholz M, Scheffzek K, Clayton C (2008) The Leishmania tarentolae exosome: purification and structural analysis by electron microscopy. *Mol Biochem Parasitol* **159:** 24-29

D'Souza V, Summers MF (2005) How retroviruses select their genomes. *Nat Rev Micro* **3:** 643-655

Darby MM, Serebreni L, Pan X, Boeke JD, Corden JL (2012) The Saccharomyces cerevisiae Nrd1-Nab3 Transcription Termination Pathway Acts in Opposition to Ras Signaling and Mediates Response to Nutrient Depletion. *Mol Cell Biol* **32:** 1762-1775

Das B, Das S, Sherman F (2006) Mutant LYS2 mRNAs retained and degraded in the nucleus of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **103:** 10871-10876

Das B, Guo Z, Russo P, Chartrand P, Sherman F (2000) The Role of Nuclear Cap Binding Protein Cbc1p of Yeast in mRNA Termination and Degradation. *Mol Cell Biol* **20:** 2827-2838

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

Davidson L, Kerr A, West S (2012) Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *EMBO J* **31:** 2566-2578

Davis CA, Ares M, Jr. (2006) Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **103:** 3262-3267

de Hoon MJL, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* **20:** 1453-1454

de la Cruz J, Kressler D, Tollervey D, Linder P (1998) Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae. *EMBO J* **17:** 1128-1140

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei C-L, Natoli G (2010) A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. *PLoS Biol* **8:** e1000384

Decker CJ, Parker R (1993) A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes & Development* **7:** 1632-1643

Dengl S, Cramer P (2009) Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem* **284:** 21270-21279

Dermody JL, Dreyfuss JM, Villén J, Ogundipe B, Gygi SP, Park PJ, Ponticelli AS, Moore CL, Buratowski S, Bucheli ME (2008) Unphosphorylated SR-Like Protein Npl3 Stimulates RNA Polymerase II Elongation. *PLoS One* **3:** e3273

Deshmukh MV, Jones BN, Quang-Dang DU, Flinders J, Floor SN, Kim C, Jemielity J, Kalek M, Darzynkiewicz E, Gross JD (2008) mRNA decapping is promoted by an RNA-binding channel in Dcp2. *Mol Cell* **29:** 324-336

Dez C, Houseley J, Tollervey D (2006) Surveillance of nuclear-restricted pre-ribosomes within a subnucleolar region of Saccharomyces cerevisiae. *EMBO J* **25:** 1534-1546

Dheur S, Nykamp KR, Viphakone N, Swanson MS, Minvielle-Sebastia L (2005) Yeast mRNA Poly(A) tail length control can be reconstituted in vitro in the absence of Pab1p-dependent Poly(A) nuclease activity. *J Biol Chem* **280:** 24532-24538

Dheur S, Vo LTA, Voisinet-Hakil F, Minet M, Schmitter J-M, Lacroute F, Wyers F, Minvielle-Sebastia L (2003) Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation. *EMBO J* **22:** 2831-2840

Dichtl B, Aasland R, Keller W (2004) Functions for S. cerevisiae Swd2p in 3′ end formation of specific mRNAs and snoRNAs and global histone 3 lysine 4 methylation. *RNA* **10:** 965-977

Dichtl B, Blank D, Ohnacker M, Friedlein A, Roeder D, Langen H, Keller W (2002a) A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Mol Cell* **10:** 1139-1150

Dichtl B, Blank D, Sadowski M, Hubner W, Weiser S, Keller W (2002b) Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J* **21:** 4125-4135

Dichtl B, Keller W (2001) Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *EMBO J* **20:** 3197-3209

Dolken L, Ruzsics Z, Radle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P, Koszinowski UH (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14:** 1959-1972

Dower K, Kuperwasser N, Merrikh H, Rosbash M (2004) A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. *RNA* **10:** 1888-1899

Duncan K, Umen JG, Guthrie C (2000) A putative ubiquitin ligase required for efficient mRNA export differentially affects hnRNP transport. *Curr Biol* **10:** 687-696

Dziembowski A, Lorentzen E, Conti E, Seraphin B (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol* **14:** 15-22

Ebisuya M, Yamamoto T, Nakajima M, Nishida E (2008) Ripples from neighbouring transcription. *Nat Cell Biol* **10:** 1106-1113

El Hage A, French SL, Beyer AL, Tollervey D (2010) Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes & Development* **24:** 1546-1558

ENCODE (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799-816

Ezeokonkwo C, Zhelkovsky A, Lee R, Bohm A, Moore CL (2011) A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation. *RNA* **17:** 652-664

Fasken MB, Stewart M, Corbett AH (2008) Functional Significance of the Interaction between the mRNA-binding Protein, Nab2, and the Nuclear Pore-associated Protein, Mlp1, in mRNA Export. *J Biol Chem* **283:** 27130-27143

Fatica A, Morlando M, Bozzoni I (2000) Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3[prime]-processing apparatus. *EMBO J* **19:** 6218-6229

Faza MB, Chang Y, Occhipinti L, Kemmler S, Panse VG (2012) Role of Mex67-Mtr2 in the Nuclear Export of 40S Pre-Ribosomes. *PLoS Genet* **8:** e1002915

Faza MB, Kemmler S, Jimeno S, Gonzalez-Aguilera C, Aguilera A, Hurt E, Panse VG (2009) Sem1 is a functional component of the nuclear pore complex-associated messenger RNA export machinery. *J Cell Biol* **184:** 833-846

Finkel JS, Chinchilla K, Ursic D, Culbertson MR (2010) Sen1p performs two genetically separable functions in transcription and processing of U5 small nuclear RNA in Saccharomyces cerevisiae. *Genetics* **184:** 107-118

Fischer T, Strasser K, Racz A, Rodriguez-Navarro S, Oppizzi M, Ihrig P, Lechner J, Hurt E (2002) go. *EMBO J* **21:** 5843-5852

Flach J, Bossie M, Vogel J, Corbett A, Jinks T, Willins DA, Silver PA (1994) A yeast RNA-binding protein shuttles between the nucleus and the cytoplasm. *Mol Cell Biol* **14:** 8399-8407

Flynn RA, Almada AE, Zamudio JR, Sharp PA (2011) Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proceedings of the National Academy of Sciences* **108:** 10460-10465

Fortes P, Inada T, Preiss T, Hentze MW, Mattaj IW, Sachs AB (2000) The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. *Mol Cell* **6:** 191-196

Funakoshi Y, Doi Y, Hosoda N, Uchida N, Osawa M, Shimada I, Tsujimoto M, Suzuki T, Katada T, Hoshino S-i (2007) Mechanism of mRNA deadenylation: evidence for a molecular interplay between translation termination factor eRF3 and mRNA deadenylases. *Genes & Development* **21:** 3135-3148

Gallardo M, Luna R, Erdjument-Bromage H, Tempst P, Aguilera A (2003) Nab2p and the Thp1p-Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. *J Biol Chem* **278:** 24225-24232

Galy V, Gadal O, Fromont-Racine M, Romano A, Jacquier A, Nehrbass U (2004) Nuclear Retention of Unspliced mRNAs in Yeast Is Mediated by Perinuclear Mlp1. *Cell* **116:** 63-73

Ganem C, Devaux F, Torchet C, Jacq C, Quevillon-Cheruel S, Labesse G, Facca C, Faye G (2003) Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *EMBO J* **22:** 1588-1598

Gao Q, Das B, Sherman F, Maquat LE (2005) Cap-binding protein 1-mediated and eukaryotic translation initiation factor 4E-mediated pioneer rounds of translation in yeast. *Proc Natl Acad Sci U S A* **102:** 4258-4263

Garas M, Dichtl B, Keller W (2008) The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. *RNA* **14:** 2671-2684

Garcia-Martinez J, Ayala G, Pelechano V, Chavez S, Herrero E, Perez-Ortin JE (2012) The relative importance of transcription rate, cryptic transcription and mRNA stability on shaping stress responses in yeast. *Transcription* **3:** 39-44

Garre E, Romero-Santacreu L, De Clercq N, Blasco-Angulo N, Sunnerhagen P, Alepuz P (2012) Yeast mRNA cap-binding protein Cbc1/Sto1 is necessary for the rapid reprogramming of translation after hyperosmotic shock. *Mol Biol Cell* **23:** 137-150

Gavin A-C, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon A-M, Cruciat C-M, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier M-A, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415:** 141-147

Geisler S, Lojek L, Khalil AM, Baker KE, Coller J (2012) Decapping of Long Noncoding RNAs Regulates Inducible Genes. *Mol Cell* **45:** 279-291

Gelfand B, Mead J, Bruning A, Apostolopoulos N, Tadigotla V, Nagaraj V, Sengupta AM, Vershon AK (2011) Regulated Antisense Transcription Controls Expression of Cell-Type-Specific Genes in Yeast. *Mol Cell Biol* **31:** 1701-1709

Ghazal G, Gagnon J, Jacques P-É, Landry J-R, Robert F, Abou Elela S (2009) Yeast RNase III Triggers Polyadenylation-Independent Transcription Termination. *Mol Cell* **36:** 99-109

Ghazy MA, Gordon JM, Lee SD, Singh BN, Bohm A, Hampsey M, Moore C (2012) The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site. *Nucleic Acids Res* **40:** 1214-1225

Ghazy MA, He X, Singh BN, Hampsey M, Moore C (2009) The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing. *Mol Cell Biol* **29:** 2296-2307

Gilbert W, Guthrie C (2004) The Glc7p Nuclear Phosphatase Promotes mRNA Export by Facilitating Association of Mex67p with mRNA. *Mol Cell* **13:** 201-212

Gilbert W, Siebel CW, Guthrie C (2001) Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. *RNA* **7:** 302-313

Golding MC, Magri LS, Zhang L, Lalone SA, Higgins MJ, Mann MRW (2011) Depletion of Kcnq1ot1 non-coding RNA does not affect imprinting maintenance in stem cells. *Development* **138:** 3667-3678

Goldstrohm AC, Seay DJ, Hook BA, Wickens M (2007) PUF Protein-mediated Deadenylation Is Catalyzed by Ccr4p. *J Biol Chem* **282:** 109-114

Gomez-Gonzalez B, Garcia-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marin A, Foiani M, Aguilera A (2011) Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J* **30:** 3106-3119

Gong C, Maquat LE (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470:** 284-288

Gonzalez-Aguilera C, Tous C, Babiano R, de la Cruz J, Luna R, Aguilera A (2011) Nab2 functions in the metabolism of RNA driven by polymerases II and III. *Mol Biol Cell* **22:** 2729-2740

Gonzalez-Aguilera C, Tous C, Gomez-Gonzalez B, Huertas P, Luna R, Aguilera A (2008) The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol Biol Cell* **19:** 4310-4318

González CI, Ruiz-Echevarría MJ, Vasudevan S, Henry MF, Peltz SW (2000) The Yeast hnRNP-like Protein Hrp1/Nab4 Marks a Transcript for Nonsense-Mediated mRNA Decay. *Mol Cell* **5:** 489-499

Gordon JM, Shikov S, Kuehner JN, Liriano M, Lee E, Stafford W, Poulsen MB, Harrison C, Moore C, Bohm A (2011) Reconstitution of CF IA from overexpressed subunits reveals stoichiometry and provides insights into molecular topology. *Biochemistry (Mosc)* **50:** 10203-10214

Govind CK, Qiu H, Ginsburg DS, Ruan C, Hofmeyer K, Hu C, Swaminathan V, Workman JL, Li B, Hinnebusch AG (2010) Phosphorylated Pol II CTD Recruits Multiple HDACs, Including Rpd3C(S), for Methylation-Dependent Deacetylation of ORF Nucleosomes. *Mol Cell* **39:** 234-246

Graber JH, Cantor CR, Mohr SC, Smith TF (1999) Genomic detection of new yeast pre-mRNA 3′-end-processing signals. *Nucleic Acids Res* **27:** 888-894

Graber JH, McAllister GD, Smith TF (2002) Probabilistic prediction of Saccharomyces cerevisiae mRNA 3′-processing sites. *Nucleic Acids Res* **30:** 1851-1858

Granneman S, Kudla G, Petfalski E, Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A* **106:** 9613-9618

Granneman S, Petfalski E, Tollervey D (2011) A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *EMBO J* **30:** 4006-4019

Granovskaia M, Jensen L, Ritchie M, Toedling J, Ning Y, Bork P, Huber W, Steinmetz L (2010) High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol* **11:** R24

Grant RP, Marshall NJ, Yang JC, Fasken MB, Kelly SM, Harreman MT, Neuhaus D, Corbett AH, Stewart M (2008) Structure of the N-terminal Mlp1-binding domain of the Saccharomyces cerevisiae mRNA-binding protein, Nab2. *J Mol Biol* **376:** 1048-1059

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, Brown JB, Cherbas L, Davis CA, Dobin A, Li R, Lin W, Malone JH, Mattiuzzo NR, Miller D, Sturgill D, Tuch BB, Zaleski C, Zhang D, Blanchette M, Dudoit S, Eads B, Green RE, Hammonds A, Jiang L, Kapranov P, Langton L, Perrimon N, Sandler JE, Wan KH, Willingham A, Zhang Y, Zou Y, Andrews J, Bickel PJ, Brenner SE, Brent MR, Cherbas P, Gingeras TR, Hoskins RA, Kaufman TC, Oliver B, Celniker SE (2011) The developmental transcriptome of Drosophila melanogaster. *Nature* **471:** 473-479

Green DM, Johnson CP, Hagan H, Corbett AH (2003) The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. *Proceedings of the National Academy of Sciences* **100:** 1010-1015

Green DM, Marfatia KA, Crafton EB, Zhang X, Cheng X, Corbett AH (2002) Nab2p Is Required for Poly(A) RNA Export in Saccharomyces cerevisiae and Is Regulated by Arginine Methylation via Hmt1p. *J Biol Chem* **277:** 7752-7760

Greenberg JR (1979) Ultraviolet light-induced crosslinking of mRNA to proteins. *Nucleic Acids Res* **6:** 715-732

Gross S, Moore C (2001a) Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. *Proc Natl Acad Sci U S A* **98:** 6080-6085

Gross S, Moore CL (2001b) Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation. *Mol Cell Biol* **21:** 8045-8055

Groušl T, Ivanov P, Frydlová I, Vašicová P, Janda F, Vojtová J, Malínská K, Malcová I, Nováková L, Janošková D, Valášek L, Hašek J (2009) Robust heat shock induces eIF2α-phosphorylation-independent assembly of stress granules containing eIF3 and 40S ribosomal subunits in budding yeast, Saccharomyces cerevisiae. *J Cell Sci* **122:** 2078-2088

Grzechnik P, Kufel J (2008) Polyadenylation Linked to Transcription Termination Directs the Processing of snoRNA Precursors in Yeast. *Mol Cell* **32:** 247-258

Guan Q, Zheng W, Tang S, Liu X, Zinkel RA, Tsui K-W, Yandell BS, Culbertson MR (2006) Impact of Nonsense-Mediated mRNA Decay on the Global Expression Profile of Budding Yeast. *PLoS Genet* **2:** e203

Gudipati RK (2012) Massive degradation of RNA precursors by the exosome in wild type cells. *Mol Cell*

Gudipati RK, Neil H, Feuerbach F, Malabat C, Jacquier A (2012) The yeast RPL9B gene is regulated by modulation between two modes of transcription termination. *EMBO J* **31:** 2427-2437

Gudipati RK, Villa T, Boulay J, Libri D (2008) Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat Struct Mol Biol* **15:** 786-794

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* **130:** 77-88

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai M-C, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464:** 1071-1076

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE,

Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458:** 223-227

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477:** 295-300

Guttman M, Rinn JL (2012) Modular regulatory principles of large non-coding RNAs. *Nature* **482:** 339-346

Gwizdek C, Hobeika M, Kus B, Ossareh-Nazari B, Dargemont C, Rodriguez MS (2005) The mRNA Nuclear Export Factor Hpr1 Is Regulated by Rsp5-mediated Ubiquitylation. *J Biol Chem* **280:** 13401-13405

Gwizdek C, Iglesias N, Rodriguez MS, Ossareh-Nazari B, Hobeika M, Divita G, Stutz F, Dargemont C (2006) Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. *Proceedings of the National Academy of Sciences* **103:** 16376-16381

Haddad R, Maurice F, Viphakone N, Voisinet-Hakil F, Fribourg S, Minvielle-Sebastia L (2012) An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination. *Nucleic Acids Res* **40:** 1226-1239

Hainer SJ, Charsar BA, Cohen SB, Martens JA (2012) Identification of Mutant Versions of the Spt16 Histone Chaperone That Are Defective for Transcription-Coupled Nucleosome Occupancy in Saccharomyces cerevisiae. *G3* **2:** 555-567

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, Rode M, Conti E (2012) The crystal structure of S. cerevisiae Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *RNA* **18:** 124-134

Hamill S, Wolin SL, Reinisch KM (2010) Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proceedings of the National Academy of Sciences* **107:** 15045-15050

Hammell CM, Gross S, Zenklusen D, Heath CV, Stutz F, Moore C, Cole CN (2002) Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol* **22:** 6441-6457

Hampsey M, Singh BN, Ansari A, Laine JP, Krishnamurthy S (2011) Control of eukaryotic gene expression: gene loops and transcriptional memory. *Adv Enzyme Regul* **51:** 118-125

Haracska L, Johnson RE, Prakash L, Prakash S (2005) Trf4 and Trf5 proteins of Saccharomyces cerevisiae exhibit poly(A) RNA polymerase activity but no DNA polymerase activity. *Mol Cell Biol* **25:** 10183-10189

Harigaya Y, Parker R (2012) Global analysis of mRNA decay intermediates in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*

Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Tsutui K, Nakagawa S (2010) The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA. *Dev Cell* **19:** 469-476

Hasegawa Y, Irie K, Gerber AP (2008) Distinct roles for Khd1p in the localization and expression of bud-localized mRNAs in yeast. *RNA* **14:** 2333-2347

He F, Li X, Spatrick P, Casillo R, Dong S, Jacobson A (2003a) Genome-Wide Analysis of mRNAs Regulated by the Nonsense-Mediated and 5' to 3' mRNA Decay Pathways in Yeast. *Mol Cell* **12:** 1439-1452

He X, Khan AU, Cheng H, Pappas DL, Hampsey M, Moore CL (2003b) Functional interactions between the transcription and mRNA 3′ end processing machineries mediated by Ssu72 and Sub1. *Genes & Development* **17:** 1030-1042

He X, Moore C (2005) Regulation of yeast mRNA 3' end processing by phosphorylation. *Mol Cell* **19:** 619-629

Hector RE, Nykamp KR, Dheur S, Anderson JT, Non PJ, Urbinati CR, Wilson SM, Minvielle-Sebastia L, Swanson MS (2002) Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *EMBO J* **21:** 1800-1810

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459:** 108-112

Helmling S, Zhelkovsky A, Moore CL (2001) Fip1 regulates the activity of Poly(A) polymerase through multiple interactions. *Mol Cell Biol* **21:** 2026-2037

Henry M, Borland CZ, Bossie M, Silver PA (1996) Potential RNA Binding Proteins in Saccharomyces cerevisiae Identified as Suppressors of Temperature-Sensitive Mutations in NPL3. *Genetics* **142:** 103-115

Henry MF, Mandel D, Routson V, Henry PA (2003) The Yeast hnRNP-like Protein Hrp1/Nab4 Accumulates in the Cytoplasm after Hyperosmotic Stress: A Novel Fps1-dependent Response. *Mol Biol Cell* **14:** 3929-3941

Hentges P, Van Driessche B, Tafforeau L, Vandenhaute J, Carr AM (2005) Three novel antibiotic marker cassettes for gene disruption and marker switching in Schizosaccharomyces pombe. *Yeast* **22:** 1013-1019

Hessle V, Bjork P, Sokolowski M, Gonzalez de Valdivia E, Silverstein R, Artemenko K, Tyagi A, Maddalo G, Ilag L, Helbig R, Zubarev RA, Visa N (2009) The exosome associates cotranscriptionally with the nascent pre-mRNP through interactions with heterogeneous nuclear ribonucleoproteins. *Mol Biol Cell* **20:** 3459-3470

Hessle V, von Euler A, González de Valdivia E, Visa N (2012) Rrp6 is recruited to transcribed genes and accompanies the spliced mRNA to the nuclear pore. *RNA* **18:** 1466-1474

Hieronymus H, Silver PA (2003) Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. *Nat Genet* **33:** 155-161

Hilleren P, McCarthy T, Rosbash M, Parker R, Jensen TH (2001) Quality control of mRNA 3[prime]-end processing is linked to the nuclear exosome. *Nature* **413:** 538-542

Hilleren PJ, Parker R (2003) Cytoplasmic Degradation of Splice-Defective Pre-mRNAs and Intermediates. *Mol Cell* **12:** 1453-1465

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

Hobeika M, Brockmann C, Gruessing F, Neuhaus D, Divita G, Stewart M, Dargemont C (2009) Structural requirements for the ubiquitin-associated domain of the mRNA export factor Mex67 to bind its specific targets, the transcription elongation THO complex component Hpr1 and nucleoporin FXFG repeats. *J Biol Chem* **284:** 17575-17583

Hobor F, Pergoli R, Kubicek K, Hrossova D, Bacikova V, Zimmermann M, Pasulka J, Hofr C, Vanacova S, Stefl R (2011) Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J Biol Chem* **286:** 3645-3657

Hockensmith JW, Kubasek WL, Vorachek WR, von Hippel PH (1986) Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. *J Biol Chem* **261:** 3512-3518

Holub P, Lalakova J, Cerna H, Pasulka J, Sarazova M, Hrazdilova K, Arce MS, Hobor F, Stefl R, Vanacova S (2012) Air2p is critical for the assembly and RNA-binding of the

TRAMP complex and the KOW domain of Mtr4p is crucial for exosome activation. *Nucleic Acids Res* **40:** 5679-5693

Hongay CF, Grisafi PL, Galitski T, Fink GR (2006) Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* **127:** 735-745

Honorine R, Mosrin-Huaman C, Hervouet-Coste Ng, Libri D, Rahmouni AR (2011) Nuclear mRNA quality control in yeast is mediated by Nrd1 co-transcriptional recruitment, as revealed by the targeting of Rho-induced aberrant transcripts. *Nucleic Acids Res* **39:** 2809-2820

Houalla R, Devaux F, Fatica A, Kufel J, Barrass D, Torchet C, Tollervey D (2006) Microarray detection of novel nuclear RNA substrates for the exosome. *Yeast* **23:** 439-454

Houseley J, Kotovic K, El Hage A, Tollervey D (2007) Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. *EMBO J* **26:** 4996-5006

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

Houseley J, Tollervey D (2009) The Many Pathways of RNA Degradation. *Cell* **136:** 763-776

Hu W, Sweet TJ, Chamnongpol S, Baker KE, Coller J (2009) Co-translational mRNA decay in Saccharomyces cerevisiae. *Nature* **461:** 225-229

Huang Y-C, Chen H-T, Teng S-C (2010) Intragenic transcription of a noncoding RNA modulates expression of ASP3 in budding yeast. *RNA* **16:** 2085-2093

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL (2010) A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. *Cell* **142:** 409-419

Huber W, Toedling J, Steinmetz LM (2006) Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* **22:** 1963-1970

Huertas P, Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12:** 711-721

Huh J-W, Wu J, Lee C-H, Yun M, Gilada D, Brautigam CA, Li B (2012) Multivalent di-nucleosome recognition enables the Rpd3S histone deacetylase complex to tolerate decreased H3K36 methylation levels. *EMBO J* **31:** 3564-35744

Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, Wang Y, Kong B, Langerod A, Borresen-Dale A-L, Kim SK, van de Vijver M, Sukumar S, Whitfield ML, Kellis M, Xiong Y, Wong DJ, Chang HY (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* **43:** 621-629

Hurt E, Luo M-j, Rother S, Reed R, Straber K (2004) Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. *Proc Natl Acad Sci U S A* **101:** 1858-1862

Iglesias N, Stutz Fo (2008) Regulation of mRNP dynamics along the export pathway. *FEBS Lett* **582:** 1987-1996

Iglesias N, Tutucci E, Gwizdek C, Vinciguerra P, Von Dach E, Corbett AH, Dargemont C, Stutz F (2010) Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. *Genes & Development* **24:** 1927-1938

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324:** 218-223

Ingolia Nicholas T, Lareau Liana F, Weissman Jonathan S (2011) Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147:** 789-802

Jackson RN, Klauer AA, Hintze BJ, Robinson H, van Hoof A, Johnson SJ (2010) The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J* **29:** 2205-2216

Jamonnak N, Creamer TJ, Darby MM, Schaughency P, Wheelan SJ, Corden JL (2011) Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA* **17:** 2011-2025

Jani D, Lutz S, Marshall NJ, Fischer T, Kohler A, Ellisdon AM, Hurt E, Stewart M (2009) Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export. *Mol Cell* **33:** 727-737

Jenks MH, O'Rourke TW, Reines D (2008) Properties of an Intergenic Terminator and Start Site Switch That Regulate IMD2 Transcription in Yeast. *Mol Cell Biol* **28:** 3883-3893

Jensen TH, Patricio K, McCarthy T, Rosbash M (2001) A Block to mRNA Nuclear Export in S. cerevisiae Leads to Hyperadenylation of Transcripts that Accumulate at the Site of Transcription. *Mol Cell* **7:** 887-898

Jeon Y, Lee Jeannie T (2011) YY1 Tethers Xist RNA to the Inactive X Nucleation Center. *Cell* **146:** 119-133

Jeon Y, Sarma K, Lee JT (2012) New and Xisting regulatory mechanisms of X chromosome inactivation. *Curr Opin Genet Dev* **22:** 62-71

Jia H, Wang X, Anderson JT, Jankowsky E (2012) RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proceedings of the National Academy of Sciences* **109:** 7292-7297

Jia H, Wang X, Liu F, Guenther UP, Srinivasan S, Anderson JT, Jankowsky E (2011) The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. *Cell* **145:** 890-901

Jiao X, Xiang S, Oh C, Martin CE, Tong L, Kiledjian M (2010) Identification of a quality-control mechanism for mRNA 5'-end capping. *Nature* **467:** 608-611

Jimeno-González S, Haaning LL, Malagon F, Jensen TH (2010) The Yeast 5′-3′ Exonuclease Rat1p Functions during Transcription Elongation by RNA Polymerase II. *Mol Cell* **37:** 580-587

Jimeno S, Luna R, Garcia-Rubio M, Aguilera A (2006) Tho1, a novel hnRNP, and Sub2 provide alternative pathways for mRNP biogenesis in yeast THO mutants. *Mol Cell Biol* **26:** 4387-4398

Jinek M, Coyle Scott M, Doudna Jennifer A (2011) Coupled 5' Nucleotide Recognition and Processivity in Xrn1-Mediated mRNA Decay. *Mol Cell* **41:** 600-608

Johnson AW (1997) Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Mol Cell Biol* **17:** 6122-6130

Johnson R, Richter N, Jauch R, Gaughwin PM, Zuccato C, Cattaneo E, Stanton LW (2010) The Human Accelerated Region 1 noncoding RNA is repressed by REST in Huntington's disease. *Physiol Genomics*

Johnson SA, Cubberley G, Bentley DL (2009) Cotranscriptional Recruitment of the mRNA Export Factor Yra1 by Direct Interaction with the 3' End Processing Factor Pcf11. *Mol Cell* **33:** 215-226

Johnson SA, Kim H, Erickson B, Bentley DL (2011) The export factor Yra1 modulates mRNA 3' end processing. *Nat Struct Mol Biol* **18:** 1164-1171

Jona G, Choder M, Gileadi O (2000) Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mRNA in yeast. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1491:** 37-48

Kadaba S, Wang X, Anderson J (2006) Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA* **12:** 508-521

Kallehauge Thomas B, Robert M-C, Bertrand E, Jensen T (2012) Nuclear Retention Prevents Premature Cytoplasmic Appearance of mRNA. *Mol Cell*

Kanhere A, Jenner R (2012) Noncoding RNA localisation mechanisms in chromatin regulation. *Silence* **3:** 2

Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW, Pombo A, Fisher AG, Young RA, Jenner RG (2010) Short RNAs Are Transcribed from Repressed Polycomb Target Genes and Interact with Polycomb Repressive Complex-2. *Mol Cell* **38:** 675-688

Kaplan CD, Laprade L, Winston F (2003) Transcription Elongation Factors Repress Transcription Initiation from Cryptic Sites. *Science* **301:** 1096-1099

Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296:** 916-919

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316:** 1484-1488

Kapranov P, St Laurent G, Raz T, Ozsolak F, Reynolds CP, Sorensen P, Reaman G, Milos P, Arceci R, Thompson J, Triche T (2010) The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biology* **8:** 149

Kawauchi J, Mischo H, Braglia P, Rondon A, Proudfoot NJ (2008) Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev* **22:** 1082-1092

Kebaara BW, Atkin AL (2009) Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. *Nucleic Acids Res* **37:** 2771-2778

Kelly SM, Corbett AH (2009) Messenger RNA Export from the Nucleus: A Series of Molecular Wardrobe Changes. *Traffic* **10:** 1199-1208

Kelly SM, Leung SW, Apponi LH, Bramley AM, Tran EJ, Chekanova JA, Wente SR, Corbett AH (2010) Recognition of polyadenosine RNA by the zinc finger domain of nuclear poly(A) RNA-binding protein 2 (Nab2) is required for correct mRNA 3'-end formation. *J Biol Chem* **285:** 26022-26032

Kelly SM, Pabit SA, Kitchen CM, Guo P, Marfatia KA, Murphy TJ, Corbett AH, Berland KM (2007) Recognition of polyadenosine RNA by zinc finger proteins. *Proceedings of the National Academy of Sciences* **104:** 12306-12311

Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W (2012) The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat Cell Biol* **14:** 659-665

Kerr SC, Azzouz N, Fuchs SM, Collart MA, Strahl BD, Corbett AH, Laribee RN (2011) The Ccr4-Not Complex Interacts with the mRNA Export Machinery. *PLoS One* **6:** e18302

Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467:** 103-107

Kervestin S, Li C, Buckingham R, Jacobson A (2012) Testing the faux-UTR model for NMD: Analysis of Upf1p and Pab1p competition for binding to eRF3/Sup35p. *Biochimie* **94:** 1560-1571

Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, Moore CL (1997) Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. *Genes & Development* **11:** 2545-2556

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA*

Kim Guisbert K, Duncan K, Li H, Guthrie C (2005) Functional specificity of shuttling hnRNPs revealed by genome-wide analysis of their RNA binding profiles. *RNA* **11:** 383-393

Kim Guisbert KS, Li H, Guthrie C (2006) Alternative 3' Pre-mRNA Processing in Saccharomyces cerevisiae Is Modulated by Nab4/Hrp1 In Vivo. *PLoS Biol* **5:** e6

Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, Megee PC, Bentley DL (2010a) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* **17:** 1279-1286

Kim HJ, Jeong SH, Heo JH, Jeong SJ, Kim ST, Youn HD, Han JW, Lee HW, Cho EJ (2004a) mRNA capping enzyme activity is coupled to an early transcription elongation. *Mol Cell Biol* **24:** 6184-6193

Kim K-Y, Levin D (2011a) Mpk1 MAPK Association with the Paf1 Complex Blocks Sen1-Mediated Premature Transcription Termination. *Cell* **144:** 745-756

Kim M, Ahn S-H, Krogan NJ, Greenblatt JF, Buratowski S (2004b) Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J* **23:** 354-364

Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S (2004c) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432:** 517-522

Kim M, Suh H, Cho E-J, Buratowski S (2009a) Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7. *J Biol Chem* **284:** 26421-26426

Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S (2006) Distinct pathways for snoRNA and mRNA termination. *Mol Cell* **24:** 723-734

Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME (2010b) Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465:** 182-187

Kim T, Buratowski S (2009b) Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* **137:** 259-272

Kim T, Xu Z, Clauder-Münster S, Steinmetz Lars M, Buratowski S (2012) Set3 HDAC Mediates Effects of Overlapping Noncoding Transcription on Gene Induction Kinetics. *Cell* **150:** 1158-1169

Kim TS, Liu CL, Yassour M, Holik J, Friedman N, Buratowski S, Rando OJ (2011b) RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. *Genome Biol* **11:** R75

Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP (2010) Noncoding RNA Gas5 Is a Growth Arrest- and Starvation-Associated Repressor of the Glucocorticoid Receptor. *Sci Signal* **3:** ra8

Kireeva ML, Komissarova N, Waugh DS, Kashlev M (2000) The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J Biol Chem* **275:** 6530-6536

Kirmizis A, Santos-Rosa H, Penkett CJ, Singer MA, Vermeulen M, Mann M, Bahler J, Green RD, Kouzarides T (2007) Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **449:** 928-932

Kobayashi T, Funakoshi Y, Hoshino S-i, Katada T (2004) The GTP-binding Release Factor eRF3 as a Key Mediator Coupling Translation Termination to mRNA Decay. *J Biol Chem* **279:** 45693-45700

Kobayashi T, Ganley AR (2005) Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. *Science* **309:** 1581-1584

Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, Tanaka F, Shibata K, Suzuki A, Komune S, Miyano S, Mori M (2011) Long non-coding RNA HOTAIR regulates Polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res*

Kopcewicz KA, O'Rourke TW, Reines D (2007) Metabolic Regulation of IMD2 Transcription and an Unusual DNA Element That Generates Short Transcripts. *Mol Cell Biol* **27:** 2821-2829

Kos M, Tollervey D (2010) Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol Cell* **37:** 809-820

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30:** 1956-1962

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128:** 693-705

Kowalczyk Monika S, Hughes Jim R, Garrick D, Lynch Magnus D, Sharpe Jacqueline A, Sloane-Stanley Jacqueline A, McGowan Simon J, De Gobbi M, Hosseini M, Vernimmen D, Brown Jill M, Gray Nicola E, Collavin L, Gibbons Richard J, Flint J, Taylor S, Buckle Veronica J, Milne Thomas A, Wood William G, Higgs Douglas R (2012) Intragenic Enhancers Act as Alternative Promoters. *Mol Cell* **45:** 447-458

Kuai L, Das B, Sherman F (2005) A nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **102:** 13962-13967

Kuehner JN, Brow DA (2008) Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. *Mol Cell* **31:** 201-211

Kuehner JN, Pearson EL, Moore C (2011) Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* **12:** 283-294

Kwak H, Fuda NJ, Core LJ, Lis JT (2013) Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science* **339:** 950-953

Kyburz A, Sadowski M, Dichtl B, Keller W (2003) The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3′ end formation. *Nucleic Acids Res* **31:** 3936-3945

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

Lahudkar S, Shukla A, Bajwa P, Durairaj G, Stanojevic N, Bhaumik SR (2011) The mRNA cap-binding complex stimulates the formation of pre-initiation complex at the promoter via its interaction with Mot1p in vivo. *Nucleic Acids Res* **39:** 2188-2209

Lange A, Mills RE, Devine SE, Corbett AH (2008) A PY-NLS Nuclear Targeting Signal Is Required for Nuclear Localization and Function of the Saccharomyces cerevisiae mRNA-binding Protein Hrp1. *J Biol Chem* **283:** 12926-12934

Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, Chu A, Davis RW, Steinmetz LM, Primig M (2011) Execution of the meiotic noncoding RNA expression

program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proceedings of the National Academy of Sciences* **108:** 1058-1063

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

Lee G, Bratkowski MA, Ding F, Ke A, Ha T (2012) Elastic Coupling Between RNA Degradation and Unwinding by an Exoribonuclease. *Science* **336:** 1726-1729

Leeper TC, Qu X, Lu C, Moore C, Varani G (2010) Novel protein-protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1. *J Mol Biol* **401:** 334-349

Lewis JD, Izaurralde E, Jarmolowski A, McGuigan C, Mattaj IW (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev* **10:** 1683-1698

Li B, Jackson J, Simon MD, Fleharty B, Gogol M, Seidel C, Workman JL, Shilatifard A (2009) Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *J Biol Chem* **284:** 7970-7976

Libri D, Dower K, Boulay J, Thomsen R, Rosbash M, Jensen TH (2002) Interactions between mRNA Export Commitment, 3′-End Quality Control, and Nuclear Degradation. *Mol Cell Biol* **22:** 8254-8266

Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL (2002) Functional Interaction of Yeast Pre-mRNA 3′ End Processing Factors with RNA Polymerase II. *Mol Cell* **9:** 1101-1111

Lin MF, Jungreis I, Kellis M (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27:** i275-i282

Liu Q, Greimann JC, Lima CD (2006) Reconstitution, Activities, and Structure of the Eukaryotic RNA Exosome. *Cell* **127:** 1223-1237

Longtine MS, McKenzie A, 3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14:** 953-961

Lorentzen E, Basquin J, Tomecki R, Dziembowski A, Conti E (2008) Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Mol Cell* **29:** 717-728

Loya TJ, O'Rourke TW, Reines D (2012) A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3. *Nucleic Acids Res*

Luke B, Panza A, Redon S, Iglesias N, Li Z, Lingner J (2008) The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae. *Mol Cell* **32:** 465-477

Lund MK, Guthrie C (2005) The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. *Mol Cell* **20:** 645-651

Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **17:** 1195-1201

Luo W, Johnson AW, Bentley DL (2006) The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* **20:** 954-965

Lykke-Andersen S, Jensen TH (2007) Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* **89:** 1177-1182

Ma Z, Atencio D, Barnes C, DeFiglio H, Hanes SD (2012) Multiple Roles for the Ess1 Prolyl Isomerase in the RNA Polymerase II Transcription Cycle. *Mol Cell Biol*

Magistri M, Faghihi MA, St Laurent Iii G, Wahlestedt C (2012) Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends Genet* **28:** 389-396

Maison C, Bailly D, Roche D, Montes de Oca R, Probst AV, Vassias I, Dingli F, Lombard B, Loew D, Quivy JP, Almouzni G (2011) SUMOylation promotes de novo targeting of HP1alpha to pericentric heterochromatin. *Nat Genet* **43:** 220-227

Malet H, Topf M, Clare DK, Ebert J, Bonneau F, Basquin J, Drazkowska K, Tomecki R, Dziembowski A, Conti E, Saibil HR, Lorentzen E (2010) RNA channelling by the eukaryotic exosome. *EMBO Rep* **11:** 936-942

Mandel C, Bai Y, Tong L (2008) Protein factors in pre-mRNA 3′-end processing. *Cell Mol Life Sci* **65:** 1099-1122

Mangus DA, Evans MC, Agrin NS, Smith M, Gongidi P, Jacobson A (2004a) Positive and Negative Regulation of Poly(A) Nuclease. *Mol Cell Biol* **24:** 5521-5533

Mangus DA, Smith MM, McSweeney JM, Jacobson A (2004b) Identification of factors regulating poly(A) tail synthesis and maturation. *Mol Cell Biol* **24:** 4196-4206

Maniatis T, Fritsch, E.F., Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual, 1st ed*: Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

Mao YS, Sunwoo H, Zhang B, Spector DL (2011) Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol* **13:** 95-101

Marfatia KA, Crafton EB, Green DM, Corbett AH (2003) Domain Analysis of the Saccharomyces cerevisiae Heterogeneous Nuclear Ribonucleoprotein, Nab2p. *J Biol Chem* **278:** 6731-6740

Mariconti L, Loll B, Schlinkmann K, Wengi A, Meinhart A, Dichtl B (2010) Coupled RNA polymerase II transcription and 3' end formation with yeast whole-cell extracts. *RNA* **16:** 2205-2217

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

Martens JA, Wu PY, Winston F (2005) Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. *Genes Dev* **19:** 2695-2704

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

Mason PB, Struhl K (2003) The FACT Complex Travels with Elongating RNA Polymerase II and Is Important for the Fidelity of Transcriptional Initiation In Vivo. *Mol Cell Biol* **23:** 8323-8333

Matsuda E, Garfinkel DJ (2009) Posttranslational interference of Ty1 retrotransposition by antisense RNAs. *Proceedings of the National Academy of Sciences* **106:** 15657-15662

Mayan M, Aragón L (2010) Cis-interactions between non-coding ribosomal spacers dependent on RNAP-II separate RNAP-I and RNAP-III transcription domains. *Cell cycle* **9:** 4328-4337

Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P (2012) CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. *Science* **336:** 1723-1725

Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* **17:** 1272-1278

Mayer SA, Dieckmann CL (1989) The yeast CBP1 gene produces two differentially regulated transcripts by alternative 3'-end formation. *Mol Cell Biol* **9:** 4161-4169

McBride AE, Cook JT, Stemmler EA, Rutledge KL, McGrath KA, Rubens JA (2005) Arginine Methylation of Yeast mRNA-binding Protein Npl3 Directly Affects Its Function, Nuclear Export, and Intranuclear Protein Interactions. *J Biol Chem* **280:** 30888-30898

McKinlay A, Araya CL, Fields S (2011) Genome-Wide Analysis of Nascent Transcription in Saccharomyces cerevisiae. *G3: Genes, Genomes, Genetics* **1:** 549-558

Meaux S, Lavoie M, Gagnon J, Abou Elela S, van Hoof A (2011) Reporter mRNAs cleaved by Rnt1p are exported and degraded in the cytoplasm. *Nucleic Acids Res* **39:** 9357-9367

Meaux S, Van Hoof A (2006) Yeast transcripts cleaved by an internal ribozyme provide new insight into the role of the cap and poly(A) tail in translation and mRNA decay. *RNA* **12:** 1323-1337

Meaux S, van Hoof A, Baker KE (2008) Nonsense-Mediated mRNA Decay in Yeast Does Not Require PAB1 or a Poly(A) Tail. *Mol Cell* **29:** 134-140

Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, Mattick JS, Rinn JL (2012) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat Biotechnol* **30:** 99-104

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP, Lee W, Mendenhall E, O/'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553-560

Miller C, Schwalb B, Maier K, Schulz D, Dumcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dolken L, Martin DE, Tresch A, Cramer P (2012) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7**

Miller MR, Robinson KJ, Cleary MD, Doe CQ (2009) TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nat Meth* **6:** 439-441

Millevoi S, Vagner Sp (2009) Molecular mechanisms of eukaryotic pre-mRNA 3â€² end processing regulation. *Nucleic Acids Res* **38:** 2757-2774

Milligan L, Decourty L, Saveanu C, Rappsilber J, Ceulemans H, Jacquier A, Tollervey D (2008) A Yeast Exosome Cofactor, Mpp6, Functions in RNA Surveillance and in the Degradation of Noncoding RNA Transcripts. *Mol Cell Biol* **28:** 5446-5457

Milligan L, Torchet C, Allmang C, Shipman T, Tollervey D (2005) A Nuclear Surveillance Pathway for mRNAs with Defective Polyadenylation. *Mol Cell Biol* **25:** 9996-10004

Minvielle-Sebastia L, Beyer K, Krecic AM, Hector RE, Swanson MS, Keller W (1998) Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. *EMBO J* **17:** 7454-7468

Minvielle-Sebastia L, Preker PJ, Wiederkehr T, Strahm Y, Keller W (1997) The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3′-end formation. *Proc Natl Acad Sci U S A* **94:** 7897-7902

Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ (2011) Yeast Sen1 Helicase Protects the Genome from Transcription-Associated Instability. *Mol Cell* **41:** 21-32

Mitchell P, Petfalski E, Houalla R, Podtelejnikov A, Mann M, Tollervey D (2003a) Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. *Mol Cell Biol* **23:** 6982-6992

Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D (1997) The Exosome: A Conserved Eukaryotic RNA Processing Complex Containing Multiple 3′→5′ Exoribonucleases. *Cell* **91:** 457-466

Mitchell P, Tollervey D (2003b) An NMD Pathway in Yeast Involving Accelerated Deadenylation and Exosome-Mediated 3′→5′ Degradation. *Mol Cell* **11:** 1405-1413

Miura F, Kawaguchi N, Sese J, Toyoda A, Hattori M, Morishita S, Ito T (2006) A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proceedings of the National Academy of Sciences* **103:** 17846-17851

Mohammad F, Pandey GK, Mondal T, Enroth S, Redrup L, Gyllensten U, Kanduri C (2012) Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. *Development* **139:** 2792-2803

Morlando M, Ballarino M, Greco P, Caffarelli E, Dichtl B, Bozzoni I (2004) Coupling between snoRNP assembly and 3' processing controls box C/D snoRNA biosynthesis in yeast. *EMBO J* **23:** 2392-2401

Morlando M, Greco P, Dichtl B, Fatica A, Keller W, Bozzoni I (2002) Functional analysis of yeast snoRNA and snRNA 3'-end formation mediated by uncoupling of cleavage and polyadenylation. *Mol Cell Biol* **22:** 1379-1389

Mueller CL, Porter SE, Hoffman MG, Jaehning JA (2004) The Paf1 Complex Has Functions Independent of Actively Transcribing RNA Polymerase II. *Mol Cell* **14:** 447-456

Munchel SE, Shultzaberger RK, Takizawa N, Weis K (2011) Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Mol Biol Cell***:** mbc.E11-01-0028

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

Murthy UM, Rangarajan PN (2010) Identification of protein interaction regions of VINC/NEAT1/Men epsilon RNA. *FEBS Lett* **584:** 1531-1535

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

Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P (2008) The Air Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin. *Science* **322:** 1717-1720

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

Nakagawa S, Naganuma T, Shioi G, Hirose T (2011) Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *The Journal of Cell Biology* **193:** 31-39

Nedea E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, Hughes T, Buratowski S, Moore CL, Greenblatt J (2003) Organization and Function of APT, a Subcomplex of the Yeast Cleavage and Polyadenylation Factor Involved in the Formation of mRNA and Small Nucleolar RNA 3'-Ends. *J Biol Chem* **278:** 33000-33010

Nedea E, Nalbant D, Xia D, Theoharis NT, Suter B, Richardson CJ, Tatchell K, Kislinger T, Greenblatt JF, Nagy PL (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol Cell* **29:** 577-587

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

Neumann S, Petfalski E, Brugger B, Groszhans H, Wieland F, Tollervey D, Hurt E (2003) Formation and nuclear export of tRNA, rRNA and mRNA is regulated by the ubiquitin ligase Rsp5p. *EMBO Rep* **4:** 1156-1162

Nishizawa M, Komai T, Katou Y, Shirahige K, Ito T, Toh-e A (2008) Nutrient-Regulated Antisense and Intragenic RNAs Modulate a Signal Transduction Pathway in Yeast. *PLoS Biol* **6:** e326

Nordick K, Hoffman MG, Betz JL, Jaehning JA (2008) Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II. *Eukaryot Cell* **7:** 1158-1167

Nourani A, Robert F, Winston F (2006) Evidence that Spt2/Sin1, an HMG-Like Factor, Plays Roles in Transcription Elongation, Chromatin Structure, and Genome Stability in Saccharomyces cerevisiae. *Mol Cell Biol* **26:** 1496-1509

Novikova IV, Hennelly SP, Sanbonmatsu KY (2012) Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res* **40:** 5034-5051

Ogawa Y, Sun BK, Lee JT (2008) Intersection of the RNA Interference and X-Inactivation Pathways. *Science* **320:** 1336-1341

Ohnacker M, Barabino SM, Preker PJ, Keller W (2000) The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex. *EMBO J* **19:** 37-47

Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R (2010) Long Noncoding RNAs with Enhancer-like Function in Human Cells. *Cell* **143:** 46-58

Owen-Hughes T, Gkikopoulos T (2012) Making sense of transcribing chromatin. *Curr Opin Cell Biol* **24:** 296-304

Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **143:** 1018-1029

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-DiNardo D, Kanduri C (2008) Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Mol Cell* **32:** 232-246

Paolo SS, Vanacova S, Schenk L, Scherrer T, Blank D, Keller W, Gerber AP (2009) Distinct Roles of Non-Canonical Poly(A) Polymerases in RNA Metabolism. *PLoS Genet* **5:** e1000555

Parker R (2012) RNA Degradation in Saccharomyces cerevisae. *Genetics* **191:** 671-702

Pascual-Garcia P, Govind CK, Queralt E, Cuenca-Bono B, Llopis A, Chavez S, Hinnebusch AG, Rodriguez-Navarro S (2008) Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2. *Genes Dev* **22:** 2811-2822

Peil K, Värv S, Lõoke M, Kristjuhan K, Kristjuhan A (2011) Uniform Distribution of Elongating RNA Polymerase II Complexes in Transcribed Gene Locus. *J Biol Chem* **286:** 23817-23822

Pelechano V, Chávez S, Pérez-Ortín JE (2010) A Complete Set of Nascent Transcription Rates for Yeast Genes. *PLoS One* **5:** e15442

Pelechano V, Jimeno-González S, Rodríguez-Gil A, García-Martínez J, Pérez-Ortín JE, Chávez S (2009) Regulon-Specific Control of Transcription Elongation across the Yeast Genome. *PLoS Genet* **5:** e1000614

Pena A, Gewartowski K, Mroczek S, Cuellar J, Szykowska A, Prokop A, Czarnocki-Cieciura M, Piwowarski J, Tous C, Aguilera A, Carrascosa JL, Valpuesta JM, Dziembowski A (2012) Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *EMBO J* **31:** 1605-1616

Peterlin BM, Price DH (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23:** 297-305

Petesch SJ, Lis JT (2012) Overcoming the nucleosome barrier during transcript elongation. *Trends Genet* **28:** 285-294

Petruk S, Sedkov Y, Riley KM, Hodgson J, Schweisguth F, Hirose S, Jaynes JB, Brock HW, Mazo A (2006) Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* **127:** 1209-1221

Pinskaya M, Gourvennec S, Morillon A (2009) H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J* **28:** 1697-1707

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122:** 517-527

Pollard KS, Salama SR, Lambert N, Lambot M-A, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A, Kern AD, Dehay C, Igel H, Ares M, Vanderhaeghen P, Haussler D (2006) An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443:** 167-172

Ponjavic J, Ponting CP, Lunter G (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res* **17:** 556-565

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

Pray-Grant MG, Daniel JA, Schieltz D, Yates JR, 3rd, Grant PA (2005) Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433:** 434-438

Preker P, Almvig K, Christensen MS, Valen E, Mapendano CK, Sandelin A, Jensen TH (2011) PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res* **39:** 7179-7193

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

Pruneski JA, Hainer SJ, Petrov KO, Martens JA (2011) The Paf1 Complex Represses SER3 Transcription in Saccharomyces cerevisiae by Facilitating Intergenic Transcription-Dependent Nucleosome Occupancy of the SER3 Promoter. *Eukaryotic Cell* **10:** 1283-1294

Qiu H, Hu C, Gaur NA, Hinnebusch AG (2012) Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *EMBO J*

Qiu H, Hu C, Hinnebusch AG (2009) Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol Cell* **33:** 752-762

Qu X, Lykke-Andersen S, Nasser T, Saguez C, Bertrand E, Jensen TH, Moore C (2009) Assembly of an Export-Competent mRNP Is Needed for Efficient Release of the 3′-End Processing Complex after Polyadenylation. *Mol Cell Biol* **29:** 5327-5338

Quan TK, Hartzog GA (2010) Histone H3K4 and K36 Methylation, Chd1 and Rpd3S Oppose the Functions of Saccharomyces cerevisiae Spt4–Spt5 in Transcription. *Genetics* **184:** 321-334

Qureshi IA, Mehler MF (2012) Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat Rev Neurosci* **13:** 528-541

Redon S, Reichenbach P, Lingner J (2010) The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res* **38:** 5797-5806

Redrup L, Branco MR, Perdeaux ER, Krueger C, Lewis A, Santos F, Nagano T, Cobb BS, Fraser P, Reik W (2009) The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development* **136:** 525-530

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

Rintala-Maki ND, Sutherland LC (2009) Identification and characterisation of a novel antisense non-coding RNA from the RBM5 gene locus. *Gene*

Rodriguez-Gil A, Garcia-Martinez J, Pelechano V, Munoz-Centeno Mde L, Geli V, Perez-Ortin JE, Chavez S (2010) The distribution of active RNA polymerase II along the transcribed region is gene-specific and controlled by elongation factors. *Nucleic Acids Res* **38:** 4651-4664

Rodriguez-Navarro S, Fischer T, Luo M-J, Antúnez O, Brettschneider S, Lechner J, Pérez-Ortin JE, Reed R, Hurt E (2004) Sus1, a Functional Component of the SAGA Histone Acetylase Complex and the Nuclear Pore-Associated mRNA Export Machinery. *Cell* **116:** 75-86

Rondón AG, Jimeno S, García-Rubio M, Aguilera A (2003) Molecular Evidence That the Eukaryotic THO/TREX Complex Is Required for Efficient Transcription Elongation. *J Biol Chem* **278:** 39037-39043

Rondon AG, Mischo HE, Kawauchi J, Proudfoot NJ (2009) Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. *Mol Cell* **36:** 88-98

Roth KM, Byam J, Fang F, Butler JS (2009) Regulation of NAB2 mRNA 3'-end formation requires the core exosome and the Trf4p component of the TRAMP complex. *RNA* **15:** 1045-1058

Roth KM, Wolf MK, Rossi M, Butler JS (2005) The Nuclear Exosome Contributes to Autogenous Control of NAB2 mRNA Levels. *Mol Cell Biol* **25:** 1577-1585

Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C, Boulay J, Jensen TH, Stutz F, Devaux F, Libri D (2008) THO/Sub2p Functions to Coordinate 3′-End Processing with Gene-Nuclear Pore Association. *Cell* **135:** 308-321

Rougemaille M, Gudipati RK, Olesen JR, Thomsen R, Seraphin B, Libri D, Jensen TH (2007) Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. *EMBO J* **26:** 2317-2326

Ruiz-Echevarría MJ, Peltz SW (2000) The RNA Binding Protein Pub1 Modulates the Stability of Transcripts Containing Upstream Open Reading Frames. *Cell* **101:** 741-751

Russnak R, Nehrke KW, Platt T (1995) REF2 encodes an RNA-binding protein directly involved in yeast mRNA 3'-end formation. *Mol Cell Biol* **15:** 1689-1697

Sachs AB, Davis RW, Kornberg RD (1987) A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol Cell Biol* **7:** 3268-3276

Sadowski M, Dichtl B, Hubner W, Keller W (2003) Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *EMBO J* **22:** 2167-2177

Saguez C, Schmid M, Olesen JR, Ghazy MAE-H, Qu X, Poulsen MB, Nasser T, Moore C, Jensen TH (2008) Nuclear mRNA Surveillance in THO/sub2 Mutants Is Triggered by Inefficient Polyadenylation. *Mol Cell* **31:** 91-103

Saha S, Murthy S, Rangarajan PN (2006) Identification and characterization of a virus-inducible non-coding RNA in mouse brain. *J Gen Virol* **87:** 1991-1995

Sambrook J, and D. W. Russell (2001) *Molecular cloning: a laboratory manual, 3rd ed.*: Cold Spring Harbor Laboratory Press, Cold Spring Harbour, 3rd ed.

Sasaki YT, Ideue T, Sano M, Mituyama T, Hirose T (2009) MENe/b noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci U S A* **106:** 2525-2530

Sayani S, Janis M, Lee CY, Toesca I, Chanfreau GF (2008) Widespread Impact of Nonsense-Mediated mRNA Decay on the Yeast Intronome. *Mol Cell* **31:** 360-370

Schaeffer D, Tsanova B, Barbas A, Reis FP, Dastidar EG, Sanchez-Rotunno M, Arraiano CM, van Hoof A (2009) The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat Struct Mol Biol* **16:** 56-62

Schaeffer D, van Hoof A (2011) Different nuclease requirements for exosome-mediated degradation of normal and nonstop mRNAs. *Proceedings of the National Academy of Sciences* **108:** 2366-2371

Scherrer T, Mittal N, Janga SC, Gerber AP (2010) A Screen for RNA-Binding Proteins in Yeast Indicates Dual Functions for Many Enzymes. *PLoS One* **5:** e15499

Schiestl RH, Gietz RD (1989) High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet* **16:** 339-346

Schmid M, Poulsen MB, Olszewski P, Pelechano V, Saguez C, Gupta I, Steinmetz LM, Moore C, Jensen TH (2012) Rrp6p controls mRNA poly(a) tail length and its decoration with poly(a) binding proteins. *Mol Cell* **47:** 267-280

Schmitz K-M, Mayer C, Postepska A, Grummt I (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes & Development* **24:** 2264-2269

Schneider C, Anderson JT, Tollervey D (2007) The Exosome Subunit Rrp44 Plays a Direct Role in RNA Substrate Recognition. *Mol Cell* **27:** 324-331

Schneider C, Kudla G, Wlotzka W, Tuck AC, Tollervey D (2012) Transcriptome-wide Analysis of Exosome Targets. *Mol Cell*

Schneider C, Leung E, Brown J, Tollervey D (2009) The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. *Nucleic Acids Res* **37:** 1127-1140

Schroeder SC, Zorio DA, Schwer B, Shuman S, Bentley D (2004) A function of yeast mRNA cap methyltransferase, Abd1, in transcription by RNA polymerase II. *Mol Cell* **13:** 377-387

Schwabish MA, Struhl K (2006) Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol Cell* **22:** 415-422

Schwer B, Erdjument-Bromage H, Shuman S (2011) Composition of yeast snRNPs and snoRNPs in the absence of trimethylguanosine caps reveals nuclear cap binding protein as a gained U1 component implicated in the cold-sensitivity of tgs1Δ cells. *Nucleic Acids Res* **39:** 6715-6728

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

Seol Y, Skinner GM, Visscher K, Buhot A, Halperin A (2007) Stretching of Homopolymeric RNA Reveals Single-Stranded Helices and Base-Stacking. *Physical Review Letters* **98:** 158103

Sheldon KE, Mauger DM, Arndt KM (2005) A Requirement for the Saccharomyces cerevisiae Paf1 Complex in snoRNA 3′ End Formation. *Mol Cell* **20:** 225-236

Shen EC, Henry MF, Weiss VH, Valentini SR, Silver PA, Lee MS (1998) Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes Dev* **12:** 679-691

Shen EC, Stage-Zimmermann T, Chui P, Silver PA (2000) The Yeast mRNA-binding Protein Npl3p Interacts with the Cap-binding Complex. *J Biol Chem* **275:** 23718-23724

Shoemaker CJ, Eyler DE, Green R (2010) Dom34:Hbs1 Promotes Subunit Dissociation and Peptidyl-tRNA Drop-Off to Initiate No-Go Decay. *Science* **330:** 369-372

Siebrasse JP, Kaminski T, Kubitscheck U (2012) Nuclear export of single native mRNA molecules observed by light sheet fluorescence microscopy. *Proc Natl Acad Sci U S A* **109:** 9426-9431

Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, Kingston RE (2011) The genomic binding sites of a noncoding RNA. *Proceedings of the National Academy of Sciences* **108:** 20497-20502

Singh N, Ma Z, Gemmill T, Wu X, DeFiglio H, Rossettini A, Rabeler C, Beane O, Morse RH, Palumbo MJ, Hanes SD (2009) The Ess1 Prolyl Isomerase Is Required for Transcription Termination of Small Noncoding RNAs via the Nrd1 Pathway. *Mol Cell* **36:** 255-266

Skaar DA, Greenleaf AL (2002) The RNA Polymerase II CTD Kinase CTDK-I Affects Pre-mRNA 3′ Cleavage/Polyadenylation through the Processing Component Pti1p. *Mol Cell* **10:** 1429-1439

Skourti-Stathaki K, Proudfoot Nicholas J, Gromak N (2011) Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Mol Cell* **42:** 794-805

Soares LM, Buratowski S (2012) Yeast Swd2 is essential because of antagonism between Set1 histone methyltransferase complex and APT (associated with Pta1) termination factor. *J Biol Chem* **287:** 15219-15231

Souquere S, Beauclair G, Harper F, Fox A, Pierron G (2010) Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies. *Mol Biol Cell* **21:** 4020-4027

Sparks KA, Mayer SA, Dieckmann CL (1997) Premature 3'-end formation of CBP1 mRNA results in the downregulation of cytochrome b mRNA during the induction of respiration in Saccharomyces cerevisiae. *Mol Cell Biol* **17:** 4199-4207

Stead JA, Costello JL, Livingstone MJ, Mitchell P (2007) The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein. *Nucleic Acids Res* **35:** 5556-5567

Steinmetz EJ, Brow DA (2003) Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Mol Cell Biol* **23:** 6339-6349

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

Steinmetz EJ, Ng SB, Cloute JP, Brow DA (2006a) cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. *Mol Cell Biol* **26:** 2688-2696

Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA (2006b) Genome-Wide Distribution of Yeast RNA Polymerase II and Its Control by Sen1 Helicase. *Mol Cell* **24:** 735-746

Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R, Hurt E (2002) TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417:** 304-308

Straszer K, Hurt E (2000) Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J* **19:** 410-420

Straszer K, Hurt E (2001) Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* **413:** 648-652

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

Suntharalingam M, Alcazar-Roman AR, Wente SR (2004) Nuclear export of the yeast mRNA-binding protein Nab2 is linked to a direct interaction with Gfd1 and to Gle1 function. *J Biol Chem* **279:** 35384-35391

Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL (2009) MENe/b nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* **19:** 347-359

Swiatkowska A, Wlotzka W, Tuck A, Barrass J, Beggs J, Tollervey D (2012) Kinetic analysis of pre-ribosome structure in vivo. *RNA* **In press**

Synowsky SA, Heck AJR (2008) The yeast Ski complex is a hetero-tetramer. *Protein Sci* **17:** 119-125

Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, Irvine K, Arakawa T, Nakamura M, Kubosaki A, Hayashida K, Kawazu C, Murata M, Nishiyori H, Fukuda S, Kawai J, Daub CO, Hume DA, Suzuki H, Orlando V, Carninci P, Hayashizaki Y, Mattick JS (2009) Tiny RNAs associated with transcription start sites in animals. *Nat Genet* **41:** 572-578

Taft RJ, Pheasant M, Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29:** 288-299

Taft RJ, Simons C, Nahkuri S, Oey H, Korbie DJ, Mercer TR, Holst J, Ritchie W, Wong JJL, Rasko JEJ, Rokhsar DS, Degnan BM, Mattick JS (2011) Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. *Nat Struct Mol Biol* **17:** 1030-1034

Takahashi S, Araki Y, Sakuno T, Katada T (2003) Interaction between Ski7p and Upf1p is required for nonsense-mediated 3'-to-5' mRNA decay in yeast. *EMBO J* **22:** 3951-3959

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

Tani H, Mizutani R, Salam KA, Tano K, Ijiri K, Wakamatsu A, Isogai T, Suzuki Y, Akimitsu N (2012) Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res* **22:** 947-956

Tani H, Nakamura Y, Ijiri K, Akimitsu N (2010) Stability of MALAT-1, a nuclear long non-coding RNA in mammalian cells, varies in various cancer cells. *Drug discoveries & therapeutics* **4:** 235-239

Terzi N, Churchman LS, Vasiljeva L, Weissman J, Buratowski S (2011) H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. *Mol Cell Biol* **31:** MCB.05590-05511

Thebault P, Boutin G, Bhat W, Rufiange A, Martens J, Nourani A (2011) Transcription regulation by the non-coding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNAP II. *Mol Cell Biol***:** MCB.01083-01010

Thiebaut M, Colin J, Neil H, Jacquier A, Seraphin B, Lacroute F, Libri D (2008) Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in S. cerevisiae. *Mol Cell* **31:** 671-682

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

Thompson DM, Parker R (2007) Cytoplasmic Decay of Intergenic Transcripts in *Saccharomyces cerevisiae*. *Mol Cell Biol* **27:** 92-101

Thomson E, Tollervey D (2010) The Final Step in 5.8S rRNA Processing Is Cytoplasmic in Saccharomyces cerevisiae. *Mol Cell Biol* **30:** 976-984

Thomson R, Libri D, Boulay J, Rosbash M, Jensen TH (2003) Localization of nuclear retained mRNAs in Saccharomyces cerevisiae. *RNA* **9:** 1049-1057

Tirosh I, Sigal N, Barkai N (2010) Widespread remodeling of mid-coding sequence nucleosomes by Isw1. *Genome Biol* **11:** R49

Toesca I, Nery CR, Fernandez CF, Sayani S, Chanfreau GF (2011) Cryptic Transcription Mediates Repression of Subtelomeric Metal Homeostasis Genes. *PLoS Genet* **7:** e1002163

Tran EJ, Zhou Y, Corbett AH, Wente SR (2007) The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. *Mol Cell* **28:** 850-859

Truant R, Fridell RA, Benson RE, Bogerd H, Cullen BR (1998) Identification and functional characterization of a novel nuclear localization signal present in the yeast Nab2 poly(A)+ RNA binding protein. *Mol Cell Biol* **18:** 1449-1458

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

Tsuboi T, Kuroha K, Kudo K, Makino S, Inoue E, Kashima I, Inada T (2012) Dom34:Hbs1 Plays a General Role in Quality-Control Systems by Dissociation of a Stalled Ribosome at the 3′ End of Aberrant mRNA. *Mol Cell* **46:** 518-529

Tsvetanova NG, Klass DM, Salzman J, Brown PO (2010) Proteome-Wide Search Reveals Unexpected RNA-Binding Proteins in Saccharomyces cerevisiae. *PLoS One* **5:** e12671

Tuck AC, Tollervey D (2011) RNA in pieces. *Trends Genet* **27:** 422-432

Tuck AC, Tollervey D (2012) An RNA reset button. *Mol Cell* **45:** 435-436

Tucker M, Staples RR, Valencia-Sanchez MA, Muhlrad D, Parker R (2002) Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in Saccharomyces cerevisiae. *EMBO J* **21:** 1427-1436

Tutucci E, Stutz F (2011) Keeping mRNPs in check during assembly and nuclear export. *Nat Rev Mol Cell Biol* **12:** 377-384

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

Ursic D, Chinchilla K, Finkel JS, Culbertson MR (2004) Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res* **32:** 2441-2452

Valentini SR, Weiss VH, Silver PA (1999) Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. *RNA* **5:** 272-280

van den Bogaart G, Meinema AC, Krasnikov V, Veenhoff LM, Poolman B (2009) Nuclear transport factor directs localization of protein synthesis during mitosis. *Nat Cell Biol* **11:** 350-356

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

van Hoof A, Frischmeyer PA, Dietz HC, Parker R (2002) Exosome-Mediated Recognition and Degradation of mRNAs Lacking a Termination Codon. *Science* **295:** 2262-2264

van Hoof A, Staples RR, Baker RE, Parker R (2000) Function of the ski4p (Csl4p) and Ski7p proteins in 3'-to-5' degradation of mRNA. *Mol Cell Biol* **20:** 8230-8243

van Werven Folkert J, 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

Vaňáčová Š, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W (2005) A New Yeast Poly(A) Polymerase Complex Involved in RNA Quality Control. *PLoS Biol* **3:** e189

Vasiljeva L, Buratowski S (2006) Nrd1 Interacts with the Nuclear Exosome for 3' Processing of RNA Polymerase II Transcripts. *Mol Cell* **21:** 239-248

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

Vasiljeva L, Kim M, Terzi N, Soares LM, Buratowski S (2008b) Transcription Termination and RNA Degradation Contribute to Silencing of RNA Polymerase II Transcription within Heterochromatin. *Mol Cell* **29:** 313-323

Venkatesh S, Smolle M, Li H, Gogol MM, Saint M, Kumar S, Natarajan K, Workman JL (2012) Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **advance online publication**

Vinciguerra P, Iglesias N, Camblong J, Zenklusen D, Stutz F (2005) Perinuclear Mlp proteins downregulate gene expression in response to a defect in mRNA export. *EMBO J* **24:** 813-823

Viphakone N, Voisinet-Hakil F, Minvielle-Sebastia L (2008) Molecular dissection of mRNA poly(A) tail length control in yeast. *Nucleic Acids Res* **36:** 2418-2433

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457:** 854-858

Vitaliano-Prunier A, Babour A, Herissant L, Apponi L, Margaritis T, Holstege FC, Corbett AH, Gwizdek C, Dargemont C (2012) H2B ubiquitylation controls the formation of export-competent mRNP. *Mol Cell* **45:** 132-139

Vo LT, Minet M, Schmitter JM, Lacroute F, Wyers F (2001) Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. *Mol Cell Biol* **21:** 8346-8356

Wagschal A, Rousset E, Basavarajaiah P, Contreras X, Harwig A, Laurent-Chabalier S, Nakamura M, Chen X, Zhang K, Meziane O, Boyer F, Parrinello H, Berkhout B, Terzian C, Benkirane M, Kiernan R (2012) Microprocessor, Setx, Xrn2, and Rrp6 Co-operate to Induce Premature Termination of Transcription by RNAPII. *Cell* **150:** 1147-1157

Wan Y, Qu K, Ouyang Z, Kertesz M, Li J, Tibshirani R, Makino Debora L, Nutter Robert C, Segal E, Chang Howard Y (2012) Genome-wide Measurement of RNA Folding Energies. *Mol Cell*

Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, Glass CK, Rosenfeld MG, Fu X-D (2011a) Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474:** 390-394

Wang HW, Wang J, Ding F, Callahan K, Bratkowski MA, Butler JS, Nogales E, Ke A (2007) Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing. *Proc Natl Acad Sci U S A* **104:** 16844-16849

Wang KC, Chang HY (2011b) Molecular Mechanisms of Long Noncoding RNAs. *Mol Cell* **43:** 904-914

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY (2011c) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472:** 120-124

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 X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454:** 126-130

Wang Z, Castaño IB, De Las Peñas A, Adams C, Christman MF (2000) Pol κ: A DNA Polymerase Required for Sister Chromatid Cohesion. *Science* **289:** 774-779

Wasmuth Elizabeth V, Lima Christopher D (2012) Exo- and Endoribonucleolytic Activities of Yeast Cytoplasmic and Nuclear RNA Exosomes Are Dependent on the Noncatalytic Core and Central Channel. *Mol Cell*

Weir JR, Bonneau F, Hentschel J, Conti E (2010) Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proceedings of the National Academy of Sciences* **107:** 12139-12144

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

Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **309:** 1570-1573

Wilson SM, Datar KV, Paddy MR, Swedlow JR, Swanson MS (1994) Characterization of nuclear polyadenylated RNA-binding proteins in Saccharomyces cerevisiae. *J Cell Biol* **127:** 1173-1184

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

Wong CM, Qiu H, Hu C, Dong J, Hinnebusch AG (2007) Yeast cap binding complex impedes recruitment of cleavage factor IA to weak termination sites. *Mol Cell Biol* **27:** 6520-6531

Wyers F, Rougemaille M, Badis G, Rousselle J-C, Dufour M-E, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, Libri D, Jacquier A (2005) Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New poly(A) Polymerase. *Cell* **121:** 725-737

Xu C, Henry MF (2004) Nuclear export of hnRNP Hrp1p and nuclear export of hnRNP Npl3p are linked and influenced by the methylation state of Npl3p. *Mol Cell Biol* **24:** 10742-10756

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**

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 Y, Bai X, Lee I, Kallstrom G, Ho J, Brown J, Stevens A, Johnson AW (2000) Saccharomyces cerevisiae RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. *Mol Cell Biol* **20:** 4006-4015

Yadon AN, Van de Mark D, Basom R, Delrow J, Whitehouse I, Tsukiyama T (2010) Chromatin Remodeling around Nucleosome-Free Regions Leads to Repression of Noncoding RNA Transcription. *Mol Cell Biol* **30:** 5110-5122

Yang L, Duff M, Graveley B, Carmichael G, Chen L-L (2011a) Genomewide characterization of non-polyadenylated RNAs. *Genome Biol* **12:** R16

Yang L, Lin C, Liu W, Zhang J, Ohgi Kenneth A, Grinstein Jonathan D, Dorrestein Pieter C, Rosenfeld Michael G (2011b) ncRNA- and Pc2 Methylation-Dependent Gene Relocation between Nuclear Structures Mediates Gene Activation Programs. *Cell* **147:** 773-788

Yao W, Roser D, Köhler A, Bradatsch B, Baßler J, Hurt E (2007) Nuclear Export of Ribosomal 60S Subunits by the General mRNA Export Receptor Mex67-Mtr2. *Mol Cell* **26:** 51-62

Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou M-M (2010) Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Mol Cell* **38:** 662-674

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

Yin Q-F, Yang L, Zhang Y, Xiang J-F, Wu Y-W, Carmichael Gordon G, Chen L-L (2012) Long Noncoding RNAs with snoRNA Ends. *Mol Cell*

Yoon J-H, Abdelmohsen K, Srikantan S, Yang X, Martindale Jennifer L, De S, Huarte M, Zhan M, Becker Kevin G, Gorospe M (2012) LincRNA-p21 Suppresses Target mRNA Translation. *Mol Cell*

Yoon OK, Brem RB (2010) Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA* **16:** 1256-1267

Yu MC, Bachand F, McBride AE, Komili S, Casolari JM, Silver PA (2004) Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors. *Genes Dev* **18:** 2024-2035

Zenklusen D, Vinciguerra P, Strahm Y, Stutz F (2001) The Yeast hnRNP-Like Proteins Yra1p and Yra2p Participate in mRNA Export through Interaction with Mex67p. *Mol Cell Biol* **21:** 4219-4232

Zenklusen D, Vinciguerra P, Wyss JC, Stutz F (2002) Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol Cell Biol* **22:** 8241-8253

Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, Xiao X, Booth CJ, Wu J, Zhang C, Spector DL (2012a) The lncRNA Malat1 Is Dispensable for Mouse Development but Its Transcription Plays a cis-Regulatory Role in the Adult. *Cell reports* **2:** 111-123

Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ (2012b) Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *J Biol Chem* **287:** 8541-8551

Zhang Z, Fu J, Gilmour DS (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev* **19:** 1572-1580

Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT (2010) Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Mol Cell* **40:** 939-953

Zhao J, Sun BK, Erwin JA, Song J-J, Lee JT (2008) Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. *Science* **322:** 750-756

Zhelkovsky A, Tacahashi Y, Nasser T, He X, Sterzer U, Jensen TH, Domdey H, Moore C (2006) The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in Saccharomyces cerevisiae. *RNA* **12:** 435-445

Zheng C, Fasken MB, Marshall NJ, Brockmann C, Rubinson ME, Wente SR, Corbett AH, Stewart M (2010) Structural basis for the function of the Saccharomyces cerevisiae Gfd1 protein in mRNA nuclear export. *J Biol Chem* **285:** 20704-20715

Review

RNA in pieces

Alex C. Tuck and David Tollervey

Wellcome Trust Centre for Cell Biology, King's Buildings, Mayfield Road, Edinburgh, EH9 3JR, UK

Eukaryotic genomes accommodate numerous types of information within diverse DNA and RNA sequence elements. At many loci, these elements overlap and the same sequence is read multiple times during the production, processing, localization, function and turnover of a single transcript. Moreover, two or more transcripts from the same locus might use a common sequence in different ways, to perform distinct biological roles. Recent results show that many transcripts also undergo post-transcriptional cleavage to release specific fragments, which can then function independently. This phenomenon appears remarkably widespread, with even well-documented transcript classes such as messenger RNAs yielding fragments. RNA fragmentation significantly expands the already extraordinary spectrum of transcripts present within eukaryotic cells, and also calls into question how the 'gene' should be defined.

The modular gene

Initial analyses of genes envisaged a simple reading whereby different types of genetic information were physically separate: transcription was driven by promoter elements located outside the transcribed region and the transcript would either specify an amino acid sequence or adopt a particular fold as a structural RNA. Transcripts themselves also appeared to be modular, comprising assemblies of distinct sequence elements. As analyses became more sophisticated, they increasingly revealed the use of alternative sites for transcription initiation, termination and splicing [1–3], which are now known to be widespread. The resultant transcripts were, however, seen as related sequence variants that modified the functions of the basic gene. Moreover, the modular notion of genes and transcripts largely assumed that each sequence element had a single function, with diversity arising from its inclusion or exclusion. For example, exons could be present or absent from a transcript, and promoter elements could be bound or unbound.

The genomic palimpsest

It is now apparent that multiple layers of information are superimposed within eukaryotic genetic sequences (Box 1). For example, a protein-coding transcript can concomitantly carry sequence-specific and structural information, governing its folding, protein binding, processing, localization and decay, as well as specifying an amino acid sequence. Similarly, at the DNA level, numerous layers of regulatory information pervade the transcribed region (Box 1), blurring the distinction between regulatory and transcribed sequences and refuting the notion of a modular gene.

Even the boundary between whether information is read from DNA or RNA is becoming blurred. Information initially assumed to be read from DNA is sometimes read at the RNA level, and vice versa: plant small interfering RNAs (siRNAs) recognize DNA targets via binding to specialized long non-coding RNAs (lncRNAs) synthesized by a dedicated RNA polymerase (Pol V) [4] and the nuclear mRNA cap-binding complex can bind to genomic coding regions to promote transcription initiation [5]. Furthermore, RNA signals can be recognized, in part, by proteins bound to chromatin, with recent examples provided by analyses of the effects of histone modifications on patterns of alternative splicing [6].

Thus, genetic information is arranged in an interleaved, overlapping fashion in both DNA and RNA. Rather than being modular, the genome resembles a 'palimpsest', an ancient parchment on which the original text has been overwritten numerous times (discussed further in Box 1). Multiple layers of genetic information can be embedded within a single sequence and, consequently: (i) a single sequence can perform multiple functions; and (ii) genetic information is not restricted in where it resides (transcribed or regulatory regions, DNA or RNA).

The many ways to use a sequence

This overlapping arrangement of genetic information contributes greatly to transcript diversity and complexity. A single locus can produce multiple transcripts that use shared sequences in distinct ways to fulfill a spectrum of fundamentally different biological functions (Figure 1).

The simplest incarnation of this concept is the handful of 'dual-functional' transcripts (Figure 1a). Here, transcripts identical in sequence and length perform alternative functions. For example, some have overlapping reading frames and encode two different proteins [7]. Other examples include the U1 small nuclear RNA (snRNA), a spliceosome component that can also protect pre-mRNAs from premature cleavage [8], and 7SK, a noncoding RNA that regulates multiple transcription factors [9]. In addition, it was known that introns within eukaryotic mRNAs (and mRNA-like species that lack protein-coding capacity) could encode small nucleolar RNAs (snoRNAs) and miRNAs, which are excised by processing [10].

However, recent high-throughput transcriptome sequencing studies reveal that multifunctional sequences occur very frequently. Eukaryotic genomes turn out to be pervasively transcribed, with many loci producing ensembles of interleaved transcripts [11,12]. Diverse functions for these are emerging; for example, lncRNAs, which

Corresponding author: Tollervey, D. (d.tollervey@ed.ac.uk).

 0168-9525/$ – see front matter doi:10.1016/j.tig.2011.06.001

frequently overlap or run antisense to protein-coding genes, can direct nucleosome assembly or provide a scaffold to recruit chromatin-modifying enzymes [13,14]. Even enhancer and promoter regions are transcribed [15], further eroding the dichotomy between regulatory and transcribed regions. Notably, the overlapping arrangement of genetic information at many genomic loci results in cohorts of transcripts that share common sequences but nonetheless show distinct functions (Figure 1b).

The most recent revelation is that direct RNA cleavage further expands the spectrum of functionally distinct transcripts [16–21] (Figure 1c). Stable fragments are derived from well-characterized classes of parent RNAs, notably tRNAs, mRNAs and snoRNAs. A growing body of evidence indicates that fragmentation does not simply reflect RNA degradation, but generates a *bona fide* class of transcripts that perform functions distinct from their parents. The current surge in deep-sequencing studies has been instrumental in the discovery of many RNA fragments and will undoubtedly lead to further progress in this field.

Here, we review the diverse origins and potential functions of these fragments. We conclude by considering how the interleaved arrangement of genetic information and the enormous complexity of the transcriptome impact upon the definition of a gene, and the implications for future research.

Small RNA fragments

Recent genome-wide studies have identified numerous large and small fragments arising from within annotated genes [17,19,22–24] (Figure 2). The small fragments [<30 nucleotides (nt)] have received particular attention, because they resemble miRNAs and, in some cases, bind the Argonaut (Ago) protein family (Box 2) [25]. Small fragments are generated by post-transcriptional cleavage of diverse parent transcripts by endonucleases (outlined in Box 2) and at least some have been shown to perform functions distinct from those of their parents.

snoRNA fragments

Well-characterized fragmentation-derived miRNA-like RNAs are generated from snoRNA-like precursors. The snoRNAs select sites of covalent RNA modification, with box H/ACA snoRNAs directing pseudouridylation and box C/D snoRNAs directing 2′-*O*-methylation. Many eukaryotes possess snoRNA-derived fragments (sdRNAs), with > 60% of snoRNAs represented (Figure 2a) [19,24,26–33]. The sdRNAs range from 15 to 35 nt and map to 5′, 3′ and central regions of snoRNAs. However, the lengths and distribution across the parent snoRNAs vary between species and even among mice subjected to a training regime [30,31]. Numerous factors apparently determine sdRNA abundance, including the poly(A) polymerase Cid14 (a cofactor for the exosome nuclease complex) and the RNA-binding proteins DiGeorge syndrome critical region gene 8 (DGCR8) and Loquacious (Box 2) [31,32].

The relationship between snoRNAs and miRNA-like RNAs has been investigated from both 'ends': identified human miRNAs were found to be derived from snoRNA-like precursors (termed primary miRNAs, 'pri-miRNAs'), whereas human snoRNAs were found to be fragmented into miRNA-like sdRNAs [26,28,29,33,34]. The snoRNA-like pri-miRNAs fall into both the H/ACA and C/D classes and can adopt folds resembling other pri-miRNAs; they also show typical snoRNA properties, including binding of characteristic proteins. Several miRNA-like sdRNAs, derived from both classes, have been shown to bind the silencing machinery and exhibit *trans*-silencing activity on endogenous targets [26,28,34].

However, whereas pri-miRNAs are cleaved by the Microprocessor complex (Drosha and DGCR8; Box 2), only a subset of H/ACA-derived sdRNAs require Microprocessor components for synthesis [26,31] and accumulation of box C/D sdRNA can be independent of DGCR8 and Dicer [31]. Thus, sdRNAs are diverse, but probably include many species that function as miRNAs.

Box 1. The layered arrangement of genetic information

Genetic information is arranged in an interleaved, overlapping fashion in eukaryotes (Figure I), resembling ancient parchments on which the original text was overwritten multiple times ('palimpsests'). DNA packaging into chromatin is partly influenced by local sequence. Sequence-specific binding recruits barrier elements against which nucleosomes are packed, chromatin remodeling factors and other non-histone proteins [85]. Moreover, the intrinsic sequence-dependent curvature of DNA directly influences nucleosome organization. Higher-order chromatin architecture is also sequence dependent. For example, certain DNA elements are tethered to nuclear pores and 'insulator' sequences interact to partition chromatin into domains [86,87]. Notably, some promoters can also act as insulators.

During transcription, information stored within DNA is transferred to RNA. Classical models envisaged that the transcription machinery primarily assembles at promoters. However, genome-wide profiling of transcription factors revealed that many bind transcribed regions [88,89]. Thus, regulatory information is not restricted to distinct sites, but pervades most genomic sequences.

Genetic information in the RNA transcripts is similarly interleaved and overlapping, but with additional capacity provided by complex features of secondary and tertiary structure. The superimposed layers of information within a transcript govern most aspects of its existence, exemplified by mRNAs as shown in Figure I. Primary mRNA transcripts undergo extensive processing, directed by sequence or structural elements. Cleavage, polyadenylation and splicing factors bind specific sequences and the conversion of adenosine to inosine is directed by three-dimensional folds [90,91]. mRNA localization is guided by structural elements in the protein-coding sequence as well as 3′ untranslated regions.

The primary role of the mRNA protein-coding sequence is as a template for translation, but additional layers of information regulate this process: (i) structures around the start codon can impede initiation; (ii) elongation might initially be slowed by codons with rare tRNAs, preventing downstream ribosomal traffic jams; and (iii) secondary structures and biased codon usage coordinate elongation rate with protein folding [92–95].

Transcripts are eventually degraded, with stabilities directed by numerous sequence elements. For example, AU- and GU-rich elements recruit decay factors and many sites are targets for RNA-guided silencing [96]. RNAi-related pathways use small miRNA guides, whereas cleavage in Staufen 1-mediated mRNA decay is programmed by lncRNAs [97,98]. Elements modulating stability reside in protein-coding as well as untranslated regions, for example Nrd1/Nab3-binding motifs, which promote decay [99]. Additionally, functionally distinct mRNA fragments are generated. Thus, many layers of information are superimposed within a transcript.

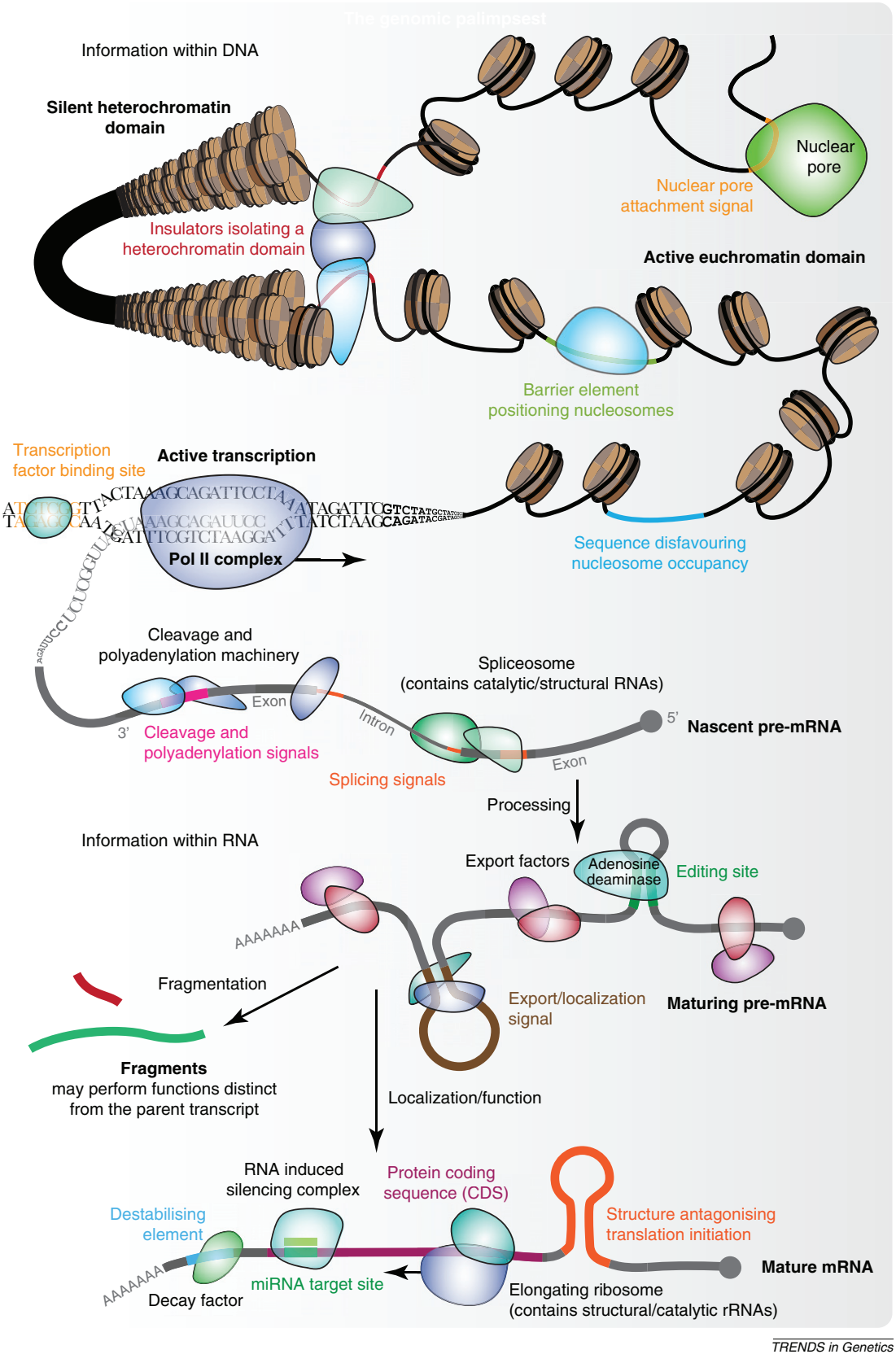


Figure I. DNA and RNA contain multiple overlapping and interleaved layers of information.

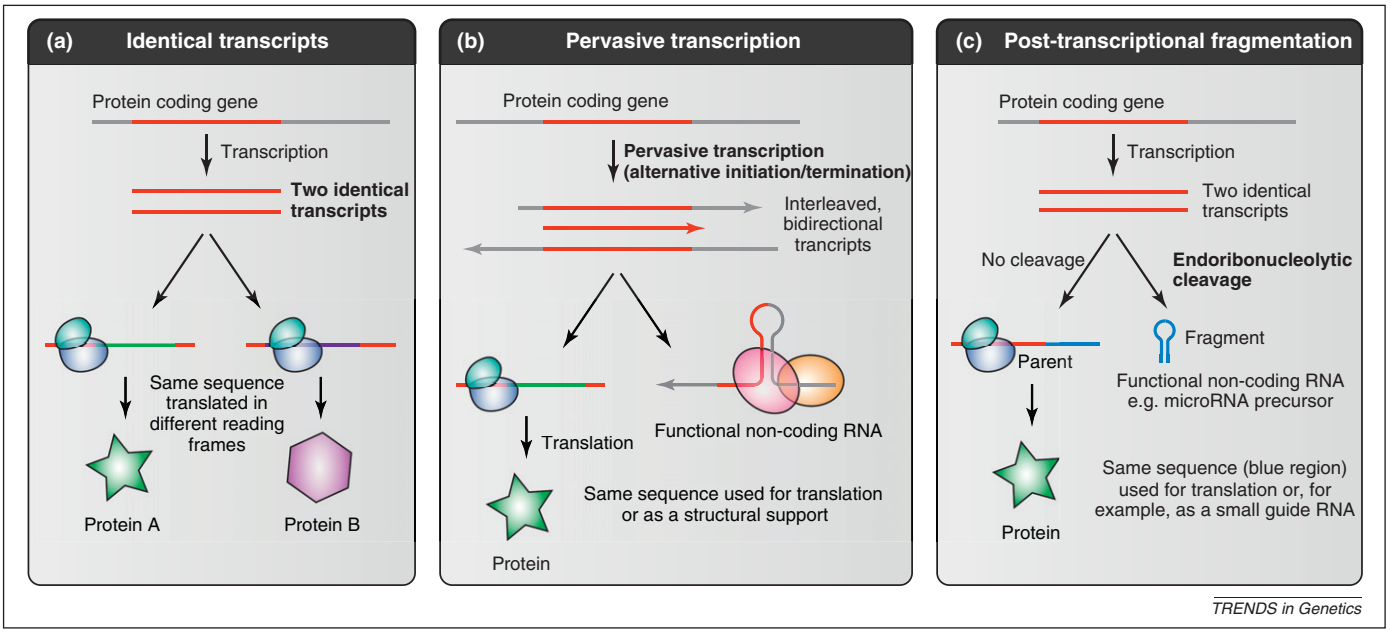


Figure 1. Two transcripts from the same locus can use the same sequence to different functional effects. The overlapping arrangement of genetic information enables a single sequence to encode multiple functions. This principle is embodied at many genomic loci, which generate ensembles of transcripts with shared sequences but disparate functions. This raises questions about how a specific function is assigned to a transcript, given the numerous possibilities. There are several explanations, illustrated by the various ways in which overlapping transcripts are generated: **(a)** Two transcripts identical in sequence and length might function differently, perhaps being translated in alternative reading frames (green or purple) to generate distinct proteins. Here, extrinsic factors are responsible for specifying which reading frame should be used. **(b)** Alternative transcription initiation and/or termination generate an ensemble of interleaved transcripts from a single genomic locus. Within this ensemble, a shared sequence (red) can perform distinct functions, perhaps contributing to an open reading frame (green) in one transcript and a structural feature in another. Here, the function of a sequence is governed by its context, with the different lengths and orientations of transcripts perhaps affecting their folding or recruitment of binding factors. **(c)** Many classes of transcript might act as precursors to shorter fragments, excised by post-transcriptional cleavage. These fragments might function in ways distinct from those of their parents. Thus, within the context of the shorter fragment, a shared sequence (blue) can perform an alternative role. This indicates that the length of a transcript might contribute to specifying which of several possible functions is performed by a particular sequence. Other post-transcriptional processes (such as splicing) can also generate alternative transcripts, but are beyond the scope of this review.

tRNA fragments

High-throughput sequencing detected small fragments of human and mouse tRNAs, some of which are highly abundant [16–20,24,35–38]. The tRNA regulatory fragments (tRFs) are classified as tRF-1, tRF-3 or tRF-5, depending on their origin [20] (Figure 2b). These probably arise from specific processing, because: (i) their abundance varies between cell lines and is distinct from parent tRNAs; (ii) tRF ends are precisely defined; (iii) there are sequence preferences for cleavage; and (iv) some specific functions have been uncovered [19,20]. Although we discuss the tRF classes separately, their biogenesis and probable functions overlap considerably.

tRF-3. tRF-3 species are 13–22 nt fragments derived from the 3′ end of mature tRNAs cleaved within the T-loop [16,19,20,24,35]. Examples studied in detail resemble miRNAs in exhibiting Dicer-dependent processing, association with Ago1–4 and *trans*-silencing activity (Box 2) [18,19,35]. Dicer binds dsRNA with a 2-nt 3′ overhang, cutting at a specific distance from the end of the duplex. One mouse tRF-3-like fragment arises when an isoleucine tRNA forms a long hairpin rather than the standard cloverleaf [16]. Other tRNAs can form duplexes with complementary RNAs. For example, a tRF-3 fragment arises by cleavage of $tRNA^{Lys}$ bound to the HIV-1 primer binding site. This tRF-3 species reduces HIV-1 replication in infected cells and might reflect a more general retroviral defence mechanism [19].

However, although tRF-3 RNAs resemble miRNAs, there are clear distinctions: Dicer-null mouse cells retain many tRFs [16] and tRF-3 RNAs do not associate with some miRNA-binding factors, such as Mov10 [18].

tRF-5. tRF-5 fragments arise from mature tRNA 5′ ends via D-loop cleavage. Similar to tRF-3 RNAs, some are Dicer dependent, cytoplasmic and able to associate with Ago proteins, albeit weakly [17,24]. The detection of long (31 nt) tRF-5 fragments suggests that compact tRNA folding allows Dicer to generate unusually long products [24]. However, *trans*-silencing activity has not yet been reported. One possibility is that tRF-5 and tRF-3 species resemble ‘mature’ and ‘star’ strands of precursor miRNAs (pre-miRNAs), respectively [17]. Selective stabilization of the ‘mature’ strand could explain why tRF-5 and tRF-3 fragments are not always either detected or functional [19]. Human miRNAs lack terminal modifications, whereas characterized fragments of $tRNA^{Gln}$ are 3′ modified, potentially explaining the weak association of these and other tRFs with Ago proteins [17].

tRF-1. tRF-1 fragments correspond to precursor (pre-)tRNA 3′ trailers, with 5′ ends generated by the endonuclease tRNase Z and 3′ ends matching Pol III termination sites (Box 2) [20]. tRF-1 species are cytoplasmic, whereas pre-tRNAs are 3′ matured in the nucleus, suggesting that tRF-1 RNAs are rapidly exported [39]. Fragments from a particular tRNA vary in length and precise 5′ end, so factors other than tRNase Z might contribute to tRF-1 production [18]. Indeed, some mouse small RNAs arise from 3′ trailers resembling pre-miRNAs, potentially processed by Dicer [40]. Additionally, Ago proteins might participate in processing, as their overexpression specifically enriches shorter

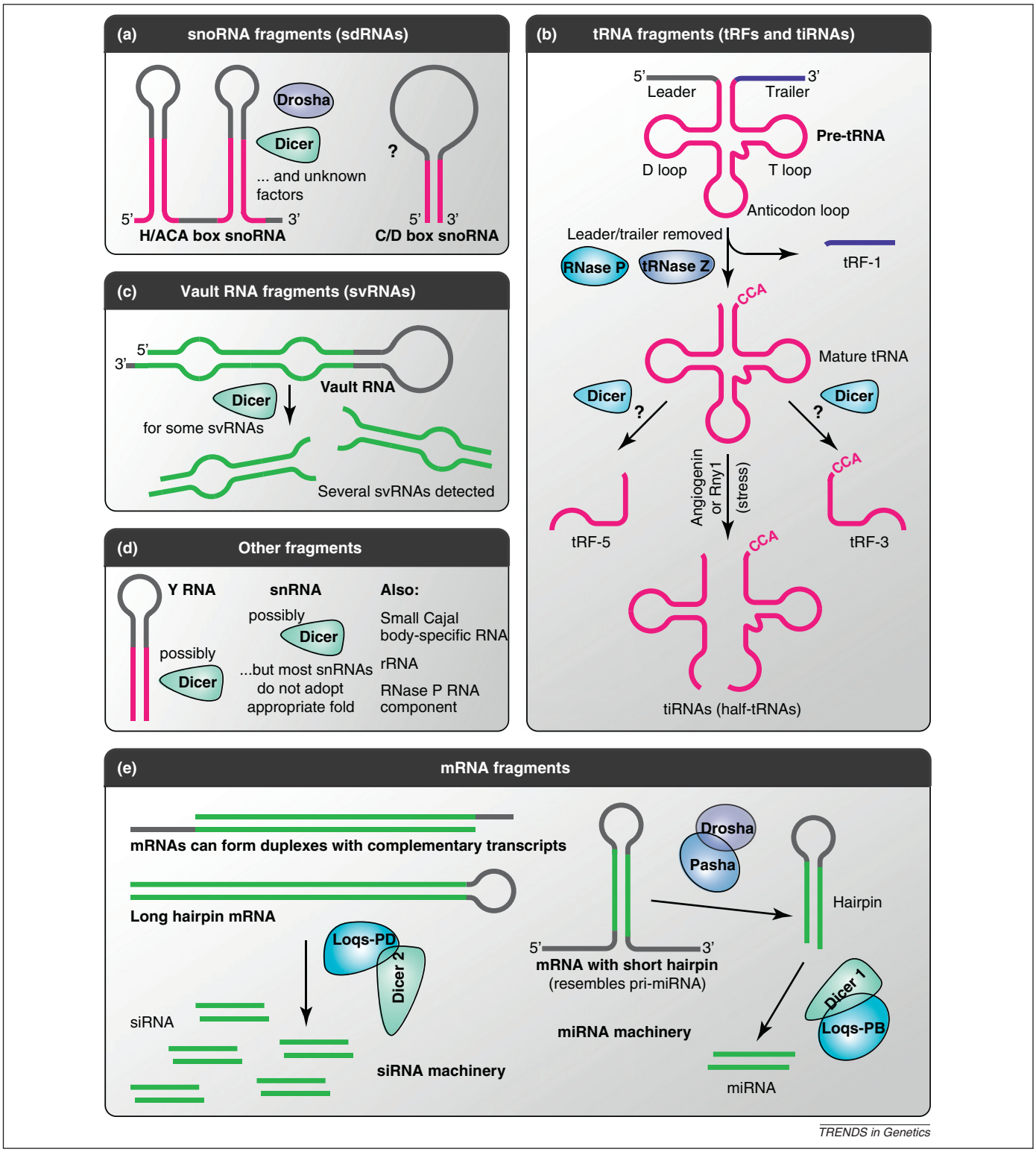


Figure 2. Products generated by post-transcriptional cleavage of diverse transcript classes. Recent studies have found that many well-documented classes of transcript can be post-transcriptionally cleaved to liberate shorter fragments, a phenomenon known as RNA fragmentation. To date, the participating endonucleases are thought to include the tRNA processing enzymes RNase P and tRNase Z, and the RNase III family members Drosha and Dicer (with their cofactors Pasha and Loqs), which generate small *trans*-silencing RNAs. Fragments from the following transcript classes have been documented: **(a)** small nucleolar RNAs (snoRNAs) target modifying enzymes to specific sites on transcripts, and fall into two classes (box H/ACA and box C/D). Fragmentation of some box H/ACA snoRNAs is catalyzed by Drosha and Dicer, but endonucleases excising fragments from other box C/D snoRNAs and other box H/ACA snoRNAs are currently unknown [31]. **(b)** Single-stranded, short regulatory tRNA fragments (tRFs) are generated by precursor (pre)-tRNA processing, which releases tRF-1 fragments (trailers), and mature tRNA cleavage, which releases tRF-3/5 fragments. Additionally, stress-induced cleavage by angiogenin or Rny1 releases longer tRNA halves (tiRNAs) [18,37,67]. **(c)** Vault RNAs (vRNAs) are a component of vault particles, which are ribonucleoprotein complexes linked to drug resistance. Several short fragments of vRNAs (svRNAs) have been detected, some of which are dependent on Dicer [44]. **(d)** Short fragments of other stable non-protein-coding RNAs have also been detected and, again, some are dependent on Dicer [19,24,27,30]. **(e)** mRNAs can adopt structures with long or short duplexes that are processed to small RNAs by the canonical small interfering RNA (siRNA) or miRNA biogenesis machinery. Abbreviations: pri-miRNA, primary miRNA; sdRNA, snoRNA-derived fragment; snRNA, small nuclear RNA.

Box 2. Endoribonucleases involved in RNA fragmentation

Analyses of eukaryotic RNA surveillance and turnover have largely focused on the roles of exonucleases. However, endonucleases also play important roles in RNA metabolism [100,101]. This is exemplified by RNA fragmentation, which involves the endonucleolytic cleavage of a parent transcript to release specific fragments, potentially with distinct functions. To date, most endonucleases implicated in RNA fragmentation also function in well-characterized RNA processing pathways.

RNAi and related pathways

miRNA biogenesis involves cropping of a pri-miRNA by Drosha and Pasha/DGCR8 to liberate a pre-miRNA, then excision of the mature miRNA by Dicer 1 and Loqs-PB (Figure Ia) [98]. Drosha and Dicer are RNase III endonucleases that make staggered cuts within a duplex. Any transcript could potentially provide a non-canonical substrate for cleavage by adopting a fold resembling pri-miRNAs or pre-miRNAs. This probably underlies the fragmentation of some tRNAs, snoRNAs, mRNAs and other structured RNAs.

siRNA biogenesis involves the processing of long duplexes by a complex between Dicer 2 and Loqs-PD (Figure Ib) [98]. Originally thought to derive exclusively from exogenous sources, siRNAs from endogenous intermolecular and intramolecular mRNA duplexes are now documented, revealing another source of RNA fragments.

The effectors of RNAi-related pathways are the Ago proteins, of which humans possess four. Each contains a PIWI domain (resembling an RNase H fold), but only Ago2 retains endonuclease activity. Ago2 cleavage can generate RNA fragments, which might function via Ago-binding to direct further cleavages, modify mRNA translation and stability or perturb the Ago-association of canonical miRNAs or siRNAs by competition.

pre-tRNA processing

Pre-tRNAs are processed by RNase P and tRNase Z to remove the 5′ leader and 3′ trailer, then a 3′ CCA triplet is added (Figure Ic) [68]. The liberated trailers constitute an additional class of tRNA-derived fragments. Both RNase P and a cytoplasmic isoform of tRNase Z can be directed by small guide RNAs to cleave specific targets, so might represent effectors programmed by RNA fragments.

Emerging endonucleases

Other endonucleases also participate in RNA fragmentation. For example, angiogenin and Rny1 cleave tRNAs during stress. Additional candidates include: (i) the PIN-domain proteins Rrp44, Swt1 and Smg6, for which roles in RNA surveillance are emerging; (ii) RNase MRP, which processes pre-rRNA and at least one cell cycle-regulated mRNA; (iii) Rnt1, the sole RNase III enzyme in *Saccharomyces cerevisiae*, which 3′ processes pre-rRNA, pre-snRNAs and pre-snoRNAs, but also cleaves pre-mRNAs; (iv) G3BP, which cleaves specific human mRNAs; (v) Ire1, involved in the unfolded protein response and decay of endoplasmic reticulum-targeted mRNAs; and (vi) RNase H, which cleaves nascent transcripts at sites of R-loop formation [100,101].

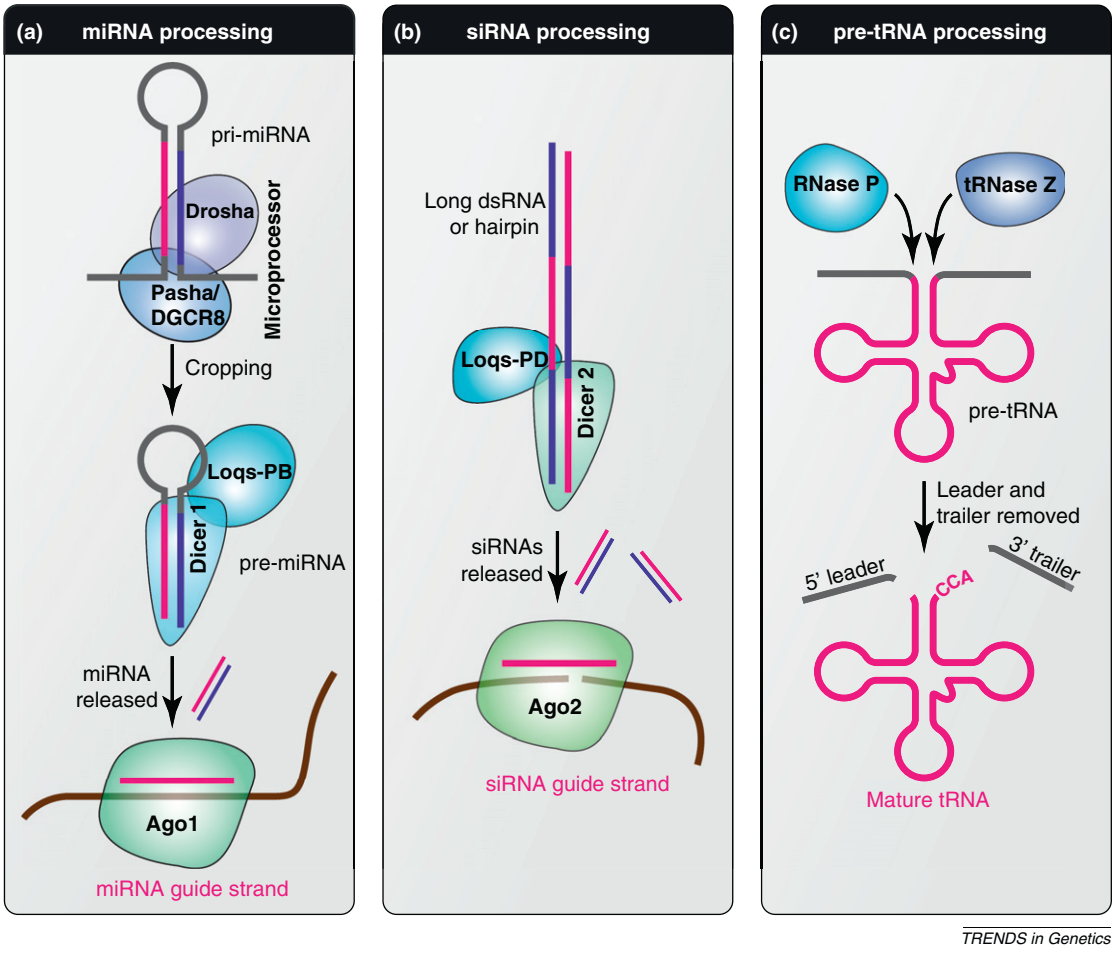


Figure I. Classical RNA processing pathways contribute to RNA fragmentation. Abbreviations: DGCR8, DiGeorge syndrome critical region gene 8; dsRNA, double-stranded RNA; pre-miRNA, precursor miRNAs; pri-miRNA, primary miRNA; pre-tRNA, precursor tRNA.

(20–21 nt) tRF-1 species [18]. This would resemble Ago2-mediated processing of the miR-451 precursor [41].

Tested tRF-1 RNAs show little *trans*-silencing activity [18,20], perhaps because they associate with Ago3/4, rather than with the slicing-competent Ago2, which preferentially binds duplex RNAs [18]. However, when duplexed with an antisense oligoribonucleotide, a tRF-1 RNA (cand45) bound Ago2 and elicited *trans*-silencing [18].

Rather than directing Ago proteins to specific targets, tRF-1 fragments might normally compete with miRNAs for Ago binding and/or alter miRNA distribution among Ago proteins, thus indirectly perturbing silencing activity [18]. Consistent with this, upregulating tRF-1 expression reduced the efficacy of miRNAs [18]. Conversely, loss of DGCR8 and Dicer increased tRF abundance, suggesting that miRNAs also antagonize tRFs [16]. In proliferative diseases such as cancer, tRNAs are overexpressed and miRNA profiles perturbed, whereas increased Pol III transcription can promote transformation [42,43]. Specifically, a tRF-1 RNA (tRF-1001) augments proliferation of a colon cancer cell line [20], and we speculate that this might be a more general property of tRFs.

Fragments of other structural RNAs

Fragments have also been detected for several other classes of RNA, including rRNA, snRNA, the RNA component of RNase P and the small, cytoplasmic vault and Y RNAs (Figures 2c and 2d) [19,24,27,30,38,44]. Some can be generated by Dicer or Drosha cleavage: Y RNAs adopt pre-miRNA-like folds; some regions of rRNAs form stable duplexes; and processing of human vault RNA 1 is Dicer dependent [19,27,30,44]. However, alternative pathways must exist as snRNAs that generate small RNAs generally lack a clear propensity to form Dicer-compatible hairpins [24].

Expression of these small RNAs is both regulated and distinct from the parent RNAs, supporting functional roles [30,44]. Small RNA fragments from snRNAs, vault RNAs (vRNAs) and rRNAs associate with Ago proteins and might therefore participate in RNA silencing [24]. Targets identified for one vRNA fragment include the mRNA encoding the drug metabolizing enzyme CYP3A4, perhaps explaining the chemotherapy resistance associated with vault particles [44]. However, no silencing activity was observed for other vRNA or Y RNA fragments [27,44].

mRNA fragments

Small mRNA fragments have been reported in diverse eukaryotes (Figure 2e) [16,19,24,45–53]. Some mRNAs with long duplexes or hairpins are processed by Dicer to generate endogenous siRNAs, often from repeat regions. For example, the *Drosophila CG4068* 3′ UTR produces siRNAs that silence the *mus308* gene [50]. Similarly, overlapping, oppositely oriented transcripts can generate '*cis*-natural antisense transcript siRNAs' (*cis*-NAT-siRNAs) from duplexed, complementary regions. This was discovered in *Arabidopsis*, where stress-induced antisense transcripts recruit Ago to their partner [54]. *Drosophila* produces *cis*-NAT-siRNAs from approximately 25% of convergent transcripts and, in mammals, they arise from convergent or divergent pairs [19,49,51,55]. In *Schizosaccharomyces pombe*, cell cycle-dependent read-though of transcription terminators on convergent genes leads to overlapping transcripts. These provoke siRNA-dependent heterochromatin formation, providing an autoregulatory system for siRNA components [56]. Bidirectional promoters, which are very common and produce divergent transcripts with a short overlap, are another source of double-stranded RNA [19,57], as are complementary transcripts from distant loci (such as gene–pseudogene pairs) [51]. Still more generally, pervasive transcription might provide antisense partners for the abundant mRNAs that generate siRNAs in the absence of hairpins or annotated antisense partners [19,49,58].

Endogenous miRNAs are processed from short hairpin structures, which are frequently located within mRNA introns. Some introns harbor full pri-miRNAs processed by Drosha and Dicer [59], whereas short 'mirtrons' produce a pre-miRNA directly by splicing and debranching (sometimes aided by exosome-mediated trimming), bypassing the need for Drosha [16,52,53,60,61]. Small Ago-associated RNAs are also excised from an exonic, pre-miRNA-like fold within human CYP46A1 [24] and Drosha apparently generates small RNAs from pre-miRNA-like folds in mouse mRNAs [45].

Together, these studies identify small RNA fragments that arise from specific post-transcriptional cleavage and interact with Ago proteins to function as mRNA silencers, perturb miRNA profiles, or stimulate proliferation. However, many small fragments are generated independently of Drosha and Dicer, do not associate with Ago and lack apparent silencing ability [16,62]. These might possess as yet untested functions.

Longer RNA fragments

In addition to the short fragments discussed above, longer (>30 nt) fragments are also generated from diverse transcripts [19,21,31,38,63–66]. These too are excised by post-transcriptional cleavage and can perform functions distinct from those of their parents.

Stress-induced fragments

The best-studied longer fragments are stress-induced, tRNA-derived RNAs (tiRNAs), found in diverse eukaryotes (Figure 2b). These are generated by endonucleolytic cleavage in the anticodon loop of mature tRNAs, catalyzed by Rny1 in yeast and angiogenin in mammals [19,64,65,67–69]. Under normal conditions, angiogenin is inhibited by binding to RNH1, whereas Rny1 is sequestered in the vacuole away from its substrates [65,67]. These enzymes have limited specificity, because cleavage occurs at various positions around the anticodon loop. Most tRNAs are susceptible to cleavage [21] although not all are equally cleaved [64]. Rny1 is an RNase T2 family member, and is predicted to leave 5′ hydroxyl and 2′-3′ cyclic phosphate groups. This is unusual for intracellular cleavage and significant because the 5′ hydroxyl group confers resistance to degradation by the 5′ exonucleases Xrn1 and Rat1/Xrn2. Moreover, the tRNA splicing ligase requires 5′ hydroxyl and 2′-3′ cyclic phosphate groups, suggesting that tRNA anticodon cleavage might be reversible under some circumstances [70].

Cleavage of tRNAs might help adaptation to stress, but this probably does not simply reflect reduced tRNA

availability as only a minority are cleaved [21]. Indeed, too much cleavage appears deleterious: *Dnmt2* mutants fail to methylate tRNAs with C38 and show increased tRNA cleavage but reduced stress tolerance [71]. Treating mammalian cells with angiogenin or synthetic tiRNAs inhibits translation [65] and promotes formation of stress granules (SGs); sites where translationally silenced mRNAs are sorted for re-initiation, decay or storage [72,73]. Denatured tiRNAs cannot inhibit translation, suggesting that structural features are important [66].

tiRNAs might also warn neighboring cells of imminent stress, giving them time to prepare. Angiogenin (a secreted enzyme) and tiRNAs (found in phloem sap) could both act as messengers.

Functional long fragments

Additional parent transcript classes produce long fragments. Deep-sequencing detected 30–40 nt fragments derived from human snoRNAs, snRNAs, rRNAs and mRNAs, whereas CAGE tags (cap analysis of gene expression; short sequences representing the 5′ ends of cDNAs) identified many transcripts of >200 nt [1,19,22,23,38,62]. Consistent with post-transcriptional cleavage, many CAGE tags do not coincide with hallmarks of transcription initiation and are derived from both protein-coding regions and 3′ untranslated regions (3′ UTRs). Some of these transcripts contain 5′ exons too short to have undergone splicing, confirming they were cleaved from already spliced transcripts and then capped [22,23,38,62]. Some mRNAs are cleaved by Ago2, targeted by miRNAs, or directly by Drosha [38,45,74,75]. However, the remainder must be generated by other endonucleases (Box 2).

Accumulating evidence suggests these mRNA fragments are functional. Some are capped, perhaps by the cytoplasmic capping complex, suggesting that they are stable, and many cleavages are conserved between humans and mice [23,62,76]. Inspection of CAGE tags and *in situ* hybridization reveals that expression of exonic fragments is tissue and developmental stage specific, and different from the parent mRNA, suggesting that they function in distinct ways [23,62]. Although some fragments potentially encode truncated proteins, many lack coding capacity and presumably function as non-coding RNAs. A few possibilities are illustrated by studies on the functions of 3′ UTRs, which by definition are not protein-coding.

Signals present in 3′ UTRs frequently control mRNA localization and stability, but separate functions have been found for several 3′ UTRs. The 3′ UTR of the *oskar* mRNA is expressed independently and is sufficient to restore Staufen accumulation to *oskar* mutant oocytes, perhaps providing a scaffold to localize regulatory proteins [62]. Other 3′ UTRs are tumor suppressors, whereas the 3′ UTRs of several muscle structural genes enhance myogenic gene expression [62]. In general, 3′ UTRs harbor protein-binding motifs and could therefore potentially sequester regulatory factors away from other targets.

Silencing by long fragments

Long fragments could target sequence-specific cleavage by RNase P or tRNase ZL, both of which can be programmed with guides resembling tRNA fragments [77,78]. However, such guides need not be derived from tRNAs, because tRNase ZL also binds rRNA and snRNA fragments. This mechanism potentially regulates many targets: tRNase ZL overexpression results in downregulation of 41 mRNAs, some of which were validated as targets for tRNase ZL primed with half-tRNAGlu or an rRNA fragment [77].

Despite being longer than canonical miRNAs/siRNAs, 30–40 nt fragments might still bind Ago to elicit *trans*-silencing. Ago2 binds pre-miRNAs or long RNAs, and the exosome and/or Ago might trim extended siRNAs [41,79,80]. Additionally, long initial fragments appear to be processed to smaller fragments [22,23,38]. Indeed, the discovery that even miRNAs generate smaller fragments hints at the existence of complex fragmentation hierarchies with many levels [81]. Many exon-derived small RNAs are generated independently of siRNA and/or miRNA processing factors, and sequential fragmentation of mRNAs to long then short species might provide a distinct pathway for small RNA generation [23]. An additional link between long fragments and RNA silencing is the observation that, in cells subjected to stress, Argonaute proteins accumulate in SGs, the structures induced by tiRNAs [82].

Overall, long fragments appear to be excised by specific, post-transcriptional cleavage of diverse transcripts. Accumulating evidence suggests that some function as scaffolds, translational inhibitors, tumor suppressors, transcriptional activators or RNA silencers, whereas others might be precursors to smaller fragments.

Functional fragments or pointless pieces?

We conclude that post-transcriptional fragmentation is widespread throughout eukaryotic transcriptomes and can generate fragments that function independently of the parent. However, a key question is how many of these fragments are functionally important? Are most just spurious degradation intermediates or evolutionary leftovers, maintained simply because counter selection is not strong enough? This resembles the debate over lncRNAs, with initial estimates of how many are functionally important ranging from almost all to almost none; and the truth appearing to be somewhere in between.

Many RNA fragments are generated by precise cleavage, conserved from mammals to protozoa, such as *Tetrahymena* [83]. They are expressed in a tissue- and condition-specific manner, their abundance is uncoupled from that of their parents and, within some transcript classes (e.g. snoRNAs and tRNAs), fragmentation appears to be a near ubiquitous phenomenon [21,31]. However, although these observations are strongly indicative of function, biological roles have only been directly demonstrated for a handful of examples. A global assessment of functionality is hampered by the fact that fragments might play widely disparate roles. Furthermore, many of these roles might only be apparent under certain conditions. This is analogous to the difficulties faced when investigating yeast genes, 80% of which are essential for viability only when tested in combination (synthetic lethal interactions).

Currently, the broadest functional studies of RNA fragments are those estimating the extent to which they participate in RNAi-related *trans*-silencing. However, within sequenced pools of Argonaute-bound small RNAs, reported

abundances of tRNA fragments range from approximately 0.01 to 10% of hits, with snoRNA fragments comprising approximately 0.2–1% [24,26,32,84]. The documented abundance of RNA fragments in whole-cell extracts similarly varies widely, from approximately 0.1 to 10% of total reads for snoRNA and tRNA fragments alike [19,24,31]. Furthermore, some fragments (such as snoRNA-derived RNAs) have known Ago-dependent functionality but are not always enriched in Ago immunoprecipitates [24]. Another complication is that, in *S. pombe* strains lacking the activities of the TRAMP and exosome RNA surveillance complexes, diverse RNA degradation products accumulate and bind Ago1 [32]. Thus, Ago association does not necessarily demonstrate biological function. A precise assessment of the participation of RNA fragments in *trans*-silencing therefore awaits the high-throughput identification of guide–target pairs, and a comprehensive list of RNA fragments. However, even if fragments do not have direct targets, the very fact that they are occupying RNA binding sites on proteins will perturb the binding of other transcripts, with potentially widespread consequences [18].

The multifunctional gene

The unexpected discovery of fragments derived from the best-characterized classes of transcripts illustrates the powerful and far-reaching consequences of the interleaved and overlapping arrangement of genetic information. Sequences with a single function might, in fact, be the exception rather than the rule, and the ability of a single sequence to encode multiple layers of information permits an almost unimaginable overall level of regulatory complexity and transcriptome diversity. These findings pose many pressing questions (Box 3). Notably, current understanding of RNA fragmentation is largely based on just a handful of high-throughput sequencing studies. Significant advances might therefore be made simply by reanalyzing the many additional data sets already in existence.

The gene as a system

Our initial 'modular' notion of a gene has been challenged by the realization that: (i) multiple layers of regulatory information permeate the transcribed region; (ii) eukaryotic genomes are pervasively transcribed, generating an ensemble of transcripts from any given locus; (iii) each of these transcripts might in turn undergo multiple rounds of cleavage to generate even greater complexity; and (iv) this panoply of transcripts can perform diverse biological roles.

Box 3. Outstanding questions

- What proportion of RNA fragments is functional, and what functions do they perform?
- Which endoribonucleases excise the many fragments for which the cleavage activity is currently unknown?
- How does the cell distinguish functional fragments from unwanted degradation products?
- For any given transcript, what determines the functions that an RNA region performs, out of the multiple possibilities?
- Has the limit of transcriptome complexity now been reached? Or do additional transcript classes await discovery and what might their nature be?

The overlapping nature of the genetic information and transcripts associated with a single locus limits the value of studies of any component in isolation. We therefore suggest that each gene must now be regarded as a system, comprising a genomic region with the corresponding network of control regions and ensemble of transcripts.

Acknowledgments

We thank Grzegorz Kudla and Agata Swiatkowska for critical reading of the manuscript. This work was funded by the Wellcome Trust.

References

1 Ni, T. *et al.* (2010) A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Meth.* 7, 521–527
2 Ozsolak, F. *et al.* (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* 143, 1018–1029
3 Wang, E.T. *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476
4 Wierzbicki, A.T. *et al.* (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135, 635–648
5 Lahudkar, S. *et al.* (2011) The mRNA cap-binding complex stimulates the formation of pre-initiation complex at the promoter via its interaction with Mot1p *in vivo*. *Nucleic Acids Res.* 39, 2188–2209
6 Luco, R.F. *et al.* (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16–26
7 Xu, H. *et al.* (2010) Length of the ORF, position of the first AUG and the Kozak motif are important factors in potential dual-coding transcripts. *Cell Res.* 20, 445–457
8 Kaida, D. *et al.* (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668
9 Eilebrecht, S. *et al.* (2011) 7SK small nuclear RNA directly affects HMGA1 function in transcription regulation. *Nucleic Acids Res.* 39, 2057–2072
10 Brown, J.W.S. *et al.* (2008) Intronic noncoding RNAs and splicing. *Trends Plant Sci.* 13, 335–342
11 Costa, F.F. (2010) Non-coding RNAs: meet thy masters. *Bioessays* 32, 599–608
12 Chen, L-L. and Carmichael, G.G. (2010) Decoding the function of nuclear long non-coding RNAs. *Curr. Opin. Cell Biol.* 22, 357–364
13 Thebault, P. *et al.* (2011) Transcription regulation by the non-coding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNAP II. *Mol. Cell. Biol.* 31, 1288–1300
14 Hainer, S.J. *et al.* (2011) Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev.* 25, 29–40
15 Kim, T-K. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187
16 Babiarz, J.E. *et al.* (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* 22, 2773–2785
17 Cole, C. *et al.* (2009) Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* 15, 2147–2160
18 Haussecker, D. *et al.* (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* 16, 673–695
19 Kawaji, H. *et al.* (2008) Hidden layers of human small RNAs. *BMC Genomics* 9, 157
20 Lee, Y.S. *et al.* (2009) A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* 23, 2639–2649
21 Thompson, D.M. *et al.* (2008) tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* 14, 2095–2103
22 Fejes-Toth, K. *et al.* (2009) Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature* 457, 1028–1032
23 Mercer, T.R. *et al.* (2010) Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome. *Genome Res.* 20, 1639–1650
24 Burroughs, A.M. *et al.* (2011) Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol.* 8, 158–177

25 Miyoshi, K. *et al.* (2010) Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. *Mol. Genet. Genomics* 284, 95–103
26 Ender, C. *et al.* (2008) A human snoRNA with microRNA-like functions. *Mol. Cell* 32, 519–528
27 Meiri, E. *et al.* (2010) Discovery of microRNAs and other small RNAs in solid tumors. *Nucleic Acids Res.* 38, 6234–6246
28 Saraiya, A.A. and Wang, C.C. (2008) snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog.* 4, e1000224
29 Scott, M.S. *et al.* (2009) Human miRNA precursors with box H/ACA snoRNA features. *PLoS Comput. Biol.* 5, e1000507
30 Smalheiser, N.R. *et al.* (2011) Endogenous siRNAs and noncoding RNA-derived small RNAs are expressed in adult mouse hippocampus and are up-regulated in olfactory discrimination training. *RNA* 17, 166–181
31 Taft, R.J. *et al.* (2009) Small RNAs derived from snoRNAs. *RNA* 15, 1233–1240
32 Buhler, M. *et al.* (2008) TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces pombe* siRNA pathway. *Nat. Struct. Mol. Biol.* 15, 1015–1023
33 Ono, M. *et al.* (2011) Identification of human miRNA precursors that resemble box C/D snoRNAs. *Nucleic Acids Res.* 39, 3879–3891
34 Brameier, M. *et al.* (2011) Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic Acids Res.* 39, 675–686
35 Yeung, M.L. *et al.* (2009) Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. *Nucleic Acids Res.* 37, 6575–6586
36 Calabrese, J.M. *et al.* (2007) RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18097–18102
37 Pederson, T. (2010) Regulatory RNAs derived from transfer RNA? *RNA* 16, 1865–1869
38 Bracken, C.P. *et al.* (2011) Global analysis of the mammalian RNA degradome reveals widespread miRNA-dependent and miRNA-independent endonucleolytic cleavage. *Nucleic Acids Res.* DOI: 10.1093/nar/gkr110
39 Liao, J-Y. *et al.* (2010) Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3′ Trailers. *PLoS ONE* 5, e10563
40 Reese, T.A. *et al.* (2010) Identification of novel microRNA-like molecules generated from herpesvirus and host tRNA transcripts. *J. Virol.* 84, 10344–10353
41 Cifuentes, D. *et al.* (2010) A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* 328, 1694–1698
42 Lu, J. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature* 435, 834–838
43 Marshall, L. and White, R.J. (2008) Non-coding RNA production by RNA polymerase III is implicated in cancer. *Nat. Rev. Cancer* 8, 911–914
44 Persson, H. *et al.* (2009) The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nat. Cell Biol.* 11, 1268–1271
45 Chong, M.M.W. *et al.* (2010) Canonical and alternate functions of the microRNA biogenesis machinery. *Genes Dev.* 24, 1951–1960
46 Ghildiyal, M. *et al.* (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320, 1077–1081
47 Jin, H. *et al.* (2008) Small RNAs and the regulation of *cis*-natural antisense transcripts in *Arabidopsis*. *BMC Mol. Biol.* 9, 6
48 Kawamura, Y. *et al.* (2008) *Drosophila* endogenous small RNAs bind to Argonaute-2 in somatic cells. *Nature* 453, 793–797
49 Okamura, K. *et al.* (2008) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat. Struct. Mol. Biol.* 15, 581–590
50 Okamura, K. *et al.* (2008) The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 453, 803–806
51 Watanabe, T. *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453, 539–543
52 Berezikov, E. *et al.* (2007) Mammalian mirtron genes. *Mol. Cell* 28, 328–336
53 Ruby, J.G. *et al.* (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature* 448, 83–86
54 Borsani, O. *et al.* (2005) Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 123, 1279–1291
55 Czech, B. *et al.* (2008) An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453, 798–802
56 Gullerova, M. *et al.* (2011) Autoregulation of convergent RNAi genes in fission yeast. *Genes Dev.* 25, 556–568
57 Wei, W. *et al.* (2011) Functional consequences of bidirectional promoters. *Trends Genet.* 27, 267–276
58 Hartig, J.V. *et al.* (2009) Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *EMBO J.* 28, 2932–2944
59 Kim, V.N. *et al.* (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* 10, 126–139
60 Okamura, K. *et al.* (2007) The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* 130, 89–100
61 Flynt, A.S. *et al.* (2010) MicroRNA biogenesis via splicing and exosome-mediated trimming in *Drosophila*. *Mol. Cell* 38, 900–907
62 Mercer, T.R. *et al.* (2010) Expression of distinct RNAs from 3′ untranslated regions. *Nucleic Acids Res.* 39, 2393–2403
63 Jöchl, C. *et al.* (2008) Small ncRNA transcriptome analysis from *Aspergillus fumigatus* suggests a novel mechanism for regulation of protein synthesis. *Nucleic Acids Res.* 36, 2677–2689
64 Fu, H. *et al.* (2009) Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett.* 583, 437–442
65 Yamasaki, S. *et al.* (2009) Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J. Cell Biol.* 185, 35–42
66 Zhang, S. *et al.* (2009) The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation. *Plant Physiol.* 150, 378–387
67 Thompson, D.M. and Parker, R. (2009) The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in *Saccharomyces cerevisiae*. *J. Cell Biol.* 185, 43–50
68 Phizicky, E.M. and Hopper, A.K. (2010) tRNA biology charges to the front. *Genes Dev.* 24, 1832–1860
69 Thompson, D.M. and Parker, R. (2009) Stressing out over tRNA cleavage. *Cell* 138, 215–219
70 Schutz, K. *et al.* (2010) Capture and sequence analysis of RNAs with terminal 2′,3′-cyclic phosphates. *RNA* 16, 621–631
71 Schaefer, M. *et al.* (2010) RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* 24, 1590–1595
72 Anderson, P. and Kedersha, N. (2009) RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nat. Rev. Mol. Cell Biol.* 10, 430–436
73 Emara, M.M. *et al.* (2010) Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *J. Biol. Chem.* 285, 10959–10968
74 Karginov, F.V. *et al.* (2010) Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon MicroRNAs, Drosha, and additional nucleases. *Mol. Cell* 38, 781–788
75 Shin, C. *et al.* (2010) Expanding the MicroRNA targeting code: functional sites with centered pairing. *Mol. Cell* 38, 789–802
76 Otsuka, Y. *et al.* (2009) Identification of a cytoplasmic complex that adds a cap onto 5′-monophosphate RNA. *Mol. Cell. Biol.* 29, 2155–2167
77 Elbarbary, R.A. *et al.* (2009) Modulation of gene expression by human cytosolic tRNase ZL through 5′-half–tRNA. *PLoS ONE* 4, e5908
78 Lundblad, E.W. and Altman, S. (2010) Inhibition of gene expression by RNase P. *New Biotechnol.* 27, 212–221
79 Tan, G.S. *et al.* (2009) Expanded RNA-binding activities of mammalian Argonaute 2. *Nucleic Acids Res.* 37, 7533–7545
80 Halic, M. and Moazed, D. (2010) Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. *Cell* 140, 504–516
81 Li, Z. *et al.* (2009) Characterization of viral and human RNAs smaller than canonical microRNAs. *J. Virol.* 83, 12751–12758
82 Leung, A.K.L. *et al.* (2006) Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18125–18130

83 Couvillion, M.T. *et al.* (2010) A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo. *Genes Dev.* 24, 2742–2747
84 Leung, A.K.L. *et al.* (2011) Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat. Struct. Mol. Biol.* 18, 237–244
85 Kaplan, N. *et al.* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458, 362–366
86 Light, W.H. *et al.* (2010) Interaction of a DNA zip code with the nuclear pore complex promotes H2A.Z. incorporation and INO1 transcriptional memory. *Mol. Cell* 40, 112–125
87 Raab, J.R. and Kamakaka, R.T. (2010) Insulators and promoters: closer than we think. *Nat. Rev. Genet.* 11, 439–446
88 Farnham, P.J. (2009) Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* 10, 605–616
89 Ganapathi, M. *et al.* (2011) Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. *Nucleic Acids Res.* 39, 2032–2044
90 Chen, M. and Manley, J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754
91 Wulff, B-E. *et al.* (2011) Elucidating the inosinome: global approaches to adenosine-to-inosine RNA editing. *Nat. Rev. Genet.* 12, 81–85
92 Kertesz, M. *et al.* (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107
93 Ingolia, N.T. *et al.* (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223
94 Lee, Y. *et al.* (2010) Translationally optimal codons associate with aggregation-prone sites in proteins. *Proteomics* 10, 4163–4171
95 Tuller, T. *et al.* (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354
96 Lee, J.E. *et al.* (2010) Systematic analysis of *cis*-elements in unstable mRNAs demonstrates that CUGBP1 is a key regulator of mRNA decay in muscle cells. *PLoS ONE* 5, e11201
97 Gong, C. and Maquat, L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* 470, 284–288
98 Czech, B. and Hannon, G.J. (2011) Small RNA sorting: matchmaking for Argonautes. *Nat. Rev. Genet.* 12, 19–31
99 Wlotzka, W. *et al.* (2011) The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J.* 30, 1790–1803
100 Tomecki, R. and Dziembowski, A. (2010) Novel endoribonucleases as central players in various pathways of eukaryotic RNA metabolism. *RNA* 16, 1692–1724
101 Li, W.M. *et al.* (2009) Endoribonucleases – enzymes gaining spotlight in mRNA metabolism. *FEBS J.* 277, 627–641

An RNA Reset Button

Alex C. Tuck[1] and David Tollervey[1,*]
[1]Wellcome Trust Centre for Cell Biology, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh, EH9 3JR, Scotland
*Correspondence: d.tollervey@ed.ac.uk
DOI 10.1016/j.molcel.2012.02.001

In this issue of *Molecular Cell*, single-cell analyses by Bumgarner et al. (2012) reveal how two antagonistic long noncoding RNAs at the *FLO11* locus define a toggle responsible for morphological heterogeneity in genetically identical populations of budding yeast.

Many microbes form multicellular assemblies, such as biofilms, which are medically and industrially significant. These structures provide protection from environmental insults and facilitate the extraction of nutrients via invasion of the substrate. Within such communities, genetically identical cells often display phenotypic heterogeneity at the individual level, which may enhance their ability to anticipate environmental fluctuations. In the budding yeast *Saccharomyces cerevisiae,* nutrient limitation or stress induces some cells to form adhesive filaments that forage locally for nutrients, while others wash away to more distal locations. This morphological variation reflects variegated expression of flocculin genes such as *FLO11*, which encodes a cell wall glycoprotein conferring adhesion (reviewed in Brückner and Mösch, 2012). Induction of *FLO11* by environmental stimuli proceeds via classical signaling cascades, whereas cell-to-cell variation involves a metastable epigenetic toggle; cells slowly (less than once per cell division) and stochastically switch between two heritable states of *FLO11* (competent for induction versus silent) (Halme et al., 2004). In this issue, Bumgarner and colleagues analyze this switch at single-cell resolution, revealing the critical role played by two *cis*-interfering long noncoding RNAs (lncRNAs).

Previous microarray data revealed reciprocally expressed noncoding transcripts spanning the *FLO11* promoter (Bumgarner et al., 2009). *ICR1* (interfering Crick RNA 1) transcribes through the *FLO11* and *PWR1* promoters and inhibits their expression, whereas *PWR1* (promoting Watson RNA 1) transcribes through the *ICR1* promoter and is expressed when *FLO11* is active (Bumgarner et al., 2009). Computational modeling suggested that the *FLO11* promoter occupies three transcriptional states: silent, basal, or active (Octavio et al., 2009). The current study (Bumgarner et al., 2012) employed fluorescent in situ hybridization (FISH) microscopy to count *FLO11*, *ICR1*, and *PWR1* transcripts in single cells, detecting the three predicted promoter states as cells with 0, 1–5, or >5 *FLO11* transcripts. Combined with mutational analyses, ChIP determination of protein binding and modeling this provided a clearer understanding of how protein and lncRNA regulators elicit redistribution between the states (Figure 1). In the basal state, the *FLO11* promoter is initially unresponsive to activators, and minimal transcription occurs. The *FLO11* activator (Flo8) and repressor (Sfl1) compete for binding, culminating in either stable silencing, mediated by Sfl1 together with the histone deacetylase (HDAC) Hda1, or Flo8-dependent competency for induction. In the latter case, Flo8 binding promotes *PWR1* transcription, suppresses transcription of *ICR1,* and renders the *FLO11* promoter competent—i.e., responsive to conventional activators and repressors. This allows rapid transcriptional regulation to establish an active state with high *FLO11* expression. However, infrequent Sfl1-promoted *ICR1* transcription ejects regulators from the promoter and "reboots" it to the unresponsive state, allowing competition between Sfl1 and Flo8 to determine its fate anew. Unusually, the HDAC complex Rpd3L activates *FLO11* transcription. This occurs because Rpd3L facilitates Flo8 binding at the expense of Sfl1 and acts together with *PWR1* transcription in blocking *ICR1* expression.

A key feature of the differentiation system described here is the intersection of fast and slow responses. All competent cells rapidly respond to conventional transcription factors that signal nutrient availability. However, the slow "epigenetic" switch to and from the silent state generates a population of cells that behave differently, and populations exchange only on a timescale comparable to the generation time. The progress made by Bumgarner et al. (2012) in understanding the complex protein and lncRNA interactions within this switch stems from their use of single-cell techniques. These allowed direct analysis of variants within a clonal population, rather than the averaged view provided by population data. However, some questions remain, such as causality within the toggle. Without kinetic data or time-resolved analyses, at this stage it is difficult to know what acts first—the proteins or lncRNAs.

Both the process of noncoding transcription and the resulting lncRNA products perform many functions (Wang and Chang, 2011). This and the centrality of lncRNAs in the *FLO11* toggle suggest that they may offer advantages over regulatory proteins. The generation of heterogeneity via lncRNA production is potentially more precise, robust, and economical than regulation via dedicated proteins. Transcription of lncRNAs is a simple and direct means to remove transcription factors, effectively resurfacing the chromatin over the promoter (Martens et al., 2004). The digital nature of transcription (a transcript is either made or not) produces clean binary decisions in response to stochastic signals, making lncRNAs particularly suitable for toggles. This clarity may underlie their proposed roles in cell-cycle regulation (Lardenois et al., 2011). Phenotypic switching is advantageous only at the appropriate, context-dependent rate (Acar et al.,

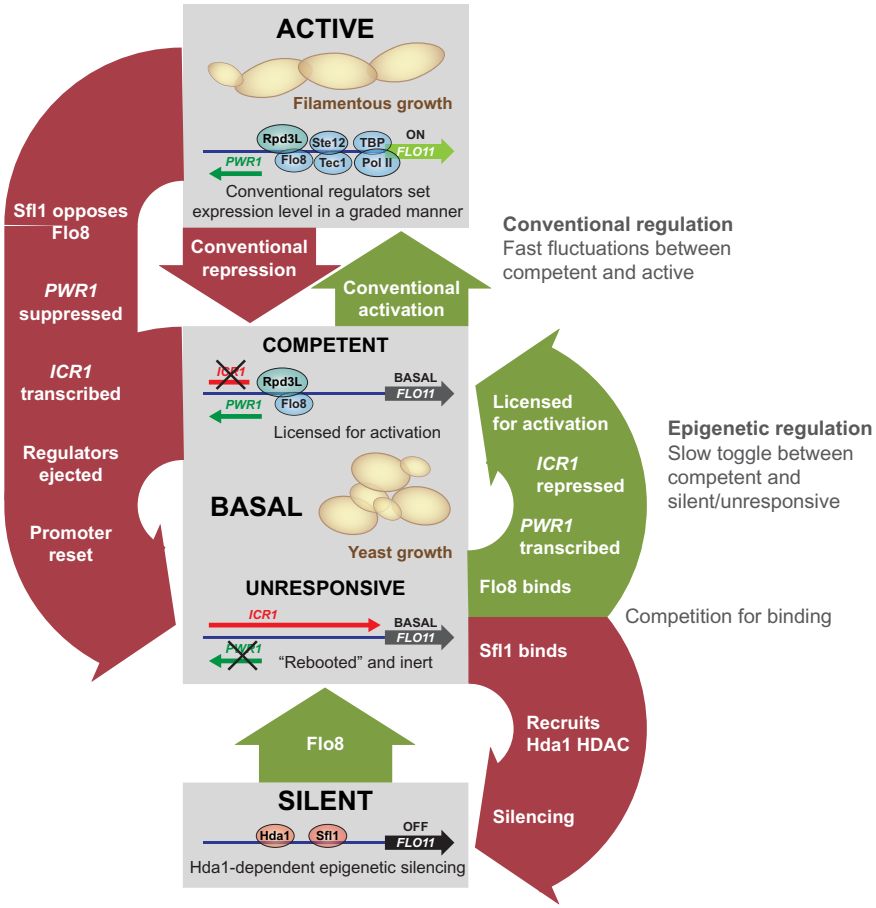


Figure 1. Key Transitions and Factors Involved in Modulating the Activation State of the *FLO11* Promoter

From the basal state, the promoter can either be silenced by Sfl1 binding together with the HDAC Hda1 or rendered activation competent by Flo8 binding and *PWR1* expression. Conventional activators such as Ste12 and Tec1 and repressors such as Nrg1/2 then drive cells in the competent state to and from full activity. Transcription of *PWR1* through the *ICR1* promoter represses its expression; however, rare Sfl1-promoted *ICR1* transcription events can eject promoter-bound factors, resetting the system to its basal, unresponsive state. The histone deacetylase complex Rpd3L favors *FLO11* expression, by facilitating Flo8 binding and suppressing *ICR1* transcription and Sfl1 recruitment.

2008), and lncRNA transcription frequency can readily be tuned via transcription start sites that are distinct from target genes. Furthermore, since they act in three dimensions, the effects of chromatin-bound proteins can become promiscuous and dilute with distance. In contrast, *cis*-acting lncRNAs are intrinsically gene specific and can potentially transmit a signal over long distances with little attenuation (Xu et al., 2011). This may be particularly useful within extended promoters such as that of *FLO11*, which binds upwards of 20 regulatory proteins over a 3.4 kb region. Thus, lncRNA transcription provides toggles, information transmission, and the occasional reboot for molecular circuit boards.

More generally, the study from Bumgarner et al. (2012) exemplifies clonal heterogeneity. Under many circumstances it is advantageous for microorganisms to develop pathways conferring phenotypic variegation upon clonal populations. This is particularly true for microorganisms inhabiting complex natural environments, which can be subject to rapid changes. Heterogeneity both increases the chance that some of the population will survive sudden adverse changes and potentially allows specialization to optimize the use of complex but limited resources. We speculate that lncRNA-dependent variegated expression may be widespread. Genes involved in metabolism, morphology, signaling, and stress responses are likely candidates, as they respond to environmental changes and are commonly associated with lncRNAs (Yassour et al., 2010). Consistent with this idea, lncRNAs are preferentially associated with genes exhibiting greater cell-to-cell expression variability (Xu et al., 2011). Future developments in single-cell analyses, such as single-cell transcriptome sequencing, should reveal the full extent of lncRNA-dependent heterogeneity.

REFERENCES

Acar, M., Mettetal, J.T., and van Oudenaarden, A. (2008). Nat. Genet. *40*, 471–475.

Brückner, S., and Mösch, H.-U. (2012). FEMS Microbiol. Rev. *36*, 25–58.

Bumgarner, S.L., Dowell, R.D., Grisafi, P., Gifford, D.K., and Fink, G.R. (2009). Proc. Natl. Acad. Sci. USA *106*, 18321–18326.

Bumgarner, S.L., Neuert, G., Voight, F.R., Symbor-Nagrabska, A., Grisafi, P., van Oudenaarden, A., and Fink, G.R. (2012). Mol. Cell *45*, this issue, 470–482.

Halme, A., Bumgarner, S., Styles, C., and Fink, G.R. (2004). Cell *116*, 405–415.

Lardenois, A., Liu, Y., Walther, T., Chalmel, F., Evrard, B., Granovskaia, M., Chu, A., Davis, R.W., Steinmetz, L.M., and Primig, M. (2011). Proc. Natl. Acad. Sci. USA *108*, 1058–1063.

Martens, J.A., Laprade, L., and Winston, F. (2004). Nature *429*, 571–574.

Octavio, L.M., Gedeon, K., and Maheshri, N. (2009). PLoS Genet. *5*, e1000673.

Wang, K.C., and Chang, H.Y. (2011). Mol. Cell *43*, 904–914.

Xu, Z., Wei, W., Gagneur, J., Clauder-Munster, S., Smolik, M., Huber, W., and Steinmetz, L.M. (2011). Mol. Syst. Biol. *7*, 468.

Yassour, M., Pfiffner, J., Levin, J., Adiconis, X., Gnirke, A., Nusbaum, C., Thompson, D.A., Friedman, N., and Regev, A. (2010). Genome Biol. *11*, R87.

Please cite this article in press as: Schneider et al., Transcriptome-wide Analysis of Exosome Targets, Molecular Cell (2012), http://dx.doi.org/10.1016/j.molcel.2012.08.013

Molecular Cell
Article

Transcriptome-wide Analysis of Exosome Targets

Claudia Schneider,[1,2,*] Grzegorz Kudla,[1,3] Wiebke Wlotzka,[1] Alex Tuck,[1] and David Tollervey[1,*]
[1]Wellcome Trust Centre for Cell Biology, The University of Edinburgh, Edinburgh UK
[2]Present address: Institute for Cell and Molecular Biosciences (ICaMB), Newcastle University, Newcastle upon Tyne, UK
[3]Present address: MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh, UK
*Correspondence: claudia.schneider@ncl.ac.uk (C.S.), d.tollervey@ed.ac.uk (D.T.)
http://dx.doi.org/10.1016/j.molcel.2012.08.013

SUMMARY

The exosome plays major roles in RNA processing and surveillance but the in vivo target range and substrate acquisition mechanisms remain unclear. Here we apply in vivo RNA crosslinking (CRAC) to the nucleases (Rrp44, Rrp6), two structural subunits (Rrp41, Csl4) and a cofactor (Trf4) of the yeast exosome. Analysis of wild-type Rrp44 and catalytic mutants showed that both the CUT and SUT classes of non-coding RNA, snoRNAs and, most prominently, pre-tRNAs and other Pol III transcripts are targeted for oligoadenylation and exosome degradation. Unspliced pre-mRNAs were also identified as targets for Rrp44 and Rrp6. CRAC performed using cleavable proteins (split-CRAC) revealed that Rrp44 endonuclease and exonuclease activities cooperate on most substrates. Mapping oligoadenylated reads suggests that the endonuclease activity may release stalled exosome substrates. Rrp6 was preferentially associated with structured targets, which frequently did not associate with the core exosome indicating that substrates follow multiple pathways to the nucleases.

INTRODUCTION

Gene expression generates an enormous variety of stable or unstable, protein-coding or non-coding RNA species produced by all three RNA polymerases. RNA abundance and integrity are closely monitored by nuclear and cytoplasmic surveillance systems (reviewed in (Houseley and Tollervey, 2009)). A key player in RNA metabolism is the exosome, which participates in 3′ end maturation and/or quality control of almost every RNA molecule in the cell. In *Saccharomyces cerevisiae*, nuclear and cytoplasmic forms of the exosome share the RNase II homolog Rrp44/Dis3, which contains two distinct catalytic sites. The RNB domain exhibits 3′-5′ exonuclease activity, whereas the N-terminal PINc domain plays a dual role in harboring endonuclease activity and tethering Rrp44 to the core, nine subunit exosome (Lebreton et al., 2008; Lorentzen et al., 2008; Schaeffer et al., 2009; Schneider et al., 2009). In addition to Rrp44, the nuclear form of the yeast exosome is associated with a second active 3′-5′ exonuclease, Rrp6 (Briggs et al., 1998).

Structural studies have shown that the nine catalytically inert subunits of the core exosome form a two-layered barrel-like structure (Liu et al., 2006). The upper layer is composed of a "cap" of three S1 or KH domain RNA binding proteins (Csl4, Rrp4, Rrp40), which rests on a ring of six proteins with homology to RNase PH (Rrp41, Rrp45, Rrp43, Rrp46, Rrp42 and Mtr3). Rrp44 is located at the base of the core exosome barrel, and in vitro data show that substrates can be threaded through the lumen of the exosome barrel to reach the exonuclease site in Rrp44 (Bonneau et al., 2009; Malet et al., 2010). However, it is not known what fraction of natural substrates follow this path. Rrp6 has distinct targets (Callahan and Butler, 2008) and associates with the exterior of the exosome complex.

Vital functions of the exosome include the processing of ribosomal RNA (rRNA), small nuclear and nucleolar RNAs (sn(o)RNAs) in the nucleus, mRNA turnover in the cytoplasm and surveillance of aberrant RNAs throughout the cell (reviewed in (Houseley and Tollervey, 2009)). It also plays key roles in the regulated degradation of pervasive transcripts that are generated all over the yeast genome. These include cryptic unstable transcripts (CUTs), which were originally identified in strains lacking Rrp6, stable un-annotated transcripts (SUTs) and many short, promoter-associated RNAs (PARs) (Davis and Ares, 2006; Neil et al., 2009; Wyers et al., 2005; Xu et al., 2009). Distinct classes of RNA substrates are likely assigned to individual nuclease activities in the exosome, but substrate specificities and targeting mechanisms for this process are largely unclear. Microarray analyses have been applied to distinguish substrate specificities of Rrp6, Rrp44/Dis3, and core exosome subunits in *Drosophila*, but this was limited to mRNAs (Kiss and Andrulis, 2010).

Most functions of the exosome are dependent on cofactors, including the Trf-Air-Mtr4 polyadenylation (TRAMP) complex and the Nrd1/Nab3 heterodimer, but direct interactions between individual exosome subunits and some specific targets have been reported (Kadaba et al., 2006; Schneider et al., 2007). Transcriptome-wide maps of RNA substrates of the TRAMP-associated poly(A) polymerase Trf4 or the Nrd1/Nab3 heterodimer have been published based on UV crosslinking (Jamonnak et al., 2011; Wlotzka et al., 2011) or RNA coprecipitation (Hogan et al., 2008; San Paolo et al., 2009). These analyses identified many surveillance targets, including a notable number of RNA polymerase III transcripts.

Here we report a transcriptome-wide map of exosome substrates and their interactions with individual exosome subunits in living cells.

Please cite this article in press as: Schneider et al., Transcriptome-wide Analysis of Exosome Targets, Molecular Cell (2012), http://dx.doi.org/10.1016/j.molcel.2012.08.013

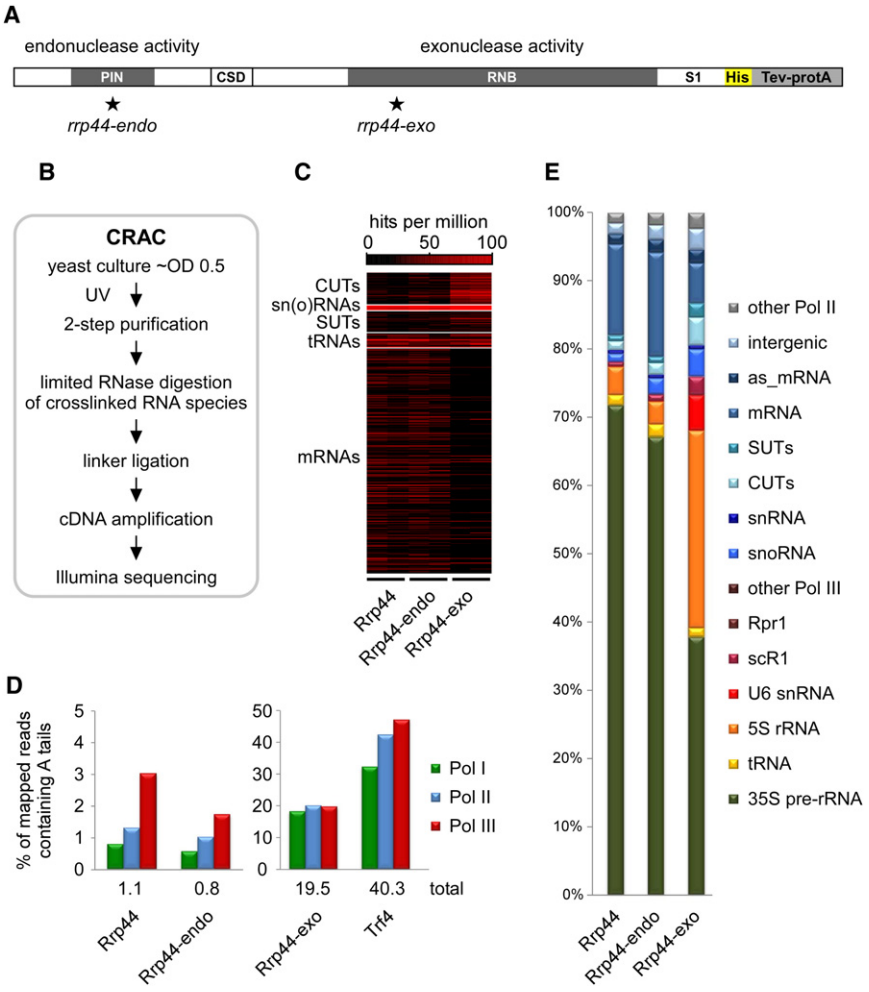


Figure 1. Comparison of Targets of Wild-Type and Mutant Rrp44

(A) Domain structure of *S. cerevisiae* Rrp44, including a C-terminal His-TEV protease-protein A (HTP) tag for purification. Point mutations inactivating the endonuclease (*rrp44-endo*) or exonuclease (*rrp44-exo*) activity of Rrp44 are indicated.

(B) Outline of the CRAC crosslinking technique.

(C–E) Illumina high-throughput sequencing of cDNA libraries generated from crosslinked RNAs recovered with purified wild-type Rrp44 and the Rrp44-endo and Rrp44-exo mutants, as well as the exosome cofactor Trf4. Here, and in all other illustrations, sequencing data of individual biological replicate experiments was mapped to the yeast genome using Novoalign and normalized to hits per million mapped sequences (hpm).

(C) Heat maps for main substrate groups. Numbers of reads mapped to individual RNAs are shown in shades of red.

(D) Frequencies of non-templated terminal oligo(A) sequence reads in data sets for wild-type Rrp44 and catalytic mutants, and the exosome cofactor Trf4. Data sets are filtered either for total reads, or for Pol I, Pol II and Pol III transcripts, that contain 2 or more non-templated As.

(E) Transcriptome-wide binding profiles. Bar diagrams illustrate the percentage of all sequences mapped to the functional RNA classes indicated on the right of the figure.

RESULTS

Comparison of Targets for Wild-Type and Mutant Forms of Rrp44

To identify targets for the core exosome nuclease Rrp44, we applied in vivo RNA-protein crosslinking (CRAC) (Granneman et al., 2009) to the wild-type protein, or *rrp44* mutants carrying point mutations in catalytic residues of the RNB exonuclease domain (*rrp44-exo* mutant, $D_{551}N$) or PIN endonuclease domain (*rrp44-endo* mutant, $D_{91}N$, $E_{120}Q$, $D_{171}N$, $D_{198}N$) (Figure 1A). HTP-tagged forms of Rrp44 were expressed from a plasmid in yeast strains derived from BY4741, in which the genomic *RRP44* ORF was precisely deleted. Growth rates and RNA processing phenotypes of strains expressing either wild-type or mutant Rrp44 were as previously reported (Schneider et al., 2009). Cells actively growing in minimal SD medium were UV-irradiated as described (Granneman et al., 2011) and RNA fragments crosslinked to Rrp44 were identified by the CRAC technique as outlined in Figure 1B. At least two independent experiments were performed in each case and analyzed separately. The primary sequence data have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE40046. Mapped reads are presented in Table S3.

Transcriptome-wide binding profiles of Rrp44 are shown in Figures 1C–1E. Wild-type and mutant forms of Rrp44 were predominately associated with classes of RNA corresponding to known exosome targets. Analysis of individual, functionally grouped RNAs (Figures 1C, S1, and Table S3) revealed similar patterns for wild-type Rrp44 and Rrp44-endo data sets. The *rrp44-endo* mutation does therefore not appear to significantly alter or interfere with Rrp44 substrate binding. In contrast, the Rrp44-exo data set was significantly enriched for sequences derived from CUTs, SUTs, snRNAs, snoRNAs and, most prominently, a subset of Pol III RNAs (5S rRNA, U6 snRNA, scR1), whereas recovery of mRNAs and the 35S pre-rRNA was relatively reduced.

The initial identification of CUTs in strains lacking only Rrp6 (Davis and Ares, 2006; Wyers et al., 2005) had suggested that Rrp6 was the major nuclease responsible for their degradation. However, the enrichment for CUTs in Rrp44-exo data sets strongly indicates that CUTs are also targeted for degradation by Rrp44.

The presence of non-templated, 3′ terminal oligo(A) tails is a characteristic of nuclear RNA surveillance targets (reviewed by (Houseley and Tollervey, 2009)). The Trf4-HTP data set generated here from actively growing cells contained a high fraction (40.3%) of reads with $\geq$ 2 non-templated adenosines

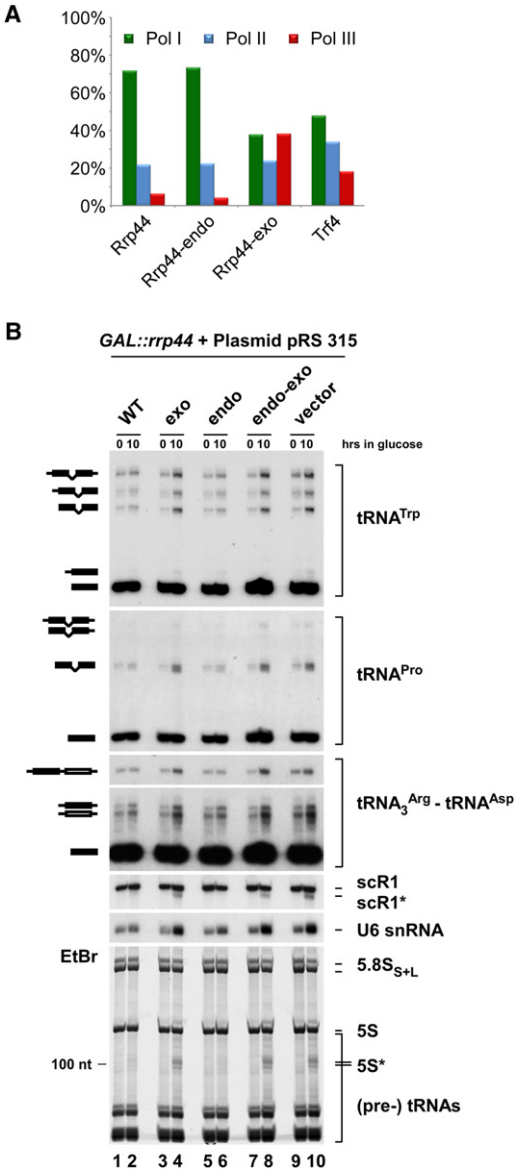


Figure 2. The Exosome and Cofactors Target RNA Pol III Transcripts

(A) Proportion of reads mapped to products of RNA Polymerases I, II and III recovered with wild-type Rrp44 and catalytic mutants, and with the exosome cofactor Trf4.

(B) Northern analyses showing pre-tRNA and other Pol III RNA accumulation in strains expressing Rrp44 mutants. A *GAL::rrp44* strain was transformed with plasmids expressing either wild-type or mutant Rrp44 protein, or an empty vector pRS315. The mutants analyzed are Rrp44-exo, Rrp44-endo and Rrp44-endo-exo (see Figure 1A). RNA was isolated from strains grown at 30°C under permissive conditions (0) and 10 hr after transcriptional repression (10). RNA was separated on an 8% polyacrylamide/ 8M urea gel and either detected by northern hybridization with oligonucleotide probes (Table S1) or by staining with EtBr. A schematic representation of the identified species is shown.

at the 3′ end (Figure 1D). In contrast, few oligoadenylated reads were recovered in wild-type Rrp44 (1.1%) or Rrp44-endo (0.8%) data sets, and such reads were predominately derived from Pol III transcripts (Figure 1D). However, for the Rrp44-exo mutant 19.5% of mapped sequences derived from all three polymerases carried an oligo(A) tail, indicating that Rrp44-exo becomes trapped on degradation intermediates of the targets of nuclear RNA surveillance. To characterize RNA targets associated with wild-type and mutant forms of Rrp44, we initially compared the distribution of mapped sequences among different substrate classes (Figure 1E). All three data sets contain a large percentage of sequences mapped to the Pol I transcribed 35S pre-rRNA, reflecting the prominent roles of Rrp44 and the exosome in ribosome biogenesis and pre-rRNA surveillance. Both stable and unstable non-coding RNAs transcribed by RNA polymerases II and III, as well as a large pool of (pre-)mRNAs, were also crosslinked to all Rrp44 variants.

A striking feature of the Rrp44-exo data set was the abundant recovery of Pol III RNAs (Figures 2A and S2A). While such transcripts represent only ~5% of all RNAs recovered with wild-type Rrp44 or Rrp44-endo, almost 40% of all RNAs crosslinked to Rrp44-exo are transcribed by Pol III. RNAs transcribed by Pol III also comprised a substantial proportion (18.2%) of the Trf4 data set (Figures 2A and S2A), consistent with crosslinking (San Paolo et al., 2009; Wlotzka et al., 2011) and experimental data implicating Trf4 in the surveillance of 5S rRNA, U6 snRNA and pre-tRNAs (Copela et al., 2008; Kadaba et al., 2004; Kadaba et al., 2006; Schneider et al., 2007; Schneider et al., 2009; Vanácová et al., 2005; Wlotzka et al., 2011).

The prominent association of Pol III targets with Rrp44-exo and nuclear exosome cofactors indicates that these are major targets for the nuclear RNA surveillance machinery. The reduced recovery of Pol III transcripts with Rrp44 and Rrp44-endo suggests that they are substrates for the Rrp44 exonuclease activity, but are normally degraded efficiently. The high proportion of oligoadenylated sequences derived from Pol III RNAs in the Rrp44-exo data sets shows that the crosslinked RNAs had already been targeted and marked for surveillance (shown for the U6 snRNA, 5S rRNA and $tRNA^{Pro}$ in Figures S2B–S2D). Persistent binding of these RNAs to the exosome in the absence of Rrp44 exonuclease activity may contribute to the impaired growth and strong RNA processing phenotypes in *rrp44-exo* strains.

To further assess the role of the distinct catalytic activities of Rrp44 in Pol III RNA surveillance in vivo, levels of pre-tRNAs and other Pol III transcripts were assessed in Rrp44 mutant strains (Figure 2B). For this, the endogenous Rrp44 was expressed as HA-fusion under the control of a repressible P_{GAL10} promoter. The strain was then transformed with plasmids expressing Rrp44, Rrp44-endo, Rrp44-exo or Rrp44-endo-exo with a C-terminal Protein A tag and under the control of the P_{RRP44} promoter (Schneider et al., 2009), or with the empty cloning vector (pRS315). Changes in Pol III RNA levels were observed in strains expressing Rrp44-exo, whereas the Rrp44-endo mutation alone had no effect. Rrp44-exo phenotypes included pre-tRNA accumulation and the appearance of 3′ truncated fragments (5S*, scR1*). Higher levels of mature U6 snRNA were seen and 3′ extended (~3nt) U6 was observed in the sequence data (data not shown), consistent with a gel mobility shift (Figure 2B). We conclude that the exonuclease activity of

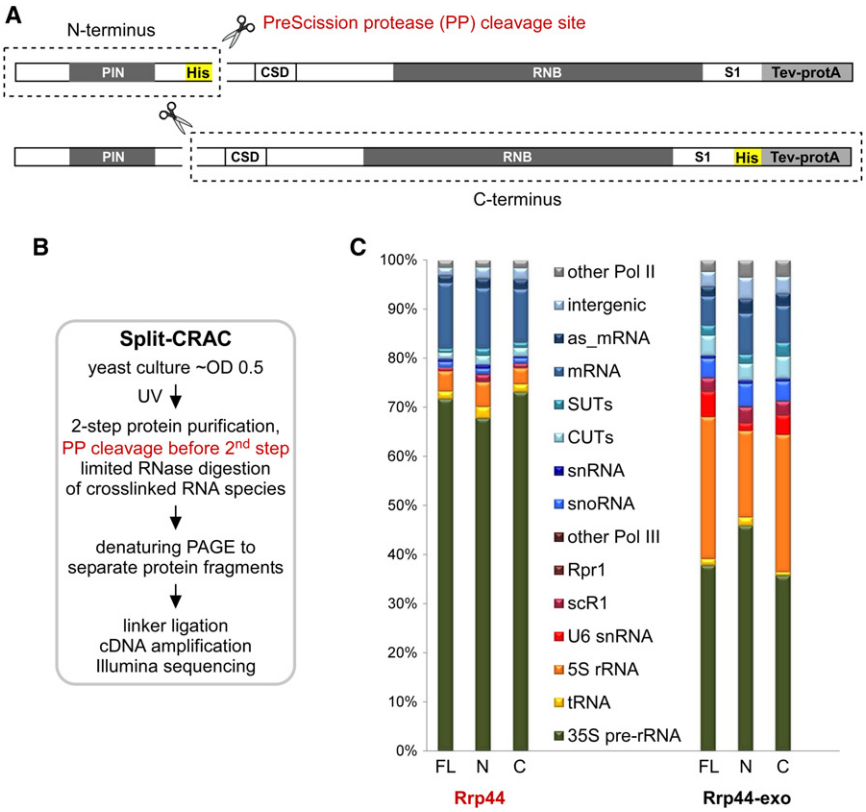


Figure 3. Split-CRAC Allows the Targets of the N-Terminal and C-Terminal Regions of Rrp44 To Be Distinguished

(A) Cleavable Rrp44-expression constructs used for split-CRAC. The location of the PreScission protease (PP) cleavage site, which allows the separation of N- and C-terminal domains (NTD and CTD) in vitro, and the purification tags are indicated.

(B) Outline of the split-CRAC crosslinking technique.

(C) Distribution of reads recovered with full-length and cleaved wild-type Rrp44 (left) and the Rrp44-exo mutant (right). Sequencing data were analyzed as in Figure 1.

Rrp44 plays multiple roles in the surveillance and/or maturation of Pol III transcribed RNAs.

Split-CRAC Separates Targets for the Exonuclease and Endonuclease Domains of Rrp44

Many proteins that function in RNA metabolism contain more than one RNA interacting domain, but determining their relative contributions in vivo can be experimentally challenging. In the case of Rrp44 we wanted to assess the possibility that the PIN and RNB domains might specifically and independently contribute to RNA target recognition in vivo. To identify RNAs preferentially associated with each of the two domains, we developed a modified CRAC protocol, termed split-CRAC (Figures 3A and 3B). In this, the intact protein is crosslinked in vivo in actively growing cells, followed by in vitro cleavage during protein purification.

Rrp44 expression plasmids were constructed in which a PreScission protease (PP) cleavage site was inserted between aa 241 and aa 242 of the *RRP44* or *rrp44-exo* ORF. This insertion site was chosen because structural studies on Rrp44 had previously shown that a C-terminal fragment truncated at aa 242 was stable (Lorentzen et al., 2008), indicating that the protein domain structure was unlikely to be perturbed by the short insert. The constructs also carry a His_6 tag, located on either the N-terminal or C-terminal side of the cleavage site (Figure 3A). Cleavage of crosslinked Rrp44-RNA complexes during purification was shown to allow selective recovery of either the N-terminal domain (NTD) or C-terminal domain (CTD) dependent on the location of the His_6 tag (Figure S3).

The transcriptome-wide interaction profiles were strikingly similar for the full-length (FL) Rrp44 protein and for both the NTD and CTD fragments (Figure 3C). Using the Rrp44-exo mutant in split-CRAC, the analyses also returned very similar overall patterns of hits for the full-length protein compared to either fragment. The only exceptions were decreased recovery of the U6 snRNA and the 5S rRNA with the NTD and mildly decreased pre-tRNAs with the CTD. However, more detailed analyses of the hit distribution across individual target RNAs with high sequence coverage revealed differences in binding profiles. This is illustrated for the U6 snRNA (Figure 4A) and the pre-rRNA 5′-External Transcribed Spacer (5′-ETS) region (Figure 4B).

In CRAC and related techniques, microdeletions are often introduced at the site of crosslinking during reverse transcription and can be used to precisely map protein binding sites (Wlotzka et al., 2011; Zhang and Darnell, 2011). To distinguish the relative positions of the NTD and CTD on target RNAs at higher resolution, we compared the mapped reads (hits) of Rrp44 and Rrp44-exo split-CRAC data sets with the positions of microdeletions (dels) (Figure 4A–4C). These analyses were performed using the complete data set (Total reads) and also following filtering for sequences that contain oligo(A) tails (A-tailed reads), to identify RNAs that the TRAMP complexes have marked for degradation.

In the U6 snRNA, different binding profiles are seen for the NTD PINc and CTD exonuclease regions of Rrp44 (Figure 4C, left panel). The NTD mainly binds at the 5′ end of the RNA, whereas CTD hits are shifted toward the 3′ end. The same pattern is seen for total and A-tailed reads, indicating that the recovered fragments are largely derived from RNAs that were targeted for surveillance. The distribution of reads and deletions in the Rrp44-exo mutant generally matches this pattern but the higher recovery of U6 snRNA sequences with microdeletions allows better mapping of the crosslinking sites for the NTD around nt 45, for the CTD around nt 90, and for all constructs

around nt 69. Notably, the number of deletions in the A-tailed reads of the NTD increased relative to the CTD (Figure 4C). These transcripts must have been released from the exosome complex, and then reinserted, in order for oligo(A) addition by TRAMP to have occurred.

The highly structured 5′-A_0 fragment of the 5′-ETS is a well-characterized exosome substrate, which is very rapidly and "constitutively" degraded as part of the pre-rRNA processing pathway (Lebreton et al., 2008; Schaeffer et al., 2009; Schneider et al., 2009). In the 5′-ETS, both wild-type and mutant full-length Rrp44 and fragments bound in the region around nt 120, but differences in the distribution of hits and deletions were seen for the fragments along the whole RNA (Figure 4C, right panel). For Rrp44-exo, the differences in the distribution of reads were more pronounced and the coverage of NTD reads was increased relative to the (catalytically inactive) CTD. Coverage is expressed in hits/dels per million mapped reads, so this reflects a change in the relative distributions of the Rrp44-exo domains over all substrates. While absolute crosslinking efficiencies cannot be reliably inferred, the yield of crosslinked RNAs was reproducibly higher in the Rrp44-exo strain than in Rrp44 (see Figure S3), consistent with prolonged interactions leading to increased crosslinking efficiency.

The major peaks across the 5′-ETS represent structured regions, where exosome pausing may occur (Lebreton et al., 2008). We postulate that endonuclease cleavage acts to release stalled RNAs that are tightly bound by the exonuclease site of Rrp44 (Figure 4D). Oligoadenylation of the released substrate by Trf4 may allow the TRAMP-associated helicase Mtr4 to unwind the structured RNA, which can then be threaded back through the exosome channel.

More generally, split-CRAC can distinguish domain-specific interactions and should be widely applicable to resolve the targets of multi-domain RNA-binding proteins – which are common.

Comparison of RNA Targets for Core Exosome Subunits and Rrp6

Yeast Rrp44 is present in the exosome throughout the cell, whereas Rrp6 is only associated with the nuclear complex. To assess the relationship between Rrp6 and the core exosome, we constructed yeast strains expressing C-terminal tagged Rrp6-HTP, Rrp41-HTP or Csl4-HTP, each expressed from the endogenous genomic locus under the control of the endogenous promoter. Rrp41 forms part of the exosome barrel, which is composed of six RNase PH-like proteins, whereas the S1 RNA-binding domain protein Csl4 is one of the three "cap" proteins at the top of the barrel (Liu et al., 2006). All strains showed wild-type growth rates, indicating that the fusion proteins were functional. CRAC was performed as for Figure 1 and crosslinking patterns of core exosome subunits (Rrp44, Rrp41, Csl4) are compared to Rrp6 in Figure 5. Core exosome targets showed a high degree of overlap, with similar distributions of hits on most RNA classes, although some variation in recovery of pre-mRNA and mRNA was observed (Figures 5A, 5B, and S1). In contrast, the Rrp6 data sets were relatively enriched in small, structured RNAs including tRNAs, snRNAs and snoRNAs (Figures 5A and 5B). These were analyzed in more detail and representative results are presented for three examples; the U2 snRNA, the intron-containing pre-$tRNA^{Pro}_{UGG}$ and the box C/D snoRNA snR40 (Figure 5C). All three targets show higher coverage for Rrp6 than for Rrp44, and very low for Csl4 or Rrp41. In the case of snR40, the reads detected for Rrp6 are distributed over the body of the RNA, although some 3′ extended reads were observed in one experiment. We interpret this as indicating a major role for Rrp6 in surveillance and degradation, rather than in 3′ end processing of pre-snR40.

The crosslinking data also provided examples of functional overlap between Rrp6 and Rrp44, but not the remaining core exosome components, on structured RNAs. This is shown for the box C/D snoRNA snR13 (Figure S4). Rrp6 and Rrp44 are both required for the 3′ end processing of this RNA, while *Δrrp6* strains also accumulate snR13 read-through transcripts (Grzechnik and Kufel, 2008; Schneider et al., 2009). These aberrant RNAs are hardly visible in *rrp44-exo* strains, but strongly enriched in *Δrrp6 rrp44-exo* double mutants compared to the *Δrrp6* strain (Schneider et al., 2009). Interestingly, Rrp6 was mainly crosslinked to the highly structured mature snR13 RNA region, whereas sequences recovered with Rrp44 and, in particular, Rrp44-exo were often derived from downstream regions (Figure S4). As seen for other structured RNAs (Figure 5), few sequences were recovered for Rrp41 or Csl4 on snR13 (Supp. Data set). The in vivo analyses and crosslinking data both suggest that transcriptional read-though mainly occurs following defective 3′ end formation on snR13, with the resulting 3′ extended transcripts being targeted to Rrp44 for degradation.

The core exosome including Rrp44 plays major roles in surveillance and turnover of cytoplasmic mRNAs as well as surveillance of defective nuclear pre-mRNAs, whereas the activity of Rrp6 is expected to be nuclear-specific (reviewed in (Houseley and Tollervey, 2009)). Nuclear and cytoplasmic mRNAs cannot be distinguished in sequence, other than by the presence of introns specifically in nuclear pre-mRNAs. We therefore analyzed spliced mRNAs for relative read coverage over introns and exons. Analysis of reads mapped to intron-exon boundaries (IE) in unspliced pre-mRNA to exon-exon boundaries (EE) in spliced mRNA (Figure 6A; IE/EE) showed substantially higher binding to pre-mRNAs for Rrp6 than for intact Rrp44 or Rrp44-endo, whereas Rrp44-exo showed strongly enhanced pre-mRNA binding. Total binding over introns relative to all mRNAs was also notably higher for Rrp6 than for intact Rrp44 (Figure 6A; Introns/Total mRNA) or the other core exosome components (data not shown), again with strongly enhanced binding by Rrp44-exo. The differences for total introns was less marked than for the IE boundary, probably because Rrp44 and Rrp6 also degrade excised introns following debranching. Comparison of total binding at the 3′ and 5′ splice sites (Figure 6A; 3′SS/5′SS) showed preferential binding to the 5′SS for both Rrp6 and Rrp44, which was particularly marked for Rrp44-exo.

Comparison of individual spliced genes (Figure 6B), confirmed the preferential association of Rrp44 with 5′ regions including the 5′SS. The Rrp44-exo mutant was particularly strongly enriched there and reads frequently extended into the intron, giving rise to the high IE/EE ratio (Figure 6A). In contrast, Rrp6 showed the highest numbers of reads across introns, possibly reflecting a major role in degradation of both excised introns and

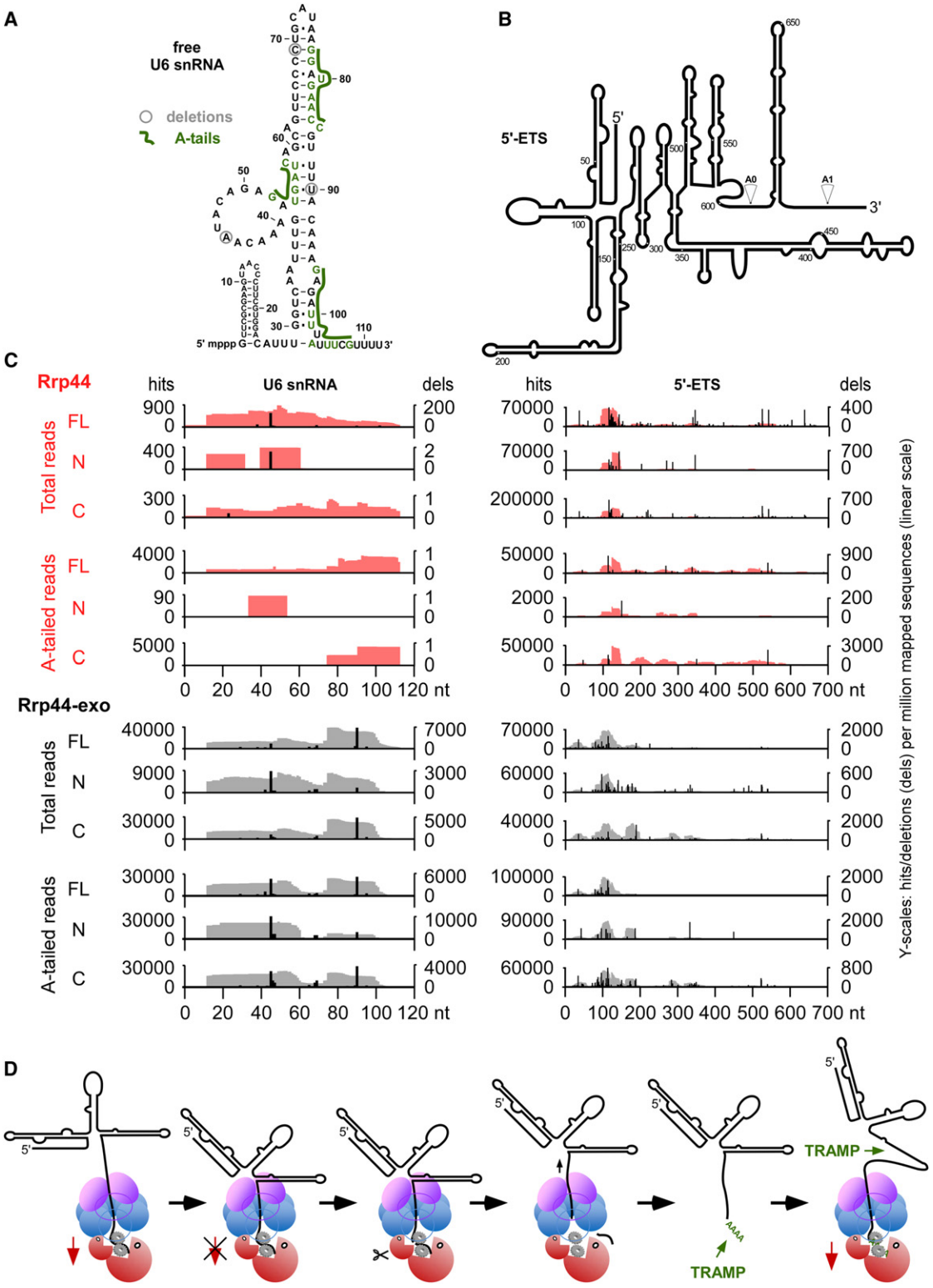


Figure 4. Comparison of Rrp44 Domains in Split-CRAC

(A) Secondary structure of U6 snRNA (112 nt) from *S. cerevisiae*. Major sites of microdeletions are circled. These are due to reverse transcriptase stops at the crosslinked nucleotide. Prominent sites of oligoadenylation are indicated in green and are located 3′ to the major crosslinking sites. The positions of the first non-templated adenosines in A-tailed reads are indicated in green.

pre-mRNAs. The coverage of Rrp44-exo over pre-mRNAs and mRNAs was lower (in hits per million) than for wild-type Rrp44, due to the enrichment of Pol III transcripts in the Rrp44-exo data set (Figures 1C and 6C). This may reflect a relative lack of strong secondary structure in the (pre-)mRNAs relative to highly structured Pol III RNAs that are potentially less readily degraded.

A notable feature of the alignments in Figure 6B was the bimodal distribution of intron lengths, coupled with notably higher sequence coverage on the genes with long introns compared to shorter introns in each of the data sets. This was particularly seen over exon 2 sequences, showing that it arises from targeting of the exosome to the pre-mRNAs or, conceivably, the spliced mRNAs. The long intron gene set was dominated by ribosomal proteins, which are more highly transcribed than most genes in the short intron set. These differences in expression levels are probably largely responsible for increased sequence coverage. However, regulated splicing of yeast ribosomal protein genes has been reported (Pleiss et al., 2007), which may be related to their targeting by the surveillance system.

Clustering of mRNAs by pattern of interactions with Rrp44, Rrp44-exo and Rrp6 revealed a class of transcripts preferentially bound by Rrp6. This subclass is enriched for intron-containing genes ($p < 1e\text{-}4$, chi-square test), as well as ribosomal protein genes ($p < 1e\text{-}10$) and ribosome synthesis factors ($p < 1e\text{-}6$) identified by the DAVID functional annotation tool (Dennis et al., 2003) (Figure 6C).

Together these results are consistent with core-independent functions and nuclear localization of Rrp6 and highlight a substantial role for the nuclear exosome in pre-mRNA surveillance and degradation.

Binding profiles over the entire 35S pre-ribosomal RNA also showed distinctions between Rrp6 and the core exosome (Figure 7A). Over the 5′-ETS, the core exosome components showed similar binding, whereas Rrp6 was clearly distinct. The Rrp44 and Rrp6 hits partially overlap in the ITS2 region, where these proteins function at different steps in the 3′ processing of 7S pre-rRNA (a 3′ extended form of 5.8S rRNA; highlighted in gray in Figure 7A) (Allmang et al., 1999). In contrast, no crosslinking was seen here for Rrp41 or Csl4.

Within the 18S rRNA region (Figure 7B), Rrp6 showed prominent peaks that were absent from the Rrp41 and Csl4 data sets, whereas some of the sites were recovered at lower levels with Rrp44. Previous analyses showed that RNA Pol I is prone to transcription pausing in this region leading to cotranscriptional cleavage of the nascent transcript, which is degraded by Rrp6 and the TRAMP polyadenylation complex (El Hage et al., 2010). Consistent with this, crosslinking to the Trf4 component of the TRAMP complex showed an overlapping peak in 18S (Figure 7B).

To assess the function of the exosome in the surveillance of RNA Pol III transcripts, hit distributions of Rrp44, Rrp6, Rrp41 and Csl4 were compared for each tRNA species (Figure 7C). Sequence reads extended beyond both ends of the mature tRNA (solid lines in Figure 7C) and included intronic sequences (dashed lines in Figure 7C) showing that pre-tRNAs are targets. This can also be seen for combined sets of all spliced and non-spliced tRNAs (Figure S5). Comparison of hit densities confirms the preferential association of pre-tRNAs with Rrp6 relative to the core exosome components (Figure S5). However, increased coverage on 5′ and 3′ extended regions seen for the Rrp44-exo mutant relative to Rrp44 (Figure 7C) demonstrates that many pre-tRNAs are targets for the exonuclease activity of Rrp44. Pre-tRNA recovery was comparable between Rrp44, Rrp41 and Csl4, indicating that these RNAs make direct contacts with the exosome core, consistent with threading through the lumen of the exosome. Oligoadenylated reads recovered with Rrp44-exo were distributed toward the 5′ end relative to total hits. These RNAs must represent truncation products that have been tailed with oligo(A) during degradation. The exosome is heavily dependent on cofactors, including the TRAMP complex. These can only bind the target RNA 5′ to the degrading exosome or Rrp6, since the 3′ end of the RNA is in the active site of the nuclease. We speculate that the processivity of degradation may be impaired when there is insufficient RNA available 5′ to the exosome for cofactors to remain bound. This may lead to stalling, substrate release and oligoadenylation by TRAMP, as outlined in Figure 4D.

DISCUSSION

The exosome targets a huge variety of RNA substrates, but in most cases it remains unclear how RNAs are specifically targeted to the distinct enzymatic activities associated with the complex. To better understand exosome targeting in budding yeast, we performed highly sensitive in vivo RNA-protein crosslinking studies (CRAC), coupled with deep sequencing, on the exosome-associated nucleases Rrp44 and Rrp6, two structural subunits Rrp41 and Csl4, and the exosome cofactor Trf4.

Increased relative crosslinking to pre-tRNA was seen for the Rrp44-exo mutant relative to wild-type Rrp44. This strongly indicates that the highly structured pre-tRNAs are substrates for the

(B) Predicted secondary structure for the pre-rRNA 5′-ETS region (699 nt) from *S. cerevisiae*. Processing sites A0 and A1 are indicated by arrowheads.
(C) Read coverage for full-length and cleaved Rrp44 (red) and Rrp44-exo (black) in the U6 snRNA (left) and the pre-rRNA 5′-ETS region (right). Mapped reads (hits) are depicted in red (Rrp44) or gray (Rrp44-exo); positions of microdeletions (dels) are indicated in black. Data sets used for analysis were either unfiltered (Total) or filtered for reads containing 2 or more non-templated As (A-tailed). FL – full-length protein; N – NTD; C – CTD.
(D) Proposed model for the cooperative action of the endonuclease and exonuclease activities of Rrp44 and the TRAMP complex on structured RNA substrates. Many substrates are threaded through the exosome barrel to reach the active sites of Rrp44, which interacts with the exosome core via the NTD (Bonneau et al., 2009; Malet et al., 2010; Schaeffer et al., 2009; Schneider et al., 2009). Proteins of the RNase II family, which includes the exonuclease domain of Rrp44, are strongly processive and bind substrates tightly in the active site cleft (Zuo et al., 2006). This presumably allows Rrp44 to actively pull substrate RNAs in through the complex (1). However, only single stranded RNAs can enter the lumen of the exosome, so stable RNA-RNA or RNA-protein structures in the substrate potentially lead to stalled complexes, in which the 3′ end is tightly but non-productively bound by Rrp44 (2). We postulate that under these circumstances, the PIN domain cleaves the RNA (3), allowing substrate release (4). The substrate could then be re-adenylated by the TRAMP complex (5) and reloaded into the exosome (6), probably with the assistance of TRAMP.

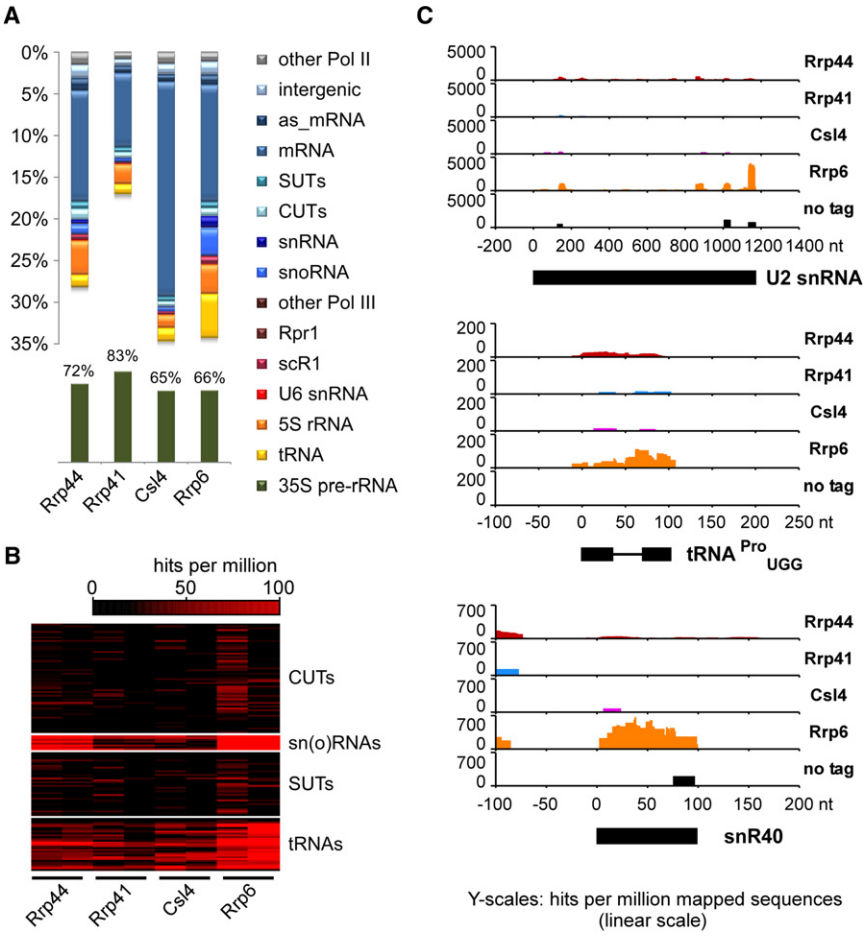


Figure 5. Comparison of Targets of the Core Exosome and Rrp6

(A) Proportion of sequences corresponding to functional RNA classes recovered with core exosome subunits (Rrp44, Rrp41, Csl4), and Rrp6. Sequencing data was analyzed as in Figure 1.

(B) Heat maps for main non-protein coding RNA substrate groups recovered with the indicated IP. Numbers of reads mapped to individual RNAs are shown in shades of red.

(C) Read coverage along highly structured RNAs (top) the U2 snRNA (*LSR1*; 1175nt); (middle) pre-$tRNA^{Pro}_{UGG}$ (102nt, intron 37-66nt) and (bottom) the box C/D snoRNA snR40 (97nt).

exonuclease activity of Rrp44, which become "stuck" in the mutant. In contrast, pre-mRNAs and mRNAs, which are expected to be relatively unstructured compared to Pol III transcripts, were under-represented in Rrp44-exo data sets. Northern analysis revealed the accumulation of pre-tRNAs in Rrp44-exo strains, whereas levels of mature tRNAs were not clearly affected. This is consistent with reduced surveillance, rather than impaired processing of pre-tRNAs (Wlotzka et al., 2011). Pre-tRNAs did not clearly accumulate in *rrp6Δ* single mutant strains (data not shown and (Copela et al., 2008)), but combinatorial loss of Rrp6 and Trf4 strongly amplified the accumulation of pre-tRNAs relative to the absence of Trf4 alone (Copela et al., 2008). We predict that Rrp6 plays a major role in pre-tRNA surveillance in vivo, but this is redundant with the exonuclease activity of Rrp44 and the core exosome.

Other Pol III transcripts, 5S rRNA, U6 snRNA and scR1 were also preferentially crosslinked to Rrp44-exo, as well as to Rrp6 and Trf4 (Wlotzka et al., 2011). This suggests that Rrp44 and Rrp6 directly cooperate to degrade these RNAs, aided by the TRAMP complex. Supporting this idea, the 3′ truncated form of the 5S rRNA (5S*) seen in Rrp44-exo strains also accumulated in strains lacking Trf4 or Rrp6 (Kadaba et al., 2006) and when interactions between Rrp6 and the core exosome were impaired (Callahan and Butler, 2010). Rrp44-exo strains accumulated mature and 3′ extended (up to 3nt) spliceosomal U6 snRNA, together with 3′ truncated forms of the 5S rRNA (5S*) and scR1 (scR1*) (Copela et al., 2008; Kadaba et al., 2006). Oligoadenylated fragments derived from these RNAs were strongly enriched among Rrp44-exo targets, consistent with Trf4 crosslinking (this study and Wlotzka et al., 2011), strongly indicating that these are surveillance rather than processing targets.

Together, the data suggest that wild-type cells produce excess Pol III transcripts, which are normally turned over by Rrp44 and other nuclear 3′ exonucleases including Rrp6 (Callahan and Butler, 2008; Copela et al., 2008; Kadaba et al., 2006; Schneider et al., 2007). Recent transcriptome-wide tiling microarrays and pulse-chase labeling of pre-tRNAs indicate that more than 50% of tRNA gene transcription fails to generate mature, functional tRNAs (Gudipati et al., 2012). A major pathway of exosome-mediated pre-tRNA turnover that competes with tRNA maturation would be consistent with our crosslinking results. Persistent binding of pre-tRNAs to the exosome in the absence of Rrp44 exonuclease activity very likely contributes to the impaired growth and RNA processing in Rrp44-exo strains. The recent finding that ~10% of patients suffering from multiple myeloma carry Rrp44-exo mutations (Chapman et al., 2011) suggests that either increased synthesis of RNA Pol III products, or the resulting impaired RNA surveillance can induce malignant transformation in human cells.

Nuclear pre-mRNAs and cytoplasmic mRNAs are both targets for the core exosome, whereas the activity of Rrp6 is predicted to be specific for the nuclear RNAs (reviewed in (Houseley and Tollervey, 2009)). However, these species cannot readily be distinguished in short sequence reads, other than by the presence of the intron. The clearest distinction is therefore the comparison between intron-exon boundaries (IE), which must be part of the unspliced pre-mRNA, and exon-exon boundaries (EE), which can only be present in the spliced mRNA. Among Rrp6 targets, IE hits were around 2 fold more numerous than EE hits, strongly supporting a role for Rrp6 in pre-mRNA

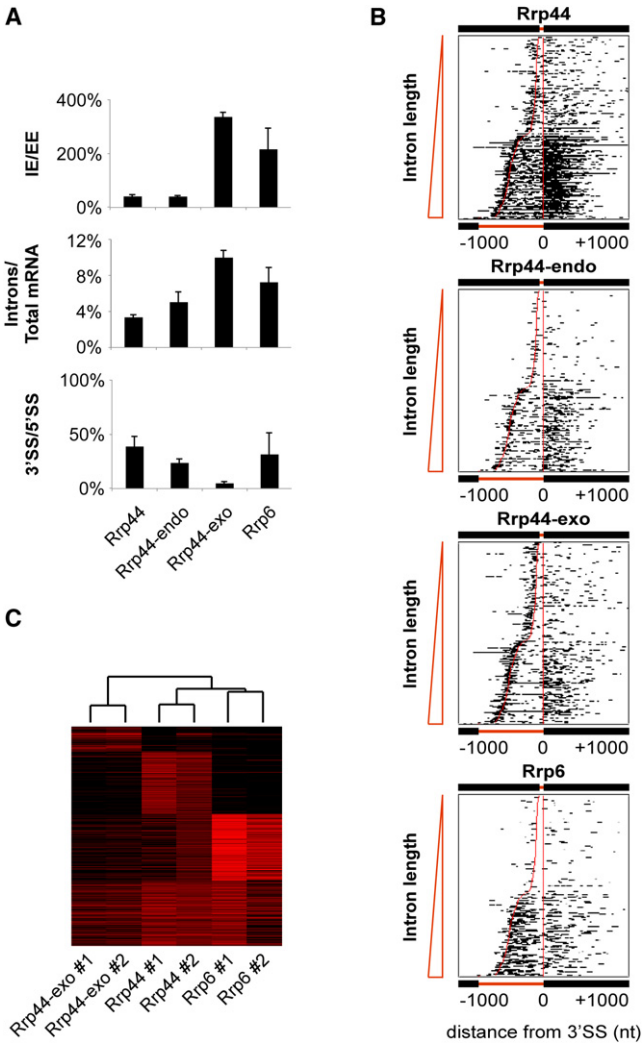


Figure 6. Interactions of Rrp44 and Rrp6 with Pre-mRNA

(A) Frequencies of reads mapped to pre-mRNAs and mature mRNAs. IE/EE: Relative numbers of reads mapped to intron-exon junctions (IE) in pre-mRNAs relative to exon-exon junctions (EE) in mature mRNAs. Introns/Total mRNA: Numbers of reads mapped to mRNA introns relative to the total number of reads mapped to mRNAs. 3′SS/5′SS: Relative numbers of reads mapped to 3′ splice sites (3′SS) in pre-mRNAs, relative to 5′ splice (5′SS) junctions. Bars indicate the standard error.

(B) Rrp44 and Rrp6 binding profiles (black) along 219 intron-containing pre-mRNAs. Pre-mRNAs are aligned at their 3′ splice sites, and ordered by intron length. Intron boundaries are shown as red lines.

(C) Grouping of 4849 mRNAs by pattern of interactions with exosome proteins. Experiments were clustered by complete linkage using the correlation distance metric. Replicate experiments clustered together confirming the reproducibility of the data. Numbers of reads mapped to individual RNAs are shown in shades of red.

surveillance. Consistent with this, analysis of the distribution of Rrp6 reads across spliced genes shows clear enrichment over introns. Cluster analyses of mRNAs showing preferential enrichment in the Rrp6 data sets identified spliced pre-mRNAs but, surprisingly, also found many ribosome synthesis factors. These mRNAs may undergo a significant level of nuclear degradation, possibly as a regulatory mechanism.

Fully functional Rrp44 showed a lower level of sequences over IE boundaries and lower total read coverage over introns, however, both were very substantially increased in the Rrp44-exo mutant. This indicates that Rrp44 is normally actively engaged in degradation of unspliced or partially spliced pre-mRNAs, but these are rapidly and efficiently cleared with little time for crosslinking. Rrp44 showed a high level of crosslinking at the 5′ ends of pre-mRNAs and preferential binding to 5′ splice sites relative to 3′ splice sites. Degradation by the exosome is dependent on cofactors, which must bind 5′ to the complex. Increased crosslinking in the 5′ region may therefore reflect loss of these cofactors leading to slowed degradation.

The Rrp44 sequence coverage over the exons of genes that contain long introns (mainly ribosomal protein genes) was strikingly higher than over genes with shorter introns. This is in agreement with the observation that pre-mRNAs with long introns are preferentially stabilized by loss of Rrp44 function (Gudipati et al., 2012), clearly showing that these are more subject to degradation by the exosome. Whether this is related to the regulated splicing reported for ribosomal protein pre-mRNAs (Pleiss et al., 2007) remains to be determined.

Other Pol II transcripts that are largely degraded in the nucleus include CUTs and SUTs. These transcript classes each showed similar sequence coverage for core exosome and Rrp6. SUTs were designated as "stable un-annotated transcripts" based on a lack of apparent stabilization in the absence of Rrp6 (Xu et al., 2009). However, the similar crosslinking patterns of CUTs and SUTs, and recent microarray analyses in exosome mutant strains (Gudipati et al., 2012), indicate that their degradation pathways are more closely related than their names suggest.

Close functional interactions between Rrp44 and Rrp6 presumably help explain why strains lacking Rrp44 exonuclease activity survive. Although the CTD domain of the Rrp44-exo protein may be tightly and non-productively associated with substrates, the endonucleolytic activity in the N-terminal PIN domain (NTD) of Rrp44 presumably remains competent to cleave these RNAs, providing free 3′ ends for Rrp6 and other exonucleases. Consistent with this model, the split-CRAC data revealed that exonuclease and endonuclease activities of Rrp44 usually act together to degrade RNA substrates. Mapping of the relative binding sites of the Rrp44 NTD and CTD regions combined with analyses of oligoadenylated substrates, leads us to propose a model (Figure 4D) for the role of the PIN domain in releasing stalled exosome substrates. In Rrp44-exo strains, RNAs will be degraded inefficiently, but will still be released from the core exosome by PIN domain cleavage and presented to Rrp6 or other nucleases. In the Rrp44-endo-exo double mutant these substrates may be permanently bound to Rrp44, leading to the accumulation of gridlocked exosome complexes and non-functional RNAs in the cell.

In contrast to pre-tRNAs, other highly structured RNAs that were strongly crosslinked to Rrp6 often showed very few hits in Rrp41 and Csl4 data sets, suggesting that they interact only with Rrp44 and Rrp6, with little or no contact to the remaining core exosome. This was unexpected because

Please cite this article in press as: Schneider et al., Transcriptome-wide Analysis of Exosome Targets, Molecular Cell (2012), http://dx.doi.org/10.1016/j.molcel.2012.08.013

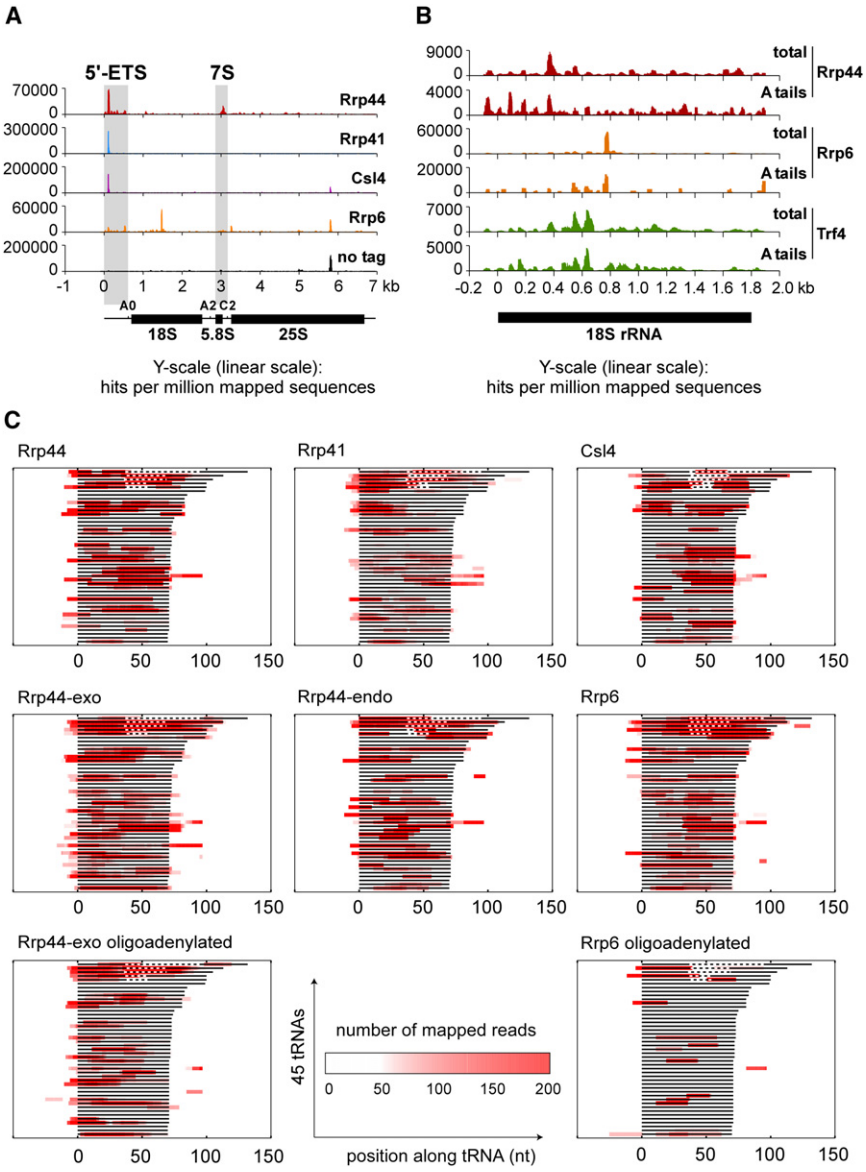


Figure 7. Distribution of High-Throughput Sequencing Reads from Core Exosome, Rrp6, and Trf4 Data Sets over the Pre-rRNA and (Pre-)tRNAs

(A) Coverage of high-throughput sequencing reads along the 35S pre-rRNA (6.9kb). The peak around 5.8kb in the 25S rRNA is a background contaminant seen in many experiments (Granneman et al., 2009; Granneman et al., 2010; van Nues et al., 2011).

(B) Coverage of reads, either unfiltered (total) or filtered for reads containing 2 or more non-templated As (A tails), from Rrp44, Rrp6 and Trf4 data sets were mapped to the 18S rRNA region of the pre-rRNA.

(C) The lines indicate 45 different yeast tRNAs (one for each anticodon family). Dashed lines indicate the presence of introns in the pre-tRNAs. The tRNAs are ranked by length (including intron if present) and aligned at the 5′ end of the mature sequence. Read coverage is indicated by color intensity.

in vitro data indicated that many substrates are channeled to Rrp44 through the catalytically inert exosome barrel (Bonneau et al., 2009). Instead, the in vivo crosslinking data on structured RNAs suggest the use of an alternative entry site to the Rrp44 catalytic center, without contacts to the exosome barrel. Such an alternative entry site can be fitted onto the Rrp44-Rrp41-Rrp45 crystal structure (Bonneau et al., 2009). We therefore hypothesize that at least some in vivo substrates are not threaded through the exosome channel. Instead, they could be docked to Rrp44 from the outside of the complex, aided by tethering to Rrp6 and other exosome cofactors. The basis for this distinction remains unclear, but a long (~33 nt) single-stranded region is required to access the exonuclease domain of Rrp44 via the exosome lumen, whereas much shorter single-stranded regions would be sufficient for direct access to the catalytic sites of Rrp44 or Rrp6.

Despite the apparent cooperation of Rrp44 and Rrp6 on many nuclear surveillance substrates, the comparison of crosslinking sites on individual core exosome subunits with Rrp6 also revealed substrates only enriched in Rrp6 data sets, revealing core-independent Rrp6 functions. One such example is the prominent Rrp6 peak in the 5′-half of the mature 18S rRNA (Figure 7B), which also coincides with a peak of crosslinking by Trf4 (Wlotzka et al., 2011). This corresponds to an RNA polymerase I pause site, at which R-loop formation leads to RNase H cleavage of the nascent transcript (El Hage et al., 2010). We conclude that Rrp6 and the TRAMP complex degrade the cotranscriptionally truncated Pol I primary transcript independently of the core exosome. Rrp6 was reported to localize to the rDNA, interacting with the Nrd1/Nab3 heterodimer and the transcription elongation factors Spt4 and Spt5 (Leporé and Lafontaine, 2011). We therefore speculate that Rrp6 is specifically recruited to the elongating Pol I to survey nascent rRNA transcripts.

EXPERIMENTAL PROCEDURES

Strains and Expression Constructs

Growth and handling of *S. cerevisiae* were by standard techniques. Strains were grown at 25°C or 30°C in synthetic dropout (SD) medium containing 0.67% nitrogen base (Difco) and either 2% glucose or 2% galactose.

Yeast strains for crosslinking studies on Trf4, Rrp41, Csl4 and Rrp6 were constructed by standard methods (Gietz et al., 1992) and expressed genomically encoded, C-terminal HTP-tagged (see below) proteins under the control of their endogenous promoter (see Tables S1 and S2). Strains expressing wild-type and mutant forms of Rrp44 were generated by plasmid shuffling of Rrp44

expression constructs into a host strain derived from BY4741, where the genomic RRP44 ORF was precisely deleted (Schneider et al., 2007; Schneider et al., 2009). Rrp44 expression plasmids comprise the RRP44 ORF under control of its endogenous promoter and different C-terminal and/or internal tags (see below). Plasmids designed for split-CRAC contain a PreScission protease cleavage site (PP) inserted between aa 241 and 242 in the RRP44 ORF to allow in vitro cleavage of purified protein, and a His_6 tag to select the respective cleaved fragment. Point mutations were created using the QuikChange kit (Stratagene). C-terminal tandem affinity purification tags used for basic CRAC and in vivo analyses: HTP: His_6 - TEV cleavage site (TEV) - two copies of the z-domain of protein A (protA); szz: Streptavidin-binding peptide (Strep-tag II) – TEV – protA. Cleavable expression constructs used for split-CRAC to purify N- and C-terminal fragments: Rrp44 N-terminus: His_6 - PP inserted at aa 241 + C-terminal TEV – protA; Rrp44 C terminus: PP inserted at aa 241 + C-terminal His6 – TEV – protA.

Crosslinking and Analysis of Illumina Sequence Data

The CRAC method was performed as previously described (Granneman et al., 2009; Granneman et al., 2011), see Figure 1B for illustration. If not stated otherwise, the same experimental procedure and bioinformatic analyses were applied to all CRAC experiments. To generate RNA-protein crosslinks, actively growing yeast cell cultures in SD medium (OD_{600} ~0.5) were UV-irradiated in a 1.2 m metal tube ("Megatron") for 100 s at 254 nm (Granneman et al., 2011). During split-CRAC on Rrp44, purified full-length proteins were first released from the IgG sepharose resin by TEV protease cleavage and then treated for 2 hr at 18°C with PreScission protease (PP). Cleaved N- and C-terminal fragments were then purified on Ni-agarose under standard CRAC denaturing conditions (Granneman et al., 2009). Illumina sequencing data was aligned to the yeast genome using Novoalign (http://www.novocraft.com). Bioinformatics analyses were performed as described (Wlotzka et al., 2011). The primary sequence data are available from the NCBI Gene Expression Omnibus (Edgar et al., 2002) through GEO Series accession number GSE40046. Mapped reads are presented in Table S3.

RNA Analyses

Yeast RNA extraction and Northern hybridization were performed as described (Tollervey, 1987). Northern signals were visualized by autoradiography or generated by a Fuji FLA-5100 PhosphorImager.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2012.08.013.

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust (077248) (C.S., G.K., A.T., D.T.), the Royal Society (UF100666) (C.S.), the Darwin Trust of Edinburgh and EC Program UNICELLSYS [201142] (W.W.). Work in the Wellcome Trust Centre for Cell Biology is supported by Wellcome Trust core funding (092076).

Received: March 26, 2012
Revised: June 11, 2012
Accepted: August 15, 2012
Published online: September 20, 2012

REFERENCES

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J. *18*, 5399–5410.

Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell *139*, 547–559.

Briggs, M.W., Burkard, K.T., and Butler, J.S. (1998). Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3′ end formation. J. Biol. Chem. *273*, 13255–13263.

Callahan, K.P., and Butler, J.S. (2008). Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p. Nucleic Acids Res. *36*, 6645–6655.

Callahan, K.P., and Butler, J.S. (2010). TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. J. Biol. Chem. *285*, 3540–3547.

Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M., et al. (2011). Initial genome sequencing and analysis of multiple myeloma. Nature *471*, 467–472.

Copela, L.A., Fernandez, C.F., Sherrer, R.L., and Wolin, S.L. (2008). Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. RNA *14*, 1214–1227.

Davis, C.A., and Ares, M.J., Jr. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *103*, 3262–3267.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. *4*, 3.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev. *24*, 1546–1558.

Gietz, D., St Jean, A., Woods, R.A., and Schiestl, R.H. (1992). Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res. *20*, 1425.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc. Natl. Acad. Sci. USA *106*, 9613–9618.

Granneman, S., Petfalski, E., Swiatkowska, A., and Tollervey, D. (2010). Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. EMBO J. *29*, 2026–2036.

Granneman, S., Petfalski, E., and Tollervey, D. (2011). A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. EMBO J. *30*, 4006–4019.

Grzechnik, P., and Kufel, J. (2008). Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. Mol. Cell *32*, 247–258.

Gudipati, R.K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012). Massive degradation of RNA precursors by the exosome in wild type cells. Mol. Cell *48*. Published online September 20, 2012. http://dx.doi.org/10.1016/j.molcel.2012.08.013.

Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D., and Brown, P.O. (2008). Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol. *6*, e255.

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

Jamonnak, N., Creamer, T.J., Darby, M.M., Schaughency, P., Wheelan, S.J., and Corden, J.L. (2011). Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. RNA *17*, 2011–2025.

Kadaba, S., Krueger, A., Trice, T., Krecic, A.M., Hinnebusch, A.G., and Anderson, J. (2004). Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae. Genes Dev. *18*, 1227–1240.

Kadaba, S., Wang, X., and Anderson, J.T. (2006). Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. RNA *12*, 508–521.

Kiss, D.L., and Andrulis, E.D. (2010). Genome-wide analysis reveals distinct substrate specificities of Rrp6, Dis3, and core exosome subunits. RNA *16*, 781–791.

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

Leporé, N., and Lafontaine, D.L. (2011). A functional interface at the rDNA connects rRNA synthesis, pre-rRNA processing and nucleolar surveillance in budding yeast. PLoS ONE *6*, e24962.

Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell *127*, 1223–1237.

Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008). Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. Mol. Cell *29*, 717–728.

Malet, H., Topf, M., Clare, D.K., Ebert, J., Bonneau, F., Basquin, J., Drazkowska, K., Tomecki, R., Dziembowski, A., Conti, E., et al. (2010). RNA channelling by the eukaryotic exosome. EMBO Rep. *11*, 936–942.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038–1042.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M., and Guthrie, C. (2007). Rapid, transcript-specific changes in splicing in response to environmental stress. Mol. Cell *27*, 928–937.

San Paolo, S., Vanacova, S., Schenk, L., Scherrer, T., Blank, D., Keller, W., and Gerber, A.P. (2009). Distinct roles of non-canonical poly(A) polymerases in RNA metabolism. PLoS Genet. *5*, e1000555.

Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G., Sanchez-Rotunno, M., Arraiano, C.M., and van Hoof, A. (2009). The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. Nat. Struct. Mol. Biol. *16*, 56–62.

Schneider, C., Anderson, J.T., and Tollervey, D. (2007). The exosome subunit Rrp44 plays a direct role in RNA substrate recognition. Mol. Cell *27*, 324–331.

Schneider, C., Leung, E., Brown, J., and Tollervey, D. (2009). The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. Nucleic Acids Res. *37*, 1127–1140.

Tollervey, D. (1987). A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. EMBO J. *6*, 4169–4175.

van Nues, R.W., Granneman, S., Kudla, G., Sloan, K.E., Chicken, M., Tollervey, D., and Watkins, N.J. (2011). Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing. EMBO J. *30*, 2420–2430.

Vanácová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol. *3*, e189.

Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J. *30*, 1790–1803.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725–737.

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

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat. Biotechnol. *29*, 607–614.

Zuo, Y., Vincent, H.A., Zhang, J., Wang, Y., Deutscher, M.P., and Malhotra, A. (2006). Structural basis for processivity and single-strand specificity of RNase II. Mol. Cell *24*, 149–156.