A new level of gene regulation: establishing a genome-wide role for antisense transcription

Doctorate of Philosophy, University of Oxford

Struan C. Murray

St John's College, Michaelmas Term 2013

Approximate word count: 45,000

A new level of gene regulation: establishing a genome-wide role for antisense transcription

A thesis submitted to the University of Oxford for the degree of Doctor of Philosophy

Struan C. Murray, St John's College, Michaelmas Term 2013.

Abstract

Transcription lies at the centre of gene expression. In eukaryotes, transcription occurs not only at genes but also across the non-coding portion of the genome, an apparently pervasive process that gives rise to a wide array of different transcripts. In recent years, it has emerged that genes themselves are frequently subject to non-coding transcription of their antisense strand. This antisense transcription is evident in eukaryotes from yeast to mammals; however its general genome-wide role, if indeed it has one, remains elusive. Here, the nature of antisense transcription in the budding yeast *Saccharomyces cerevisiae* is explored on a genome-wide scale. Antisense transcription is ubiquitous and often abundant, and appears to be driven by a promoter architecture at the 3' end of genes, one which shows evidence of regulation, and which mirrors that found at the 5' end. Furthermore, antisense transcription shows evidence of changing gene behaviour. It is associated with a drastically altered chromatin environment at the 5' promoter and across the gene body; however it is *not* associated with a change in the level of gene transcription itself. Rather, these chromatin changes appear to enforce a change in the *mode* of gene transcription, promoting rapid bursts of transcription re-initiation that result in noisier gene expression – a hitherto unknown role of antisense transcription. It is proposed that antisense transcription represents a fundamental layer of gene regulation, and that it should be considered a canonical feature of eukaryotic genes.

Acknowledgements

There are a great many people I would like to thank for helping me, whether knowingly or unknowingly, in the making of this thesis. Special thanks go to Jane Mellor, my supervisor, who has provided me throughout with guidance and invaluable discussions, Michael Youdell; a constant source of wisdom and encouragement, and Françoise Howe, for her advice, tremendous aid, and for performing many of the experiments that feature in this work. I would also like to thank Tania Nguyen, Ana Serra Barros, Simon Hänni and Anitha Nair for performing further experiments, and the rest of the Mellor lab for providing such a cheerful environment in which to work.

I'd like to thank the following people for making my time at Oxford what it was: Ian Ashpole, Edward Lent, Izzy Lever, Duncan Palmer, Katie Stileman, Gregory Ross, Aimee Zisner, Jessica McGillen, James Sleigh, Jonas Schlüter, Lydia le Page, Ant Hibbs, Norbert Linke, Michael Osborne, Isaac Turner, Chloé Omissi, Aidan Russell, Matt Lewis, Madura Jayatunga and Adrastos Omissi.

Finally, I'd like to than my family for being unfailingly supportive.

A note on experimental data

Though the bulk of this thesis is computational in nature, it also contains experimental data intended to compliment the computational work and to validate hypotheses developed along the way. This work was performed by Françoise Howe, Tania Nguyen, Ana Serra Barros, Simon Hänni, Karolina Chocian and Anitha Nair. Françoise Howe performed RNA-FISH and ChIP experiments, Tania Nguyen and Simon Hänni constructed strains and performed northern blots and RNA-FISH experiments, Ana Serra Barros constructed strains and performed northern blots, Karolina Chocian performed ChIP experiments, and Anitha Nair constructed strains.

Where shown, figure legends accompanying this data shall specify who performed particular experiments. For the RNA-FISH data, all image analysis was performed by the author.

Publications

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 Research* **40**: 2432-2444.

Abbreviations

Abbreviation	Term in full
ac	Acetylation
bp	Base-pair
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation-sequencing
COMPASS	Complex of proteins associated with Set1
CTD	C-terminal domain
CUT	Cryptic unstable transcript
CFP	Cyan fluorescent protein
DBD	DNA binding domain
FACT	Facilitates Chromatin Transcactions [complex]
GFP	Green fluorescent protein
me	Methylation
me1	Monomethylation
me2	Dimethylation
me2a	Asymmetric dimethylation
me3	Trimethylation
MNase	Micrococcal nuclease
mRNA	Messenger RNA
NET-seq	Native elongating transcript sequencing
ncRNA	Non-coding RNA
NDR	Nucleosome depleted region
NFR	Nucleosome free region
OD	Optical density
ORF	Open reading frame
PIC	Pre-initiation complex
RACE	Rapid amplification of cDNA ends
rRNA	Ribosomal RNA
RNA FISH	RNA fluorescence *in situ* hybridisation
RNA-seq	RNA sequencing
RNAP	RNA polymerase
RNAPI	RNA polymerase I

Abbreviation	Term in full
RNAPII	RNA polymerase II
RNAPIII	RNA polymerase III
SAGA	Spt-Ada-Gcn5 acetyltransferase
SAGE	Serial analysis of gene expression
Ser	Serine residue
SGD	Saccharomyces Genome Database
snRNA	Small nuclear RNA
snoRNA	Small nucleolar RNA
SUT	Stable unannotated transcript
tRNA	Transfer RNA
TBP	TATA-binding protein
TFBS	Transcription factor binding site
TSS	Transcription start site
TTS	Transcription termination site
ub	Ubiquitination
XUT	Xrn1-senstive unstable transcript
YFP	Yellow fluorescent protein
YPD	Yeast extract – bactopeptone – glucose
YPG	Yeast extract – bactopeptone – galactose

Throughout this thesis, 'yeast' will be used interchangeably with *Saccharomyces cerevisiae*.

Table of Contents

Table of Figures

Chapter One

Introduction

1. Introduction

1.1. Transcription lies at the centre of gene expression

Transcription is a fundamental cellular process in which the sequence of nucleotide residues within a region of DNA is used to direct the synthesis of a single strand of RNA and, crucially, to define that RNA's sequence. Genes transcribed in this manner give rise to a particular class of transcript that are then used to inform the amino acid sequences of proteins through the process of translation. For this reason such transcripts are referred to as *messenger* RNAs (mRNAs), or else coding RNAs. Through this gene transcription and subsequent translation does an organism's genetic information become utilised, and hence the sum of these processes is referred to as gene expression. However, as well as forming a direct component of this transfer of information, transcription also plays other crucial roles in gene expression, through the transcription of non-genic regions of the genome, and the non-coding RNAs (ncRNAs) that result from this process. ncRNAs, as their name suggests, do not encode proteins, but rather perform important structural and regulatory roles, and have been implicated in organismal development (Conte et al. 2013), ageing (Dimmeler and Nicotera 2013) and also diseases including cancer (Cheng et al. 2013), muscular dystrophy (Batista and Chang 2013) and Alzheimer's (Smith-Vikos and Slack 2013). Though some of their cellular roles are well established, other ncRNAs are emerging whose functions are less well described. Furthermore, the process of transcription itself is beginning to appear more complex and widespread than previously anticipated, and consequently our views on its role within a genomic context are changing. Here, questions are addressed regarding non-coding transcription in the genome of the budding yeast *Saccharomyces cerevisiae*. To begin with, however, it is necessary to review the current understanding of transcription more generally.

1.2. The mechanism of transcription comprises three stages

Transcription requires an enzyme known as RNA polymerase (RNAP), which can bind and then process along a region of DNA, recognise the DNA sequence, and synthesise a corresponding RNA molecule as it processes. RNA polymerase cannot synthesise RNA independent of DNA and so is said to be DNA dependent. Briefly, the process of transcription involves three stages: initiation, in which the polymerase goes from being freely dissociated to being DNA bound and synthesising the initial base pairs of RNA; elongation, in which the polymerase processes along the DNA whilst synthesising RNA, assuming transcription is not aborted or arrested prematurely (see below); and termination, in which the polymerase disengages from the DNA.

Transcription initiation occurs within, or immediately downstream of, a region of DNA called a promoter. The sequences of promoters vary extensively, as will be discussed in chapter 3, but are generally considered to be any sequence sufficient to support some level of initiation. Promoters are found immediately upstream of genes, where they direct production of mRNAs, but are also present upstream of the sequences of the non-coding transcripts described below. In eukaryotes, transcription initiation requires the ordered assembly of general transcription factors at the promoter, which direct the recruitment of the polymerase and its subsequent passage into elongation. Together, these factors comprise the pre-initiation complex (PIC). The composition of the PIC is dependent upon which RNA polymerase is being recruited – in eukaryotes there are three types: RNAPI, II and III. RNAPI and III are involved in production of non-coding RNAs as will be discussed below, while RNAPII is involved in the production of mRNAs as well as non-coding RNAs.

In vitro studies have shown that transcription initiation of RNAPII minimally requires TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, while TFIIA serves to enhance the level of transcription (Roeder 1996). Binding of TFIID, a complex containing the TATA-binding

protein (TBP, discussed in chapter 3) and at least fourteen other factors, is required before subsequent assembly can occur. TFIID introduces a kink in the promoter DNA, which allows for recognition and binding by subsequent factors. TFIIB then binds and stabilizes TFIID at the promoter, binding both TFIID and sequences in the promoter flanking the TFIID binding site. TFIIB then recruits the complex of RNAPII and TFIIF, subsequent to which TFIIE and TFIIH bind. TFIIH catalyses the 'melting' (unwinding) of the DNA downstream, making it accessible for the procession of RNAPII into elongation (Dvir et al. 2001). Following this, certain PIC components (TFIID, TFIIA, TFIIH, TFIIE) have been shown to remain bound to the DNA, where they can support subsequent rounds of initiation (termed re-initiation) (Yudkovsky et al. 2000). Note that at approximately 10% of yeast genes initiation involves a TBP-containing complex known as SAGA in place of TFIID and the other general transcription factors, which is thought to play important roles in the stress response (Huisinga and Pugh 2004). Note also that many promoters contain a sequence element called the TATA-box, which, when present, is bound by TBP, though it is not necessary for TBP binding. Deletion of the TATA box results in a reduction of transcription, as shall be discussed in chapter 3.

The passage into elongation is not an inevitable consequence of transcription initiation. Successful passage of the polymerase complex into elongation is referred to as promoter escape (Dvir 2002). Failure to escape results in aborted initiation – Luse and Jacob (1987) found that in an *in vitro* system RNAPII produced trinucleotides in reactions primed by dinucleotides, while Holstege et al., (1997) found that transcripts accumulated of up to 10 nucleotides following initiation on an adenovirus promoter. Crystal structures of a TFIIB-RNAPII complex show that the N-terminal region of TFIIB forms a loop (called the "B-finger) that reaches into the catalytic centre of the polymerase, contacting both DNA and nascent RNA (Bushnell et al. 2004). Once the nascent RNA transcript reaches 7 nucleotides in length it clashes sterically with this

finger, necessitating TFIIB displacement or else abortive transcription (Liu et al. 2010). Suppression of abortive transcription has been shown to be dependent upon the promoter itself, as well as other factors such as phosphorylation of RNAPII, and modification of the histone proteins, both of which shall be discussed below, which are thought to contribute towards the stability of the RNAPII-DNA-RNA ternary complex. (Shandilya and Roberts 2012).

Upon entering into elongation, the RNAPII complex begins to undergo certain modifications, required for the recruitment of factors that aid ultimately in termination but also in processing of the nascent RNA transcript – chemical modifications that ensure the transcript's export from the nucleus into the cytoplasm. The largest subunit of RNAPII, Rpb1, has a C-terminal domain (CTD) comprising multiple copies (52 in mammals and 26 in yeast) of the heptad repeat YSPTSPS, which can undergo phosphorylation. Prior to elongation, the fifth and seventh serine residues of the CTD (Ser5 and Ser7) are phosphorylated by the kinase Kin28, part of the TFIIH complex (Akhtar et al., 2009; Svejstrup et al., 1994). During transcription elongation, this pattern of modification changes, with levels of Ser5 phosphorylation dropping due to the action of the Rtr1 and Ssu72 phosphatases (Mosley et al. 2009; Krishnamurthy et al. 2004; Bataille et al. 2012; Komarnitsky 2000) and with phosphorylation of the second serine of the repeat (Ser2) rising towards the end of the gene, catalysed by Ctk1 and Bur1 (Cho et al. 2001; Liu et al. 2009; Qiu et al. 2009). Following termination, the Fcp1 phosphatase removes Ser2 phosphorylation prior to further rounds of transcription. Ser5 phosphorylation recruits factors involved in *capping* the RNA transcript as it is being synthesised – the cap being a 5' guanine residue bound by a 5' to 5' triphosphate linkage – which prevents the transcript from being prematurely degraded (Rodriguez et al. 2000; Schroeder 2000). Further downstream, Ser2 phosphorylation directs the recruitment of factors involved in termination and 3'-end processing, as described

below. CTD modifications are also involved in the recruitment of factors that regulate transcription, as will be discussed in chapter 5 of this work.

Termination involves two intertwined processes: the disengagement of RNAP and the processing of the nascent transcript's 3' end (Mischo and Proudfoot 2013). Processing involves cleavage of the 3' end of the nascent transcript and subsequent synthesis of a poly(A) tail by a poly(A) polymerase. This process is dependent upon a sequence within the 3' region of a gene called the poly(A) signal, which in mammals takes on the sequence AAUAAA (Proudfoot 2011), the mutation of which leads to aberrant transcription (Fitzgerald and Shenkt 1981). Mutation of this sequence impairs both processing *and* termination, demonstrating the connectivity of these processes (Edwalds-Gilbert et al., 1993). As well as the poly(A) signal, a downstream, GU rich sequence enhances 3' end formation (Conway and Wickens 1985; Green and Hart 1988), as does an upstream sequence (Danckwardt et al. 2007; Venkataraman et al. 2005). Several downstream elements have been defined in mammals, though they all share the ability to create a region of high RNAPII density – a phenomenon referred to as polymerase pausing – which is believed to slow down the RNAPII and promote termination (Gromak et al. 2006). Yeast terminator sequences also appear to require such a tripartite arrangement of sequences, however these elements are less well defined and show a distinct lack of sequence homology (Guo et al. 1996; Humphrey et al. 1994, 1991). Nevertheless, the poly(A) site has been shown to be necessary for termination in yeast: Osborne and Guarente (1989) used mutational analysis to define an 83bp region necessary for correct polyadenylation and termination in the *CYC1* gene, while downstream pausing sites have been shown to enhance the process (Hyman and Moore 1993). Such a sequence, alone sufficient to direct 3' end processing and termination, is referred to as a *terminator.* The mechanism of processing and termination in mammals involves five proteins (poly(A) polymerase, cleavage and

polyadenylation specificity facto, cleavage stimulation factor, and cleavage factors I and II) that direct cleavage of the nascent mRNA's 3' end followed by polyadenylation, all of which have homologous factors in yeast (Proudfoot 2004).

Together, promoters and terminators define what are known as transcription units – self-contained regions undergoing transcription from a transcription start site (TSS) to a transcription termination site (TTS). This is complicated, however, by the fact that some genes can be initiated and/or terminated from distinct sites resulting in alternative transcription units being used and leading to transcript heterogeneity (Nagalakshmi et al. 2008). Nevertheless, promoters and terminators can generally be considered both the boundaries and the necessary constituents of a transcription unit. Promoters can be experimentally positioned upstream of exogenous open reading frames (ORFs), such as that of green fluorescent protein (GFP), and in doing so promote expression (Stagoj et al. 2005), while it is shown in chapter 3 that introduction of a terminator into the ORF of a gene redefines its transcription unit. Clearly, promoters and terminators represent critical elements in defining the transcriptional landscape of the genome. Furthermore, recent evidence would seem to suggest that promoters and terminators might in some way be functionally analogous, as shall be discussed below.

1.3. Transcription is a highly regulated process

All genes are not transcribed equally. Depending on external conditions or a given cell's role within an organism, signalling programs define which genes are being transcribed at any given time, and to what extent, establishing the overall complement of transcripts in a cell – the so-called transcriptome. Transcription can vary in terms of the rate at which polymerases initiate at the promoter, or else elongate across the gene, or alternatively a given gene might be actively transcribed in some cells within a population but not in others. Generally, mechanisms that serve to increase transcription

are termed *activatory*, and those that reduce it termed *repressive.* Numerous mechanisms exist to regulate transcription in eukaryotes, all of which are integrated in order to define the overall transcription state of a given gene.

1.3.1. Transcription factors direct changes in the transcriptome

Transcription factors are proteins that regulate transcription, typically by altering the stability of the transcription complex itself (the assembly of RNAP and the general transcription factors), or through other mechanisms described below. Transcription factors respond to changes in environmental signals and consequently bind to DNA, and in doing so modulate gene activity. For example, the cell cycle in yeast is dependent upon the periodic transcription of distinct clusters of genes, which are in turn controlled by periodic changes in the activation states of certain transcription factors; the SBF transcription factor, for instance, is repressed by binding of the Whi5 protein, but activated in the late G1 phase when the Whi5 protein is phosphorylated and subsequently degraded, leading to activation of the relevant genes (Wittenberg and Reed 2005). Chromatin immunoprecipitation (ChIP) experiments have shown that, upon stimulation of transcription factors, the levels of PIC components crosslinking to promoter regions significantly increases (Weake and Workman 2010). Transcription factors are one of the most abundant classes of gene in the yeast genome (Teixeira et al. 2006), each recognising and binding to its own specific binding site – short stretches of DNA sequence recognised by the protein's DNA binding domain (DBD). Attempts to map transcription factor binding sites (TFBSs) genome-wide in yeast have met with wildly different reports of the total number of binding sites, but they are generally thought to be most enriched within the region 50-150bp upstream of the TSS of a gene (Hughes et al. 2000). Transcription factors can promote transcription simply by helping to tether the transcription machinery to the DNA – crosslinking experiments show that the Gcn4

transcription factor binds to multiple different PIC components (Fishburn et al. 2005), while fusion of TBP to the DBD of the transcription factor LexA dramatically increased transcription, without the need for the transcriptional activation domain of LexA (Chatterjee and Struhl, 1995). Alternatively, transcription factors can bind to the transcription complex indirectly via Mediator, a 25-subunit assembly that serves as an interface between transcription factors and the transcription complex (Biddick and Young 2005), or else recruit coactivators as shall be described below. Transcription factors can themselves be regulated by changes in activity, as described above, or by changes in their levels or localization. For example, Gcn4's cellular concentration is regulated by nutrient availability and ubiquitination-dependent turnover (Hinnebusch 2005), while Pho4 can be exported to the cytoplasm following its phosphorylation by a cyclin-dependent kinase complex (Raser and O'Shea, 2004). Through such regulation, transcription factors can respond to changes in the cellular environment and in turn modify the transcriptome of the cell accordingly.

1.3.2. An additional layer of regulation: histones, nucleosomes and the NDR

The DNA of a eukaryotic genome does not exist in isolation. It is bound by all manner of proteins, such as the transcription factors, polymerases and PIC components described above, constantly associating and disassociating from the DNA and directing gene expression. In addition, eukaryotic DNA is bound extensively by a core of basic histone proteins, an octameric structure consisting of two H2A, H2B, H3 and H4 subunits, around which 146bp of DNA is wound in a superhelix to produce a structure called a nucleosome (Luger et al. 1997a; Richmond et al. 1997; Phillips and Johns 1965). Estimates suggest that approximately 55,000-70,000 nucleosomes are bound to yeast DNA – thus covering approximately 80% of the genome (Lee et al., 2007; Mavrich et al., 2008; Shivaswamy et al., 2008). That DNA should be bound in such a way creates an

issue of topology – other proteins should be impeded in binding to histone-bound DNA, and so processes such as transcription should also be impeded. Indeed, *in vitro* studies showed that histones inhibited transcription initiation, and that elongation led to displacement of histones from the DNA (Lorch et al. 1987), while TFBSs tend to be depleted of nucleosomes *in vivo* (Bernstein et al. 2004). Furthermore, genome-wide studies show that nucleosome occupancy is generally high at the promoters of repressed genes (Lee et al. 2007). Most genes, however, display a distinctive nucleosome arrangement at their 5' end, whereby the TSS is flanked by two nucleosomes (termed the -1 and +1 nucleosomes, upstream and downstream respectively), bordering a region of around 50-150bp (Yuan et al. 2005). This region is referred to as a Nucleosome Free Region (NFR), or else a Nucleosome Depleted Region (NDR), and is thought to serve as a platform for the recruitment of the PIC and accompanying transcription factors – indeed, it has been proposed that DNA signals and regulatory factors tightly define nucleosome positioning at the promoter, and that this regulates the positioning of adjacent nucleosomes further up- and downstream (Mavrich et al. 2008). At the promoter, the size of the NDR is thought to play a role in defining the activation level of the gene (Figure 1). Active genes more frequently have open NDRs (~150bp), while at inactive genes they are more frequently closed (~50bp) (Zaugg and Luscombe 2012). Note that these states represent extremes: in a population of cells the same gene might have a closed NDR in some cells and an open one in others, resulting in intermediate levels of transcription. NDRs might also be partially closed, potentially resulting in intermediate transcription, though Zaugg and Luscombe (2012) found that NDRs form a bimodal distribution genome-wide, suggesting a preference for either the closed or open state. They also found that, despite this preferance, there are still inactive genes that have open NDRs and vice versa, thus other factors can clearly still regulate transcription independent of nucleosomes. Interestingly, NDRs have also

been shown to be present at the 3' ends of genes, where they are thought to be permissive to termination (Fan et al. 2010). Though nucleosomes would appear to play a generally repressive role in transcription, they have also been shown to aid transcription by helping to juxtapose TFBS (Stünkel et al. 1997). Nucleosomes clearly represent an additional level of genomic complexity, all the more because of their capacity to be covalently modified, as shall be discussed below.

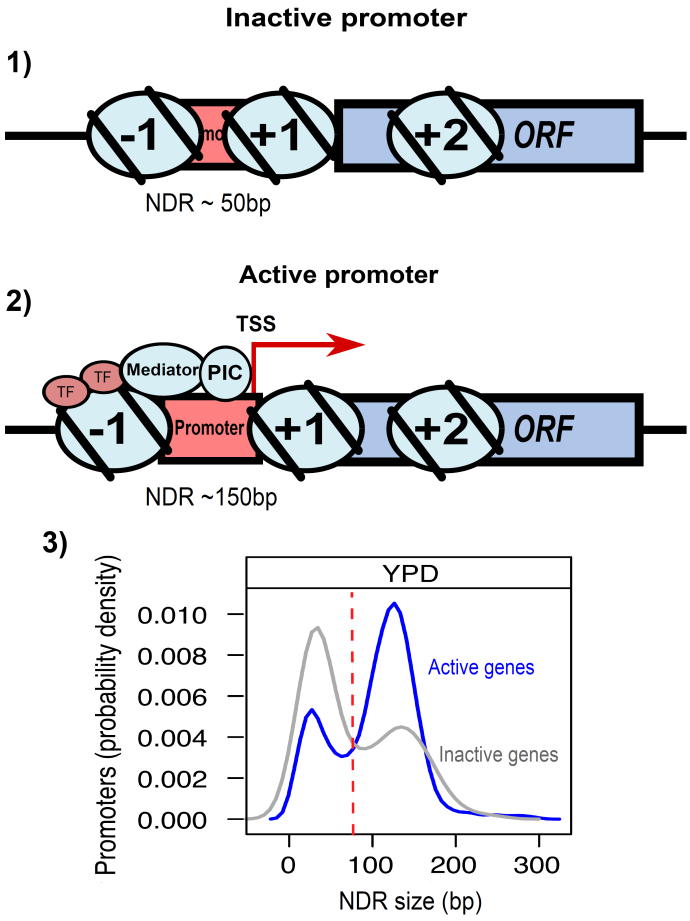


Figure 1: A model of nucleosome-mediated promoter activation

A schematic showing the arrangement of nucleosomes around promoters, as established by genome-wide studies (Yuan et al., 2005). **1)** Inactive promoters are frequently marked by high nucleosome occupancy, with small NDRs thought to prevent access of the transcription machinery to the promoter sequence and so blocking transcription initiation. **2)** Repositioning of the -1 and +1 nucleosomes allows access of transcription machinery and potentially of activatory transcription factors, resulting in transcription initiation. **3)** A probability density curve showing the distributions of NDR sizes in both active and inactive genes in YPD medium, taken from Zaugg and Luscombe, (2012). A bimodal distribution of NDR sizes is evident in both classes, suggesting a preference for genes to adopt one of these two extremes. The red dotted line shows the approximate minima of the bimodal distribution for all genes.

1.3.3. Histone modification and regulation

All four core histones can undergo extensive covalent modification, including acetylation, methylation, phosphorylation, ubiquitination and sumoylation, which are thought to have important gene regulatory consequences (Berger 2007), or at least to denote certain gene states (Henikoff and Shilatifard 2011). Histone modifications form a major focus of chapter 5, and shall be discussed in more comprehensive detail there. Suffice to say that different histone modifications are associated with different genes, gene states, and gene processes – for example, lysine acetylation of H3 at gene promoters is associated with active genes (Pokholok et al. 2005), trimethylation of lysine 79 on H3 (H3K79me3) is typically found at the 3' end of long genes (Schulze et al. 2011), and acetylation of lysine 56 on H3 (H3K56ac) is associated with the deposition of new histones on the DNA. Histone modifications are governed by histone modifying enzymes, which can in turn be recruited by transcription factors to the promoters of genes (Narlikar et al. 2002). Furthermore, histone modifications can themselves recruit proteins via certain domains – the bromodomain, for instance, binds to acetylated regions (Zeng and Zhou 2002), which can lead to further regulatory events. Histones can also be enzymatically moved along DNA, whilst remaining bound, by the chromatin remodelling enzymes, which again will be discussed in more detail in chapter 5. These enzymes can alter the accessibility of DNA, can be recruited both by transcription factors (Narlikar et al. 2002) and by histone modifications (Kasten et al. 2004), and are involved in the opening and closing of the NDR (Yen et al. 2012). The gene promoter therefore plays host to a complex interplay of transcription factors, nucleosomes, histone modifications and remodelling enzymes, defining the overall level of transcription.

1.4. Certain non-coding RNAs have well defined roles

A number of endogenous ncRNAs have been well understood for many years, at least compared to the more recently emerging transcripts discussed below. The transfer RNAs (tRNAs), for example, function as the adaptor between mRNA and protein, recognising the RNA codon and bringing the correct amino acid to the active site of the ribosome (Hopper and Phizicky 2003). The ribosome itself contains many ribosomal RNAs (rRNAs), indeed, the ribosomal RNA genes in yeast comprise approximately 10% of the genome, and much of the transcriptome as well (Warner 1999). The active site of the ribosome – the peptidyl-transferase centre where peptide bonds are formed – is itself formed entirely by rRNAs, not by amino acids, demonstrating that RNAs can catalyse chemical reactions (Ban et al. 2000; Rodnina et al. 2007). Other RNAs include small nuclear RNAs and small nucleolar RNAs, which are involved in the processing of other RNAs. Finally, microRNAs are a class of RNA involved in the regulation of gene expression in multicellular organisms (Novina and Sharp 2004). microRNAs are 19-25bp in length, and form complimentary hairpin structures that are then processed by the dicer machinery and used to supress translation by binding to mRNAs with a complimentary sequence (Novina and Sharp 2004). Short interfering RNAs are similar save that they arise from double stranded RNAs and can impede transcription itself by recruiting inhibitory proteins to genes or through the modification of chromatin or DNA. Notably, the mechanism through which this silencing is directed is absent in *S. cerevisiae*. Clearly these distinct classes of ncRNAs perform a broad array of functions; however, there is also a newly emerging class of ncRNA, one that appears to be produced pervasively genome-wide, at sites overlapping the transcription units of genes themselves, which hints at an even wider ncRNA complexity than previously thought.

1.5. Eukaryotic genomes are pervasively transcribed

A movement towards studying genomes on a large scale, through techniques such as microarrays and high-throughput sequencing, has revealed a hitherto-unexpected level of transcriptional complexity within eukaryotes. Rather than being tightly confined to the coding strands of ORFs (and their accompanying 5' and 3' UTRs) and various well-established non-coding RNAs (tRNAs, rRNAs, snoRNAs etc.), these studies point towards a transcriptional landscape in which transcription occurs, at least at low levels, in much of the remaining genome, with non-coding transcripts arising from within intergenic regions, and, as discussed below, the antisense strands of genes.

Even prior to genome-wide studies, evidence for widespread transcription existed as early as the 1970s. Milcarek et al., (1974) found an abundance of RNAs in a poly(A) minus fraction from HeLa cells (30% of the cytoplasmic mRNA abundance), while another study in sea urchin found that the sequence diversity of heterogenous nuclear RNA was ten times that of the cytoplasmic mRNA population (Hough et al. 1975). With the advent of microarray technologies, numerous studies in mammals provided evidence for a more pervasive transcription not restricted to open reading frames, supporting the observed transcript diversity. Studying transcription of human chromosomes 21 and 22 with microarrays, Kapranov et al., (2002) found that nearly ten times as much of the genomic sequence was transcribed as expected based on the location of 770 previously predicted and/or characterised genes. A subsequent study of the human genome, in which rapid amplification of cDNA ends (RACE) was combined with microarray analysis, identified a vast array of transcripts, the majority of which were previously unannotated, corresponded to non-coding regions, were not polyadenylated, and were retained within the nucleus (Cheng et al. 2005). Further work assessing the locations of PIC components within human fibroblasts found a wealth of promoter structures that did not correspond to the locations of known genes (Kim et al.

2005), suggesting that non-coding transcription might be driven by a similar mechanism to gene transcription.

In yeast, similar approaches have also revealed that transcription is pervasive. The yeast genome, unlike mammalian genomes, is highly compact; chromosome II, for example, was found to consist mostly of open reading frames, comprising 72% of the total sequence (Feldmann et al. 1994). It is perhaps not so surprising, then, that the yeast genome is pervasively transcribed. Indeed, through use of a genome-wide tiling array in yeast, it was found that 85% of the genome is transcribed in rich media (David et al. 2006). Here, a region was defined as transcribed if it showed transcription significantly above background *on either strand* (such that if significant expression was detected one strand but not the other at a given position it was still considered towards this percentage). However, there is still evidence that much of this is non-coding or else in regions not previously thought to be transcribed. A genome-wide map of RNAPII occupancy (Steinmetz et al. 2006) found that RNAPII is present in heterochromatic regions – previously thought to be transcriptionally silent, and in intergenic regions, suggesting a role in non-coding transcription.

The presence of such pervasive transcription raises the question as to its possible function. In recent years, numerous studies have identified a class of transcript in eukaryotes that may represent a major source of pervasive transcription. Here, these *antisense transcripts* are introduced, and their possible genome-wide role discussed.

1.6. Antisense transcription is a major source of pervasive transcription

Here, an antisense transcript is defined as a non-coding transcript whose transcription unit overlaps a protein-coding gene in the antisense orientation. The coding transcript is referred to as the *sense* transcript. Note that some studies have made the distinction

between *cis* antisense transcripts – those described above – and *trans* antisense transcripts, which are those transcribed at a distinct loci but which share at least a partial sequence complementarity with that of a coding gene (this is not to be confused with the *cis*-acting and *trans*-acting regulatory roles for antisense transcription, which are described below). Here the discussion is restricted entirely to the *cis* class. Furthermore, there are instances in which coding genes overlap one another's transcription units in a convergent manner, such that the genes can be said to be *antisense* to one another, and both can give rise to coding genes. Again, however, this discussion is limited strictly to *non-coding* transcripts. Individual examples of antisense transcripts have been identified in viruses, prokaryotes and eukaryotes (Lapidot and Pilpel 2006; Knee and Murphy 1997), however, their true prevalence did not become evident until genome-wide studies were performed, as will be discussed below.

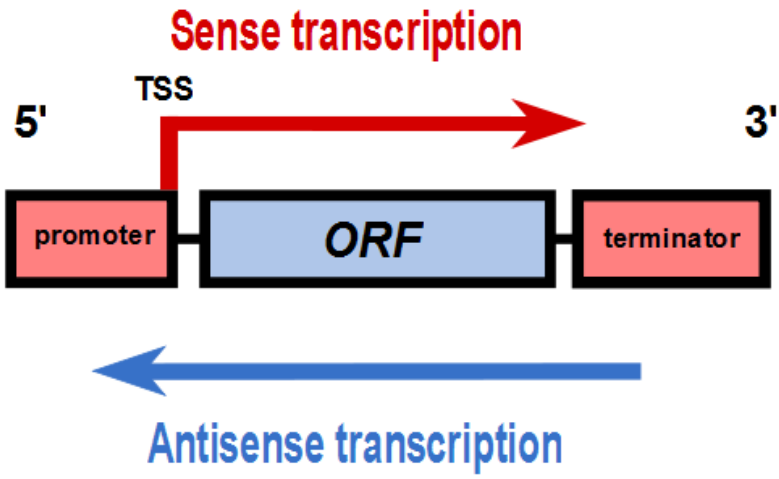


Figure 2: A definition of antisense transcription and the antisense transcript.

A schematic demonstrating how antisense transcription is described in this work. The transcription unit of a coding sense transcript is flanked by a promoter at the sense transcript's 5' end (the TSS) and a terminator at its 3' end (the TTS). Throughout this work, the 5' and 3' end of a gene refers to the end at which the 5' or 3' end of the sense transcript is found. An antisense transcript is one which (at least partially) overlaps the sense transcript on the non-coding strand of the gene. The TSS of the antisense transcript is left undefined in this schematic, though as discussed below, it does show a preference for a site upstream of the TTS of the sense transcript.

1.6.1. Measuring antisense transcription in yeast

Due to the inability of early genome-wide approaches to distinguish transcription on one strand from transcription on another, it was at first difficult to identify antisense transcription on a genome-wide scale. Indeed, in determining levels of gene expression via such microarrays, antisense transcription is still sometimes assumed to be negligible, the level of sense transcription being described by the total level of transcription on *both* strands. Even with the advent of strand-specific methods, it was found that in generating cDNA libraries for tiling arrays, spurious synthesis of second strand DNAs occurred during the reverse transcription reaction, resulting in an over-estimation of antisense transcription (Perocchi et al. 2007). However, by treating samples with actinomycin D, Perocchi et al., (2007) were able to reduce such artefacts, and were still able to observe around 1,500 different antisense transcripts. It was later shown that such non-coding transcripts are generally clustered in the vicinity of coding genes, rather than distributed randomly across the genome – thus promoting the idea that antisense transcripts are associated with a corresponding sense transcript (Neil et al. 2009; Xu et al. 2009). These studies suggested that approximately a third of genes in the yeast genome have an associated antisense transcript, but as discussed later, this fraction may be much higher.

1.6.2. Measuring antisense transcription in mammals

Antisense transcripts have also been identified in mammals, with evidence that they are a widespread genomic phenomenon Kapranov et al., (2007) found that ~1.1% of the human genome comprises short transcripts found predominantly at the 5' and 3' ends of annotated protein coding genes. Further to this, He et al., (2008) developed a modification of the serial analysis of gene expression (SAGE) technique in which RNA is first treated with bisulphite (changing cytosine residues to uridine), allowing for

unambiguous assignment of RNA to a specific DNA strand, and so enabling them to identify antisense transcripts in human cells genome-wide. Looking across five different cell types, they found between 2900 and 6400 genes that possess antisense transcripts, representing as much as 11% of the human transcriptome. What was more, these genes with antisense transcripts changed between cell types, suggesting that antisense transcripts can be regulated. In support of this, antisense transcript tags were distributed non-randomly. In keeping with their being non-coding, fewer than 1% of antisense transcripts showed evidence of splicing. Intriguingly, while the gene promoters generally showed only very low levels of sense transcription within them (as one would expect), levels of antisense transcription over the gene promoter were high, suggesting that antisense transcribing polymerases are not prevented from entering the gene promoter, and raising the possibility that the arriving polymerase could affect promoter behaviour.

1.6.3. Sense and antisense transcript pairs

Xu et al., (2011) described the distribution of antisense transcripts in yeast with respect to their associated sense transcripts, using a previous genome-wide, strand-specific tilling array (Xu et al. 2009; Neil et al. 2009). The mode distance between the antisense TTS and the sense TSS was 98bp while the distance between the antisense TSS and the sense TTS was 77bp, as shown in Figure 3, with the two overlapping transcripts skewed with respect to one another, such that the antisense transcription unit verges upon the gene promoter. By combining tiling arrays with 3' SAGE analysis – in order to accurately identify 3' ends (Xu et al. 2009; Neil et al. 2009) – it was shown that antisense transcripts generally initiate in the vicinity of the 3' end of their associated gene, often within an NDR and adjacent to the 5' end of a downstream, tandemly oriented gene, as discussed in greater detail below (Xu et al. 2009; Neil et al. 2009; Xu et al. 2011). That

antisense transcripts generally terminate within the gene promoter mimics what is observed in humans (He et al., 2008), suggesting they might have similar roles.

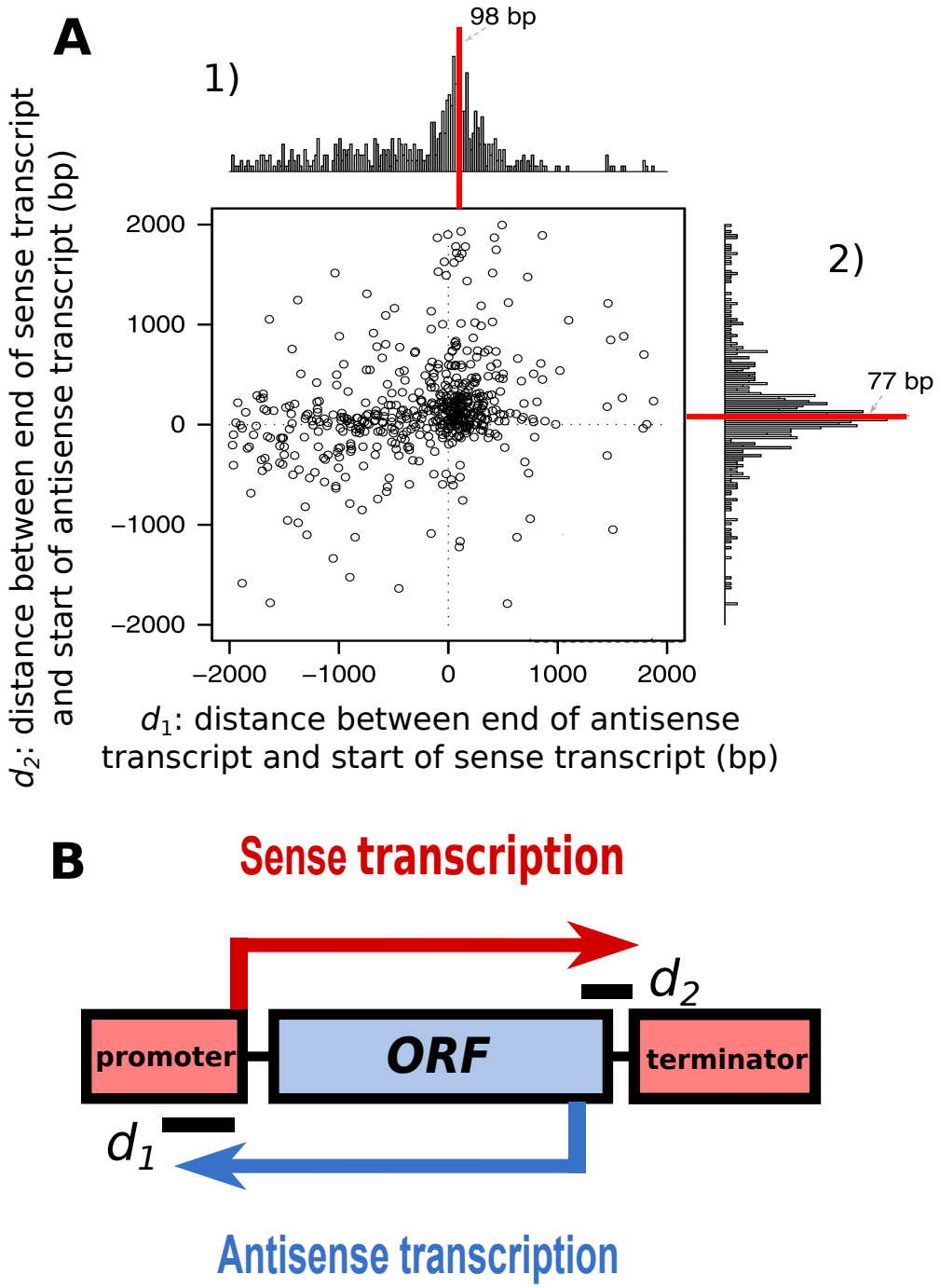


Figure 3: The distribution of sense and antisense transcript start and ends.

(A) A scatter plot with two accompanying historgrams, each showing the distribution of distances between 1) antisense transcript ends and sense transcript start sites (***d*₁**), and 2) sense transcript ends and antisense transcript start sites (***d*₂**), adapted from (Xu et al. 2011). Red lines show the mode distance. **(B)** A model of sense and antisense transcription based on these distributions, with black rectangles indicating the sites of measurement of ***d*₁** and ***d*₂.**

1.6.4. Antisense transcripts can be subdivided on the basis of their degradation pathway

Antisense transcripts appear to be particularly subject to degradation. A number of studies have found, at least for a subset of antisense transcripts, that transcript levels are increased following disruption of RNA exonuclease activities, meaning that they are often undetectable in wild-type strains. Wyers et al., (2005) identified a number of regions, thought to be transcriptionally silent, in which RNAPII-dependent transcription occurred. Despite being both capped and polyadenylated, the resultant transcripts were unstable in wild-type cells, however their levels accumulated following mutation of Rrp6, a component of the nuclear exosome complex. Termination of such cryptic unstable transcripts (CUTs) was later shown to rely on the Nrd1-Nab3 pathway (Arigo et al. 2006; Thiebaut et al. 2006), which is thought to recruit the TRAMP complex, resulting in polyadenylation of the CUT and its subsequent degradation by the nuclear exosome complex (Vanácová et al. 2005; LaCava et al. 2005; Wyers et al. 2005). CUTs appear to be degraded immediately following their synthesis (Kuehner and Brow 2008). That these transcripts should be so quickly degraded raises the question of why synthesis should happen at all, and whether it is simply an aberrant, unwanted process (as discussed below), or whether, in fact, it is the act of transcription itself that performs the biological function, rendering persistence of the transcript unnecessary.

1,496 CUTs were identified genome-wide using an *RRP6* deletion strain as part of a wider study into the yeast transcriptome (Xu et al. 2009; Neil et al. 2009), comprising 13% of the total transcripts. Additionally, these same studies identified another class of non-coding transcript that were readily detectable without the need for *RRP6* deletion. These stable unnanotated transcripts (SUTs) comprised 12% of the total transcripts, and, like the CUTs, were almost entirely oriented antisense to their associated sense mRNA (>95%). Thus antisense transcripts show variable rates of

degradation, in a similar manner to sense transcripts (Miller et al. 2011), with a major subset undergoing rapid degradation following synthesis. Whether these different classes of antisense transcript have different functional consequences has yet to be determined. A further class of antisense transcripts have been shown to be degraded cytoplasmically (Thompson and Parker 2007). van Dijk et al., (2011) used RNA sequencing (RNA-seq) in a yeast strain in which the cytoplasmic exonuclease Xrn1 (Kem1) was deleted, and so identified over 1,600 transcripts whose levels were increased. Strikingly, over 75% of the previously identified SUTs were in fact Xrn1-sensitive unstable transcripts (XUTs). Thus antisense transcripts are more sensitive to degradation than previously thought, and certainly more extensively degraded than sense transcripts. Studies in mammals have shown similar trends – the depletion of exosome subunits in human cells results in the accumulation of short, polyadenylated gene-associated transcripts (Preker et al. 2008). This instability has broad implications. Measuring the true extent of antisense transcription – both its extent and abundance – will be crucial to understanding its function; however, attempts to measure this through steady state transcript levels alone might result in substantial underestimation, thus necessitating alternative approaches to measure actual *nascent* transcription itself, as shall be discussed in chapter 4.

1.7. Proposed functions of antisense transcription

The existence and prevalence of antisense transcripts and transcription is well established, with a large number of such transcripts identified in yeast, plants and mammals (amongst others). What is less clear, however, is their physiological function. A number of studies have ascribed roles to antisense transcripts, generally on a gene-by-gene basis. Here the available evidence for some of these proposed functions is discussed.

1.7.1. Antisense transcripts have no function

One possible explanation for antisense transcription is that they are neither a desired process nor a functional one. That antisense transcripts are so rapidly degraded could support this. Indeed it has been proposed that pervasive transcription, and by extension antisense transcription, might be the consequence not of regulated, directed transcriptional processes, but rather a kind of widespread biological error, an inevitable by-product of active transcription that serves no biological purpose (Struhl 2007). Here, the author defines infidelity or 'promiscuity' in a biological process as "the frequency of an 'incorrect' event as compared with the 'correct' event". Based on available data, they hypothesise that the majority of RNAPII molecules in yeast are engaged in the production of 'junk' RNA. Because RNAPII elongation is linked with histone depletion (Schwabish and Struhl 2004; Kristjuhan and Svejstrup 2004), high levels of transcription will increase DNA accessibility and so the frequency of 'incorrect' initiation within genes. This does not explain why antisense transcripts still tend to have well defined TSSs (Xu et al. 2009; Neil et al. 2009), however it is possible that certain hotspots of initiation that are prone to transcription initiation exist within genes – *cryptic* promoters – from which such apparently aberrant or spurious transcripts initiate. However, unless these cryptic promoter sequences perform some useful function it is hard to imagine why they might have been retained in the genome. Furthermore, one might expect 'functionless' promoters to direct transcription in the sense direction as frequently as they do in the antisense direction, whereas SUTs and CUTs are transcribed almost exclusively in the antisense direction (Xu et al. 2009; Neil et al. 2009). Ultimately, that so many polymerases are not engaged in mRNA production does not serve as proof that antisense transcription is innocuous, but rather would seem to suggest the opposite. Indeed, it has recently been shown that the expression levels of antisense transcripts are conserved across different mammalian species (Ling et al.

2013), and that regions of the yeast genome that are antisense-transcribed are more highly conserved than those that are not (Goodman et al. 2013), suggesting that antisense transcription has itself been selected for during evolution.

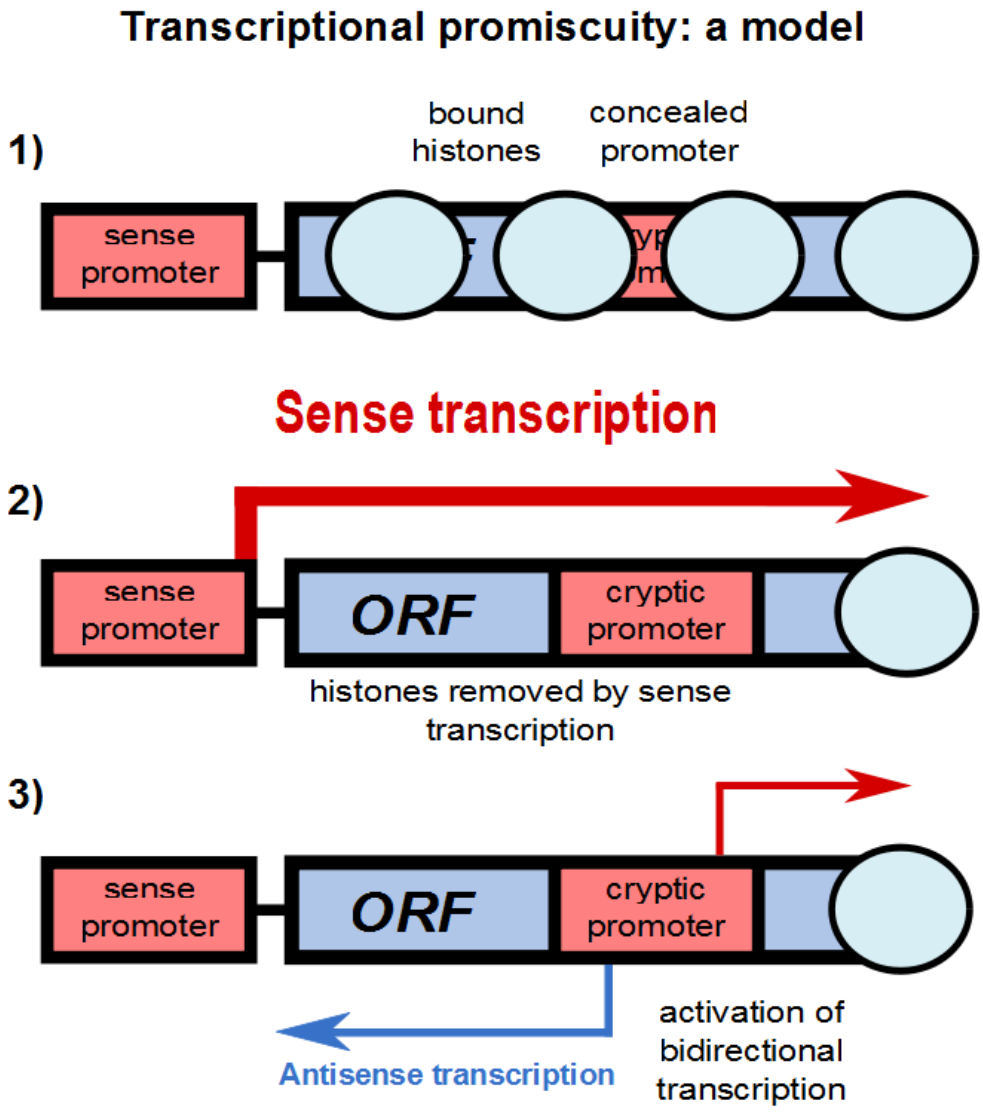


Figure 4: A model of transcriptional promiscuity.

A schematic demonstrating how pervasive, non-coding transcription might result as a by-product of high levels of coding sense transcription, as propounded by Struhl (2007). Here, a gene is bound across its ORF by histones, which conceal a sequence capable of binding polymerase and directing transcription in both directions. High levels of sense transcription remove or redistribute these histones, unveiling the cryptic promoter and resulting in cryptic transcription in both directions.

1.7.2. Antisense transcription as a repressor of sense transcription

A number of studies have suggested that antisense transcripts – or antisense transcription –might act to repress sense transcription. Houseley et al., (2008) found evidence of a long non-coding RNA that initiates at the 3' end of the *GAL10* gene, running antisense into its promoter region, where it continues along the sense strand of the adjacent *GAL1* gene. Mutation of a Reb1 binding site at the 3' end of *GAL10* was sufficient to reduce this antisense transcription. Antisense transcription was found to slow the rate of gene induction – ablation of antisense transcription led to higher mRNA levels earlier in the induction time course. Thus antisense transcription appears to *enhance* repression of *GAL10*. Intriguingly, antisense transcription was evident at *GAL10* irrespective of conditions. By comparison of hybridization intensity of the antisense transcript to *in vitro* transcribed RNA, the authors estimated that the transcript was present in approximately one in every 14 cells, not just under glucose but also galactose conditions – when *GAL10* sense transcription is active. Thus the antisense transcription does not continually limit sense transcription, but rather appears to modulate the change in gene behaviour between conditions. Furthermore, by utilising a heterozygous mutant in which antisense transcription was ablated at only one copy of the *GAL10* gene, Houseley et al., (2008) found that the induction of this gene was not repressed. This suggests that the antisense transcript cannot operate in *trans,* and that it may indeed be the act of transcription conferring downstream effects on gene behaviour, not the transcript. They went onto show that antisense transcription can modulate the chromatin at *GAL10*, suggesting a possible mechanism by which antisense transcription could act – perhaps continually – to modulate not the transcript levels at steady state but rather the transition from one state to the next.

Xu et al., (2011) were also able to demonstrate a role of antisense transcription in gene repression. They investigated the *SER7* gene, which sits adjacent to the *GAL80*

gene promoter in a tandem fashion, and which has a SUT that initiates from within the vicinity of the *GAL80* promoter (Figure 5). Deletion of a binding site for Gal4 – a transcriptional activator present under galactose conditions – resulted in reduction of the SUT levels and an increase in *SER7* transcript levels, again suggesting that the antisense transcript is repressing gene transcription. This also presents an instance whereby antisense transcription can be environmentally regulated – in this case by Gal4 – and so suggests that it could be tuned in response to changing conditions in order to alter gene expression patterns.

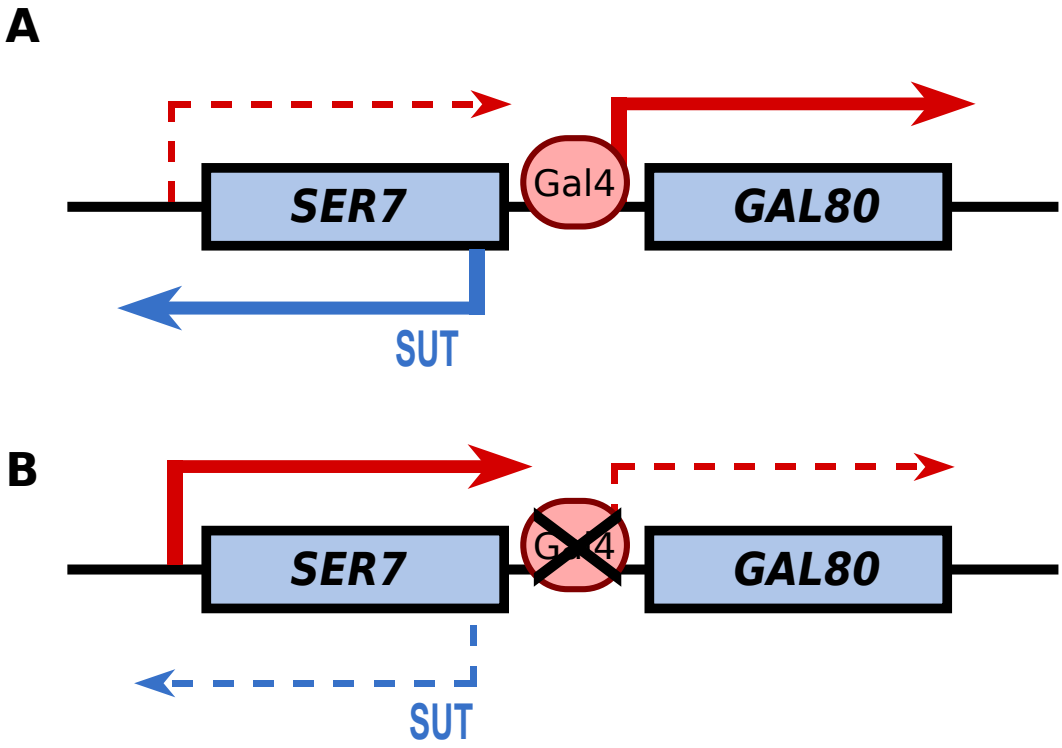


Figure 5: Regulation of *SER7* transcription by antisense transcription.

(A) The *SER7* gene sits in a tandem arrangement with *GAL80,* where it is transcribed at low levels under galactose conditions. A SUT is transcribed from its 3' end. Under galactose conditions, the Gal4 transcription activator binds to the *GAL80* promoter and promotes *GAL80* sense transcription. **(B)** Following deletion of the Gal4 binding site, *GAL80* transcription is reduced, as is transcription of the SUT. Coincident with this is an *increase* in transcription of the *SER7* sense transcript, suggesting that the SUT is acting normally to repress sense transcription.

Some studies suggest that antisense transcription might behave in *trans* (through the transcript itself) as well as or instead of in *cis* (through the act of transcription). Camblong et al., (2007) identified two CUTs at the *PHO84* gene, which as described above were stabilised upon deletion of *RRP6*. Stabilisation in this instance resulted in repression of the *PHO84* gene, together with recruitment of the histone deacetylase Hda1 and consequent repression. Antisense transcripts were still less abundant than sense transcripts following deletion of *RRP6*, with a correspondingly low level of RNAPII detectable at the 3' end. They propose that the stabilisation of antisense transcript recruits Hda1, though they do not propose a possible mechanism. Interestingly, the exosome is known to bind to chromatin (Andrulis et al. 2002; Hieronymus et al. 2004), and indeed Camblong et al., (2007) found evidence that Rrp6 itself binds to *PHO84*. Current work in our lab suggests that Rrp6 has a role in transcription regulation independent of its role in degradation (Mellor lab, unpublished), thus it is possible that by deleting Rrp6 they are simply relieving Rrp6-mediated repression. It is still possible that antisense transcription recruits Rrp6 to the *PHO84* promoter, and indeed, ablating the antisense transcripts by insertion of a *HIS5* cassette lead to the relief of *PHO84* repression even following the deletion of Rrp6. Thus by virtue of their instability, antisense transcripts may direct transcription regulators to the promoter. A specific mutation at Rrp6 led to a decrease in *PHO84* repression, a stabilisation of antisense transcription *and* a decrease in binding of Rrp6 to *PHO84*, suggesting a *cis* acting mechanism for antisense transcription mediated silencing. Later work showed that *RRP6* deletion resulted in repression due to enhanced antisense transcription *elongation*, suggesting that it is antisense transcription into the *PHO84* promoter that mediates repression, and not the accumulation of transcript (Castelnuovo et al. 2013). However, a further study at *PHO84* did identify a *trans* role for antisense transcription at *PHO84* (Camblong et al. 2009). Here, insertion of an additional *PHO84* gene on a plasmid

resulted in repression of both copies of the gene. By varying antisense transcription via the *GAL* promoter they showed that this repression was dependent upon the antisense transcript, demonstrating a *trans* effect. Unlike the *cis* effect, this repression was not altered by deletion of histone deacetylases, suggesting two distinct mechanisms. Analysis of three additional antisense transcribed genes (*VTC3*, *YJR129c* and *GYP5*) showed that although the showed evidence of repression in an *RRP6* delete, transforming strains with additional copies of the same genes did not lead to repression, suggesting that repression is mediated exclusively via *cis*-acting mechanism in these genes (Camblong et al. 2009). Another study of a XUT at the Ty1 retrotransposon, however also found evidence of a *trans* acting mechanism (Berretta et al. 2008). Here, deletion of the exonuclease Xrn1 resulted in stabilization of the XUT and reduction of Ty1 sense transcription, with expression of the XUT from a plasmid again sufficient to promote repression. This mechanism was also dependent upon histone deacetylation, thus antisense transcription regulation could be generally dependent on changes in histone modification. In support of this, Pinskaya et al., (2009) found that repression of the *GAL1* gene by cryptic transcription relied upon the Rpd3 histone deacetylase, which was tethered to the *GAL1* promoter by H3K4me2/3 upon repression.

Hongay et al., (2006) identified a regulatory role for antisense transcription with important physiological implications. Antisense transcription was shown to regulate entry into the meiotic cell cycle; the process in which a single diploid cell produces four haploid cells. Specifically, the *IME4* gene, required for entry into meiosis, was found to be either antisense transcribed or sense transcribed but not both; diploid cells produced a sense transcript while haploid cells produced a polyadenylated antisense transcript that terminated within the *IME4* promoter. Haploid mutants that showed increased sense transcription and reduced antisense transcription showed several diploid-like characteristics, such as the meiosis-specific transcript *IME2*. Thus the effects of

antisense transcription can have crucial biological consequences, arguing against the non-functional role described above.

1.7.3. Antisense transcription as an activator of sense transcription

As discussed above, it has been suggested that sense transcription might promote antisense transcription through the removal of histones (Struhl 2007), thus it is reasonable to hypothesise that the reverse might also be true; that antisense transcription might promote sense transcription. Uhler et al., (2007) found evidence for an activatory role for antisense transcription at the *PHO5* gene. Upon induction, *PHO5* loses four nucleosomes from its promoter (Almer et al. 1986), resulting in the exposure of the necessary TFBSs for activation. Intriguingly, mutation of the TATA box at the *PHO5* did *not* affect nucleosome eviction upon induction, suggesting that sense transcription initiation does not play a role in its direction (Uhler et al., 2007). However, by impeding RNAPII elongation (through a variety of approaches including mutation of the Rpb2 subunit of RNAPII) they found that nucleosome eviction *was* diminished. Thus, although *sense* transcription does not seem to be necessary for eviction, another form of transcription does. A CUT was identified, initiating in the vicinity of the *PHO5* 3' end, and terminating upstream of the *PHO5* promoter. Deletion of the 3' end was sufficient to abrogate this antisense transcript, including in an *RRP6* deletion strain, indicating that antisense transcription itself had been abrogated. Abrogation resulted in a slower nucleosome eviction and recruitment of RNAPII to the *PHO4* promoter. Thus, even though antisense-transcribing RNAPII enters the promoter, it does not interfere with sense transcription but rather modulates promoter chromatin creating a more suitable platform for initiation and facilitating the change from one cell state to another. A role in such environmental responsiveness has been proposed genome-wide by (Xu et al. 2011), who determined that antisense transcripts are significantly enriched at stress

response genes – i.e. genes for which a timely response to changing conditions is imperative.

1.8. Clues to the mechanism of antisense transcription initiation

The mechanism of pervasive transcription, and antisense transcription more generally, is poorly understood, as is the extent to which it mimics that of coding sense transcription. For example, if non-coding, pervasive transcription is the result of uncoordinated, promiscuous polymerase activity, does it still necessitate the ordered assembly of PIC components that precedes normal initiation? Indications as to how antisense transcription is initiated can be gleaned from genome-wide analyses of antisense transcript positions as well as maps of the transcription machinery. Xu et al., (2009) found that the majority of the 847 SUTs and 925 CUTs they identified originated from NDRs associated with the promoters of other transcripts, or else from the NDRs at the 3' ends of genes. Thus antisense transcripts would appear to require a nucleosome depleted platform for polymerase assembly and subsequent initiation. That so many of these NDRs are positioned at the TSSs of genes led to these promoters being termed 'bidirectional', i.e. capable of initiating transcription in either direction. It should be noted, however, that because of the compact nature of the yeast genome, the 5' end of a gene is often adjacent to the 3' end of another gene, in instances where they are arranged in a tandem fashion. Thus an antisense transcript arising from an NDR at the 5' end of one gene could also be said to be arising from an NDR at the 3' end of another. Nevertheless, it has been proposed that antisense transcription is a consequence of gene promoters being inherently bidirectional – perhaps a consequence of the polymerase promiscuity described above (Struhl 2007), and indeed Core et al., (2008) found that 77% of active genes in human cells had divergently transcribing RNAPII bound at their

promoters, while Seila et al., (2008) identified short (often divergent) transcripts upstream of over 50% of genes in the mouse genome. What remains unclear is whether or not these bidirectional promoters are single promoters – in which a single PIC can form but from which transcription in both directions can occur – or whether they are in fact two distinct promoters which can direct sense and antisense transcription independently. Neil et al., (2009) found at the *TPI1*, *GPM1* and *FBA1* genes, all of which had divergent CUTs initiating upstream of their TSSs, that upon switching from glycerol to glucose the levels of the antisense transcript went up while mRNA levels went down, indicative of inverse regulation, while Yadon et al., (2010) found that mutation of the chromatin remodelling enzyme *ISW2* resulted in increased antisense transcription from bidirectional promoters, suggesting it is something that must be actively repressed. These results would seem to support a model whereby transcription from a single PIC can be 'funnelled' into either sense or antisense transcription, but not both, perhaps because formation of one transcription complex precludes formation of the other. The different possible models of bidirectional transcription are show in Figure 6.

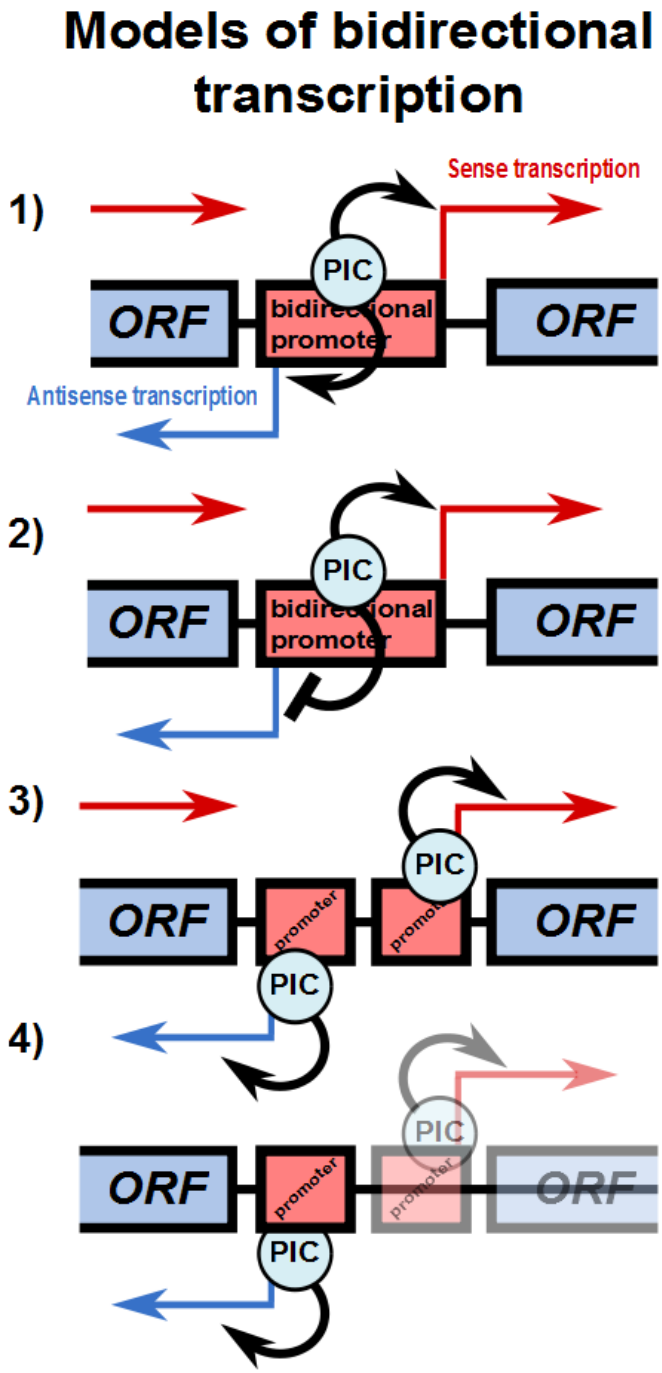


Figure 6: Models of bidirectional transcription.

A schematic showing possible mechanisms of how antisense transcription might arise from bidirectional promoters. **1)** Bidirectional promoters involve either a single PIC or a single shared transcription complex that initiates antisense transcription at a rate proportional to the rate of sense transcription initiation – i.e. the two are co-regulated. **2)** As 1, except that sense transcription precludes antisense transcription, and vice versa, such that the two are inversely regulated. **3)** Bidirectional promoters are in fact two separate promoters, independently regulated from one another and able to recruit their own PICs. **4)** Removal of the sense promoter in 3 has no effect on the transcription of antisense.

Like protein coding genes, antisense transcripts also appear to be transcribed by RNAPII. Berretta et al., (2008) showed that antisense transcription at the Ty1 retrotransposon was reduced by thermosensitive mutations of RNAPII components (RNAPI and III-dependent transcripts were unaffected), while Churchman and Weissman (2011) demonstrated using a combination of an RNAPII pull-down and RNA-seq that antisense-reads were prevalent genome-wide and corresponded to previously identified SUTs and CUTs (Neil et al. 2009; Xu et al. 2009). This would seem to suggest that the process of antisense transcription is similar to sense transcription. Consequently, one would expect to see PIC components at the sites of initiation, and indeed, Mavrich et al., (2008) found TFIIB at the 3' ends of genes, even after they removed from consideration any genes whose 3' ends bordered the 5' ends of other genes. That antisense transcription might be initiated in the same fashion as sense transcription raises the possibility that it might also be regulated similarly, and indeed there is evidence that environmentally responsive transcription factors might regulate antisense transcription. Cawley et al., (2004) determined the binding sites of transcription factors cMyc, Sp1, and p53 in human chromosomes 21 and 22, and found that 36% of the sites occurred adjacent to the 3' end of annotated genes, whereas only 22% were found at the 5' end. As reported above, deletion of Gal4 and Reb1 binding sites at antisense TSSs in yeast resulted in a loss of transcription (Houseley et al. 2008; Xu et al. 2011). Furthermore, there is evidence of histone modifications involved in the regulation of sense transcription also being involved in antisense transcription. H3K4me3 was identified near the 3' end of *GAL10* when the gene was repressed, at the site of initiation of the antisense transcript (Houseley et al. 2008). Disruption of the antisense transcript *also* reduced H3K4me3. As discussed in Chapter 5, H3K4me3 is a marker of active genes, thus it is possible that antisense transcription has similar associations with histone modifications compared with sense transcription, while also

being directed by transcription factors in the same manner. That transcription factors should be directing antisense transcription in this way argues against the non-functional role for antisense transcription described above.

Another important consideration with regards to antisense transcription is whether it can occur contemporaneously with sense transcription at the same gene. By this it is not necessarily meant that both polymerases pass across the gene contemporaneously – something which could lead to polymerase collisions and a block to transcription (Hobson et al. 2012) – but rather that both promoters are simultaneously active. Could the two processes prevent one another's occurrence, or else inhibit the opposing promoter, whether through chromatin modulation, or indeed collision? Hongay et al., (2006) found that antisense transcription did not occur at the same time as sense transcription within the same cells, however, given that antisense transcription might have repressive effects upon sense transcription they may be mutually exclusive due to regulatory concerns rather than because it is physically impossible for both promoters to be simultaneously active. Indeed, genome-wide studies show instances where sense and antisense transcription can occur across the same transcription unit (Xu et al. 2011; Churchman and Weissman 2011), while other studies using RNA fluorescence *in situ* hybridisation (RNA FISH) have demonstrated that nascent *PHO84* sense and antisense transcripts can be found simultaneously within the nucleus of the same cells, though this is only observed rarely (Castelnuovo et al., 2013; less than 5%). Mathematical models describing transcriptional interference (where transcription in one direction physically occludes transcription in the other) have shown that strong interference requires particularly strong convergent transcription (Sneppen et al. 2005), thus it is possible that sense and antisense transcription only begin to occlude one another at high levels.

If antisense transcription is directed in a similar fashion to sense transcription, then one would expect to find PIC components, transcription factors and chromatin features at the site of antisense transcription initiation. Furthermore, one would expect the levels of these features to reflect the level of antisense transcription at that site, as is the case for sense transcription. If antisense transcripts tend to initiate from the 3' ends of genes, and/or from the bidirectional promoters of adjacent downstream genes, then one would expect to see such features at these locations. The implication of this is that *terminators* themselves would contain these features, and would therefore be behaving not only as terminators of sense transcription but also as promoters of antisense transcription (or perhaps as bidirectional promoters of antisense transcription and downstream sense transcription). If antisense transcription were pronounced genome-wide, then one would expect correspondingly pronounced levels of PIC components and transcription factors etc. It is hypothesised that genes with antisense transcripts will possess an inherent functional symmetry, with promoters at either end of the open reading frame, a sense promoter and an antisense promoter, and that terminators themselves might be capable of acting as promoters (Figure 7). These hypotheses will be tested in chapter 3, where it shall first be established whether terminators show promoter-like attributes, followed by an experimental validation to assess whether terminators are sufficient to direct antisense transcription, and thus behave as promoters.

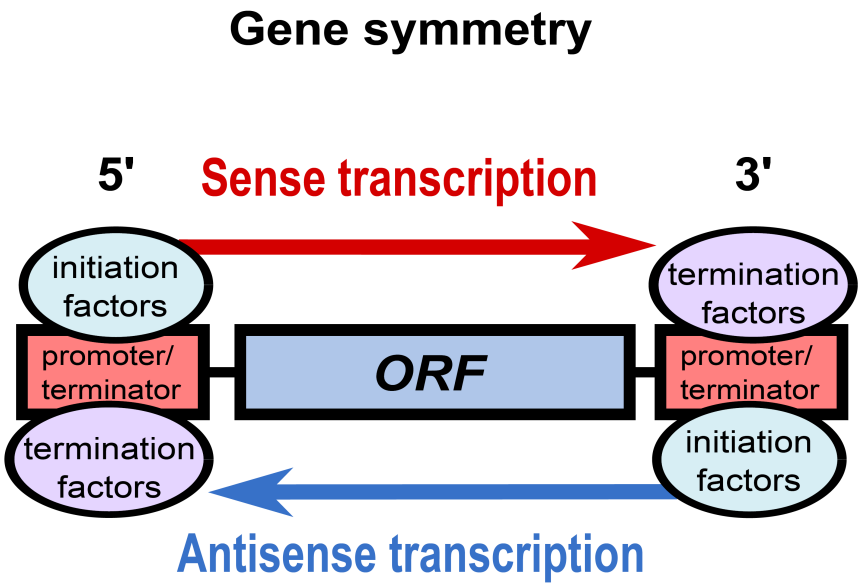


Figure 7: The gene symmetry hypothesis

A schematic demonstrating how the presence of antisense transcripts in yeast suggests there is an inherent symmetry to genes. Antisense transcripts generally start and end transcription within the vicinity of their corresponding sense transcripts end and start respectively, as shown. If antisense transcription initiation and termination are dependent upon similar mechanisms to sense transcription, then one would expect to find factors involved in both processes at *both* ends of the gene. If antisense transcription occurs at only very low levels genome-wide, then this symmetry would be very slight. However the more abundant antisense transcription is genome-wide, the more pronounced this symmetry would become. If the mechanisms of antisense transcription initiation and/or termination are identical to their sense counterparts, then it would also be expected that the canonical sense promoter and terminator would themselves be symmetrical, i.e. would share sequence elements.

1.9. Aims

The overall aim of this work is to investigate the nature of antisense transcription in the budding yeast *Saccharomyces cerevisiae*; to further understanding of this emerging phenomenon by addressing questions as to its origins and function. Firstly, the initiation of antisense transcription is explored; both in terms of where it occurs and to what extent it is regulated. Additionally, the relationship between sense and antisense transcription is itself assessed – to explore how the latter might affect the former – by

determining how genes with antisense transcription differ from those without. A detailed description of these aims is described below and in Figure 8.

1. To explore gene symmetry on a genome-wide level by assessing levels of canonical promoter features at the 3' ends of genes.
2. To assess whether antisense transcription shows evidence of regulation, and whether antisense transcription can be directed and regulated in the absence of adjacent, bidirectional transcription.
3. To explore how genes differ at their 5' sense promoters in the presence of varying levels of antisense transcription, and how it might alter gene behaviour.
4. To assess whether sense and antisense transcription are correlated genome-wide, to test whether antisense transcripts function generally to repress genes.
5. To build a hypothesis based upon these observations, describing a genome-wide role for antisense transcription, which can then be tested experimentally.

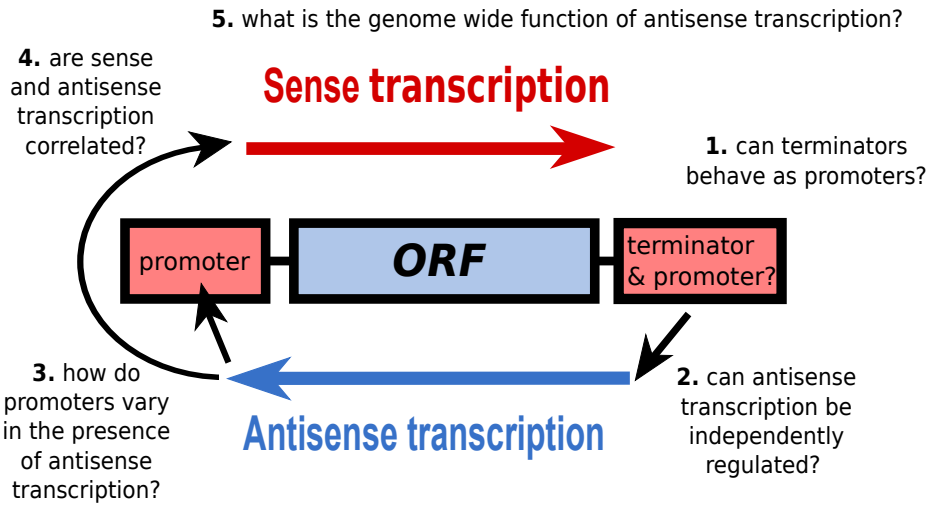


Figure 8: Thesis aims

A schematic outlining the aims of this work, numbered as described in the text. Black arrows represent possible relationships.

Chapter Two

Materials and Methods

2. Materials and Methods

The computational and experimental methods used in this work will be described on a chapter-by-chapter basis. Where mentioned, named persons other than the author carried out procedures. Otherwise, procedures were carried out by the author.

2.1. List of published datasets used in this work

Below is a list of the published datasets used in this work, together with a reference to the paper in which they were published.

Dataset	Source
Transcript map of wild-type yeast grown to exponential phase, YPD medium, determined by RNA-seq. 4534 unique transcripts.	Nagalakshmi et al., (2008)
Transcript map of wild-type and *rrp6*Δ yeast strains, grown to exponential phase in YPD, determined by tiling array. 7272 unique transcripts.	Neil et al., (2009); Xu et al., (2009)
Transcript map of wild-type yeast strains arrested at G1, determined by tiling array. 657 unique transcripts.	Granovskaia et al., (2010)
Transcript map of wild-type yeast grown to exponential phase in YPD, determined by RNA-seq. 8778 unique transcripts.	Yassour et al., (2010)
Transcript map of wild-type yeast grown to exponential phase in YPD, determined by RNA-seq. 6160 unique transcripts.	Yassour et al., (2009)
Genome-wide levels of transcription, determined using NET-seq in order to assess *nascent* transcription specifically. Yeast grown to exponential phase in YPD. Briefly, NET-seq involves immunoprecipitating tagged RNAPII, isolating the nascent RNA, and subjecting this to RNA-seq. Hence this approach could also be said to assess genome-wide levels of elongating RNAPII.	Churchman and Weissman, (2011)
Genome-wide levels of transcription, determined using RNA-seq. Yeast grown to exponential phase in YPD.	Churchman and Weissman, (2011)
A genome-wide map of binding sites for 124 different transcription factors, determined using a combined approach that first identifies binding sites based on sequence, and then determines whether a given site is a 'true' binding site based on available ChiP data, and by assessing the degree of conservation of the site between four different yeast species.	MacIsaac et al., (2006)
Transcriptional plasticity data for 4,486 different yeast genes. Plasticity is a measure of the variation of the expression of a given gene across a range of different conditions. Here, values were obtained using over 1500 different gene expression arrays.	Tirosh and Barkai (2008)

Estimations of gene expression noise for over 2,500 yeast genes, determined using high-throughput flow cytometry in conjunction with a library of GFP-tagged yeast strains.	Newman et al. (2006)
Lists of TATA-box containing and TATA-less genes. Determined by assessing which gene promoters contained the TATA-box consensus TATA(A/T)A(A/T)(A/G).	Basehoar et al. (2004)
Histone turnover data. Briefly, yeast were engineed carrying two differently tagged H3 proteins – a Myc-tagged H3 under a constitutive promoter, and a Flag-tagged H3 under the inducible *GAL* promoter. The ratio of Flag/Myc was determined at several time points following induction to estimate rate of turnover. Results were obtained in wild-type and *set2*Δ strains.	Dion et al., (2007); Venkatesh and Workman (2012)
All subsequent data collected using wild-type yeast grown to exponential phase in YPD	
Genome wide levels of TBP, determined by ChIP	Venters and Pugh, (2009)
Genome wide levels of TFIIB, determined by ChIP	Venters and Pugh, (2009)
Genome wide levels of Swr1, determined by ChIP	Venters and Pugh, (2009)
Genome wide levels of Rsc9, determined by ChIP	Venters and Pugh, (2009)
Genome wide levels of H3K36me3, H3K79me3, H3R2me and H3R2me2a, determined by ChIP.	Kirmizis et al. (2009)
Genome wide levels of H3K4me1, H3K4me2 and H3K4me3, determined by ChIP.	Kirmizis et al., (2007)
Genome wide levels of H3K4me3, determined by ChIP.	Guillemette et al., (2011)
Genome wide levels of H3K56ac, determined by ChIP.	Xu et al., (2007)
Genome wide levels of H3K9ac and H3K14ac, determined by ChIP.	Pokholok et al., (2005)
Genome wide levels of H3K79me2 and H2B normalised levels of H2BK123 ubiquitination, determined by ChIP.	Schulze et al., (2011)
Genome wide levels of RNAPII CTD Ser2 phosphorylation, determined by ChIP.	Kim et al., (2010)
Nucleosome occupancy levels, determined using a method that does not utilise MNase, but rather involves the immunoprecipitation of histone H3.	Fan et al., (2010)
Nucleosome occupancy levels determined by MNase treatment of genomic DNA, followed by deep sequencing.	Kaplan et al. (2009)

Table 1: Published datasets used in this work

2.2. Chapter Three Methods

2.2.1. Classification of genes

Throughout this work, 'genes' refers specifically to protein coding genes with non-dubious ORFs. ORF start and end sites were obtained from the Saccharomyces Genome Database (SGD; http://www.yeastgenome.org/). The transcription start and transcript termination sites (TSS and TTS respectively), as determined by Nagalakshmi et al., (2008), were then mapped to the ORFs. Throughout chapter 3, reference is made to the '3' region' of a gene, which is hypothesised to behave as both a canonical terminator and also a promoter of antisense transcription. The 3' region of a gene was defined as being delimited by a point 100bp upstream of a gene's translation stop codon and by the border of the nearest downstream ORF (or 600bp downstream of the stop codon otherwise). For the purposes of the analysis it was important to define genes on the basis of whether their 3' regions bordered the 5' ends of any downstream, protein coding sense transcripts. Thus any genes whose stop codons were not within 1000bp of the 5' end of a protein coding sense transcript were defined in this way. Genes were also defined on the basis of whether their 3' regions were adjacent to any 'modestly' transcribed genes. A gene was defined as not having a modestly transcribed gene in the vicinity of its 3' region if there was no other ORF border (start or stop codon) within 1000bp of the gene's TSS, belonging to a gene with a transcription value above 80 reads (See below for an explanation of how a gene's transcription value was calculated). Genes labelled as dubious were considered for the purposes of whether a gene's 3' region abutted any other genes, but were otherwise excluded from all analyses. The full set of genes used in this analysis is available as a supplementary table in Murray et al., (2012).

2.2.2. Collated transcript map

Transcript coordinates were obtained from Nagalakshmi et al., (2008), Neil et al., (2009); Xu et al., (2009), Yassour et al., (2009), Granovskaia et al., (2010) and Yassour et

al., (2010), and mapped to the yeast genome. The aim was to combine multiple sources in order to obtain a more comprehensive map of both sense and antisense transcripts. Combining the maps from the five sources gave a total list of 27,401 transcripts, however it was expected many of the transcripts between the different sources would in fact be the same species. Transcripts were defined here as being the same between two different sources if both their start *and* end sites fell within 75 bp of one another and they were of the same orientation. To decide which of the two transcripts would be discarded from such a pairing, priorities were assigned to the five different sources. In order of descending priority, this was: Granovskaia et al., (2010), Neil et al., (2009)/Xu et al., (2009), Yassour et al., (2010), Yassour et al., (2009) and Nagalakshmi et al., (2008). By removing transcripts deemed to be the same, the list was reduced from 27,401 down to 19,877 transcripts. Of the transcript maps, 100% of the transcripts from Granovskaia et al., (2010) was retained (657 transcripts), 98.5% of the transcripts from Neil et al., (2009)/Xu et al., (2009) were retained (7163 transcripts out of a total of 7272), 50.5% of the transcripts from Yassour et al., (2010) were retained (4420 transcripts out of a total of 8778), 99.9% of the transcripts from Yassour et al., (2009) were retained (6156 transcripts out of a total of 6160) and 32.4% of the transcripts from Nagalakshmi et al., (2008) were retained (1471 transcripts out of a total of 4534). Antisense transcripts were defined as those transcripts whose TSS lay within the 3' region of a gene (as defined above) and which ran antisense to that gene. Such genes were in turn defined as those possessing an antisense transcript. The collated list of transcripts is available as a supplementary table in Murray et al., (2012).

2.2.3. Transcriptional plasticity data

Transcriptional plasticity data was obtained from Tirosh and Barkai (2008), who in turn estimated these values using a range of microarray expression data conducted under

different experimental conditions from Ihmels et al., (2002). Briefly, they estimated plasticity as the sum of squared deviation scores of the log_2-transformed expression ratios. Genes were ranked in terms of transcriptional plasticity in order to obtain the high and low transcriptional plasticity genes.

2.2.4. Genome-wide ChIP data

Genome-wide levels of TBP (Spt15), TFIIB (Sua7), Swr1 and Rsc9 were obtained from Venters and Pugh, (2009). Normalised nucleosome occupancy values were obtained from Kaplan et al., (2009) and Fan et al., (2010) to obtain the MNase-dependent and independent occupancy respectively. All genome-wide ChIP data was obtained under approximately similar growth conditions. Average levels around either the TSS or the TTS of a particular gene set were calculated by aligning the genes by either the TSS (for the 5' region) or the TSS (for the 3' region) and averaging the occupancy at each base. The resultant average occupancy was then smoothed with a 41bp window. When aligning the protein-coding genes by their TSS or TTS ends the mapped ends determined by Nagalakshmi et al., (2008) were used. When aligning the antisense transcripts by their TSS the annotated ends from the collated transcript map described above were used. To determine the levels of TBP and TFIIB in the 3' region as a percentage of those in the 5' region the difference between the maximum and minimum average level in the 3' region was divided by the difference between the maximum and minimum average level in the 5' region.

2.2.5. Calculating AT richness

The AT richness for a specific base was defined as the average frequency of A or T nucleotides within a 21bp window centred on that base. The average AT richness around the TTS was determined by aligning all the genes in a given set by the TTS of

their protein-coding transcript and averaging the AT richness at each base. The resultant average AT richness was then smoothed with a 41bp window. For comparative purposes, displayed AT richness data has been inverted, to highlight the similarity between AT richness and nucleosome occupancy as determined using an MNase-dependent assay.

2.2.6. Nascent transcription levels

Genome-wide levels of elongating RNAPII (nascent transcription) were obtained from Churchman and Weissman (2011). To quantify the level of transcription of a given transcript a similar method was used to that described previously (Churchman and Weissman, 2011). In summary, the sum of the read density in a sliding 500bp window was calculated, starting 50bp upstream of the transcript start site and ending 700bp downstream of the start site, or else the end of the transcript if it was shorter than 700bp. The transcription level was defined as the maximum value obtained in this way. This sliding window method was performed due to concerns that annotated TSSs might be inaccurate (Churchman and Weissman, 2011). To determine the level of transcription of a gene's antisense strand for those genes without an antisense transcript the TTS of its protein-coding sense transcript was used in place of the TSS. For the purposes of determining if a gene was modestly transcribed, for reasons described above, it was necessary to determine the transcription level of all genes, even those without mapped TSSs (including dubious ORFs). For these genes a point 100bp upstream of the start codon was used to calculate the transcription level. To obtain the average RNAPII level across the antisense strand – for the purposes of comparing antisense transcription between genes with defined antisense transcripts and those without – all genes were aligned by the TTS of their protein-coding sense transcript, as determined by Nagalakshmi et al., (2008), and calculated the average number of reads at each

nucleotide relative to the anchor point. It was also necessary to determine the transcription levels of the antisense transcripts themselves – distinct from determining the level of RNAPII on the antisense strand of coding genes. The transcription level in the 5' regions of the antisense transcripts was determined using the same sliding window approach as the sense transcripts above. Additionally, it was necessary to determine the average RNAPII density upstream of the antisense TSS and on the opposite strand, in order to assess divergent sense transcription. The highly transcribed group of antisense transcripts were defined as the top 20% of the data, based upon the transcription values obtained above. Before defining the lowly transcribed group the bottom 10% of the data was masked from consideration, to remove potential false positive antisense transcripts for which one might expect no PIC formation to be present. The lowly transcribed group was then defined as the 20% above this. To determine antisense transcripts with low levels of divergent sense transcription the sum of the sense NET-seq reads was determined in an 800bp window placed 50bp upstream of the antisense TSS. From this it was determined which antisense transcripts had low divergent sense transcription (the bottom 35% of the data). From this group, antisense transcripts were selected for that were not adjacent to the 5' end of any non-RNAPII transcribed elements, to ensure an absence of divergent transcription. This group included rRNAs, tRNAs, snRNAs, snoRNAs, retrotransposons and transposable elements. Their coordinates were obtained from the SGD browser. The highly and lowly transcribed groups were selected as described above.

2.2.7. Transcription factor binding sites

The transcription factor binding site map was downloaded from the Fraenkel lab's website (http://fraenkel.mit.edu/improved_map/; MacIsaac et al., 2006). A binding site was defined as a 'true' site based upon the confidence of the ChIP-chip data ($p < 0.005$).

Genes were defined as having 3' TFBSs if one or more sites were present within its 3' region, as defined above. Distributions of TFBSs around the 3' end were determined using the same approach applied by Lin et al., (2010) to the 5' end. The frequency at a given bp relative to the TTS was determined by determining the number of genes which have a TFBS at that site by the total number of genes. If a particular bp position was not assessed in the ChIP-chip analysis by MacIsaac et al., (2006) at a given gene then it was not considered towards this total number.

2.2.8. Analysing bidirectional promoters

Classes of bidirectional promoters were selected as described in the Results section of Chapter 3. The antisense transcripts considered were those defined above as part of the collated map.

2.2.9. Error bars and p-values

Where present, error bars represent the standard deviation determined from 1000 bootstrap pseudoreplicates of the data. p-values were determined using the Wilcoxon rank sum test, by comparing the distributions of data at the maximum points, i.e. when comparing the average TBP level in genes with or without antisense transcripts. This was performed for all genomic features of interest, except for nucleosome occupancy, in which case the minimum points were compared.

2.2.10. Construction of yeast strains and experimental methods

(Performed by Ana Serra Barros)

Experiments were done in yeast strain BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0). DNA for yeast transformation was generated by PCR using primers with sequences homologous to the regions flanking the insertion site. The ADH1 and TEF terminator

regions ($ADH1_T$ and TEF_T) used here are the sequences present in vectors used to construct yeast gene deletions (Longtine et al. 1998), and as such were inserted initially as part of a selective KanMX marker. $ADH1_T$ was present upstream of the marker while TEF_T was present downstream of it. The $ADH1_T$:TEF_P:KanMX:TEF_T cassette was inserted at +757 in GAL1 with respect to the ATG, while the $ADH1_T$:TEF_P:HIS:TEF_T cassette was inserted into the 3' region of GAL10, resulting in loss of 554 residues between +1453 and +2007 with respect to the ATG. This also resulted in the removal of a bi-directional promoter responsible for the GAL10 antisense transcripts produced in glucose, initiating around +1611 (Houseley et al. 2008).

The Cre-lox recombination system (Güldener et al. 1996) was then used to remove undesired sequence elements after transformation, utilising loxP sites flanking the marker sequence. The production of *ADH_T*.loxP via a Leu^+ intermediate (*$ADH1_T$*.loxP-*LEU2*-loxP) was achieved through transformation with his^+ pSH62; leaky expression of the cre recombinase behind a *GAL1* promoter drove excision of the leu^+ marker. Kanamycin sensitive, leucine auxotrophs were identified by replica plating and sequenced. The complete removal of the *kanMX6* cassette – leaving the *$ADH1_T$* sequence followed by a 106 bp crelox recombination signature – did not affect the length or abundance of sense or antisense transcripts in the *$ADH1_T$*.loxP strain (data not shown). Excision of *$ADH1_T$* together with the *TEF_P* and KanMX protein coding sequences left just the *TEF* terminator downstream of a loxP site.

Yeast cultures were incubated at 30°C oscillating with a 2 cm radius at 200 rpm. Overnight cultures were grown to stationary phase in rich media (YPD) consisting of 1% bactopeptone, 1% Difco yeast extract and 2% glucose. Cells were grown to exponential phase in conical flasks by inoculating 50 ml YPD with overnight cultures to an optical density (OD) of 0.2. *GAL* genes were induced by incubation of exponential cells (OD of 0.4-0.6) with 50 ml (YPG/ 1% bactopeptone, 1% Difco yeast extract and 2%

Galactose) for the time indicated after two washes with YP. Total RNA was obtained by hot phenol extraction and normalised to 1 µg/µl using a nanodrop spectrophotometer and stored at -20°C. RNA samples (15 µg) were separated by electrophoresis on a formaldehyde 1% agarose gel containing ethidium bromide at 100 V for 3 h (unless stated otherwise). rRNA loading controls were photographed under ultra-violet light using a 0.5 s exposure.

2.2.11. Preparing strand-specific probes

(Performed by Ana Serra Barros)

Asymmetric PCR on PCR amplified dsDNA fragments was used to create ssDNA radiolabelled probes to allow detection of the sense or antisense transcripts at *GAL1* or *GAL10.* The 50µl reaction contained 50ng of template DNA, 2µl of a nested primer at 50µM (50X molar excess), 10µl betaine (5M), 5µl 10X taq buffer, 4µl dGAT (2.5mM each), 1.5µl dCTP (0.1mM) 1µl Taq Pol (5U/µl), H_2O to 47.5µl, 2.5µl [α-^{32}P]dCTP 370MBq/ml, and subjected to 25-30 cycles in the PCR machine, 50µl of T.E. added and cleaned using a Sephadex G-50 spin column. 1 x 10^6 cpm/ml is added to the northern blot membrane after pre-hybridization without heat denaturation.

2.3. Chapter Four Methods

2.3.1. Determining sense and antisense ratios

RNA-seq and NET-seq data was obtained from Churchman and Weissman (2011), in which yeast strains were grown in YPD at 30°C with shaking from an initial optical density (OD) of 0.1 to mid-log phase with an OD of 0.6–0.8. The average number of reads per base was determined within the windows described in chapter 4, for both sense and antisense reads, and from both NET-seq and RNA-seq experiments. Ratios were obtained by dividing average antisense reads by average sense reads. Gene promoters

were considered as being part of a bidirectional promoter directing divergent coding transcription if there was the TSS of a divergently oriented, non-dubious ORF within 600bp of its own TSS. Gene promoters were considered as being part of a bidirectional promoter directing sense and antisense transcription if there was no TSS of a divergently oriented, ORF (dubious or non-dubious) within 1000bp of its own TSS. p-values were obtained by comparing the distribution of ratios for a given window between NET-seq and RNA-seq antisense/sense ratios.

2.3.2. Determining sense and antisense transcription correlations

Spearman rank correlation coefficients were obtained by comparing sense and antisense NET-seq reads in the windows described in chapter 4.

2.4. Chapter Five Methods

2.4.1. Classification of genes

Transcription start (TSS) and end sites were annotated as described above. If no site was available, a hypothetical TSS was assigned using a genome-wide consensus. Genes defined as dubious or which overlapped with other genes were removed from all analyses. The remaining genes (5,183) were divided into five classes, each comprising approximately 20% of the total, based on levels determined previously by NET-seq (Churchman and Weissman, 2011), in a 300bp window downstream of the sense TSS. Genes were defined as TATA-box containing or TATA-less using the classifications by (Basehoar et al. 2004).

2.4.2. Genome-wide ChIP data

Genome-wide levels of Spt15 (TBP) and Sua7 (TFIIB) were obtained from Venters and Pugh (2009). Levels of H3K36me3, H3K79me3, H3R2me and H3R2me2a were obtained

from (Kirmizis et al. 2009). Levels of H3K4me1, H3K4me2 and H3K4me3 were obtained from Kirmizis et al., (2007). Levels of H3K4ac were obtained from Guillemette et al., (2011). Levels of H3K56ac were obtained from Xu et al., (2007). Levels of H3K9ac and H3K14ac were obtained from Pokholok et al., (2005). Levels of H3K79me2 and H2B normalised levels of H2BK123 ubiquitination were obtained from Schulze et al., (2011). Levels of Htz1 were obtained from Guillemette et al., (2005). RNAPII CTD Ser2 phosphorylation data was obtained from Kim et al., (2010). All H3 modifications were normalised to H3 levels before subsequent analysis. All genome-wide ChIP data was obtained under approximately similar growth conditions. To assess the correlation of a given feature with sense and, separately, antisense transcription, the level of the feature inside a 10bp window slid across the TSS from -1000 to +1500 was determined for the 5183 genes described in the main text, and the Spearman's correlation coefficient determined for each window step.

2.4.3. Transcription machinery at promoters

Levels of 202 different transcription-related proteins at gene promoters were determined using a comprehensive study (Venters et al. 2011). p-values were determined using the Wilcoxon rank-sum test and comparing the distribution of binding levels between genes with high and low levels of antisense transcription.

2.4.4. Nucleosome occupancy and NFR sizes

Nucleosome occupancy levels were obtained from Kaplan et al., (2009) and coordinates determined as described previously by Zaugg and Luscombe (2012). Briefly, nucleosome boundary coordinates were determined using the GeneTrack software (Albert et al. 2008). The NDR of a gene was defined as the distance between the 3' coordinate of the upstream nucleosome (-1) and the 5' coordinate of the downstream

(+1) nucleosome. NDRs of more than 500bp were excluded from consideration due to previous concerns that they are the result of missing data (Zaugg and Luscombe 2012). The distance between the -1 and +1 nucleosome was defined as the NDR size for a given gene determined using these coordinates, and the median size for a given group calculated.

2.4.5. H3K4 methylation distributions

To assess the distribution of H3K4 methylation across genes the gene body was divided into 30 equally sized bins. The level of H3K4 methylation (me, me2 and me3) in each bin was then determined as a fraction of the total level across the entire gene body.

2.4.6. Histone turnover data

Estimated levels of H3 turnover genome-wide were obtained from Dion et al., (2007), and Venkatesh and Workman (2012). At a given gene region (-1 to +4 nucleosomes), the H3 turnover rate was determined as the maximum rate of all the probes overlapping that region. To correlate histone turnover with histone modification genome-wide, the average level of a given modification was determined within every probe for which turnover estimates were available. A Spearman's correlation coefficient was then determined by comparing levels of modification with histone turnover.

2.4.7. Genome-wide correlations with nascent transcription

To determine the level of association between numerous histone modifications and both sense and antisense-transcribing RNAPII, the yeast genome was first divided into overlapping 50bp windows. For those windows that fell within a single transcription unit (i.e. that did not lie outside transcription units or within a region overlapped by two transcription units with reverse orientation), the level of nascent sense and antisense

transcription was determined, as was the level of the histone modifications. A Spearman rank correlation coefficient was then determined for each of these modifications by correlating levels of the modification with the level of nascent sense and, separately, antisense transcription in the same window.

2.4.8. Statistical tests

All p-values were calculated using the Wilcoxon rank sum test. For figures displaying the average levels of a given factor or modification around the TSS, p-values shown are those calculated by comparing the distributions at the maximum point in the vicinity of the TSS for the highest and the lowest antisense-transcribed groups, or at the minimum points for graphs of nucleosome occupancy, H3K36me3, H3K79me2, H3K79me3 and H2BK123ub.

2.5. Chapter Six Methods

2.5.1. Genome-wide expression noise data

Genes were assigned expression noise levels from Newman et al., (2006). Genes were defined as being TATA-less or TATA-box containing based on the classification by (Basehoar et al. 2004).

2.5.2. Strain construction

(Performed by Ana Serra Barros, Tania Nguyen and Simon Hänni)

The *GAL1*:*ADH1*$_T$ strain used was that described above for chapter 3. The *HMS2:ADH1*$_T$ insert strain was devised similarly, with a 10bp region at 651-660 relative to the ATG being replaced with *ADH1*$_T$ attached to a kanMX6 expression cassette. For subsequent experiments, *GAL1* wild-type and derivative strains were grown for two hours in YPG, as

described in the methods for chapter 3, unless otherwise stated. *HMS2* wild-type and derivative strains were grown for two hours in YPD, unless otherwise stated.

2.5.3. RNA FISH

(Performed by Françoise Howe, Tania Nguyen and Simon Hänni)

RNA FISH was carried out as described previously (Zenklusen and Singer 2010). For *GAL1, GAL10, HMS2* and all derivatives, four DNA probes of approximately 50 nucleotides in length and 50% in G C content were designed to be complementary to the 5' region. This would allow for detection of nascent transcripts as they emerged from the elongating RNAPII complex. Probes contained five amino-allyl dT modified bases, which were coupled to fluorescent cyanine dyes. Probes detecting *GAL10* antisense transcripts were complementary to the 3' end of *GAL10*.

Cells were grown as described in the methods for chapter 3. For *GAL1* and *GAL10*, cells were grown in YPD and induced in YPG for 2 hours after washing. For *HMS2* cells were grown in YPD. Briefly, cells were spheroplasted by digestion with a lyticase enzyme following centrifugation of cell pellets, which allows access of the probes. Spheroplasts were then immobilised on glass coverslips. Hybridization was performed using with 2 ng of probe mix in a hybridization chamber. Image acquisition was performed using a DeltaVision CORE: Wide-field fluorescence deconvolution imaging microscope. 31 "z-stacks" were obtained using a 100x objective lens, in order to obtain 3D images. Images were captured using a CoolSNAP HQ camera (Photometrics, Tucson, AZ).

2.5.4. Image analysis

Image analysis was performed in ImageJ (Schneider et al. 2012). Analysis was performed on a composite image of DAPI and probe signals, comprised from the

maximum intensities obtained from z-stacks 12-22. To define a threshold by which to determine whether a given cell was expressing or not, a knockout strain was used (an *HMS2* knockout for the *HMS2* experiments and a *GAL1*/*GAL10* double knockout for the *GAL1* and *GAL10* experiments). This threshold was determined by obtaining the average maximum intensity within cells of the knockout strains, plus two standard deviations. Cell boundaries were defined using custom macros written in ImageJ. Nuclear dots were determined manually, based on the presence of a bright dot within the DAPI stained nucleus. To calculate the number of nascent transcripts within a cell, a similar method was used to that described by Zenklusen et al., (2008). The maximum intensity of a given nuclear dot was divided by the average maximum intensity of cytoplasmic dots – thought to represent single cytoplasmic transcripts – and the result rounded to the nearest whole number to yield the number of nascent transcripts. The distributions were pooled from multiple biological repeats – two for the ada2/sgf73 experiments and three or more for the remainder.

2.5.5. Mapping of transcript initiation and termination sites

(Performed by Simon Hänni)

Initiation and termination sites for both sense and antisense transcripts at *GAL1* and the *GAL1* constructs were mapped following 15 and 90 minutes induction with galactose. Mapping was carried out as described previously (Grange 2008).

2.5.6. Chromatin immunoprecipitation

(Performed by Françoise Howe and Karolina Chocian)

ChIP was performed as described in (Morillon et al., 2005; Pinskaya et al., 2009).

Chapter Three

Gene symmetry: antisense transcription and 3’ promoters

3. Gene symmetry: antisense transcription and 3' promoters

3.1. Summary

As discussed in chapter 1, the widespread occurrence of antisense transcripts initiating from the 3' end of genes suggests that terminators of genes may also behave as promoters – i.e. that there may be an inherent functional symmetry to genes. Antisense transcription could therefore represent an additional canonical feature of a gene, tuned by regulating transcription initiation at the *de facto* terminator. An alternative to this is that antisense transcription is a consequence of unregulated transcription initiation events occurring at downstream promoters, a consequence of the polymerase 'promiscuity' discussed in chapter 1. In this chapter the idea of gene symmetry is explored, as is the possibility that antisense transcription might be an independently regulated process in its own right. Genome-wide data sets are used to assess whether genes with antisense transcripts are enriched for traditional promoter elements and components at their terminators, whether these "3' antisense promoters" show evidence of regulation, and whether they are in turn regulating the canonical sense promoter. It is also established whether these observations hold for genes whose terminators do not abut the promoters of downstream genes, to address questions regarding polymerase promiscuity, and the regulatory roles for antisense transcripts are assessed by asking if they increase transcriptional plasticity.

3.2. Introduction

The existence of antisense transcripts suggests an inherent symmetry to gene structure, in which promoters are present at either end. Here, the gene symmetry hypothesis is explored by assessing whether terminators generally behave as promoters, employing a genome-wide strategy to assess the levels of promoter features at the 3' regions of genes. To this end, the nature of promoters and 'promoter features' are discussed below.

3.2.1. Promoters direct transcription initiation

Generally, a promoter is a sequence upstream of a transcription unit that is necessary to direct transcription of that unit. One might say that a *gene* promoter is strictly one that directs transcription of the coding transcript of the gene. In antisense transcribed genes, however, both strands of a gene are subject to transcription, suggesting the existence of a promoter at the 3' end. For now, this hypothetical sequence is termed the 3' *antisense* promoter, to discriminate it from the canonical gene promoter, which is termed the 5' *sense* promoter. Note that other non-coding RNAs such as tRNAs, rRNAs and snoRNAs have their own promoters, and also that promoters can be found within genes – internal promoters, which may in fact be synonymous with the promoters of antisense transcription (Kim et al. 2012). As discussed in chapter 1, it has been proposed that antisense transcription is a consequence of polymerase promiscuity, a low level biological error. Here this idea is explored by establishing whether antisense transcription is associated with 3' promoter structures and regulatory features comparable to those at the 5' end, which would argue against such a model. Antisense transcription is performed by RNAPII (Berretta et al. 2008; Churchman and Weissman 2011), and so here the consideration is limited to factors involved in its regulation, and to the features typically found at the promoters of RNAPII-transcribed units.

3.2.2. Promoters are marked by distinct sequence elements.

The core promoter is the region of the promoter to which RNAPII binds, together with the basal transcription machinery. Core promoters are far from invariant, and indeed are highly diverse in both sequence and functionality (Butler and Kadonaga 2002). Nevertheless, certain sequence elements have been identified that, though not ubiquitous to all promoters, play important roles in recruitment of the transcription machinery and the definition of the transcription unit. The first such element of the core promoter identified was the TATA-box, the binding site of the TATA-binding protein (TBP), a highly conserved transcription factor necessary for transcription initiation by all three eukaryotic RNA polymerases (Greenblatt 1991; Cormack and Struhl 1992). The TATA-box was discovered through the comparison of intergenic sequences in *Drosophila melanogaster*, mammalian, and viral protein-coding genes (Breathnach and Chambon 1981), in which the vast majority of genes had the sequence TATAAA 25-30bp upstream of their transcription start site. In yeast, the TATA-box is much less tightly confined to a distance between 40-120bp upstream of the TSS (Smale and Kadonaga 2003). Recently, it has been shown that the sequence of the core promoter is itself closely tied to the strength of the promoter – i.e. the transcriptional activity it confers upon the downstream gene (Lubliner et al. 2013). The TATA-box has several identified sequence variants to which TBP binds with differing affinity and support varying levels of transcription (Chen and Struhl 1988; Yean and Gralla 1997; Singer et al. 1990). These findings suggest that the core promoter defines a set-point level of transcription that is independent of environmental conditions i.e. that in the absence of regulation, individual genes still have different activities.

Though higher eukaryotes possess other core elements such as the Inr and the DPE (Juven-Gershon and Kadonaga 2010), these elements are not found in yeast. Another study has identified the TA richness surrounding the TSS as being a feature of

strong core promoters, with AT rich region upstream and an A rich region downstream (Maicas and Friesen 1990). Whether the TATA-box is present at all promoters or not has been the subject of several additional studies. By comparing promoter sequences and combining this with conservation studies in four different yeast species, the consensus motif TATA(A/T)A(A/T)(A/G) was identified, and found to be present at only 19% of genes (Basehoar et al. 2004). This definition has been shown to be predictive of a number of other gene features. So called TATA-containing genes are more likely to be bound by the SAGA-complex than by TFIID (Basehoar et al. 2004), are more extensively bound by gene-specific transcription factors (Lin et al. 2010), have a higher level of gene expression noise, and are more transcriptionally plastic, a feature of genes which shall be discussed below (Tirosh and Barkai 2008). Nevertheless, using a more relaxed definition of the TATA-box, in which genes were allowed to deviate from the consensus by up to two nucleotides, almost all yeast promoters have been shown to contain the element (Rhee and Pugh 2011), and to be capable of binding to TBP, albeit more weakly. TATA-boxes and TBP therefore represent important promoter features, and below it is assessed whether they are present at the 3' ends of genes. Additionally, the levels of the PIC component TFIIB, which is required for RNAPII recruitment to promoters, and NDRs upon which the PIC forms are assessed (see chapter 1).

3.2.3. TFBS represent important elements of the proximal promoter

As discussed in chapter 1, transcription factors are involved in regulating genes in response to changing environmental conditions, binding to their cognate binding sites (TFBSs) and modulating transcription levels (Jacob and Monod 1961). Identifying the cognate sites of TFBS has involved techniques such as DNA footprinting (Hampshire et al. 2007), protein binding microarrays (Newburger and Bulyk 2009) or else the application of motif-finding algorithms to the sequences of coregulated promoters

(Neuwald et al. 1995). Attempts to identify TFBSs genome-wide can involve simply matching the consensus site to the available genome sequence – using scoring matrices and an arbitrary threshold to define positions (Stormo 2000). However, these approaches are thought to give an overwhelming number of false-positives – non-functional binding sites, often present within the coding regions themselves. Attempts to overcome this have involved the use of phylogenetic analysis (Kellis et al. 2003), determining which of these TFBSs are evolutionarily conserved, and allowing for the identification of e.g. the Gal4 binding sites in the *GAL1-10* promoter. Such approaches have been reinforced by the additional use of genome-wide ChIP approaches, assessing which of the computationally identified sites also show evidence of physical binding *and* evolutionary conservation (Harbison et al. 2004). This map was recently used to show that genes with more TFBSs at their promoters show evidence of being more extensively regulated across changing environmental conditions, as one would expect (Lin et al. 2010). The same map is used here to assess whether antisense transcripts might also be regulated.

3.2.4. Is bidirectional transcription a prerequisite for antisense transcription initiation and regulation?

Bidirectional promoters were discussed in more detail in chapter 1. Briefly, it has been suggested that antisense transcripts initiate from the promoters of downstream genes, in the vicinity of the transcription start site of the downstream gene (Neil et al. 2009). This has given rise to the idea that a single promoter, termed a bidirectional promoter, can direct transcription of both transcripts. What is unclear is whether antisense transcription is an inevitable consequence of transcription being intrinsically bidirectional, a possible consequence of transcriptional infidelity and which could perhaps explain why many antisense transcripts are rapidly degraded following

transcription (Perocchi et al. 2007; Wyers et al. 2005), or whether antisense transcription can be regulated independently of divergent sense transcription, i.e. whether their promoters, though in close proximity, can respond to distinct regulatory inputs. What also remains unclear is whether bidirectional promoters consist of a single PIC, from which RNAPII transcribes in both directions, or whether there are two divergently oriented PICs, as is the case in the model presented by Neil et al., (2009).

3.2.5. Transcriptional plasticity: the hallmark of highly regulated genes

Transcriptional plasticity is a measure of the variability of a gene's expression across a variety of experimental conditions, and reflects a gene's capacity to change its expression in response to varying environmental inputs (Lehner 2010). It has been calculated previously as the variation of the expression of a gene across a large set of different growth conditions (Tirosh and Barkai 2008). For example, a stress responsive gene would have a particularly high plasticity whilst a constitutively expressed housekeeping gene might have a plasticity value close to zero. If antisense transcripts are involved in the regulation of sense expression, as discussed in chapter 1, then one would expect genes with antisense transcripts to be more transcriptionally plastic than those without. Furthermore, if 3' TFBSs are required for antisense initiation and regulation, then one would expect genes with 3' TFBS to also be more plastic. Of course, if antisense transcripts are present at all genes, but only under certain environmental conditions, then one might expect antisense transcription to have no association with transcriptional plasticity – even if it can modulate levels of sense transcription. This is because an ideal measure of plasticity would be collected across all possible conditions. Therefore, if all genes could produce antisense transcripts then one gene might, on average, produce the same amount of antisense transcription as another gene across the same set of conditions, even if the amount of antisense transcription at these genes

under a given condition is different. Thus the overall effect upon gene regulation would be the same. It is possible, however, that whilst all genes might be capable of producing antisense transcripts, some may be more capable of supporting higher levels than others.

3.2.6. A hypothesis: genes are functionally and structurally symmetrical

If the initiation of antisense transcripts is driven by a similar mechanism to sense transcription then one would expect to find elements of the transcription initiation machinery within the 3' regions of genes, and indeed there is evidence to suggest that this true. Components of the PIC, including TBP and TFIIB, have been found within the 3' regions of certain genes, as have histone modifications thought to be specific to the 5' region (Mavrich et al. 2008; Venters and Pugh 2009). Furthermore, NDRs, another hallmark of promoters, have been reported in the 3' regions of genes (Mavrich et al. 2008; Kaplan et al. 2009), although the veracity of these studies is complicated by their use of the enzyme micrococcal nuclease (MNase), given that the 3' region is very AT rich and MNase has a strong preference for such sequences (Chung et al. 2010). It is predicted that genes with antisense transcripts should have a functional and structural symmetry – that mechanisms driving transcription initiation should be present at both the 5' and 3' ends, and that both these mechanisms should have the capacity to be regulated. Furthermore, it is predicted that regulation of the 3' end should in turn regulate the 5' end via antisense transcription. A further question to be assessed is whether this gene symmetry is merely the result of so many genes having the bidirectional promoter of another gene abutting their 3' end.

Here, a wide range of genome-wide data is integrated to assess questions regarding gene symmetry and the initiation of antisense transcription. A collated map of transcripts is used to identify the positions of antisense transcripts in the genome – to

determine which genes possess such transcripts – while a recent, genome-wide map of *nascent* transcription (Churchman and Weissman 2011) is used to quantify the actual level of transcription of these transcripts. This data is then used to establish how the presence or absence of antisense transcripts, or the presence of varying levels of antisense *transcription,* is associated with other promoter features at the 3' regions of genes.

It is shown that the 3' regions of genes are indeed marked by strikingly high average levels of PIC components, whose presence is associated with antisense initiation. Furthermore, high levels of antisense transcription are supported by a pronounced promoter architecture comprised of the PIC components TBP (Sua7) and TFIIB (Spt15), an NDR, chromatin remodelling factors and TFBSs, evidencing regulation. Evidence of two PICs being present at bidirectional promoters is shown, suggesting that sense and antisense transcription from bidirectional promoters may arise from two distinct transcription complexes. Most notably, there is evidence for antisense transcripts that are transcribed to high levels in the absence of even moderate levels of sense transcription, and that this, too, requires a distinct PIC within an NDR. 3' TFBSs and antisense transcription are both associated with transcriptional plasticity, again even in the absence of adjacent promoters. It is proposed that antisense transcription can be driven by PIC formation in the 3' regions of genes independently of adjacent, divergent sense transcription and that it does in turn regulate gene expression by enhancing transcription plasticity.

3.3. Results

3.3.1. TFBSs are a feature of the 3' region of genes

The question of whether TFBSs within the 3' region (3' TFBSs) are a common feature of protein-coding genes was addressed. The 3' region was defined as being delimited by a

point 100bp upstream of a gene's translation stop codon and by the border of the nearest downstream ORF (or 600bp downstream of the stop codon in no nearby ORF was present). TFBS positions were obtained using the combination of a motif-finding algorithm and a ChIP-chip analysis across all intergenic regions in the *S. cerevisiae* genome (Harbison et al. 2004; MacIsaac et al. 2006), comprising the genomic coordinates of over 120 different non-general transcription factors. Briefly, an intergenic region had to contain both a strong match to the consensus binding site of a particular transcription factor as well as evidence of the intergenic region being bound by the same factor in the accompanying ChIP-chip analysis. In order to demonstrate the robustness of our findings, two different maps of varying stringencies were used, one with over 16,000 different binding sites genome-wide and the other with over 27,000. Genes were defined as having a 3' TFBS is they had one or more binding site within this 3' region. Genes defined as dubious by the SGD browser were not considered in this analysis.

Due to the compact nature of the yeast genome many 3' regions abut the 5' regions (promoters) of other genes. Thus, many 3' TFBSs might be involved strictly in directing downstream sense transcription, or else might form part of a bidirectional promoter (Neil et al. 2009). To assess whether 3' TFBSs are present regardless of a gene's genomic context, genes lacking adjacent downstream sense transcripts in the same orientation as the gene i.e. those without evidence of adjacent protein-coding transcription were also selected. This comprised a set of 1456 genes.

The percentage of genes possessing 3' TFBSs was comparable between the total set of genes and those genes lacking initiating protein-coding sense transcripts downstream (37.7% and 32.7% respectively; Figure 9A). This observation remained as the stringency of the TFBS map was increased (data not shown), and suggests that

promoters at the 3' ends of genes may have a role in antisense transcription regardless of whether they are also involved in the production of protein-coding sense transcripts.

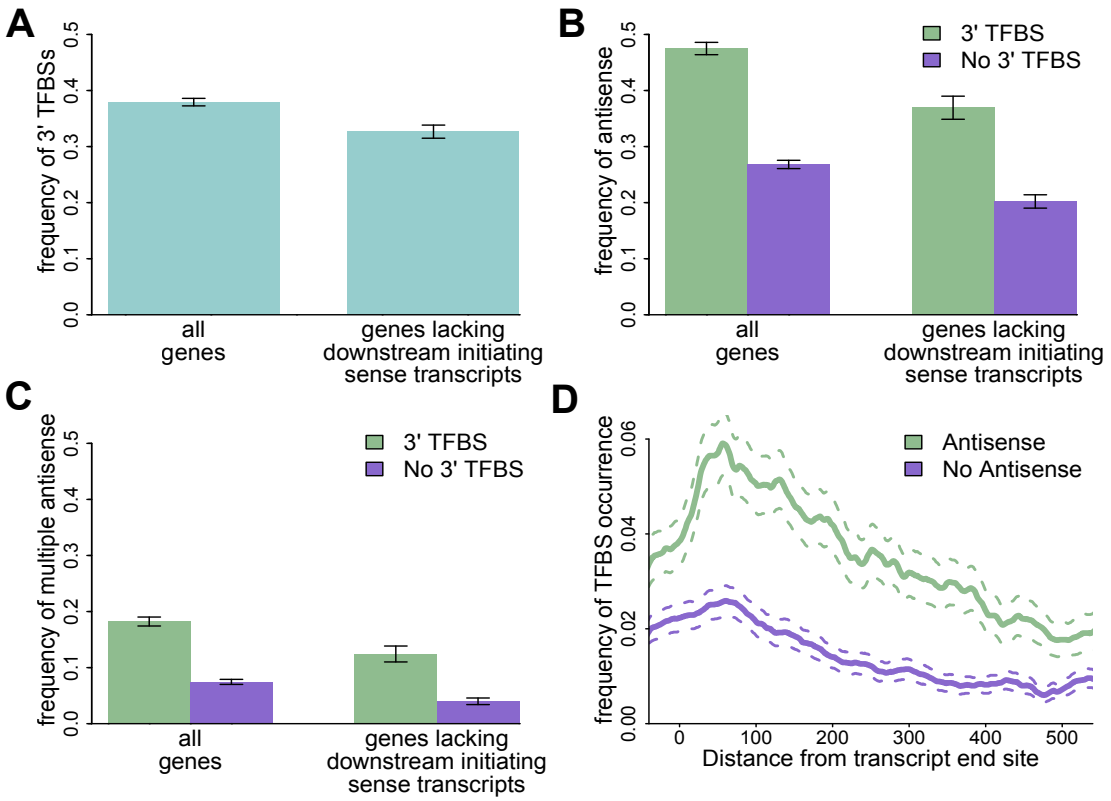


Figure 9: TFBSs at the 3' ends of genes are associated with antisense transcription

(A) The fraction of total genes, and those without downstream initiating protein-coding sense transcripts, that possess at least one TFBS within their 3' intergenic region. **(B)** The fraction of genes, both with and without 3' TFBSs, that have an antisense transcript initiating from their 3' end. **(C)** The fraction of genes, both with and without 3' TFBSs, that have two or more different antisense transcripts initiating from their 3' end. **(D)** The frequency of TFBSs downstream of the 3' end of the protein-coding transcript for genes with and without antisense transcripts initiating from their 3' end. Shown are the resultant frequencies when the total gene set was considered - similar results were observed when only those genes lacking downstream initiating protein-coding transcripts were considered. Note that the transcript end site is synonymous with the TTS described in the text. All error bars were calculated by using bootstrapping. For example, in **A**, a single bootstrap pseudoreplicates was determined from the 'all genes' group, by sampling with replacement from the group *n* times, where *n* is the number of genes in the group. This was repeated to generate 1000 distinct bootstrap pseudoreplicates, and for each of these the fraction of genes with a 3' TFBS was determined. From these 1000 different fractions the standard deviation was determined, and used as the error bar.

3.3.2. 3' TFBSs are associated with antisense transcription

Next, it was established whether the presence of 3' TFBSs is associated with antisense transcription, by assessing whether genes with 3' TFBSs more frequently possessed

antisense transcripts compared to those genes that do not. To obtain as comprehensive a map of antisense transcription as possible, a genome-wide transcript map was compiled from a range of RNA-seq and tiling array experiments performed in *S. cerevisiae* (Granovskaia et al., 2010; Nagalakshmi et al., 2008; Neil et al., 2009; Xu et al., 2009; Yassour et al., 2009, 2010). This gave a map consisting of over 20,000 protein-coding and non-coding transcripts, and allowed identification of those genes that had one or more antisense transcripts initiating in the vicinity of their 3' ends (36.6% of the total gene set; 25.7% of those genes lacking a downstream initiating protein-coding sense transcript). 36.6% (1896) of genes were defined as possessing an antisense transcript, where antisense transcript strictly refers to a non-coding transcript (ncRNA) from the compiled map defined above that initiates within the 3' region of a verified gene and runs antisense to that gene.

It was found that genes containing 3' TFBSs more frequently had 3' initiating antisense transcripts than those genes that lacked 3' TFBSs (49.1% compared with 28.2%, Figure 9B). This observation held true for both the total set of genes and for those genes lacking initiating protein-coding sense transcripts downstream (36.9% compared with 20.2%, Figure 9B), and remained as the stringency of the TFBS map was increased (data not shown). Furthermore, 3' TFBS-containing genes more frequently produced multiple different antisense transcripts (i.e. two or more antisense transcripts initiating and/or ending at distinct sites) than those that did not (20.4% compared with 8.7% for total genes; 12.4% compared with 4.0% for those genes lacking downstream initiating protein-coding sense transcripts; Figure 9C). These results implicate 3' TFBSs in the initiation of antisense transcripts and suggest that antisense transcript initiation is regulated in the same fashion as protein-coding sense transcripts, again arguing against a non-functional role for antisense transcription (Struhl 2007).

If 3' TFBSs are involved in antisense transcription, it might be expected for them to be more abundant at the 3' ends of genes that produce antisense transcripts than those that do not. To obtain the average distribution of TFBSs at the 3' region, genes were aligned by the 3' ends of their protein-coding sense transcripts (Nagalakshmi et al. 2008), and determined the average frequency of TFBS occurrence at each base relative to the 3' end, a procedure similar to that carried out previously at the 5' ends of genes (Lin et al. 2010). TFBSs were found to be much more prevalent at the 3' end of those genes that produced antisense transcripts than of those that did not, with a maximum frequency that was more than twice that of the antisense lacking genes (Figure 9D), and comparable to the average frequency of TFBSs observed previously at the 5' end (Lin et al. 2010). This finding provides further evidence for a role of transcription factors in antisense regulation. A similar enrichment was also observed for those antisense-possessing genes without downstream-initiating protein-coding sense transcripts (data not shown), suggesting that bidirectional promoters are not a prerequisite for antisense initiation.

3.3.3. The 3' regions of genes are enriched for TATA-boxes

As discussed above, the TATA-box is a sequence element that plays a role in assembling the transcription machinery at promoters. If the 3' regions of genes were to behave as promoters then one might expect them to possess TATA-boxes. To this end, two different approaches were used to identify TATA-boxes genome-wide, and assess whether they were enriched within 3' regions and how this enrichment compared to that at 5' promoters. The first approach involved using a scoring matrix obtained from the JASPAR database (Bryne et al. 2008), determined using protein binding microarrays (Newburger and Bulyk 2009). The second involved using the motif TATA(A/T)A(A/T)(A/G), determined by Basehoar et al., (2004) by identifying

conserved sequence motifs within gene promoters. Both approaches yielded similar results (Figure 10). Unexpectedly, though the 5' end was enriched compared to the ORF, the 3' end showed a much more pronounced enrichment. Assuming the identified TATA-boxes are equivalent to their promoter counterparts and are involved in directing transcription, then this provides evidence that the terminators of genes are generally equipped to direct initiation. Supporting this, the peak of this TATA-box enrichment at the 3' end occurs downstream (relative to the 3' end) of the median reported start site of antisense transcription (Xu et al. 2011). However, the peak of TATA-box enrichment is much narrower than the distribution of antisense transcript start sites reported previously (Xu et al. 2011), suggesting that the TATA-box or TATA-box-like sequences identified here may be playing some other role. The TA rich sequences may be involved in termination or 3' end formation given their close association to the 3' end (Proudfoot 2011). It is of course possible that both scenarios are true, and that the TA rich element directs termination/3' end processing of the sense transcript whilst also behaving as a TATA-box for antisense transcription. Indeed, the distance between the TATA-box and the TSS is not tightly defined (Smale and Kadonaga 2003), perhaps explaining why the antisense TSS need not be tightly defined relative to the sense TTS. It is possible that the element evolved primarily to direct downstream sense transcription events and was later subverted for the purposes of antisense transcription initiation.

It is worth noting that genes containing defined antisense transcripts are not enriched for 3' TATA-boxes compared to those that do not (Figure 10). This might suggest that antisense transcripts are not associated with TATA-boxes. However, whereas TATA-box sequences are present irrespective of growth condition, antisense transcripts are likely to vary in level across changing different conditions, particularly if they can be regulated by the presence of 3' TFBSs as suggested above. Thus all genes may in fact be capable of supporting antisense transcription, but in a regulated fashion,

such that two antisense transcript maps collected under different conditions might be very different from one another.

It is worth considering whether the 3' ends of other yeast species, such as fission yeast, might also contain TATA-boxes. The 3' ends of fission yeast genes are, like those in budding yeast, extremely AT rich (Humphrey et al. 1991), and so could also be enriched for TATA-boxes. TATA-boxes in fission yeast could direct antisense transcription, and notably, this could have consequences not observed in budding yeast. Unlike *S. cerevisiae, S. pombe* possesses components of the RNAi machinery (Harrison et al. 2009; White and Allshire, 2008). In response to double stranded RNAs, gene silencing can occur via the RNAi pathway, through the formation of the RNA-induced silencing complex (RISC; Novina and Sharp, 2004). Thus, the presence of 3' TATA boxes in fission yeast could result in antisense transcription, the formation of double stranded RNAs (formed by sense and antisense transcripts), and subsequent gene silencing. Intriguingly, by introducing three components of the human RNAi machinery into *budding* yeast, they were able to reconstitute the RNAi system (Suk et al. 2011). It would be interesting to assess whether genes with antisense transcription are more susceptible to gene silencing following the introduction of this system than those without. Such an approach could help to elucidate the possible function of 3' TATA-boxes and antisense transcription in lower eukaryotes that *do* possess the RNAi machinery.

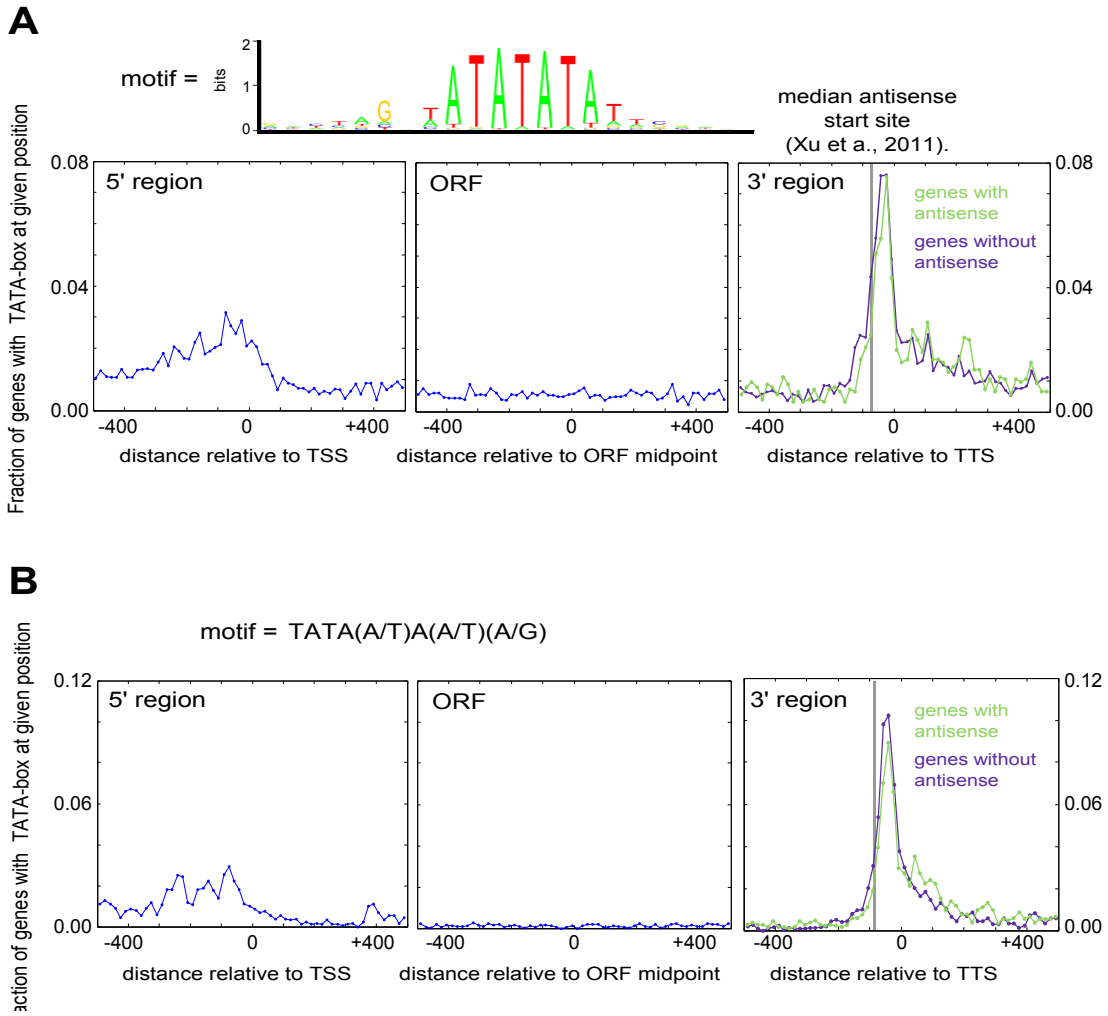


Figure 10: Sequences matching TATA-box motifs are highly prevalent at the 3' regions of genes.

(A-B) Histograms showing the fraction of genes containing sequences matching a TATA-box consensus motif, at the 5' region, ORF, and 3' region of genes. For the 3' region, genes were divided into those with defined 3' initiating antisense transcripts and those without. Genes were divided into 15bp bins. TATA-boxes were determined in two ways as follows: **(A)** Using a position-specific scoring matrix obtained using protein binding microarrays and **(B)** using a consensus motif obtained from a genome-wide study.

3.3.4. 3' end TFBSs and antisense transcripts are both associated with transcriptionally plastic genes

To explore a possible connection between antisense transcription and transcriptional plasticity, every gene was assigned a transcriptional plasticity value, using a curated data set of gene expression microarrays encompassing more than 1,500 different experimental conditions (Ihmels et al. 2002). From these groups of high and low

plasticity genes were selected (consisting of 1,209 and 1,418 genes respectively for the total genes, and 347 and 436 genes respectively for those genes without downstream initiating sense transcripts), and the frequency with which they possessed one or more 3' initiating antisense transcripts was assessed, using the same map as described above. High plasticity groups more frequently possessed one or more 3' initiating antisense transcripts than the low plasticity genes (44.9% compared with 28.6% for total genes; 36.1% compared with 15.6% for those genes lacking downstream initiating protein-coding sense transcripts; Figure 11A), and also showed a higher frequency of multiple 3' antisense transcripts (16.8% compared with 9.8% for total genes; 12.1% compared with 2.5% for those genes lacking downstream initiating protein-coding sense transcripts; Figure 11B).

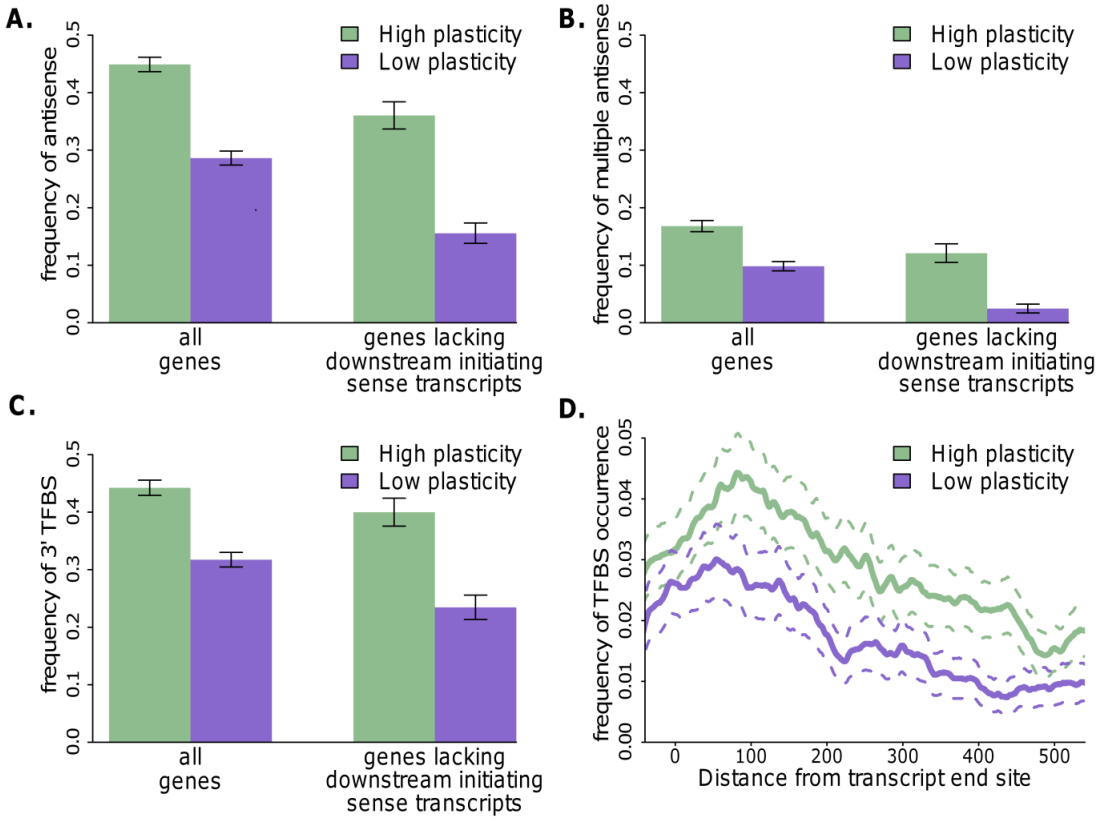


Figure 11: Association of 3' TFBS and 3' initiating antisense transcripts with transcriptional plasticity.

(A) The fraction of genes belonging to either the high or low plasticity group which possess at least one 3' initiating antisense transcript. **(B)** The fraction of genes possessing two or more 3' initiating antisense transcripts. **(C)** The fraction of genes possessing at least one 3' TFBS. **(D)** The frequency of TFBS binding sites downstream of the 3' end of the protein-coding transcript for both the high and low plasticity groups. Dotted lines indicate the error for the correspondingly coloured frequency lines. Shown are the resultant frequencies when the total gene set was considered. The results when only those genes lacking downstream initiating protein-coding transcripts were considered were similar. Note that the transcript end site is synonymous with the TTS described in the text. All error bars were calculated by bootstrapping.

TFBSs at the 3' end were also found to be associated with highly plastic genes. By assessing which genes had at least one TFBS at their 3' ends it was found that high plasticity genes more frequently possessed at least one 3' TFBS than low plasticity genes (44.3% compared with 31.7% for total genes; 40.0% compared with 23.5% for those genes lacking downstream initiating protein-coding sense transcripts; Figure 11C). Furthermore, investigation of the average TFBS distribution around the 3' ends identified a peak of TFBS frequency approximately 100bp downstream of the 3' end that

was more pronounced in the high plasticity genes than the low plasticity genes, with a maximum frequency of 0.044 compared with 0.030 (Figure 11D). All these observations held true when considering either the total set of genes or just those lacking downstream initiating protein-coding sense transcripts.

To explore the relationship between plasticity, 3' TFBS and antisense transcripts further, genes were subdivided on the basis of whether they had no 3' TFBS or 3' initiating antisense transcripts, whether they had one or the other, or whether they had both. Plasticity values were then compared between these groups. When considering the entire genome, genes with just 3' TFBSs were more plastic than those without, while those with just an antisense transcript were more plastic still (Figure 12A). Genes with both features were only marginally more plastic than genes with only an antisense transcript. When considering just those genes without downstream initiating sense transcripts, the relationship remained the same (Figure 12B), save that the genes with both features also showed a higher plasticity compared to just those with antisense transcripts. It is possible that the analysis in Figure 12B was refined compared to Figure 12A by removing 3' TFBSs that might be involved in regulating downstream sense transcription, and which might therefore not play a role in directing antisense transcription at all. Taken together, the results support a model in which TFBS-regulated antisense transcripts initiating from the 3' end enhance the transcriptional plasticity of a protein-coding sense transcript.

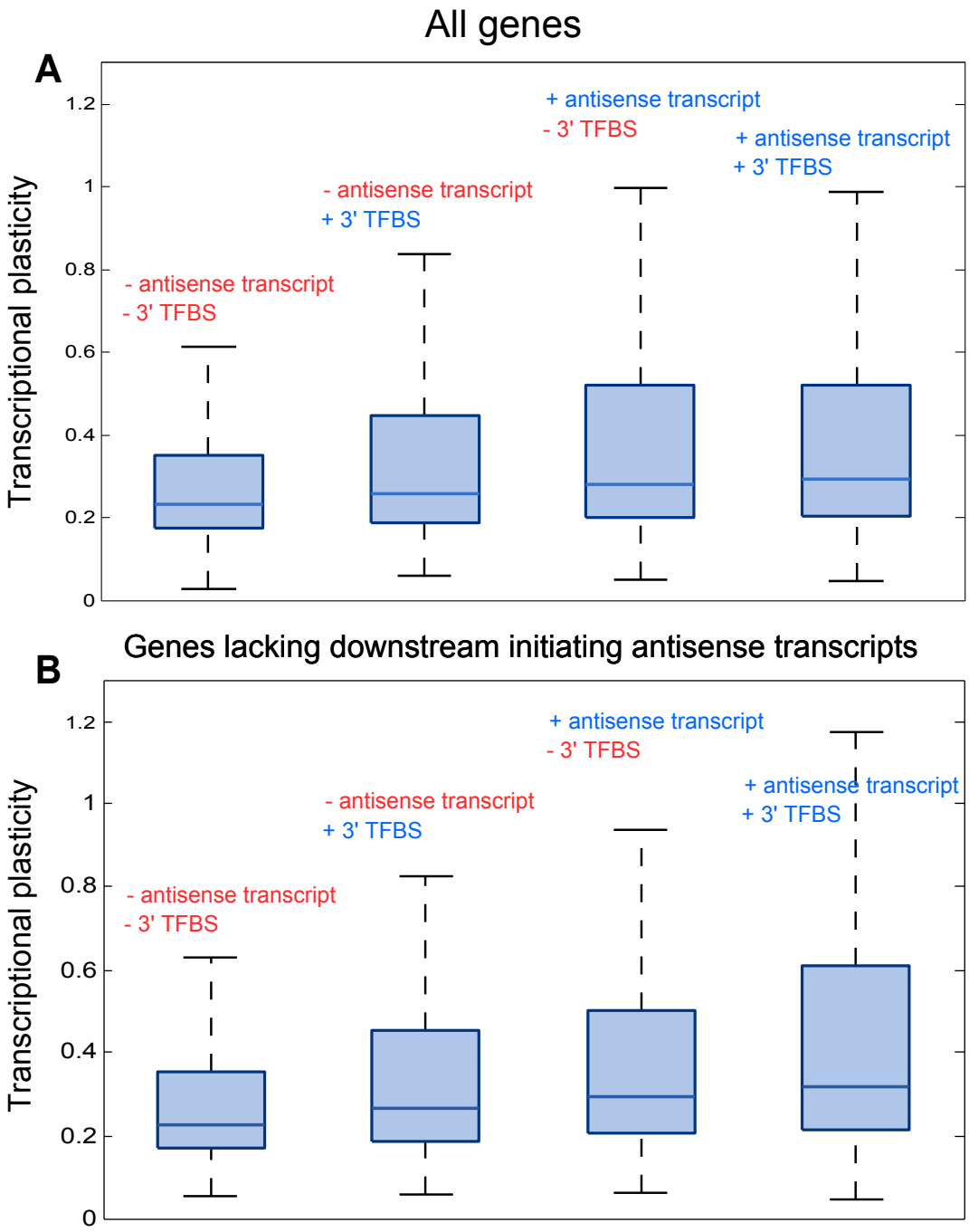


Figure 12: Genes with antisense transcripts and 3' TFBSs are more transcriptionally plastic.

(A) Boxplots showing the distribution of transcriptional plasticity values of genes divided into four groups, depending on the presence or absence of identified 3' TFBSs and/or annotated antisense transcripts, as described in the text. **(B)** The same arrangement, save that only genes without neighbouring tandemly oriented genes adjacent to their 3' end were considered.

3.3.5. A 3' NDR is associated with antisense production, 3' TFBSs and transcriptional plasticity

For an antisense transcript to initiate from the 3' end of a gene one might anticipate that an NDR would be required to recruit the initiation machinery, including any transcription factors, to the site adjacent to the antisense transcript's start site. By extension, one would anticipate that an NDR at the 3' end of genes would be associated with transcriptionally plastic genes, potentially by providing a platform for TFBS-mediated antisense initiation. To address these hypotheses, two different genome-wide analyses of nucleosome occupancy were used, one that was performed using an MNase-dependent procedure (Kaplan et al. 2009), and another that was MNase-independent (Fan et al. 2010). For each map the nucleosome occupancy surrounding the transcription end site of the different genes was averaged, to identify differences in 3' occupancy associated with antisense initiation, the presence of 3' TFBSs, and transcriptional plasticity. Additionally, the distribution of AT richness at the 3' end was determined, to assess whether any differences in the MNase-dependent occupancy could be a result of MNase's preference for cleaving at AT rich sequences (Chung et al. 2010).

Firstly, the question of whether those genes with 3' initiating antisense transcripts had a more pronounced NDR than those that did not was addressed, initially by using the MNase-independent map. For both the total set of genes (data not shown) and those lacking a downstream initiating protein-coding sense transcript (Figure 13), a substantial nucleosome depletion was observed at the 3' ends of those genes that possessed a 3' initiating antisense transcript relative to those genes that did not, with a minimum average depletion of -0.73 compared with -0.20, approximately 100bp downstream of the 3' end of the protein-coding sense transcript (Figure 13A). Interestingly, the MNase-dependent map yielded different results (Figure 13B). Though

there was some depletion downstream of the 3' end in the genes possessing antisense transcripts compared to those that did not, the extent of depletion at the 3' end itself was similar (-1.57 and -1.46 respectively). However, by assessing the AT richness around the 3' end in both sets of genes a trend was found that was similar to that of the MNase-dependent occupancy (Figure 13C), suggesting that the observed differences between the MNase-independent and MNase-dependent occupancy maps may be a function of MNase's own sequence preference.

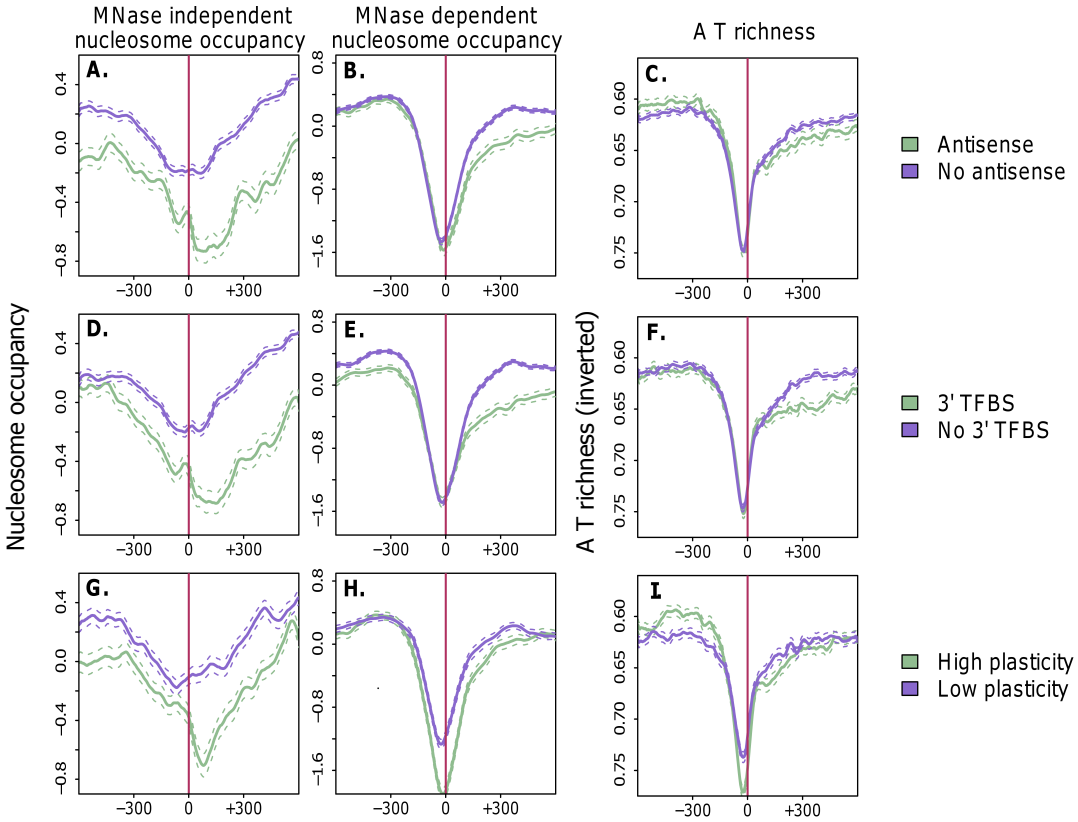


Figure 13: Nucleosome depletion at the 3' region is associated with antisense transcription, transcriptional plasticity and 3' TFBS.

(A) Comparison of average 3' nucleosome occupancy using the MNase-independent occupancy data between those genes possessing 3' initiating antisense transcripts and those that do not. **(B)** Comparison of average 3' nucleosome occupancy using the MNase-dependent occupancy data between those genes possessing 3' initiating antisense transcripts and those that do not. **(C)** Comparison of average 3' AT richness between those genes possessing 3' initiating antisense transcripts and those that do not. Note that the y-axis has been inverted for the purpose of comparison. **(D-F)** Average 3' occupancies and AT richness for those genes with at least one 3' TFBS and those without any. **(G-I)** Average 3' occupancies and AT richness for the high and low plasticity genes. Shown are the results for those genes lacking any downstream initiating protein-

coding sense transcripts. The results when the total gene set was considered were similar. All error bars (dotted lines) were calculated by bootstrapping.

To establish whether recruitment of TFBSs to the 3' end might be associated with an NDR, the difference in 3' nucleosome occupancy was assessed between those genes with at least one TFBS at their 3' end and those genes that had none. Using the MNase-independent occupancy map, it was found that the 3' ends of those genes containing 3' TFBSs were depleted for nucleosomes relative to those genes that did not (a minimum occupancy of -0.69 compared to -0.20; Figure 13D), with the most pronounced depletion being found approximately 100bp downstream of the 3' end; a pattern similar to that seen for the antisense transcript-containing genes (Figure 13A), and which also corresponds to the region at which one finds the highest frequency of 3' TFBSs (A and Figure 11A). Furthermore, it was found that the 3' ends of the high plasticity genes were more depleted for nucleosomes relative to the low plasticity genes (with a minimum occupancy of -0.71 compared to -0.18; Figure 13G). In both instances, the MNase-dependent occupancy map showed less pronounced differences between the 3' TFBS-containing and 3' TFBS-lacking genes and between the high and low plasticity genes (Figure 13E and Figure 13H), though these differences were again mirrored in the patterns of AT richness between the different sets of genes (Figure 13F and Figure 13I). It is worth noting that some depletion was still observed in highly plastic genes when using the MNase-dependent data set (Figure 13H).

Taken together, the results demonstrate that antisense initiation, 3' TFBSs and transcriptional plasticity are all associated with nucleosome depletion at the 3' end. Fan et al., (2010) proposed that increased nucleosome depletion at the 3' region is a consequence of increased RNAPII passage from the gene promoter, suggesting that antisense transcription might arise from high levels of sense transcription on the same gene. If antisense transcription levels were a function of sense transcription levels then

one might expect the two to be correlated. However, as shall be discussed in chapter 4, there is only a very small and negative correlation between the levels of sense and associated antisense transcription on the same gene (Spearman's correlation coefficient = -0.02). The results suggest that nucleosome depletion in the 3' region is associated with antisense initiation, and that this depletion is independent of gene transcription in the sense direction. The contributions of other transcription-related events (such as termination) toward depletion cannot be ruled out, however it is proposed that the additional depletion observed for those genes with antisense transcripts is itself a consequence of antisense transcription.

3.3.6. The 3' ends of genes show a pronounced promoter architecture that mirrors the 5' end, and which is associated with the presence of antisense transcripts.

The question was asked as to whether TBP and TFIIB, both components of the PIC, are a feature of the 3' region of genes, and whether their presence is associated with the presence of antisense transcripts and antisense transcription (Figure 14). To assess the average levels of PIC components TBP and TFIIB at the 3' region, and compare this to their levels in 5' promoter regions, a genome-wide data set from Venters and Pugh, (2009) was used. When considering the set of genes with antisense transcripts, a peak of both PIC components was found downstream of the 3' TTS, which mirrored what was seen upstream of the 5' TSS (Figure 14A-B). These peaks were more pronounced at genes with an antisense transcript than for those genes without (Figure 14C-D), though interestingly a smaller but similar peak was still observed for the latter group. Other factors associated with transcription, namely Swr1, which deposits the histone variant H2A.Z at gene promoters (Mizuguchi et al. 2004), and Rsc9, a component of the RSC chromatin remodelling complex (Cairns et al. 1996), were found to be significantly

enriched within the 3' regions of genes with antisense transcripts compared to those without (Figure 15A-B; $p = 2.5 \times 10^{-13}$ and 3.3×10^{-22} respectively, Wilcoxon rank sum test). The levels of Swr1 and Rsc9 at the 3' ends of genes with antisense transcription were of the same order of magnitude to the levels at the 5' ends (data not shown). Furthermore, the ratios between the levels at the 3' and those at the 5' end were similar to the ratios observed for TBP and TFIIB between the 3' and 5' end. These results demonstrate the presence of a promoter architecture at the 3' region, which is associated with the presence of an antisense transcript and which is similar to that at the promoters of protein-coding genes.

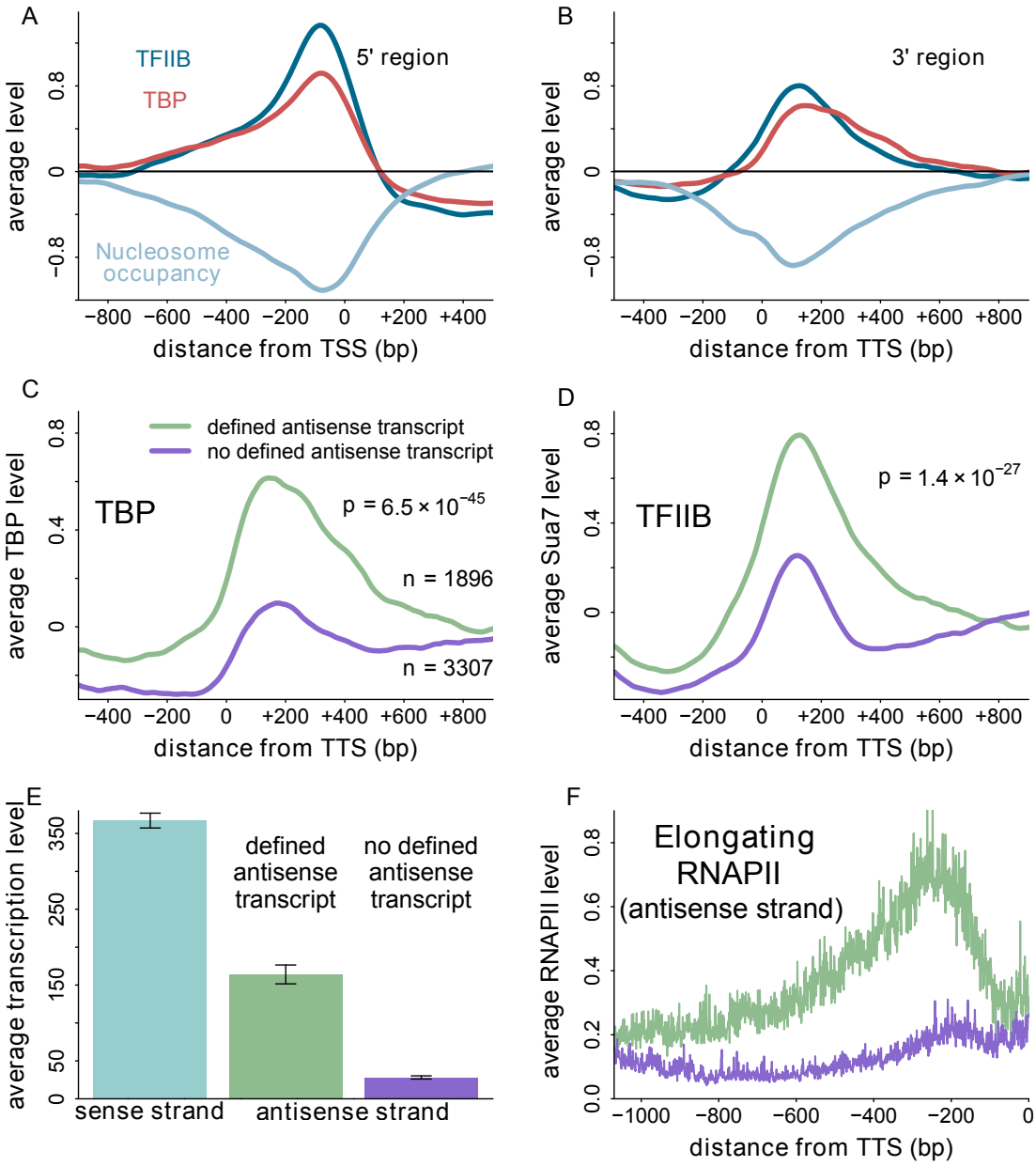


Figure 14: The 3' regions of genes show evidence of PIC formation.

(A) The average levels of TBP and TFIIB at gene 5' regions, obtained by aligning genes by their TSS. Nucleosome occupancy as determined by Fan *et al.* (2010) is included for comparative purposes, and is discussed in greater detail in the text. **(B)** The average levels of TBP and TFIIB at the 3' regions of genes with antisense transcripts, obtained by aligning genes by their TTS. **(C-D)** TBP and TFIIB levels at the 3' regions of genes, comparing genes with antisense transcripts to those without. p-values were calculated using the Wilcoxon rank-sum test, comparing the average transcription levels of the sense and antisense strands of protein-coding genes. The antisense strands of genes with antisense transcripts are considered separately from those without. Error bars were calculated by bootstrapping (the standard deviation of the mean of 1000 bootstrap pseudoreplicates). **(E)** Average levels of nascent sense and antisense transcript reads within protein-coding genes with and without defined antisense transcripts. Nascent transcript levels were determined using stand-specific, nascent transcript sequencing data from Churchman and Weissman (2011). Shown is the median number of sense reads within the transcription unit of all protein-coding genes considered (turquoise), and the median number of

antisense reads within those genes possessing 3' initiating antisense transcripts (green) compared with those genes that do not (purple). **(F)** Average levels of RNAPII density across the antisense strands of protein-coding genes aligned by their 3' ends, using the same data as in E. Shown is the average RNAPII density of those genes possessing 3' initiating antisense transcripts (green) compared with those genes that do not (purple).

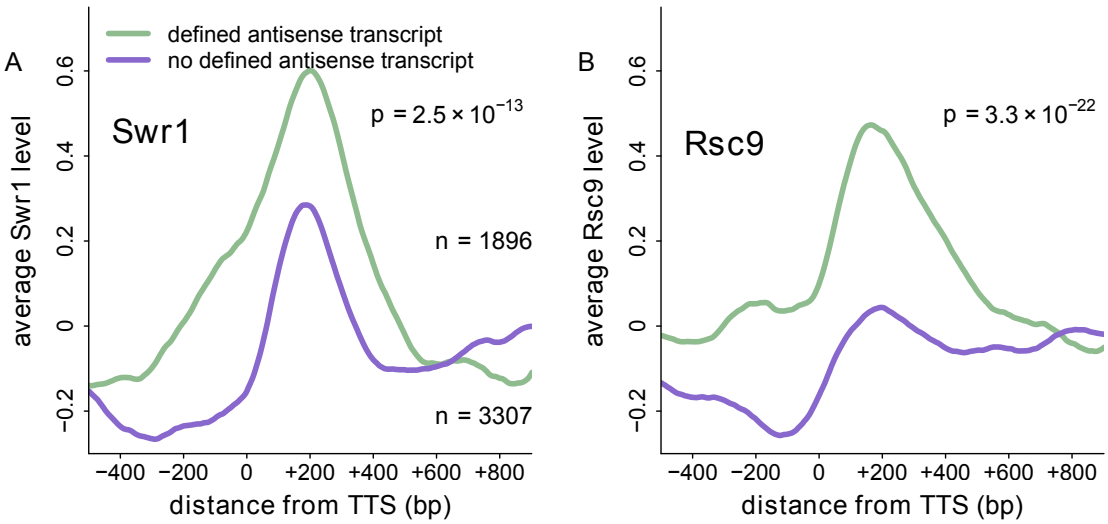


Figure 15: Levels of Swr1 and Rsc9 are more pronounced at the 3' ends of genes with 3' initiating antisense transcripts.

(A-B) Swr1 and Rsc9 levels at the 3' ends of genes, comparing genes possessing 3' initiating antisense transcripts to those that do not. p-values were calculated using the Wilcoxon rank-sum test, comparing the distribution of values at the maximum point of each averaged curve.

Next it was asked whether the PIC observed at the 3' region reflects the presence of adjacent protein-coding genes. Of the set of 1,097 genes described above, 412 (37.6%, similar to the total gene set) had an antisense transcript and were significantly enriched for TBP and TFIIB in their 3' region relative to the remaining 685 genes (Figure 16). Thus, there is an association between PIC components at the 3' region and antisense transcripts even when those genes with moderately transcribed protein-coding genes downstream are removed from the analysis, suggesting that the formation of a PIC at the 3' region can occur irrespective of the presence of an adjacent gene promoter. However, when comparing the 3' ends of all genes with antisense to just those genes with antisense but not other genes adjacent to their 3' promoters, one does see a substantial

reduction of 3' TBP and TFIIB in the latter group. Hence there does appear to be a confounding effect of adjacent genes/promoters upon 3' TBP and TFIIB levels.

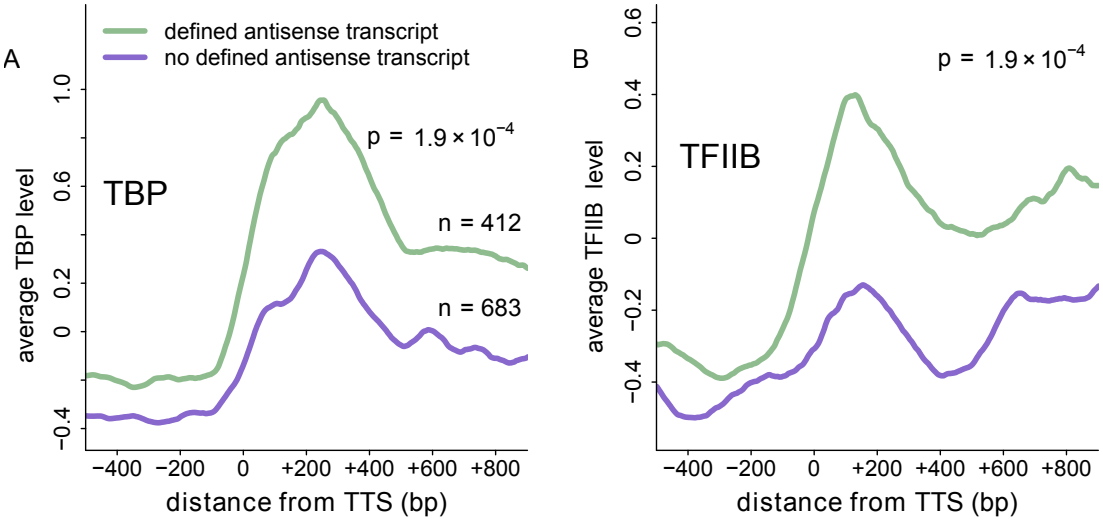


Figure 16: A relationship between 3' PIC assembly and initiation of antisense transcription remains evident when those genes whose 3' ends are close to other, moderately transcribed genes are removed from the analysis.

(A-B) TBP and TFIIB levels at the 3' ends of genes, comparing genes possessing 3' initiating antisense transcripts to those that do not. p-values were calculated using the Wilcoxon rank-sum test, comparing the distribution of values at the maximum point of each averaged curve.

3.3.7. Levels of PIC components at the 3' region reflect levels of nascent antisense transcription.

The average peaks of PIC components at the 5' regions of all protein-coding genes were compared with those at the 3' regions of genes with antisense transcripts. Surprisingly, the average peaks of TBP and TFIIB at the 3' regions of genes with antisense transcripts were more than half those at the 5' regions (TBP levels were 62% of what they were at the 5' regions, TFIIB levels were 60%; Figure 14A-B). This might reflect the fact that antisense transcripts are transcribed at higher levels than evident from steady-state levels, which are generally low due to transcript degradation (Perocchi et al. 2007; Wyers et al. 2005). To this end, the average levels of transcription at of these distinct

gene groups were determined. A strand-specific, genome-wide map of elongating RNAPII was utilised to obtain the level of nascent sense and antisense transcription and to assign a measure of transcription to each transcript within the collated map (Churchman and Weissman 2011). Remarkably, the median level of antisense transcription at genes with an antisense transcript was 45% that of sense transcription at all protein-coding genes (Figure 14E). This is strikingly similar to the difference in average levels of PIC components between the 5' region of all protein-coding genes and 3' region of genes with antisense transcripts. Thus there is an association between levels of PIC components and levels of transcription for both protein-coding sense and non-coding antisense transcripts.

The small peaks of TBP and TFIIB observed for the group of genes without antisense transcripts suggests that many of these genes may support low levels of antisense transcription which were not detected in any of the transcript maps utilised. Indeed, there is evidence for antisense transcription within these genes, though the average level of elongating RNAPII is significantly lower than for those genes with defined antisense transcripts initiating from the 3' end (Figure 14F). An alternative explanation could be that these peaks of TBP and TFIIB are a consequence of conditional long-range chromosomal juxtapositions between a promoter and a terminator, also known as gene loops (O'Sullivan et al. 2004), which are known to be TFIIB/Sua7-dependent (Singh and Hampsey 2007; Medler et al. 2011). It is certainly possible that antisense transcription could itself be a consequence of gene looping, and that in presenting a PIC to the 3' region an antisense transcript is also produced which could explain why those genes with antisense transcripts have higher levels of PIC components in their 3' region than those without.

3.3.8. Extensive antisense transcription necessitates a more pronounced promoter architecture

Next it was assessed whether high levels of antisense transcription were supported by a correspondingly pronounced promoter architecture at the 3' region, and whether the reverse was true for lowly transcribed antisense transcripts. A group of highly and lowly transcribed antisense transcripts were defined (the top 20% and bottom 20% of the data, after masking the bottom 10% from consideration to remove potentially false positive antisense transcripts) and the average levels of TBP, TFIIB and nucleosome occupancy at the 5' end of the antisense transcripts were determined (see chapter 2). Peaks of TBP and TFIIB were identified at the 3' region that were substantially higher for the genes with highly transcribed antisense transcripts (Figure 17A-B). Nucleosome depletion was also found to be more pronounced for the highly expressed antisense transcripts (Figure 17C). These results support the notion that antisense transcription is driven by PIC formation, in a manner similar to protein-coding sense transcription, and that highly transcribed antisense transcripts are driven by similarly high levels of the PIC components TBP and TFIIB.

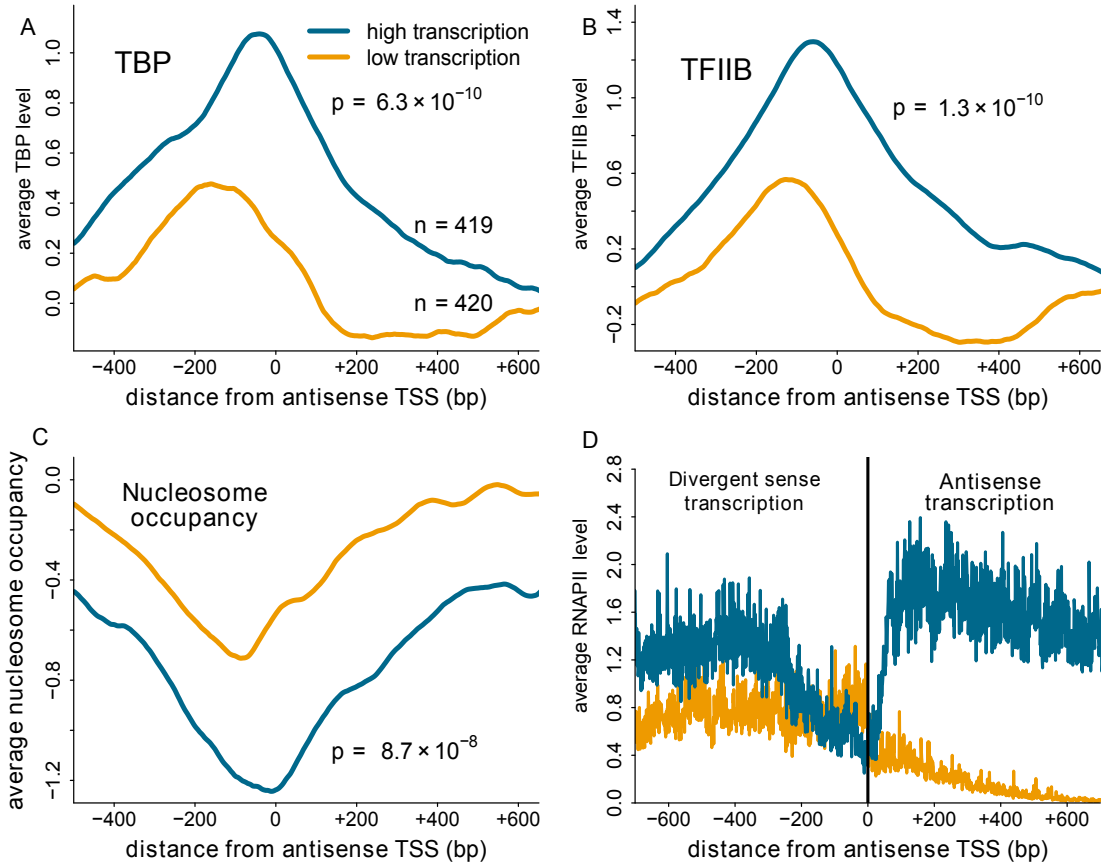


Figure 17: Initiation of antisense transcription is supported by PIC formation.

(A-C) The average levels of TBP, TFIIB and nucleosome occupancy around the TSS of both highly and lowly transcribed antisense transcripts. All p-values were calculated using the Wilcoxon rank-sum test, comparing the distribution of values at the maximum point of each averaged curve (the minimum point in the case of nucleosome occupancy). Antisense transcripts were defined as being highly or lowly transcribed using genome-wide NET-seq data obtained from Churchman and Weissman (2011). **(D)** The average level of nascent transcription around the TSS of the antisense transcripts. Shown left of the TSS is the average level of divergent sense transcription whilst on the right of the TSS is the average level of antisense transcription.

3.3.9. Bidirectional promoters show evidence of two distinct pre-initiation complexes.

It has been proposed that promoters are inherently bidirectional (Neil et al. 2009). It would be reasonable to expect that for these highly expressed antisense transcripts one would observe similarly high levels of transcription in the sense direction. Using the two groups of genes with highly or lowly expressed antisense transcripts, levels of divergent

sense transcription were assessed. Although the average level of transcription was higher in the sense direction for the highly expressed antisense group than for the lowly expressed antisense group, the level was not as high as the antisense expression itself (Figure 17D). This suggests that high levels of antisense transcription can be supported in the absence of high sense transcription, i.e. that inherently bidirectional promoters can be biased in both the sense and antisense directions, and perhaps that the two can be regulated independently of one another.

To address this, two classes of bidirectional promoter were defined to find evidence of two distinct PICs at bidirectional promoters. The classes were defined on the basis of how far the TSS of the antisense transcript was from the TSS of the sense transcript. In the first group, termed the narrow promoters, the TSSs were within 300bp. In the second group, termed the broad promoters, they were within 700bp but further than 400bp (with no other transcript TSS between them). The antisense transcripts considered were the same as defined above (i.e. they initiated within the 3' region of a gene). The average levels of TBP, TFIIB and nucleosome occupancy around the TSS of the antisense transcripts were assessed as before.

The two classes of bidirectional promoter showed distinct patterns. Overlaying the TBP profile of the broad promoters with that of the narrow promoters demonstrated a substantial enrichment of TBP upstream of the antisense TSS as well as over the TSS itself, supporting a model in which two distinct PICs form at bidirectional promoters (Figure 18A). Additionally, two partially overlapping peaks of TFIIB were observed over the broad promoters, whilst only one peak was discernible over the group of narrow promoters (Figure 18B). Of these two peaks, one corresponded to the single peak seen in the narrow promoters, both of which lay over the antisense 5' end. The second peak was further upstream, potentially representing the PIC driving the divergent sense transcript. Finally, there was a much broader nucleosome depletion in

the broad promoter group, which one would expect if the two PICs were further apart in the broad group than in the narrow group (Figure 18C). This analysis supports a model in which two PIC recruitment sites exist at bidirectional promoters, providing a mechanism to explain how sense and antisense levels could be independently regulated

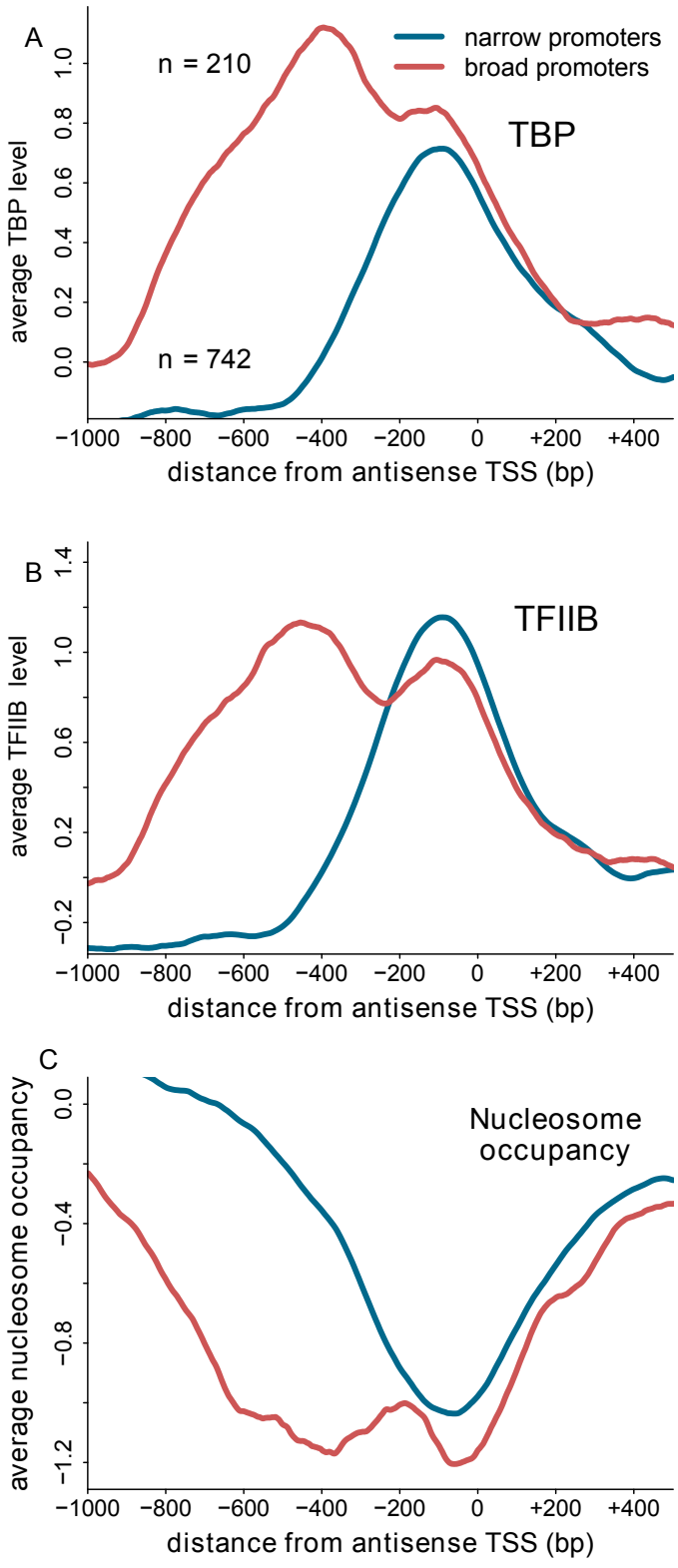


Figure 18: Bidirectional promoters show evidence of possessing two distinct PICs.

(A-C) The average levels of TBP, TFIIB and nucleosome occupancy around the TSS of antisense transcripts with an adjacent and divergent sense transcript. Shown are both the narrow and broad classes of bidirectional promoter described in the text.

3.3.10. Extensive antisense transcription can be supported by a promoter architecture in the absence of adjacent divergent sense transcription.

If antisense transcription can be regulated independently of divergent sense direction from bidirectional promoters then one might expect to find instances where high levels of antisense transcription initiation occur in the absence of adjacent sense transcription. One would also expect to find a pronounced promoter architecture at such a site of initiation, i.e. a unidirectional promoter driving antisense transcription. To this end, antisense transcripts were selected which had only very low levels of sense transcription upstream of their 5' end (see chapter 2), and which were not also adjacent to any non-RNAPII transcribed elements, including tRNAs and rRNAs. This gave a set of 804 transcripts, with an average level of nascent transcription that was much higher in the antisense direction than in the divergent sense direction (Figure 19A).

The highly and lowly transcribed subgroups (the top 25% and bottom 25% of the data, after masking the bottom 10% from consideration to remove potentially false positive antisense transcripts) were selected from these antisense transcripts, and the average levels of TBP, TFIIB and nucleosome occupancy were compared at their TSSs. The highly expressed antisense transcripts had a more pronounced peak of TBP and TFIIB than the lowly expressed antisense transcripts, and were significantly more depleted for nucleosomes (Figure 19A). These results demonstrate that high levels of antisense can be supported in a PIC-dependent manner in the absence of adjacent divergent sense transcription, indicating that antisense transcription is not a consequence of divergent sense transcription, and that the process of antisense transcription can be regulated independently of divergent sense transcription.

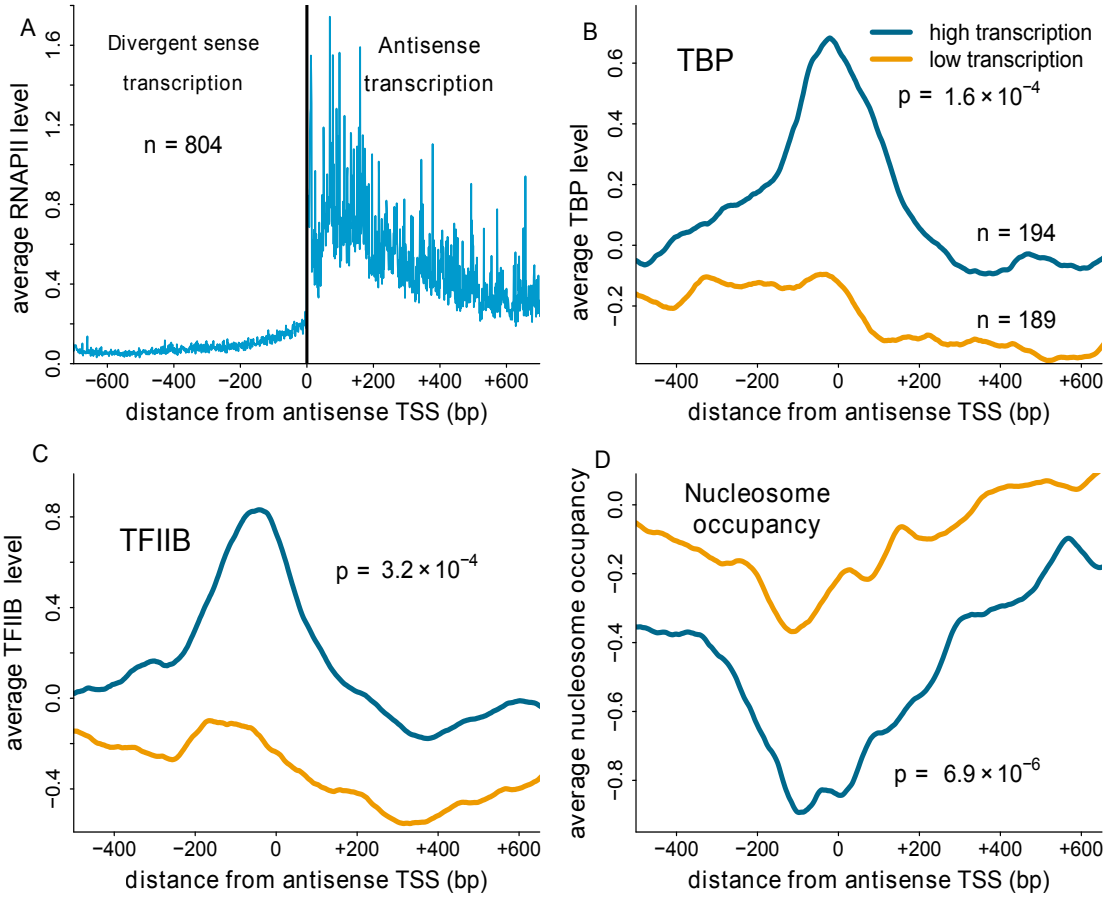


Figure 19: PIC formation supports high levels of antisense transcription in the absence of high levels of divergent sense transcription.

PIC formation supports high levels of antisense transcription in the absence of high levels of divergent sense transcription. **(*A*)** A group of antisense transcripts with low levels of divergent sense transcription were selected. Shown is the average level of nascent transcription around the TSS of antisense transcripts belonging to this group, with nascent sense transcription to the left of the TSS and nascent antisense transcription to the right. **(*B-D*)** The average levels of TBP, TFIIB and nucleosome occupancy around the TSS of both highly and lowly transcribed antisense transcripts belonging to this group. Antisense transcripts were defined as being highly or lowly transcribed using genome-wide NET-seq data obtained from Churchman and Weissman (2011). All p-values were calculated using the Wilcoxon rank-sum test, comparing the distribution of values at the maximum point of each averaged curve (the minimum point in the case of nucleosome occupancy).

3.3.11. Short 3' regions placed within *GAL1* and *GAL10* ORFs drive production of antisense transcripts whilst also behaving as terminators

The gene symmetry hypothesis requires that gene terminators possess the inherent capacity to behave in a bi-functional manner as the promoters of antisense transcription. To test this, experimental systems were devised in which yeast terminator

sequences were inserted into the ORFs of genes (Figure 20). Specifically, the *ADH1* terminator ($ADH1_T$) and the *TEF* terminator (TEF_T) were inserted, separately, into the ORF of the *GAL10* gene, while $ADH1_T$ was also inserted into *GAL1*. These sequences were short (approximately 250bp), and were inserted into *GAL10* only after deletion of its internal bidirectional promoter (Houseley et al. 2008) (Figure 20A). In three all cases, when induced, high levels of sense transcript were observed, with similar levels to the wild-type strain (Figure 20B,E and G). However, these sense transcripts were truncated compared to wild-type, implying that the inserted sequences were sufficient to direct transcription processing and termination, as was anticipated. Most strikingly, however, both the inserted terminator sequences were also sufficient to drive the production of antisense transcripts, which were of similar size to the sense transcript (Figure 20C,F and H). *GAL1* normally possesses a XUT (Mellor lab, unpublished), however insertion of $ADH1_T$ resulted in production of a stable antisense transcript that was shorter than the wild-type XUT, suggesting that insertion of the sequence has redefined both the sense and antisense transcription units of the gene.. Antisense transcripts were present both when the galactose inducible *GAL1* and *GAL10* promoters were repressed and also during induction (Figure 20C,F and H). These antisense transcripts persist through induction, although the levels of the antisense transcript produced from *GAL1*:$ADH1_T$ after 3 hours in galactose (steady-state) are barely detectable, a possible consequence of degradation (Figure 20G and H). The results demonstrate that these short sequences are capable of recapitulating functions associated with both a promoter and a terminator when placed within the middle of a gene, demonstrating that the 3' regions of genes can direct antisense transcription in addition to their expected functions in termination of the protein-coding sense transcript.

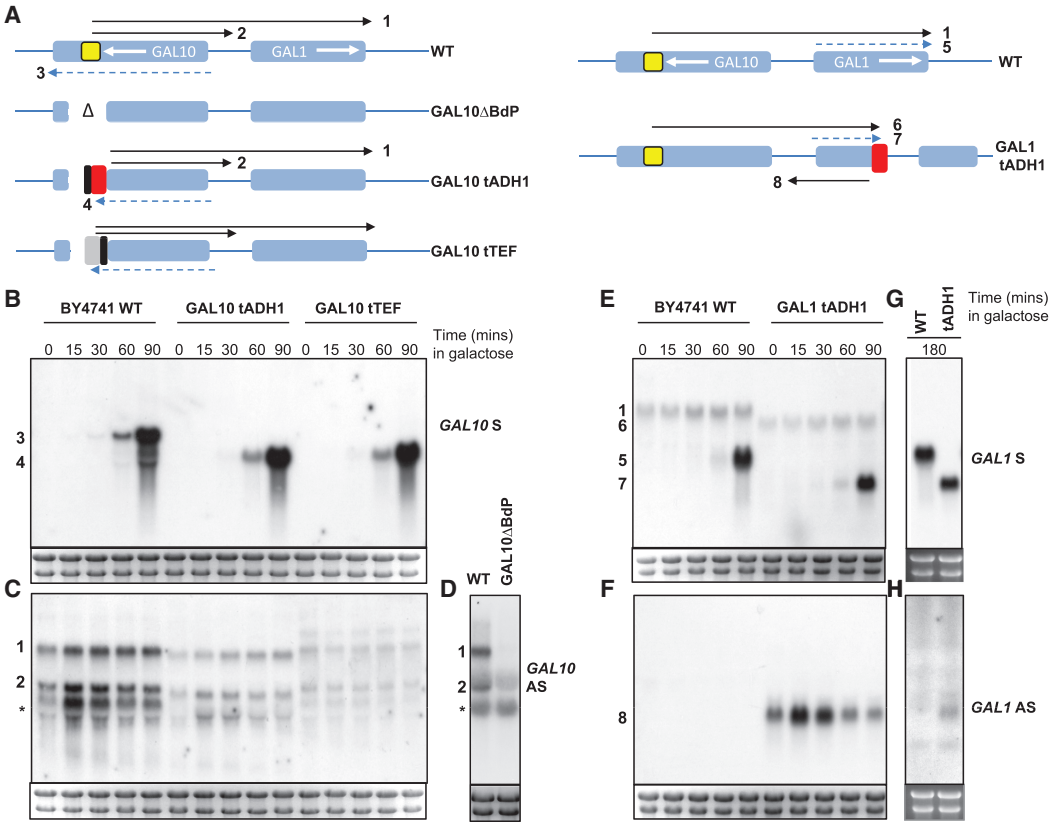


Figure 20: Short 3' regions placed within *GAL1* and *GAL10* ORFs drive production of antisense transcripts whilst also behaving as terminators.

Evidence that 3' regions can function as both promoters of antisense transcription and as terminators of protein-coding sense transcripts. **(A)** Schematics showing the derivatives of *GAL10* and *GAL1* used in this analysis. The direction of sense transcription is shown by the white arrows. Wild-type strains (BY4741) are shown, as is a strain in which the internal bidirectional promoter of *GAL10* has been deleted (*GAL10*ΔBdP). Terminator sequences were inserted into *GAL1* and *GAL10*ΔBdP. $ADH1_T$ (red box) was inserted into both genes and TEF_T (grey box) was inserted into *GAL10*ΔBdP. The black boxes represent the residual loxP sequences used in the construction of strains with terminator insertions. **(B–H)** Autoradiographs of northern blots hybridized to strand-specific probes designed to detect the sense (S) or antisense (AS) transcripts in the wild-type and derivative strains of *GAL10* (B–D) and *GAL1* (E–H) described in (A). Total RNA was prepared from cells cultured in YPD or after induction in galactose for the time shown. *GAL10* produces two major antisense transcripts (black arrows) in glucose medium, one extends over the *GAL1-10* promoter (2) and the second longer transcript (1) extends sense to the end of *GAL1* and persists for 90–100 min after induction but then is no longer detectable. Loss of the *GAL10* internal promoter causes loss of both antisense transcripts (D) which can be restored by inserting either $ADH1_T$ or TEF_T (C). There are other antisense transcripts (asterisks) that do not initiate at the bidirectional promoter. The 18S and 25S rRNAs are used to estimate loading. For (B–F) exposure times and specific activities of probes are similar, although probe lengths are three times longer for *GAL10* compared to *GAL1*. G is exposed four times shorter than B–F, which reflects accumulation of *GAL1* transcript to high levels. H is exposed 10 times longer than B–F to reveal very low levels of antisense transcript in *GAL1*:$ADH1_T$ 3 hours after induction. Note that tADH1 and tTEF as described in the figure are equivalent to $ADH1_T$ and TEF_T as described in the text. Experiments were performed by Ana Serra Barros.

3.4. Discussion

Here, questions relating to how antisense transcripts arise, how they are regulated, and what effect they might have upon gene regulation, have been addressed. Furthermore, and perhaps most importantly, the events of antisense transcription initiation at the 3' end of a gene have been compared with sense transcription initiation, to assess how similar they are and to establish the true extent of antisense transcription genome-wide. This was achieved by using two distinct collections of genome-wide maps. The first attempted to collate transcript maps genome-wide, and assign quantitative levels of transcription to each transcript using a map of strand-specific elongating RNAPII (Granovskaia et al. 2010; Yassour et al. 2010, 2009; Churchman and Weissman 2011; Neil et al. 2009; Xu et al. 2009; Nagalakshmi et al. 2008). The second attempted to collate major promoter features, namely PIC components, transcription factor binding sites and nucleosome depleted regions (Kaplan et al. 2009; Fan et al. 2010; Harbison et al. 2004; MacIsaac et al. 2006; Venters and Pugh 2009). In this way, the extent to which canonical gene terminators resemble gene promoters could be assessed. It was found that the presence of 3' promoter features is associated with antisense transcription, and that they are present even in the absence of adjacent, active sense promoters. Furthermore, at genes with annotated antisense transcripts (roughly a third of the genome) there are, on average, almost half the number of antisense transcript NET-seq reads compared to sense transcripts NET-seq reads, demonstrating the abundance of antisense transcription genome-wide. Finally, it was shown that short terminator sequences, when placed within the ORFs of different genes, are able to behave as both terminators and promoters of sense and antisense transcription respectively, redefining the transcription units of the gene. These results demonstrate an inherent symmetry to gene structure, with promoters present at either end (Murray et al. 2012).

3.4.1. Gene symmetry

As discussed in the introduction, promoters are marked by numerous features involved in directing transcription initiation. Strikingly, many of these same features are present within the terminators of genes. Sequences thought to be specific to promoter regions – namely transcription factor binding sites and TATA-boxes, are present within the 3' region. Furthermore, general transcription factors, sequence-specific transcription factors and chromatin remodelling enzymes are all enriched at the 3' regions of genes, while nucleosomes are depleted. How might such gene symmetry have arisen? As discussed in the introduction, the yeast genome is heavily interleaved, such that the promoters and terminators of some genes are often adjacent to those of others. During gene duplication events, genes might retain promoters adjacent to their terminators while losing the gene the promoter was originally attached to. If the retained promoter provided some benefit to the gene (such as through antisense transcription) then it might be maintained and possibly become integrated into the terminator itself. Alternatively, as proposed above, sequences like the TATA-box could be inherently bi-functional sequences. The findings discussed here require a change in how we view gene structure, and suggest that the intergenic sequences flanking transcription units are capable of performing multiple functions.

As well as directing the initiation of antisense transcription, another possibility is that the observed 3' PICs are a consequence of gene loops juxtaposing promoters and terminators, thus bringing promoter features into the proximity of terminators, resulting in their being detected via ChIP (O'Sullivan et al. 2004). Gene loops have been proposed to be involved in processes including transcription re-initiation (O'Sullivan et al. 2004), regulating the directionality of bidirectional promoters (Tan-Wong et al. 2012), and transcription noise (Hebenstreit 2013). As discussed in chapter 6, antisense

transcription is also associated with transcription noise, thus it is tempting to speculate that looping and antisense transcription are related processes, perhaps co-regulated by binding events at the 3' promoter.

3.4.2. Gene terminators as the promoters of antisense transcription

There is evidence presented here to support a model in which 3' promoters direct antisense transcription. Genes with defined antisense transcripts initiating from within their 3' region are more likely to have occupied 3' TFBSs than those without, and are enriched for TBP, TFIIB, Rsc9 and Swr1, whilst also being more depleted for nucleosomes (determined using a MNase-free assay). What is more, an increase in 3' PIC components correlates with an increase in the level of antisense transcription. These findings demonstrate that antisense transcription initiation shares features with sense initiation, namely the formation of a PIC, as has previously been suggested (Venters and Pugh 2009). These results provide insight into the findings discussed in the previous chapter – that a terminator region is alone sufficient to direct antisense transcription – by demonstrating that PIC binding at the terminator is associated with the presence of an antisense transcript.

Levels of PIC components are well correlated with the accompanying levels of transcription (Kim and Iyer 2004). At first sight, the high levels of PIC components associated with the 3' regions of genes with antisense transcripts appears to be at odds with the low steady state levels of many antisense transcripts. However, as discussed in chapter 1, antisense transcripts are targeted for degradation by different exonucleases (Camblong et al., 2007; Thompson and Parker, 2007; Uhler et al., 2007). Over 1,000 transcripts antisense to open reading frames are polyadenylated, exported and degraded in the cytoplasm by the 5' to 3' exonuclease Xrn1 (Kem1) (van Dijk et al. 2011), while almost 1,000 were identified as being degraded in the nucleus by the 3' to

5’ exonuclease Rrp6 (Xu et al. 2009). As a consequence, their steady state levels do not accurately reflect the levels of nascent transcription within the genome and this likely gives an impression that antisense transcription from the 3’ region of genes is less pronounced than it really is, as shall be discussed more extensively in chapter 4. In support of this, the repressed *GAL1* gene produces a Kem1-degraded antisense from its 3’ region that is not detectable in steady state (Mellor lab, unpublished). When genes with antisense transcripts are considered, average levels of antisense transcription are 45% of the average levels of protein-coding sense transcription. This value is consistent with a strong correlation between levels of PIC components and levels of nascent transcription and allows the suggestion that in other genomes, levels of PIC components at sites at which transcription is initiated will reflect levels of nascent transcription (though not necessarily steady state transcript levels). That antisense transcription is so abundant genome-wide (Figure 14E) supports a possible regulatory function, which, given that antisense transcripts are so often unstable, is likely to be exerted by the act of transcription itself, and not by the resultant transcripts. Certainly the majority of studied instances suggest that antisense transcription generally functions in *cis,* and indeed, Camblong et al., (2009) found that at all the genes they studied, antisense transcription operated in *cis,* and that only at a single gene (*PHO84*) did antisense transcription also function in *trans*. It is possible that the degradation machinery itself serves an additional function as a transcription regulator, as has been suggested for Rrp6 (Hieronymus et al. 2004; Andrulis et al. 2002; Camblong et al. 2007). More recently, Xrn1 has also been found to bind to promoters genome-wide and influence transcription (Haimovich et al. 2013), thus it is possible that rapid degradation is itself what grants antisense transcription its biological function.

3.4.3. Gene promoters as the terminators of antisense transcription?

Given that terminators possess promoter-like features, an important consideration in exploring gene symmetry is whether or not promoters also possess terminator-like features, given the tendency for the ends of sense and antisense transcripts to be found near to one another (Xu et al. 2011). If this is true, then the presence of terminator features at the promoters of genes might also be observed. In support of this, Ssu72, involved in 3' end processing, has been found at the promoters of genes (Nedea et al. 2003). Given that antisense transcripts tend to be polyadenylated (van Dijk et al. 2011) it might be prudent to assess whether there is evidence of poly(A) signals, or indeed the upstream or downstream elements typical of terminators, though this might be complicated by the poor homology of these sequences in yeast (Guo et al. 1996; Humphrey et al. 1991, 1994). A recent study in mouse embryonic stem cells, however, has shown that poly(A) signals exist upstream of protein-coding sense TSSs, and are involved in the cleavage and polyadenylation of antisense transcripts (Almada et al. 2013).

Given the extent of degradation, and the difference in processing of antisense transcripts, it would be reasonable to expect that the termination of antisense transcription would also be different, given the intertwined nature of processing and termination (Proudfoot 2004). Indeed, Arigo et al., (2006) found that termination of CUTs was directed by the RNA-binding proteins Nrd1 and Nab3, which play a role in the termination of snRNAs, snoRNAs and a subset of mRNAs including *NRD1* itself (Steinmetz et al. 2001). Mutation of *NRD1* or *NAB3* led to inefficient termination and 3' extended CUTs. Thus it may be that, although antisense-transcribed genes show symmetry in their promoter architectures, the termination processes could be quite different. This of course raises the question of what it is that might make sense and antisense transcription different, which will be discussed in chapter 5. Another possible

mechanism of antisense transcription termination is that sense promoters could possess inherent terminator activity by virtue of binding highly stable PICs, and that such complexes could dislodge incoming polymerases – the reverse of the proposed "sitting-duck" mechanism of transcriptional interference (Sneppen et al. 2005).

3.4.4. The regulation of antisense transcription

If antisense transcription represents a genome-wide mechanism of gene regulation, then understanding how antisense transcription is in turn regulated becomes crucially important. Antisense transcription is associated with PIC formation at the 3' regions of genes, a formation which does not necessarily predicate divergent sense transcription. Moreover, DNA-bound regulatory transcription factors are frequently present at sites of antisense transcription where there is little divergent sense transcription, supporting a model in which antisense transcription can be regulated independently of divergent sense transcription. It follows that if the promoters of antisense transcripts can be regulated independently of downstream gene promoters, antisense-mediated gene regulation is not merely a function of downstream gene transcription. This allows for a more exquisite level of control over the transcriptional landscape of the organism, with regulatory signals able to impact upon both the 5' and 3' promoters of a gene, influencing sense transcription and thus gene expression.

3.4.5. Evidence for initiation and regulation of antisense transcription independent of sense transcription

3' promoter structures remain pronounced at those genes that do have antisense transcripts but do not have the promoters of other, adjacent genes downstream of their 3' ends. Such genes also still show an association between antisense transcription and plasticity. Furthermore, antisense transcripts whose 5' ends abut regions of very low or

no sense transcription can still support high levels of antisense transcription and PIC components. These findings suggest that antisense transcripts can arise from their own unique 3' end promoters and do not need to 'piggyback' off the promoters of other genes, as has been suggested previously (Neil et al. 2009); that antisense transcription is not dependent upon the formation of transcription complexes driving protein-coding sense transcription, as discussed in greater detail below. These results support a regulatory model for antisense transcription, in which the presence of a TFBS-containing NDR at the 3' end allows for differential transcription of an antisense transcript across varying environmental conditions, independent of adjacent transcription events, a 'tuning' which in turn allows for differential transcription of the sense transcript.

3.4.6. Disentangling bidirectional promoters

Neil et al., (2009) proposed a model describing the initiation of bidirectional transcription, in which two independent PICs form within an NDR to direct sense and antisense transcription. This model was based on the observation in *TPI1,* that the deletion of a TATA-box abrogates sense transcription of the protein-coding mRNA whilst enhancing levels of the non-coding antisense transcript. The authors suggest that the PICs at bidirectional promoters compete for transcription factors, such that the resultant distribution of factors determines the relative PIC stability and the extent of sense and antisense transcription. This would imply an inverse correlation between antisense transcription and divergent sense transcription at bidirectional promoters; however it has also been shown that bidirectional promoters can be coregulated in a positive manner (Xu et al. 2011). Furthermore, Churchman and Weissman, (2011) have demonstrated that sense and antisense levels from bidirectional promoters are positively correlated (Spearman's correlation coefficient = 0.3). The work in this chapter

sheds light on two important features of antisense transcription from bidirectional promoters. Firstly, that antisense transcription is a consequence of directed PIC formation, rather than of transcriptional infidelity (Figure 14. Figure 17, Figure 18 and Figure 19), and secondly that high levels of antisense transcription can be supported by high levels of PIC components (implying a highly stable PIC) in both the presence and absence of high levels of divergent sense transcription (Figure 17 and Figure 19). Therefore, bidirectional promoters are in many respects two distinct but closely spaced promoters that can be coregulated in a positive manner, such as at the *SUT719/GAL80* promoter (Xu et al. 2011), an inverse manner or indeed independently from one another.

3.4.7. 3' nucleosome depletion

Fan et al., (2010) have shown that under conditions of carbon starvation, the extent of nucleosome depletion at the 3' end of genes is generally less pronounced than in the presence of glucose, something that is not observed at the 5' end. Based on the finding that genes producing antisense transcripts from their 3' ends have more pronounced 3' nucleosome depletion than those without (Figure 13A), it would be predicted that this reflects a reduction of global antisense production under such stressed conditions. This would suggest that 3' nucleosome depletion is mainly a function of the rate of antisense initiation and not a direct result of intrinsic histone-DNA interactions as is believed to be the case at the 5' end (Kaplan et al. 2009). This is supported by work presented by Fan et al., (2010), in which exclusion of RNA polymerase II from the *PMA1* gene led to an increased nucleosome occupancy at the 3' end but not at the 5'.

3.4.8. Antisense transcription and transcriptional plasticity

There is evidence to suggest that antisense transcripts initiating from the 3' regions of protein-coding genes may regulate sense gene transcription and thus gene expression. Antisense transcripts have been shown to be capable of regulating protein-coding sense transcription both by recruiting histone modifying enzymes and by transcriptional interference (Camblong et al. 2007; Berretta and Morillon 2009) and there is evidence that they can both upregulate and downregulate expression (Houseley et al. 2008; Uhler et al. 2007). Furthermore, by assessing the levels of gene expression across a number of different environmental conditions, it has been shown that genes with antisense transcripts show a higher variability of expression, i.e. they are more transcriptionally plastic than genes without (Xu et al. 2011). Here it is found that antisense transcripts are remarkably abundant genome-wide. Given that they are so frequently degraded, one might argue that the process of antisense transcription is wasteful. However, if the act of antisense transcription is involved in changing the pattern of gene expression (Xu et al. 2011), or in expediting the transition in expression in response to changing conditions (Uhler et al. 2007), then it may represent a necessary investment to ensure a rapid and appropriate change in the transcriptional landscape.

It has been shown previously that the promoters of highly plastic genes tend to have more upstream TFBSs than low plasticity genes (Lin et al. 2010). The ability to bind a greater number of TFBSs presumably allows a gene to respond to a greater number of environmental inputs and thus display a wider range of expression values. Here it is shown that the same is true of TFBSs at the 3' end, and that, further to this, a greater number of 3' TFBSs are associated with genes producing antisense transcripts from their 3' ends. Furthermore, genes with 3' TFBSs *and* antisense transcription were more plastic than those with just one or the other (Figure 12), at least when considering genes without downstream sense transcripts. This provides support for a mechanism of

3' TFBS-directed regulation of sense transcription, mediated by the production of an antisense transcript. Such a mechanism is supported by a previous study by Houseley et al., (2008), in which a Reb1 binding site was shown to be necessary for the production of a long ncRNA at the *GAL1-10* gene cluster. It may be of interest to assess which TFBSs are generally found around the TSSs of antisense transcripts, and whether they tend to be different from those sites found at the 5' end of the gene. The results in this chapter support a regulatory model for antisense transcription, in which the presence of a TFBS-containing NDR at the 3' end allows for differential expression of an antisense transcript across varying environmental conditions, which in turn allows for differential expression of the sense transcript. This model holds regardless of whether the antisense is enhancing or repressing sense transcription.

A relationship between nucleosome occupancy at the 5' end of genes and transcriptional plasticity is well established (Tirosh and Barkai 2008). The extent of depletion at the 5' end was found to correlate negatively with plasticity, suggesting that the additional requirement of overcoming nucleosomes present at the promoter prior to initiation makes a gene more transcriptionally plastic. Interestingly, the reverse was found to be true for nucleosome occupancy at the 3' end; transcriptionally plastic genes had, on average, a more pronounced depletion at the 3' end than low plasticity genes (Figure 13G). Furthermore, it was found that 3' ends containing one or more TFBS(s) had a more pronounced average depletion, in line with their role as determinants of nucleosome depletion (Bernstein et al. 2004) (Figure 13D), and that 3' ends capable of producing antisense transcripts were also more depleted (Figure 13A). Taken together, these results suggest that the presence of a 3' NDR increases a gene's plasticity by allowing access to an environmentally-responsive platform for antisense initiation. This is in accordance with a recent demonstration that nucleosomes at the 3' end supress the initiation of antisense transcripts (Yadon et al. 2010).

In the above plasticity analyses genome-wide maps of TFBSs, transcripts and nucleosome occupancies have been utilised, which were collected under a limited set of conditions. Here, they were used as approximations to hypothetical, idealised maps; ones that would capture transcripts, TFBSs and nucleosome occupancies both genome-wide and across a much wider set of conditions. Ideally it would be useful to classify those genes that are capable of making antisense transcripts – at least under certain conditions - and those which are incapable – under all conditions - and to determine if the former are more transcriptionally plastic. Here it was shown, using the collated transcript map, that high plasticity genes are more likely to contain an antisense transcript than those that do not (Figure 11A-B) - *even though* plasticity is a parameter collected across a broad range of conditions, not just those under which the antisense transcript maps were determined. It is possible, therefore, that antisense transcripts are generally present at genes capable of producing them, though at levels that vary across environmental conditions, while other, low plasticity genes, can never make antisense transcripts. Thus it is the changing levels of antisense (a consequence of changing levels of TFBS binding) that grant the gene its enhanced capacity to vary its expression with changing environmental conditions, making it more transcriptionally plastic. It would be of interest to obtain maps of antisense transcripts under a broader range of conditions, to identify which antisense transcripts vary the most i.e. which are themselves transcriptionally plastic.

That antisense transcription *is* associated with high transcriptional plasticity, and that it is an abundant process genome-wide in spite of its extensive degradation, leads to the hypothesis that antisense transcription is directly modulating the sense promoter. In the next chapter the correlations between sense and antisense transcription are assessed, while chapter 5 looks directly at how antisense transcription might be changing the sense promoter.

Chapter Four

Exploring the relationship between sense and antisense transcription genome-wide

4. Exploring the relationship between sense and antisense transcription genome-wide

4.1. Summary

Antisense transcription is prevalent genome-wide, with approximately one third of genes containing a defined antisense transcript, with the average levels of nascent antisense transcription at these genes being approximately a half of the average level of nascent sense transcription. Antisense transcription is also associated with sense transcriptional plasticity, in line with a model whereby it changes the level of sense transcription. Here, the same nascent transcription sequencing data (NET-seq) as utilised in the previous chapter is used to identify whether sense and antisense transcription are correlated genome-wide, and establish whether antisense transcription plays a predominantly repressive or activatory role. Strikingly, however, it is found that sense and antisense transcription are *not* correlated genome-wide, which has broad implications for understanding gene regulation by antisense transcription. Furthermore, NET-seq is compared with traditional RNA-seq in terms of their estimates of genome-wide antisense transcription, to assess the contribution of antisense transcript stability to genome-wide measurements. NET-seq results in higher estimates of antisense to sense transcription ratios at the same gene than RNA-seq, suggesting that the degradation machinery does indeed affect sense and antisense transcripts differentially, and that NET-seq is a better method for assessing levels of antisense transcription.

4.2. Introduction

4.2.1. Transcriptomics aims to catalogue and quantify transcripts and transcription on a genome-wide scale

As discussed in chapter 1, transcription lies at the centre of gene of gene expression. In understanding transcription on a genome-wide scale, it has been necessary to determine the complement of transcripts in a particular organism and the level to which they are transcribed. Both these features are dynamic; the precise set of transcripts and the levels of those transcripts will vary between cell types, under different environmental conditions, and even within a genetically identical population of cells (as discussed in chapter 6). Genome-wide approaches allow for the assessment of global levels and identity of transcripts, however they also suffer from certain limitations which are relevant to this work, and which will be described below.

DNA microarray technologies are based around the principle of nucleic acid hybridization, in which one can quantitatively measure levels of known sequences in a fluorescently labeled cDNA library by their binding to immobilized probes. They have been instrumental in the study of transcriptomics, and are notable for being both high throughput and relatively inexpensive compare to other technologies. This has allowed for extensive studies of gene regulatory networks in yeast, assessing how the transcriptome changes following hundreds of different mutations – Lenstra et al., (2011) assessed global transcript levels in 165 mutants to assess how the chromatin regulating machinery influenced sense expression. Microarrays that assess transcripts genome-wide at a particularly high density are called tiling arrays, and use short, overlapping probes that cover a large proportion of the genome. Tiling arrays were used to identify antisense transcripts in the yeast genome, though, importantly, a large fraction of these (the CUTS) were only identifiable following deletion of *RRP6* (Neil et al. 2009; Xu et al. 2009). Microarrays have several limitations, however, and for the purposes of this work

the most crucial of these are a low signal-to-noise ratio resulting from cross hybridization and a limited dynamic range, itself a result of high background but also signal saturation (Wang et al. 2009; Royce et al. 2007; Okoniewski and Miller 2006). With regards to antisense transcription and pervasive transcription in general, this means that low level transcripts may be hard to detect, which would result in an underestimation of the true extent of transcription.

The advent of massively parallel sequencing has led to its use in the study of transcriptomics. RNA sequencing (RNA-seq) typically involves the reverse transcription of RNA from an organism to produce a cDNA library. The cDNA or else the RNA itself is fragmented, and adaptors are attached to either one or both ends of the resulting fragments. Sequencing is then performed in one or both directions to obtain short reads that can be mapped to the genome. The number of reads corresponding to a given region of the genome is proportional to the number of transcripts containing sequences corresponding to that same region, thus RNA-seq provides a quantitative assessment of transcript levels. Indeed, RNA-seq can assess transcript levels over a much larger dynamic range than microarrays (Shendure 2008). However, RNA-seq still strictly assesses the levels of transcript rather than transcription from an organism. Steady state levels of a given transcript are a function not only of its rate of synthesis but also of its rate of degradation. As discussed in chapter 1, antisense transcripts are particularly prone to degradation. This has two important ramifications: 1) that any regulation resulting from antisense transcription is likely to be mediated by the transcription itself rather than the transcript and 2) that deep sequencing of steady state transcripts should underestimate the true extent of antisense transcription. Recently however, a technique based upon RNA-seq has been developed that should overcome this limitation with regards to antisense transcription.

4.2.2. NET-seq allows for measurement of nascent transcription

Native elongating transcripts sequencing (NET-seq) involves deep sequencing of the 3' ends of nascent transcripts purified from the DNA-RNA-RNAPII complex (Churchman and Weissman 2011) (Figure 21). In essence, this allows for a quantitative assessment of elongating RNAP across the genome, and – crucially for the purposes of this work – in a strand-specific manner. The global assessment of RNAPII levels by NET-seq was in agreement with nuclear run on experiments that showed RNAPII levels to be highest within the early coding region of genes (Rodrıguez-Gil et al. 2010). Churchman and Weissman (2011) were also able to demonstrate that purified nascent transcripts did not undergo degradation, thus NET-seq represents a means of assessing strictly transcript synthesis. In support of this, NET-seq was found to show higher levels of the CUT upstream of the *URA1* gene than RNA-seq (Figure 22). Note, however, that unlike RNA-seq, NET-seq is sensitive to polymerase pausing, as shall be discussed below.

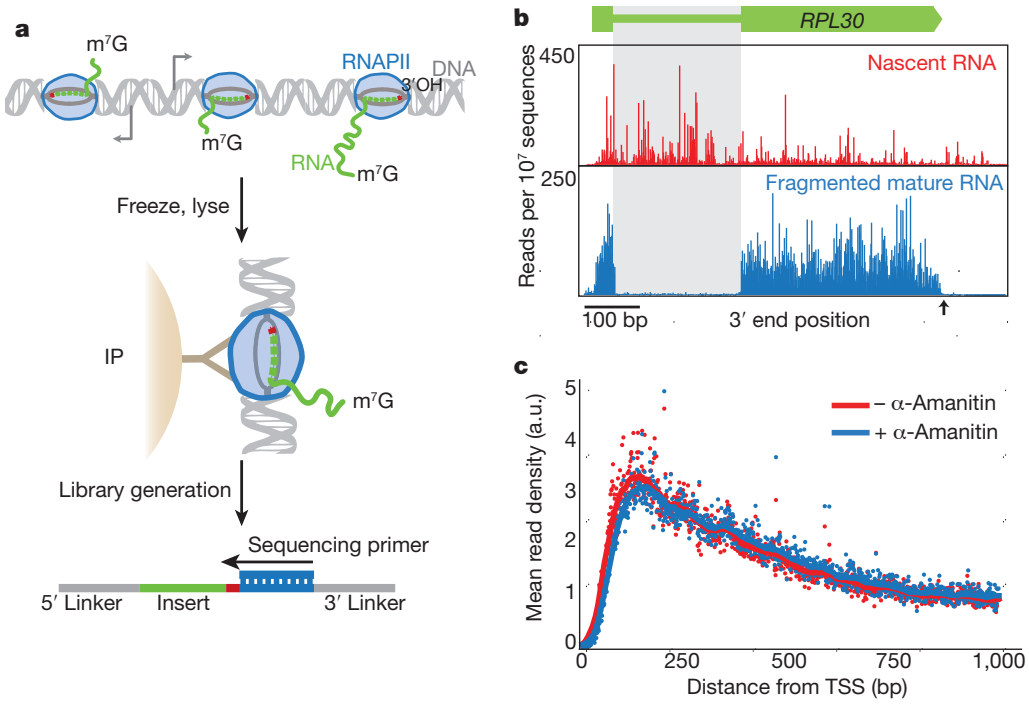


Figure 21: NET-seq allows for the assessment of nascent transcription

(A) NET-seq involves the immunoprecipitation of RNAPII following cell lysis. The nascent RNA is then purified from the DNA-RNA-RNAPII complex and subject to sequencing. **(B)** A comparison of transcription at the *RPL30* gene, as assessed by NET-seq (top) and RNA-seq (bottom). Evident are the effects of gene splicing upon mature RNA. NET-seq, however, captures transcription across the intron, and also downstream of the 3' end of the mature transcript, the observation of which is precluded in RNA-seq due to 3' end processing. **(C)** Average density of nascent transcription across a gene. The transcriptional inhibitor α-amanitin was added after cell lysis to demonstrate that transcription did not continue during processing of the cell lysate. Taken from Churchman and Weissman (2011).

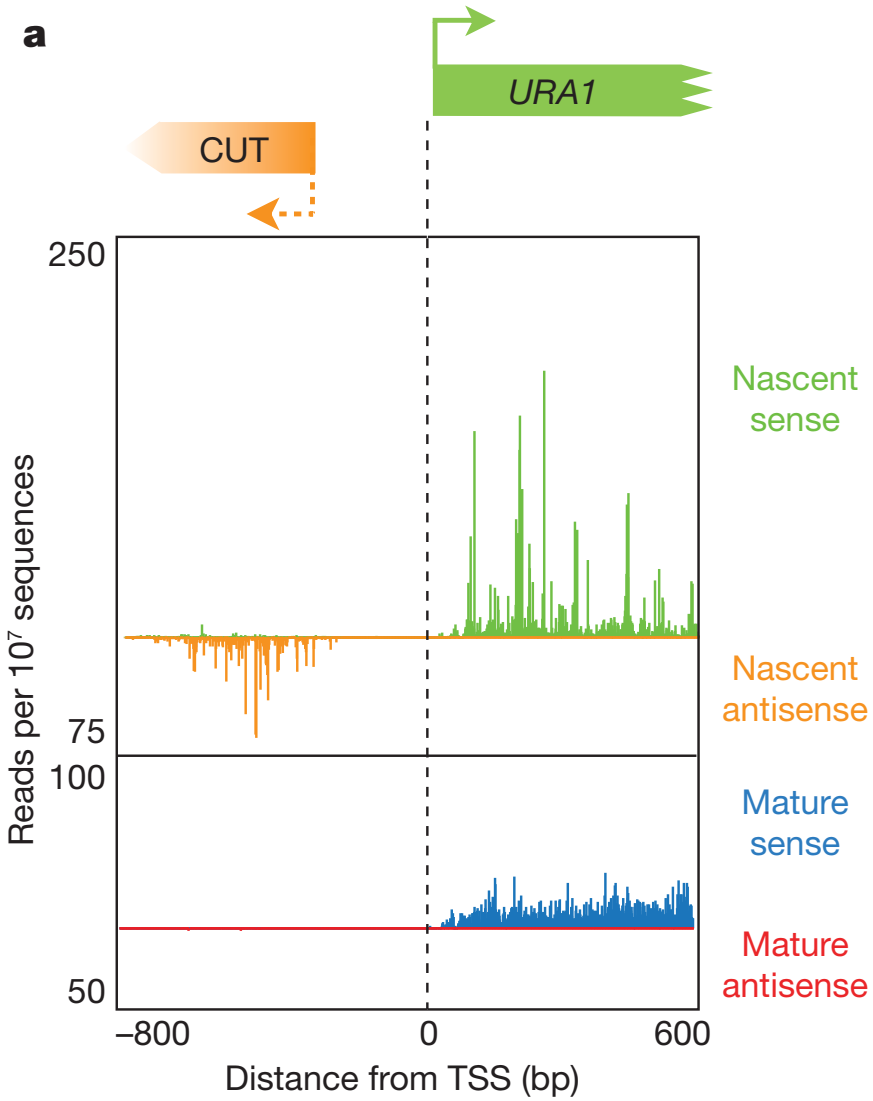


Figure 22: NET-seq can assess cryptic transcription

A comparison of transcription at the *URA1* gene, as assessed by NET-seq (top) and RNA-seq (bottom). Both methods predict similar levels of sense transcription of *URA1.* However, the upstream CUT shows much higher levels when assessed by NET-seq as opposed to RNA-seq, a consequence of it being degraded by Rrp6. Taken from Churchman and Weissman (2011).

The ability of NET-seq to measure strand-specific, nascent transcription makes it an ideal method for assessing the extent of antisense transcription genome-wide, and for accurately determining the amount of antisense transcription at a given gene. NET-seq is likely to be a more appropriate method in this respect than the use of tiling arrays or RNA-seq in conjunction with deletions of components of the degradation machinery,

especially given that such components might have roles in transcription initiation, distinct from roles in degradation (Haimovich et al. 2013).

Here, NET-seq and RNA-seq are comprehensively compared in their ability to estimate antisense transcription. The aim is to determine whether the gap between average levels of sense and antisense transcription is smaller when using NET-seq than when using RNA-seq, and so validate its use in this work over more traditional methods.

4.3. Results

4.3.1. NET-seq predicts a higher ratio of sense to antisense transcription over a gene

If antisense transcripts are more extensively degraded than sense transcripts, then one would expect the average genome-wide ratios of antisense to sense transcript levels to be lower than the ratio of antisense to sense *transcription*. This would confirm the hypothesis that using RNA-seq as a method of determining levels of antisense transcription generally underestimates the true extent of antisense transcription genome-wide. To test this, the ratios of sense and antisense reads were determined at a number of different positions, using the RNA-seq and NET-seq data from Churchman and Weissman (2011). The distributions of antisense/sense transcription ratios were then compared between experiment types. Importantly, the total number of uniquely mapped reads in the RNA-seq data and NET-seq data was similar (17 compared to 20 million), meaning that the genome-wide coverage itself was comparable. Dubious genes were excluded from the analysis, as were genes overlapping other coding genes in a convergent manner, leaving 5,412 genes in the analysis.

Sense and antisense read ratios were assessed in four different windows (Figure 23), in part to assess the robustness of the results. Initially, the ratio of antisense and sense transcription/transcript reads was compared within a 300bp window placed

immediately downstream of the TSS (Figure 23A) – i.e. the average number of antisense reads within the window divided by the average number of sense reads. For sense transcription, this was an attempt to capture the rate of sense initiation. For antisense transcription, this was an attempt to measure the amount of antisense transcription coming *in* to the sense promoter, a parameter that will be considered more extensively in chapter 5. Next, the ratio of *bidirectional* antisense and sense transcription was considered, using the same 300bp window as before to determine sense transcription but a 150bp window placed 400bp upstream to determine antisense transcription (Figure 23B). This was an attempt to approximately measure the amount of antisense transcription upstream of the -1 nucleosome, where antisense transcription initiation is thought to occur (chapter 3; Neil et al., 2009). Removed from this analysis were any divergent genes – i.e. those whose promoters direct coding transcription in both directions, hence the smaller *n* number. Next, the amount of antisense transcription in a 300bp window at the 3' end was compared to the amount of sense transcription in the 5' 300bp window described above, in an attempt to compare the ratio of antisense and sense transcription initiation at the same gene (Figure 23C). Finally, the ratio of antisense and sense transcription across the sense transcription unit was compared, in an attempt to compare total transcription on both strands (Figure 23C).

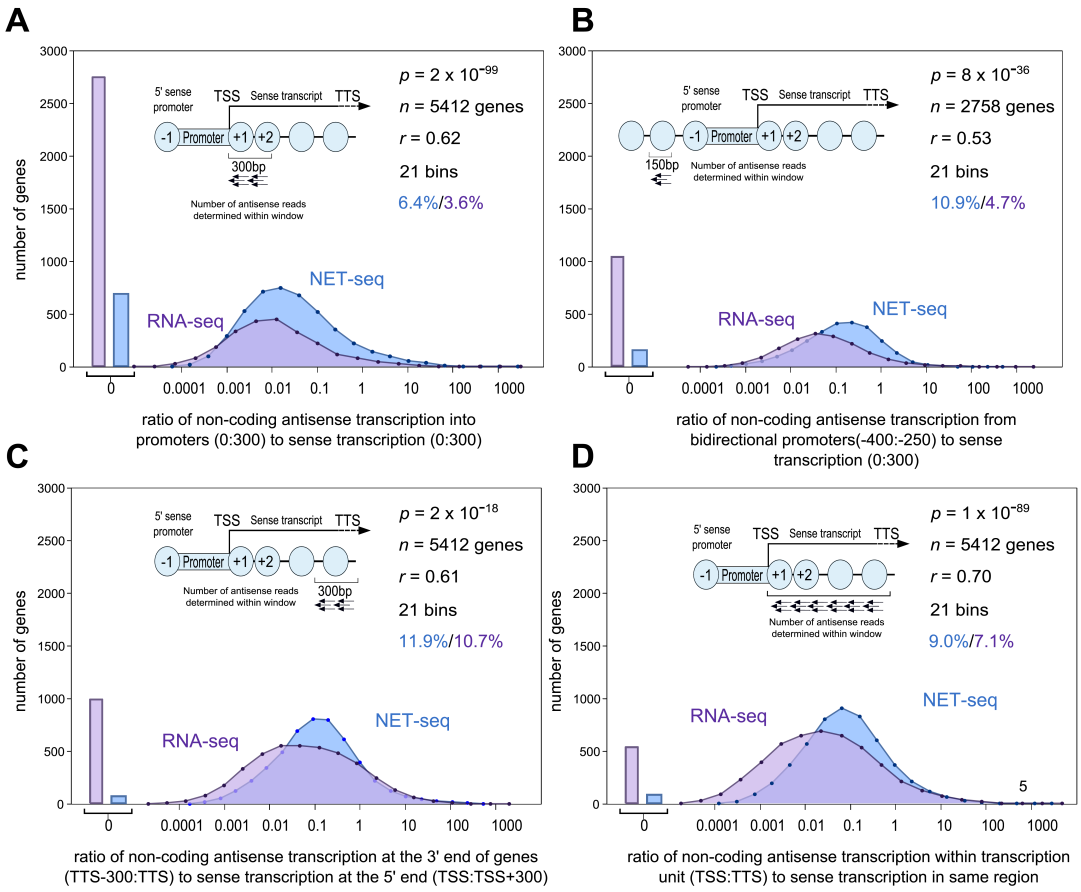


Figure 23: NET-seq estimates higher levels of antisense transcription than RNA-seq.

(A-D) Histograms showing the distributions of the ratios of antisense reads to sense reads determined by RNA-seq (purple) and NET-seq (blue) within different gene windows: **(A)** The ratio of antisense reads in a 300bp window immediately downstream of the TSS compared to sense reads within the same window. **(B)** The ratio of antisense transcription in a 150bp window 250bp upstream of the TSS compared to sense transcription within the 300bp window used in A. Divergent genes were excluded from this analysis. **(C)** The ratio of antisense transcription in a 300bp window immediately upstream of the TTS compared to sense transcription within the 300bp window used in A. **(D)** The ratio of antisense to sense transcription within the transcription unit. Shown are the p-values determined by comparing the distributions using the Wilcoxon rank-sum test (p), the number of genes under consideration (n), and the Spearman rank correlation coefficient determined by comparing the RNA-seq and NET-seq ratios (r). Data was divided into 21 bins, including a bin for those windows with no antisense reads. The percentages refer to the percentage of genes in which there is more antisense transcription than there is sense transcription in the given window. Note that the total number of reads in the RNA-seq data and NET-seq data was similar (17 compared to 20 million), thus the large differences in 0 antisense read counts observed in all cases are not simply a result of differences in total read count. It is notable that the distributions of antisense/sense NET-seq reads are similar between the bidirectional antisense transcription and the 3' antisense transcription (B-C). This is to be expected if the 3' ends of most genes border the 5' ends of others, such that the different windows will be capturing similar values.

In all cases, the average ratio of antisense to sense transcription was significantly higher when determined by NET-seq than when determined by RNA-seq (Figure 23). Furthermore, there were far fewer instances of genes having no antisense reads in NET-seq compared to RNA-seq, despite the overall coverage of both experiments being roughly similar, suggesting that NET-seq is more sensitive in detecting low levels of antisense transcription. NET-seq also consistently identified a higher fraction of genes that had more antisense reads than sense reads within the windows being compared. Intriguingly, however, the Spearman rank correlation between ratios as determined by NET-seq and RNA-seq were moderately high (with r ranging from 0.53-0.7). This means that, although the average ratios are significantly different between both experiments, the ratios themselves are still well correlated. Thus, while sense and antisense transcripts are differentially degraded – resulting in a difference in average ratios between NET-seq and RNA-seq – when one considers each transcript type individually one observes a more uniform degradation, resulting in such moderately high correlations. Evidently, however, there are still differences between antisense transcripts in terms of degradation: r^2 values of 0.28-0.49 indicate that antisense/sense transcription ratios as determined by NET-seq can only predict between 28-49% of the variation in antisense/sense transcript ratios as determined by RNA-seq, hence antisense transcripts do not all experience degradation to the same extent – in support of the existence of SUTs, CUTs and XUTs as different 'species' of antisense transcript (Neil et al. 2009; Xu et al. 2009; van Dijk et al. 2011).

Based on this analysis, it appears that NET-seq estimates higher ratios of antisense to sense ratios than RNA-seq. This leads to the conclusion that transcript degradation impairs the ability of RNA-seq to estimate the true extent of antisense transcription genome-wide. In chapter 5, how genes vary depending upon how much antisense transcription is coming into the sense promoter is assessed, to explore

possible regulatory roles for antisense transcription. Clearly, RNA-seq predicts particularly low levels of antisense transcription in this window compared to NET-seq – about half of yeast genes show no antisense RNA-seq reads within this window, compared to only a tenth when determined by NET-seq (Figure 23A). Based on these observations, NET-seq data is used exclusively in subsequent analyses to determine the amount of sense and antisense transcription at genes.

This analysis provides insight into the prevalence of antisense transcription genome-wide. Only a very small number of genes have no antisense NET-seq reads whatsoever – less than two hundred in all cases except Figure 23A. Strikingly, when considering total transcription across the sense transcription unit, 9% of genes had more antisense reads (as determined by NET-seq) than sense reads (Figure 23D). Thus antisense transcription predominates in nearly one tenth of the genome. One might consider whether – at this 9% - levels of sense transcription are low, in line with a repressive model for antisense transcription. As such, it is next considered whether sense and antisense transcription are correlated genome-wide.

4.3.2. Sense and antisense transcription are not correlated at the sense promoters of genes

Sense and antisense transcription were correlated using the same windows as those in Figure 23. Sense and antisense transcription were correlated within a 300bp window immediately downstream of the sense TSS, to attempt to assess the relationship between sense initiation *from* the sense promoter and antisense transcription coming *into* the sense promoter. Antisense transcription generally proceeds into the sense promoter (Xu et al. 2011), and one might hypothesise that this could result in a regulatory effect. If antisense transcription is generally repressive, then one would expect to see a negative correlation between sense and antisense transcription in this

300bp window. However, it was found that sense and antisense transcription were *not* correlated at this window (Figure 24A) (Spearman rank correlation coefficient = -0.023). This suggests that antisense transcription is not generally repressive of sense transcription.

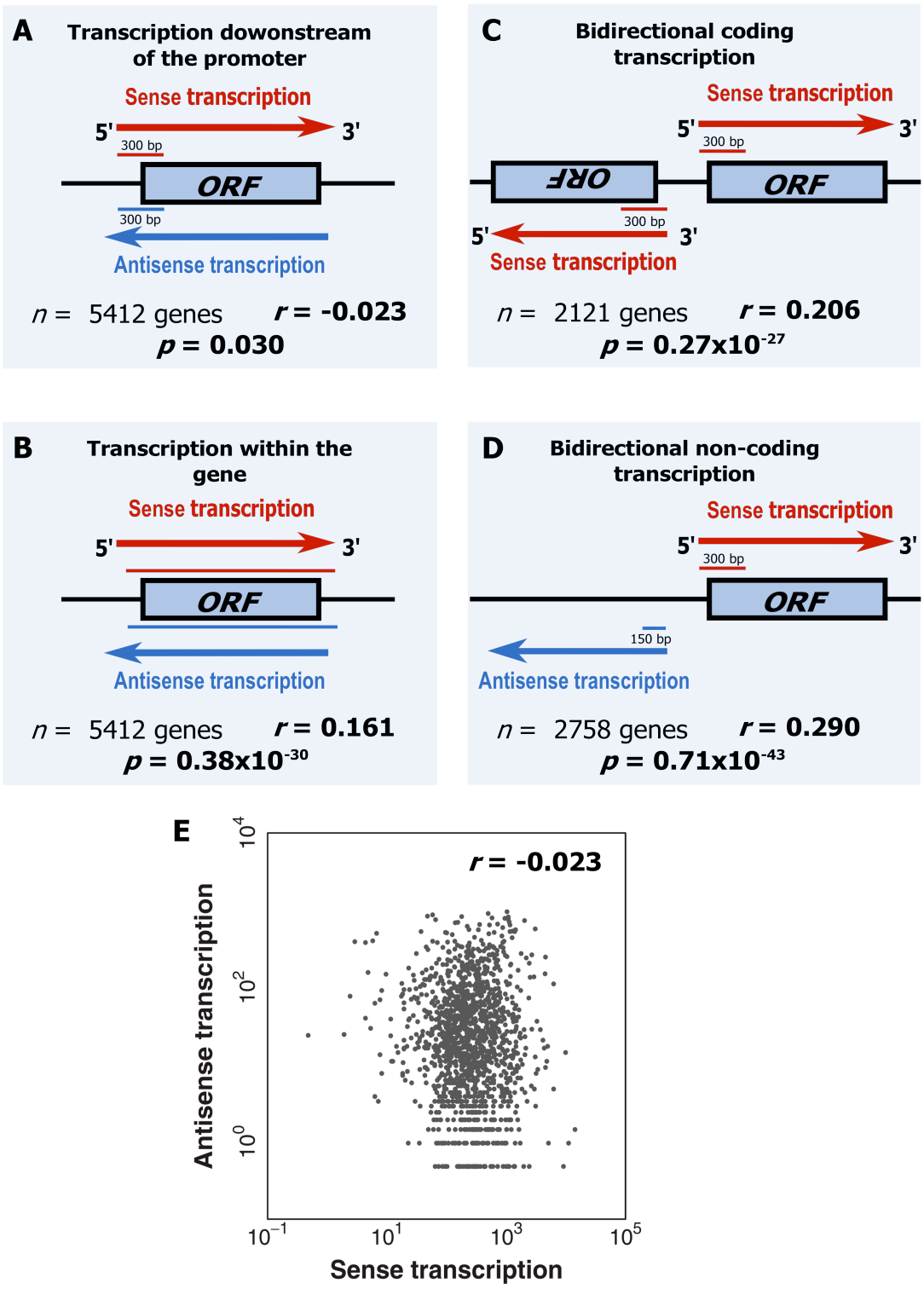


Figure 24: Correlations between different sorts of transcription at genes and gene promoters

(A-D) A series of schematics showing different windows in which the levels of nascent transcription were determined. In each case the levels of transcription in two distinct windows were correlated for a set of n genes, to obtain the Spearman correlation coefficient r. Correlations were as follows: **(A)** sense and antisense transcription within a 300bp immediately downstream

of the gene's TSS, representing sense initiation and antisense transcription entering the promoter region respectively. **(B)** Sense and antisense transcription across the whole transcription unit. **(C)** Coding gene transcription from divergently oriented gene pairs - specifically genes whose TSSs were within 600bp of another gene, using 300bp windows placed immediately downstream of the TSS. **(D)** Coding sense transcription and non-coding antisense transcription from bidirectional promoters. The window to determine antisense transcription was placed 250bp upstream of the gene TSS. Genes were excluded which had the TSS of a divergent coding gene within 1000bp upstream of its TSS. *n* reports the number of genes considered in each analysis. **(E)** Scatter plot showing the distribution of genes in terms of levels of both sense and antisense transcription within the 300bp window described in **A.**

Assessing sense and antisense transcription across genes, it was found that they were weakly positively correlated (Figure 24B) (Spearman rank correlation coefficient = 0.161). That this should differ from the lack of correlation seen in Figure 24A is curious. It is possible that high amounts of sense transcription do result in an increase in cryptic transcription from internal promoters, as suggested by Struhl, (2007). Several studies have shown that deletion of chromatin regulators, such as components of the Rpd3S and Set3 deacetylase complexes, can result in the production of short, spurious transcripts arising from cryptic, internal promoters (Carrozza et al. 2005; Kim et al. 2012). Furthermore, Mason and Struhl (2003) found that loss of the FACT complex resulted in similarly short, spurious transcription which, interestingly, was associated with an increase in RNAPII density further downstream of the 5' end of the gene, but *not* at the 5' end itself. It is possible, then, that an increase in cryptic transcription resulting from high sense transcription might result in production of short transcripts that do not typically reach the 5' 300bp window used in Figure 24A, hence explaining why no positive correlation is observed here. Alternatively, sense and/or antisense transcription might result in pausing of the opposing transcription process, resulting in the polymerase spending more time on the gene and leading to an increase in NET-seq reads. Indeed, it is shown in chapter 5 that antisense transcription is associated with a rise in nucleosome occupancy across the body of the gene, which could delay processing polymerases.

As discussed in the previous chapter, high levels of antisense transcription can be supported in the absence of sense transcription, however, it is evident that there is still a certain degree of co-regulation at bidirectional promoters (Spearman rank correlation coefficient = 0.290 and 0.206; Figure 24C-D). This could likely be due to bidirectional transcription arising from within the same NDR (Rhee and Pugh 2012; Neil et al. 2009; Murray et al. 2012) – i.e. establishing a suitable NDR for one direction inevitably sets it up for transcription in the other. Alternatively, binding of transcription factors or associated chromatin remodelling/histone modifying enzymes could establish a favourable environment for both.

4.4. Discussion

4.4.1. NET-seq predicts higher genome-wide levels of antisense transcription than RNA-seq

Here, it is shown that the ratios of antisense to sense transcription at various windows around genes are higher when one uses NET-seq to estimate sense and antisense transcription than when one uses RNA-seq. This is likely due to NET-seq's ability to measure nascent transcription, as opposed to steady state levels of RNA, which is a function of both the rate of RNA synthesis but also the rate of degradation. Because sense and antisense transcripts are differentially processed (Wyers et al. 2005; Xu et al. 2009; van Dijk et al. 2011), RNA-seq overestimates the true extent of sense transcription while underestimating antisense transcription.

There is a possible issue with using NET-seq to estimate levels of transcription generally. In determining transcription levels here it is assumed that the rate of RNAPII procession is always the same. However, polymerase procession could be slowed, for example by the chromatin template (Kirov et al. 1992). Thus while two genes might have the same overall level of RNAPII as determined by NET-seq, in one instance this

could be due to high rates of RNAPII passage, while in another it could be due to high RNAPII pausing. As discussed above, pausing might explain the weak positive correlation between sense and antisense transcription across a gene – a non-permissive chromatin template will likely slow transcription on both strands. However, given that this correlation is only small, and also that antisense transcription as determined by NET-seq appears to be well correlated with levels of the pre-initiation components TBP and TFIIB (chapter 3), it can be concluded that NET-seq remains an effective method of measuring antisense transcription levels.

It is worth noting that the total number of uniquely mapped reads in the RNA-seq data and NET-seq data was not the same (17 compared to 20 million). It was assumed that by determining ratios of sense to antisense reads across a gene, one would be able to compare ratios fairly between RNA-seq and NET-seq, without correcting for this difference. However, due to the nature of sequencing data, this is not the case. For example, the discrepancy in total read count could account for the larger number of genes with 0 antisense reads in the RNA-seq dataset compared to NET-seq. One way to correct this would be to randomly remove reads from the NET-seq data, such that the overall total mapped reads between both experiments was the same (17 million).

4.4.2. NET-seq predicts higher genome-wide levels of antisense transcription than RNA-seq

The most striking finding of this work is that sense and antisense transcription are not even modestly negatively correlated – an increase in antisense transcription is not associated with a decrease in sense transcription or vice versa. This goes against the findings of single gene studies that suggest a repressive model (Camblong et al. 2007; Xu et al. 2011; Hongay et al. 2006). In the following chapter the association of antisense transcription with gene chromatin is assessed, to consider how antisense transcription

might be altering gene structure and behaviour without necessarily altering the level of sense transcription. As such, the lack of correlation described here will be considered more extensively in the following chapter in light of these findings.

Chapter Five

Towards the function of antisense transcription: associations with gene chromatin and its modification

5. Towards the function of antisense transcription: associations with gene chromatin and its modification

5.1. Summary

Antisense transcription is abundant in the yeast genome, but strikingly is not correlated with the level of sense transcription at the same gene. This raises questions as to its function and as to why it might be associated with transcriptional plasticity, as discussed in chapter 3. Antisense transcripts are often rapidly degraded, suggesting that it is the act of transcription rather than the transcript itself that could be exerting regulatory effects upon the gene. The passage of an elongating RNAPII complex is known to modulate chromatin and chromatin modifications, and to recruit factors such as chromatin remodelling enzymes, all of which could modulate gene behaviour. In this chapter, genes with high levels of antisense transcription are compared to those with low levels of antisense transcription. The aim is to explore the hypothesis that antisense transcription might alter properties of a gene; both at the sense promoter and within the coding region. Strikingly, major differences are identified in the levels of nucleosome occupancy and turnover, as well as a broad array of histone modifications. Furthermore, a number of chromatin remodelling enzymes and histone chaperones are found to be enriched at genes subject to antisense transcription. Intriguingly, the associations between antisense transcription and chromatin are distinct from those between *sense* transcription and chromatin, suggesting the two processes may be fundamentally different. A model is developed in which antisense transcription brings about changes in histone modification not through the recruitment of modifying activities but via turnover, and it is proposed that these changes may affect gene behaviour.

5.2. Introduction

Work in the previous chapter has shown that NET-seq estimates higher levels of antisense transcription genome-wide than RNA-seq, and allows assignment of a quantitative measure of antisense transcription to genes. Here, it is hypothesised that antisense-transcribing RNAPII entering the sense promoter should alter its structure, using the number of NET-seq antisense reads immediately downstream of the sense promoter as a measure of how much antisense-transcribing RNAPII the promoter is subject to. Questions are addressed regarding how genes subject to high levels of such antisense transcription might be associated with changes in chromatin. To begin, it is important to review our understanding of chromatin, its modification and modulation, and to assess the known interactions between elongating polymerase and chromatin.

5.2.1. Promoter structure

The basic structure of the yeast promoter has been discussed in chapter 3. Briefly, promoters are flanked by two nucleosomes between which lies the NDR. Nucleosomes are not typically fixed to the same spot but experience a certain amount of movement whilst still bound to the DNA (Chakravarthy et al. 2005). Attempts to assign defined genomic positions to nucleosomes – by mapping them to genome-wide data – account for this by ascribing each nucleosome a measure of 'fuzziness' based on the standard deviation of its mapped location (Albert et al. 2007). In reality, however, nucleosomes are better described through 'occupancy' rather than position; reflecting the amount of time a given position is filled by nucleosomes. Because nucleosome positioning is dynamic the bordered region of the promoter is described here as nucleosome 'depleted' rather than 'free'. This region contains *cis*-acting sequences (i.e. transcription factor binding sites), and is where the pre-initiation complex assembles (Venters and Pugh 2009).

5.2.2. Nucleosome properties

As well as the position, occupancy and 'fuzziness' described above, nucleosomes are distinguished by a number of parameters which can be defined bioinformatically using available genome-wide data. Chief amongst these are the particular combination of post-translational modifications on the tails of the histone subunits – termed histone modifications – but also two additional features: namely subunit composition and rate of turnover (the rate at which DNA-bound histones are replaced with free histones from the nuclear pool). These nucleosome properties are under the governance of three distinct classes of enzymes that will be discussed below: position and fuzziness by chromatin remodelling enzymes, modification state by histone modifying enzymes, and turnover and composition by histone chaperones.

5.2.3. Histones are modified on their N- and C- terminal tails

Histone proteins have flexible N- and C-terminal tails that protrude through channels formed by aligned DNA minor grooves (Luger et al. 1997a), making them accessible for modification. Though these tails do not contribute towards nucleosome stability or rotational positioning of the DNA on the nucleosome (Luger et al. 1997b) it is possible that they are involved in the assembly of nucleosomes into higher order chromatin fibres, something which particular modifications could serve to enhance or mitigate (Luger and Richmond 1998).

There is a host of different types of histone modification that can be found on these tails, including acetylation, methylation, phosphorylation, ubiquitination and sumoylation. These modifications can occur at numerous sites upon histone tails, many of which shall be discussed below. The presence of these modifications can be associated to varying degrees with particular gene states, such as active or inactive, or

with active or silent chromatin. This has led to the proposal of a histone code – that combinations of different modifications can somehow be 'read', with functional consequences for the gene (Strahl and Allis 2000; Turner 2000; Jenuwein and Allis 2001). Whether or not it can truly be considered a code or indeed something more nuanced (Gardner et al. 2011), histone modifications represent an important layer of additional complexity, one that must be considered in any analysis of transcription regulation. Here, known associations and possible functions for a range of well-studied modifications will be discussed.

5.2.4. Histone acetylation at promoters is generally associated with active genes

All four histone subunits in *S. cerevisiae* can undergo acetylation at numerous different lysine residues, though this discussion will be restricted to acetylation of histone H3. The formation of lysine acetylation is brought about by histone acetyltransferases and their removal by histone deacetylases.

Histone acetylation is broadly thought to predominate at the promoters of genes (Liu et al., 2005), where it is a marker of gene activity: acetylation at H3K9, H3K14, H3K18, H4K5 and H4K12 are all associated with high levels of transcription (Pokholok et al. 2005). Indeed, there is evidence to suggest that, rather than just being a marker, acetylation can itself exert positive effects upon transcription – genes with high promoter levels of H3K4 acetylation (H3K4ac) show reduced gene expression following mutation of the H3K4 residue (Guillemette et al. 2011). How might acetylation exert such effects? Acetylation neutralises the positive charge upon lysine residues, which could potentially weaken both the interactions between histone tails and DNA and between adjacent nucleosomes, and so could conceivably disrupt chromatin compaction. In support of this, histone tail removal and acetylation both result in increased accessibility of nucleosomal DNA (Anderson et al. 2001) and nucleosome

instability (Widlund et al. 2000). Another possibility is that, rather than causing direct structural effects, acetylation recruits *trans*-acting factors that bind the acetylated residues. Such *bromodomain* proteins include, for example, Rsc4, part of the RSC chromatin remodelling complex, which is recruited to H3K14ac via tandem bromodomains (Kasten et al. 2004). A final possibility is that acetylation expedites the passage of polymerase into elongation or indeed directly increases the rate of elongation. In support of this, acetylation enhances transcription by RNA polymerase III in an *in vitro* system composed of phased nucleosomes (Tse et al. 1998). However, H4K8ac and H4K16ac, both of which are predominantly found within the coding regions of genes, have been shown to negatively correlate with transcription (Liu et al., 2005), while deacetylation by Hos2 is necessary for transcription activation (Wang et al. 2002), thus the relationship between acetylation and activity in the gene body may not be so straightforward.

5.2.5. Methylation of H3K4 is partitioned across active genes

Methylation of H3K4 occurs in three states: mono- di- and tri- (Strahl et al. 1999). The roles of these states are not equivalent: each has its own reported functional consequences and associations. Dimethylation, for example, is found in active euchromatic regions but not in silent heterochromatic sites (Noma et al., 2001), while trimethylation is associated with active genes (Santos-rosa et al. 2002). Where present, these three modifications tend to be partitioned across the gene body, with trimethylation peaking downstream of the TSS, followed by di- and monomethylation downstream (Liu et al. 2005; Pokholok et al. 2005) – itself suggestive of distinct functions for each modification. Methylation of H3K4 is carried out by the Set1-containing complex COMPASS (Briggs et al. 2001), which can be recruited via the Paf1 complex and by the Ser5-phosphorylated CTD of RNAPII (Krogan et al. 2003a; Ng et al.

2003), providing a possible explanation for why active genes are more extensively methylated.

H3K4 methylation can also effect downstream changes. H3K4me3 can bind to the PHD finger domain of Yng1, a component of the NuA3 histone acetyltransferase complex, as well as Sgf29 of the SAGA complex (Bian et al. 2011) and so bring about acetylation of H3K9/K14/K18 (Taverna et al. 2006), possibly explaining why H3K4me3 is so often found on the same H3 molecule as acetylation marks (Jiang et al. 2007). As for acetylation on H3K4, trimethylation by COMPASS limits H3K4ac (Guillemette et al. 2011), thus these modifications are to an extent mutually exclusive, though intriguingly both are associated with active genes, and given that nucleosomes have two H3 tails it is entirely possible to have both marks present on the same nucleosome. H3K4me3 can also recruit the chromatin remodelling enzyme Isw1 (Santos-Rosa et al. 2003), and so could play a direct role in gene activation. However, there is also evidence that H3K4 methylation can have a repressive effect. Deletion of *SET1* can lead to increased expression at certain genes (Carvin and Kladde 2004), while levels of H3K4me3 increase at ribosomal genes following repression (Weiner et al., 2012). Evidently, methylation of H3K4 is as nuanced as its acetylation, lending weight to the argument that the effects of a particular modification are highly context-dependent.

5.2.6. Trimethylation of H3K36 and H3K79 occurs in gene bodies and is associated with stable nucleosomes

H3K36 trimethylation is found within the body of active genes, where it is deposited by Set2 following recruitment to Ser2-phosphorylated RNAPII (Strahl et al. 2002; Krogan et al. 2003b; Li et al. 2003), and removed by the demethylases Rph1 and Jhd1 (Kim and Buratowski 2007; Tu et al. 2007). H3K36me3 is necessary for the recruitment of the Rpd3S deacetylase complex, through which it maintains the gene body in a hypoacetylated state (Carrozza et al. 2005; Govind et al. 2010). Failure to deacetylate in

this way results in spurious transcription (Li et al. 2007) – something which is not necessarily distinct from antisense transcription, thus one might predict that the promoters of antisense transcripts would be depleted of H3K36me3. Intriguingly, H3K36me3 has also been found to suppress histone turnover. Using a technique described below, Venkatesh et al., (2012) demonstrated that disruption of H3K36me3 resulted in increased histone turnover. Thus, H3K36me3 may represent a marker of nucleosome 'residency' – that is, the level of HK36me3 at a given genomic region could represent the average amount of time a histone spends bound at that region.

H3K79me3, like H3K36me3 is deposited on gene bodies (Pokholok et al. 2005), in this case by the Dot1 methyltransferase (Ng et al. 2002; van Leeuwen et al. 2002). However, unlike H3K36me3 it was not found to be strongly correlated with gene activity (Pokholok et al. 2005), but rather appears to play a role in euchromatin maintenance, by keeping the Sir2 and Sir3 silencing proteins from the modified regions and so restricting them to the telomeric regions (Ng et al. 2002; van Leeuwen et al. 2002). Like H3K36me3, H3K79me3 also has a known association with nucleosome stability. Combining mass spectrometry with mathematical models of H3K79 methylation, De Vos et al., (2011) found that H3K79me3 accumulates on histones following their assembly as nucleosomes. Thus H3K79me3 also represents a marker of nucleosome residency.

5.2.7. Ubiqutination of H2BK123

Monubiquitination of H2BK123 (the C-terminus) is associated with active transcription, and is absent from transcriptionally silent genes (Xiao et al. 2005; Kao et al. 2004). Indeed, RNAPII passage contributes to ubiquitination through interaction with the ubiquitin-conjugating enzyme Rad6 and the ubiquitin ligase Bre1, both of which are tethered by the PAF complex and the serine 5-phosphorylated CTD of RNAPII (Henry et al. 2003; Hwang et al. 2003; Xiao et al. 2005). Furthermore, H2B ubiquitination has been

shown to be required for nucleosome reassembly during RNAPII passage across a gene, working in conjunction with Spt16 (Fleming et al. 2008). Thus H2B123ub, like H3K36me3 and H3K79me3, also has known links to nucleosome stability. This stability has been shown to inhibit the assembly of the transcription machinery at inactive promoters, while promoting transcription elongation within the bodies of active genes (Batta et al. 2011).

5.2.8. H2A.Z

H2A.Z is a histone variant that has been shown to be present at approximately two-thirds of yeast genes (Guillemette et al. 2005), where it is most frequently found at the +1 nucleosome, and occasionally the -1 nucleosome (Zhang et al. 2005; Albert et al. 2007), and is deposited by the SWR1 chromatin remodelling complex, which exchanges H2A/H2B for H2A.Z/H2B dimers (Mizuguchi et al. 2004). H2A.Z is not present at both very lowly and very highly transcribed genes (Millar et al. 2006), nor at subtelomeric regions (Raisner et al. 2005). Though 74% of H2A.Z-containing nucleosomes have been reported to be located within promoters, the majority of the remaining nucleosomes are found within open reading frames, thus H2A.Z is not necessarily consigned to the +1 and -1 nucleosomes (Guillemette et al. 2005), nor is it necessary for NFR formation (Tirosh et al. 2010; Li et al. 2005). The role of H2A.Z remains unclear; its mutation leads to only very minor changes in steady-state transcript levels, though it has been proposed to play a role in preventing the ectopic spread of silent heterochromatin (Meneghini et al. 2003).

5.2.9. Histone turnover is most pronounced at promoters and can be directed by chromatin remodelling enzymes, histone chaperones and RNAPII itself.

As stated above, nucleosomes can be evicted upon transcription. At the *GAL10* gene, nucleosome occupancy across the gene body is inversely correlated with RNAPII levels and transcriptional activity, and nucleosomes are evicted and then reassembled rapidly in the wake of RNAPII (Schwabish and Struhl 2004). However, at other genes, nucleosome eviction is not related to transcription elongation (Kristjuhan and Svejstrup 2004), suggesting that RNAPII procession can still occur in the presence of bound nucleosomes, and indeed Thiriet and Hayes (2005) found evidence in *Physarum polycephalum* to suggest that histones are not necessarily replaced during RNAPII passage, something that was later demonstrated in yeast (Kulaeva et al. 2009, 2007). That RNAPII itself might be directly involved in turnover has been suggested by *in vitro* studies that show that successive rounds of polymerase passage result in loss of the H2A/H2B dimer followed by complete removal of the histone (Kulaeva et al. 2009, 2007).

Dion et al., (2007) performed an elegant study in yeast to measure the rates of histone turnover genome-wide, in the hope that this would shed light on the extent and purpose of turnover. They used yeast carrying two differently tagged H3 proteins – a Myc-tagged H3 under a constitutive promoter, and a Flag-tagged H3 under the inducible *GAL* promoter. In this way, they were able to use ChIP-chip to assess the ratio of Flag/Myc at several time points following induction, and from this estimate the rate of histone turnover at a given genomic region. Their results yielded several interesting findings: 1) that the rate of turnover of promoter-bound histones is generally greater than that of those in the open reading frame; 2) that the rate of turnover in the open reading frame was correlated with the polymerase density in the same region and 3) that the rate of turnover at the promoter was *not* well correlated with polymerase

density in the open reading frame. Thus the rate of histone turnover at the promoter would appear to be independent of the actual level of transcription across the gene. Dion et al., (2007) proposed that these 'hot' promoter nucleosomes serve a boundaries; preventing the spread of chromatin domains. Soluble histones (those not associated with DNA) tend to be hyperacetylated and hypomethylated, thus incorporation of new histones should replace the pre-existing modifications with acetylation instead, and would thus stop the persistence of extensively modified chromatin (Burgess et al. 2010; Fillingham et al. 2008; Verreault et al. 1996).

What directs nucleosome eviction, and histone turnover more generally? Such a role has been ascribed to chromatin remodelling enzymes, histone chaperones and RNAPII itself (as described above). That chromatin remodelling enzymes could direct nucleosome eviction as well as sliding was demonstrated by Boeger et al., (2004), who showed that upon activation of the *PHO5* promoter on a chromatin circle, there was a loss of nucleosomes, rather than just a redistribution. At the *PHO8* gene a similar eviction was recently shown to be dependent upon the SWI/SNF remodelling complex (Brown et al. 2011), while SWI/SNF can also direct removal of H2A/H2B dimers (Yang et al. 2007). SWI/SNF and the related RSC complex have both been shown to destabilise nucleosome structure (Chaban et al., 2008; Floer et al., 2010; Cote et al., 1998), and also to be necessary for transcription elongation, where they are proposed to remove nucleosome blocks (Schwabish and Struhl 2007; Carey et al. 2006). Deleting components of the of SWI/SNF, RSC and ISW1 complexes results in a reduction in histone loss at the promoters of heat shock genes following induction (Erkina et al. 2010).

Remodelling enzymes have also been shown to be capable of loading histones onto DNA. In an *in vitro* system, the RSC complex was sufficient to transfer histones from nucleosomal DNA to naked DNA (Lorch et al. 1999), while the SWI/SNF, ISWI and INO80

complexes could direct the exchange of bound H2A/H2B dimers with free dimers (Bruno et al. 2003; Papamichos-Chronakis et al. 2011).

Asf1 is a histone chaperone that was shown to be required for histone eviction upon activation of *PHO5* (Adkins et al. 2004) and to generally be involved in histone eviction genome-wide (Adkins and Tyler 2004). Asf1 is associated both with yeast promoter regions and with the gene body (Schwabish and Struhl 2006), where it directs H3 eviction. Interestingly, Asf1 forms a complex with Rtt109, the histone acetyltransferase that deposits H3K56ac, and is indeed necessary for this mark (Tsubota et al. 2007). In support of this, regions of high H3/H4 turnover are enriched for H3K56ac, while deleting *ASF1* results in slower turnover (Kaplan et al. 2008; Rufiange et al. 2007). H3K56ac is generally thought to be an important mark for turnover, and is necessary for nucleosome eviction and induction at *PHO5* (Williams et al. 2008). Other chaperones include the FACT complex, composed of Spt16 and Pob3, which is also necessary for nucleosome eviction at *PHO5* and can direct H2A/H2B exchange (Adkins and Tyler 2004; Ransom et al. 2009), and Spt6, which amongst a variety of ascribed functions has been shown to be necessary for the maintenance of normal histone levels across the gene body of highly transcribed genes (Kaplan et al. 2003; Ivanovska et al. 2011).

5.2.10. The possible consequences of antisense-transcribing polymerase on sense promoter chromatin

Xu et al., (2011) reported that the majority of antisense transcripts in yeast terminate upstream of the sense promoter, suggesting that the antisense-transcribing polymerase passes through the TSS of the sense transcript. Thus, the antisense-transcribing polymerase will occupy the same region that the promoter chromatin and transcription machinery would normally occupy. Based on the observations that polymerase passage

can recruit histone-modifying enzymes, and that in certain instances it can bring about nucleosome eviction, it was hypothesised that polymerase passage might disrupt or modulate promoter chromatin. In support of this hypothesis, multiple studies report that histone turnover rates are high at environmentally responsive genes – the same genes most enriched for antisense transcription (Rufiange et al. 2007; Jamai et al. 2007; Dion et al. 2007; Xu et al. 2011). Here, correlations were sought between the level of antisense transcription at sense promoters genome-wide and a variety of chromatin features: histone levels, modifications and dynamics, as well as levels of chromatin remodelling enzymes, histone chaperones, and histone modifying enzymes, to explore whether antisense transcription might be modulating gene structure and behaviour.

5.3. Results

5.3.1. Sense promoters subject to antisense transcription show increased levels of chromatin remodelling enzymes and histone chaperones

To obtain a broad overview of how antisense transcription might impact on the canonical features of the sense promoter, the enrichment or depletion of particular transcription and chromatin regulating factors were assessed at those genes subject to high levels of antisense transcription. It was anticipated that if antisense was disrupting or modulating promoter chromatin then it would conceivably alter the binding of such factors.

Genes were separated into the five classes based on the quantitative analysis laid out in chapter 4. In summary, this was based on the level of *nascent* antisense transcription in a 300bp window immediately downstream of the promoter (Figure 25) From this list any overlapping, convergent genes were excluded, to remove instances where *coding* transcription might be running antisense to a sense promoter. By selecting a window immediately downstream of the transcription start site, rather than

one overlapping the promoter, any conflation with bi-directional transcription initiation from the promoter was avoided (Neil et al. 2009). This gave us a class of 1,024 genes (20% of the total group) with high levels of antisense transcription (≥28 reads, a median of 70 reads), a class of 1,240 genes (24%) with little or no antisense transcription (0 or 1 read), and three intermediate classes (Figure 25).

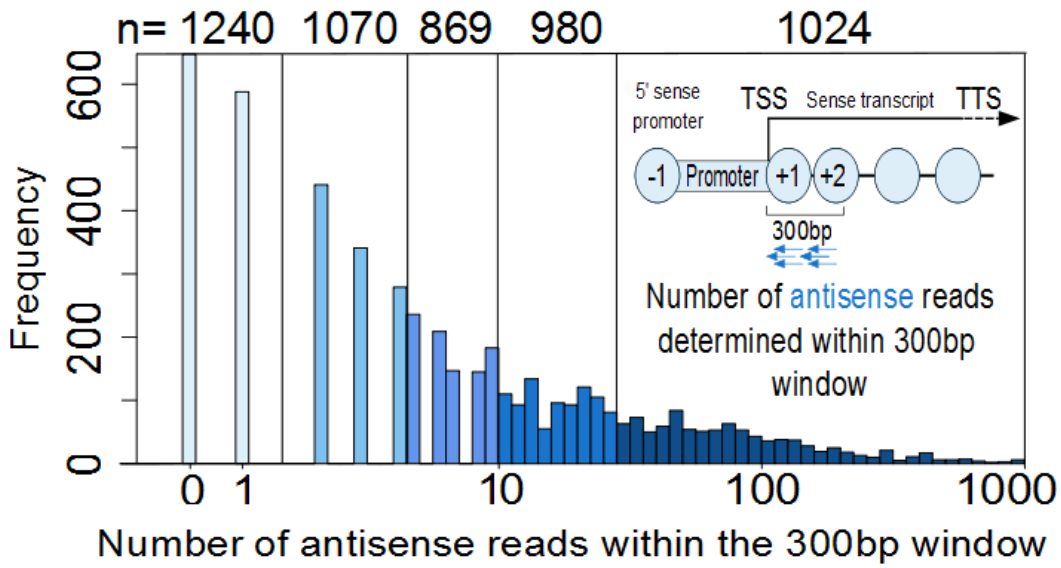


Figure 25: Splitting yeast genes into five classes based on antisense transcription

A histogram showing the distribution of nascent antisense transcription reads within a 300bp window placed directly downstream of the sense transcription start site (see inset), using NET-seq data from Churchman and Weissman, (2011). Genes were divided into five separate classes, demarcated by vertical lines, with the numbers of genes in each class shown above the corresponding class. The colour coding scheme used here is used again in figures below, with the darkest blue class being that with the most antisense transcription.

It was then assessed whether genes in the class with the highest antisense transcription were enriched for specific factors when compared to those in the lowest, utilizing an extensive analysis in which the levels of 202 factors were determined at the sense promoters of almost all the coding genes in the yeast genome (Venters et al. 2011). This list covers quite comprehensively the different classes of transcription and chromatin regulating proteins and complexes: RNAPII components, histones and variant histones,

PIC components, condition-specific transcription factors, elongation factors, RNAPII CTD kinases, histone modifying enzymes, chromatin remodelling enzymes and histone chaperones. The Wilcoxon rank sum test was used to determine which factors were significantly enriched or depleted ($p < 0.00001$).

Strikingly, it was found that the promoters of genes subject to high levels of antisense transcription were significantly enriched for factors involved in modulating the chromatin environment (Table 2). Most significantly enriched were the chromatin remodelling enzymes Isw1 and Ino80, followed by Pob3 of the FACT histone chaperone complex, and Rsc9 and Swi3, of the RSC and SWI/SNF chromatin remodelling complexes respectively. Also enriched was the deacetylase Rpd3, though this has also been shown recently to also behave as a histone chaperone (Chen et al. 2012), the Itc1 subunit of the Isw2 complex, the Spt6 histone chaperone, and its interactor Spt1. What is immediately apparent is the wealth of different chromatin remodelling enzymes and histone chaperones enriched at these enzymes. Other factors included the RNAPII components Rpb2, Rpb7 and Rpo21, as well as the CTD kinase Ctk1. It is likely that these factors are enriched due to the passage of antisense-transcribing RNAPII into the promoter, and not due to sense-transcribing RNAPII. Spt3, of the SAGA complex, was also found to be enriched, in keeping with antisense transcribed genes being environmentally regulated (Xu et al. 2011; Huisinga and Pugh 2004). Finally, it was found that the core histone H2B was also enriched at these genes, suggesting antisense transcription might result in higher histone levels at the promoter.

Enriched factor	Complex	Function	p-value*
Isw1	Isw1a, Isw1b	Chromatin remodelling enzyme	4.2 × 10 -9
Ino80	INO80	Chromatin remodelling enzyme	4.9 × 10-9
Pob3	FACT	Histone chaperone	2.5 × 10-7
Rsc9	RSC1, RSC2, RSCa	Chromatin remodelling enzyme	3.2 × 10-7
Swi3	SWI-SNF	Chromatin remodelling enzyme	4.9 × 10-7
Rpd3	RPD3	Histone deacetylase, putative histone chaperone	2.9 × 10-6
Rpb7	Pol II	Recruitment of 3'-end processing factors	3.8 × 10-6
Itc1	Isw2	Component of chromatin remodelling complex	1.2 × 10-5
Spt3	SAGA, SLIK	Subunit of the SAGA and SAGA-like transcriptional regulatory complexes	1.9 × 10-5
Ctk1	CTK	CTD phosphorylation, regulates mRNA 3' end processing	2.1 × 10-5
Spn1 (Iws1)	SPT6 interactor	Interacts with RNAP II, TBP and chromatin remodelling factors	2.2 × 10-5
Spt6	SPT6	Histone chaperone	2.9 × 10-5
Rpo21	Pol II	Largest Pol II subunit	3.8 ×10-5
Htb2	nucleosome	Core histone	6.7 × 10-5
Rpb2	Pol II	Second largest Pol II subunit	8.8 × 10-5

Table 2: Antisense-transcribed genes are enriched for distinct promoter-bound transcription related proteins.

Factors are ranked in order of p-values determined using the Wilcoxon rank sum test, and were considered enriched if their levels were higher in those genes with high antisense compared to low antisense and if they had a p-value less than 0.0001. p-values were calculated by comparing the distributions of values for a particular protein between the gene class with the lowest level of antisense transcription and the gene class with the highest (the two classes being those shown in Figure 25), and using the Wilcoxon rank sum test. The harsh p-value cut-off was selected to avoid issues associated with multiple comparisons. Factors in blue were significantly enriched in genes with antisense transcription compared to those with sense transcription but no antisense transcription (see Table 2).

To assess whether these enriched factors were a unique feature of genes subject to antisense transcription, or just transcribed genes in general, the features of antisense

transcription were compared with those of sense transcription. The gene group with the highest antisense transcription (median 70 antisense reads, 578 sense reads, n=1024) was compared with another group which had no or low levels of antisense transcription (0 or 1 read) but higher levels of sense transcription (median 948 reads) within the same 300bp window (Table 3). Of the factors in Table 2, components of the ISW1, ISW2, INO80 and FACT complexes (Burgess and Zhang, 2013; Rando and Winston, 2012) were specifically enriched at the sense promoters of genes subject to antisense transcription compared to gene promoters with high levels of sense transcription. Thus, antisense transcription is likely to be associated with specific chromatin remodelling enzymes and changes in promoter chromatin structure.

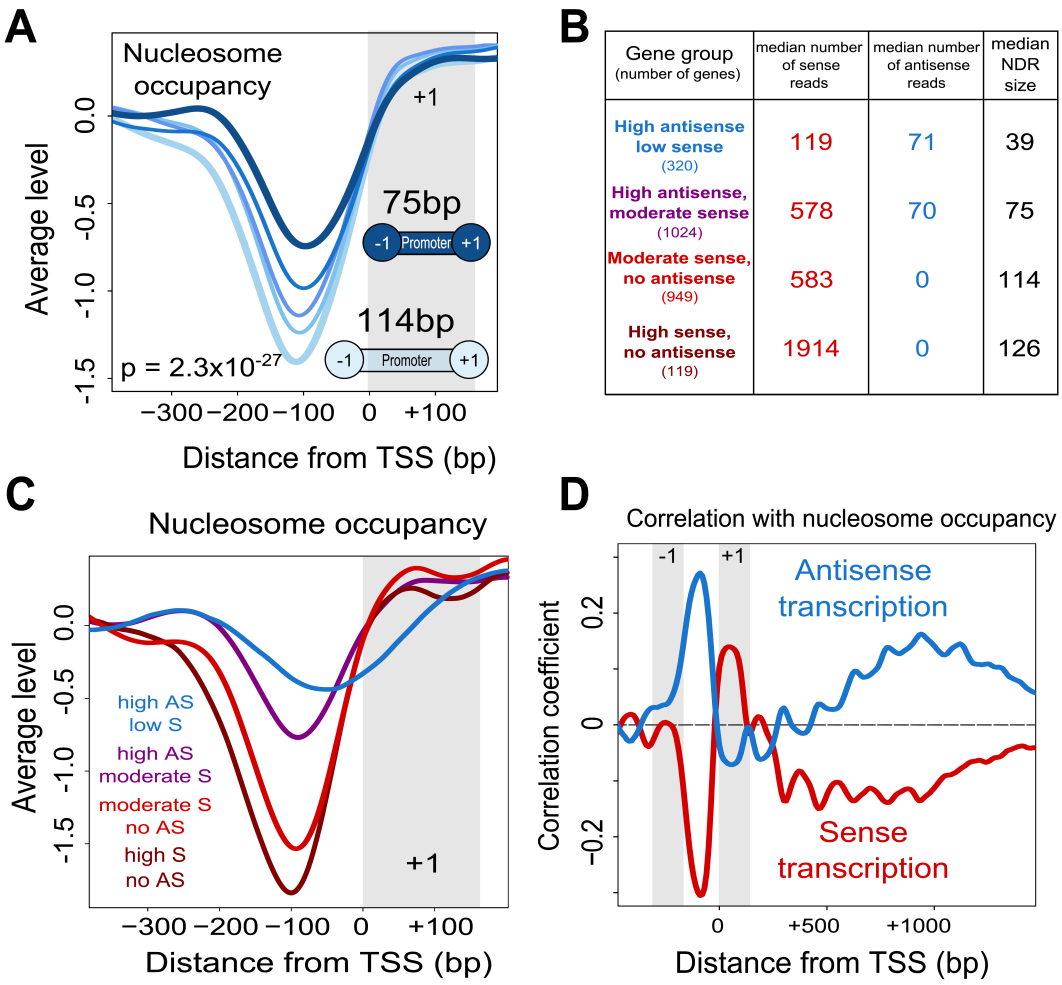
FACTOR	COMPLEX	Change in level	p-value
Ssl1	TFIIH	down	1.29E-14
Vps72	SWR-C	down	2.18E-10
Rad3	TFIIH, NER3	down	1.52E-09
Swc1	SWR-C	down	2.78E-09
Bur6	NC2	down	4.26E-09
Isw1	**ISW1a, ISW1b**	**up**	1.77E-08
Esa1	NuA4/Piccolo	down	6.22E-08
Taf13	TFIID	down	1.55E-07
Snf1	SNF1	down	2.60E-07
Tfb1	TFIIH	down	4.97E-07
Ino80	**INO80**	**up**	9.21E-07
Pob3	**FACT**	**up**	9.96E-07
Ncb2	NC2	down	1.62E-06
Rpt6	APIS	down	2.38E-06
Sir3	SIR	down	3.53E-06
Kin28	TFIIH	down	1.24E-05
Htz1	nucleosome	down	1.48E-05
Mot1	Mot1	down	3.17E-05
Tfb3	TFIIH	down	3.87E-05
Itc1	**ISWI**	**up**	6.93E-05

Table 3: Factors enriched (up) or depleted (down) at the promoters of highly antisense-transcribed genes versus highly sense-transcribed genes.

Factors shown are those deemed significantly enriched or depleted ($p < 0.00001$). Highly sense-transcribed genes were selected as those which had no antisense transcription and a median read count of 948 for sense transcription. High antisense-transcribed genes had a median read count of 70 for the antisense strand and 578 for the sense strand.

5.3.2. Antisense transcription is associated with a change in chromatin architecture at the sense promoter

Given this enrichment for chromatin remodelling enzymes, histone chaperones and H2B at genes subject to high antisense transcription, genes subject to varying levels of antisense transcription were compared in terms of positioning, dynamics, and occupancy of their promoter-bound nucleosomes. Using a genome-wide map of nucleosome occupancy (Kaplan et al., 2009), it was found that genes subject to high levels of antisense transcription showed much higher levels of nucleosome occupancy across their promoters ($p = 2.3x10-27$; Figure 26A). The NDR between the -1 and +1 nucleosomes was also shorter in genes with high antisense transcription (Figure 26A; a median length of 75bp versus 114bp for the highest and lowest classes respectively, $p=1x10-16$). Intriguingly, the increase in nucleosome occupancy occurred in a stepwise fashion, suggesting that antisense transcription can exert changes in the chromatin even at low levels. As there is only a very weak correlation between sense and antisense transcription within the 300bp window (Chapter 4, Spearman's correlation coefficient = -0.02), the increase in nucleosome occupancy associated with increasing antisense transcription is likely to be independent of sense transcription.



Gene group (number of genes)	median number of sense reads	median number of antisense reads	median NDR size
High antisense low sense (320)	119	71	39
High antisense, moderate sense (1024)	578	70	75
Moderate sense, no antisense (949)	583	0	114
High sense, no antisense (119)	1914	0	126

Figure 26: Antisense transcription is associated with an increase in nucleosome occupancy at sense promoters.

(A) Average levels of nucleosome occupancy around the TSS (0) in the five classes shown in Figure 25, together with the NDR sizes of the top and bottom classes (see inset). The grey rectangle represents the approximate position of the +1 nucleosome. **(B)** Median number of sense and antisense reads in the gene groups shown in Figure 25, together with their median NDR size. **(C)** Average nucleosome occupancy in genes with varying median numbers of sense and antisense reads in the same window, with colours referring to the gene groups described in **B**. **(D)** Correlation between nucleosome occupancy and the two types of transcription. Shown is the Spearman correlation coefficient calculated for sense transcription (red) and, separately, antisense transcription (blue) in the 300bp window with nucleosome occupancy at varying positions around the TSS. Grey rectangles represent the approximate positions of the -1 (left) and +1 (right) nucleosomes.

5.3.3. Sense and antisense transcription are associated with distinct patterns of nucleosome occupancy at the sense promoter

Next, it was assessed how nascent sense transcription in the same 300bp window influences nucleosome occupancy and NDR size at the sense promoter in the presence of varying levels of antisense transcription (Figure 26B-C). Genes with sense transcription but no antisense transcription tended towards an open promoter chromatin structure, with low nucleosome occupancy and a large NDR (126bp), while genes with high antisense transcription and low sense transcription had a closed promoter chromatin structure with high nucleosome occupancy across the promoter and a small NDR (39bp). Genes with the same amount of sense transcription but different levels of antisense transcription still showed marked differences in nucleosome occupancy (Figure 26B-C), suggesting their associations with occupancy are independent of one another, in line with the lack of correlation between sense and antisense transcription. Finally, it was found that sense transcription in the 300bp window was negatively correlated with nucleosome occupancy immediately upstream of the TSS, but positively correlated immediately downstream, in the region corresponding to the +1 nucleosome (Figure 26D), supporting a model in which sense transcription positions the +1 nucleosome (Hughes et al. 2012). In contrast, antisense transcription correlated positively with nucleosome occupancy over the promoter, directly upstream of the TSS, and in the first 1kb of the transcribed region. Thus, sense and antisense transcription appear to determine distinct properties of chromatin organization surrounding the sense TSS. Although sense and antisense transcription appear to be associated inversely with nucleosome occupancy, the very weak correlation between sense and antisense transcription suggests that they are independent processes.

5.3.4. Nascent sense and antisense transcription are associated with distinct patterns of histone modification

In order to understand the basis of the increased nucleosome occupancy at the promoters of genes subject to high nascent antisense transcription, the patterns of histone modifications were examined. Though there was no enrichment or depletion for histone modifying enzymes other than Rpd3, others have reported associations between antisense transcription and histone modifications or modifying enzymes (Kim et al. 2012; Cui et al. 2012; Houseley et al. 2008; Camblong et al. 2009). It was therefore prudent to assess how histone modifications differed between groups subject to varying levels of antisense.

Several different approaches to this were taken. In the first approach, different histone modifications were correlated generally with both nascent sense and antisense transcription genome-wide was (Table 4). That is, the regions of the genome that fall within transcribed regions were taken, divided into 30bp bins, and within each bin correlated sense and, separately, antisense with the level of multiple modifications within the same bin. This gave two values for each modification considered – a measure of how it correlates with both sense and antisense transcription genome-wide.

Secondly, the levels of a particular modification at each base pair position relative to the TSS were correlated with sense/antisense transcription in the 300bp window – i.e. it was identified where in the average gene body sense and antisense transcription were likely to be at their highest or lowest (Figure 27A). This gave two values for every position relative to the TSS, again, for both sense and antisense transcription.

Thirdly, the levels of different modifications across an average gene were assessed, after stratifying by levels of antisense transcription in the 300bp window

(Figure 27B), and finally the class of genes with the highest antisense transcription were compared to a class of genes with the highest sense transcription (Figure 27C).

This analysis revealed that different modifications show strong positive and negative correlations with both nascent sense and antisense transcription, and that the two types of transcription do not show the same associations. The observed trends remained when SAGA-regulated TATA-box containing genes (Basehoar et al. 2004) were excluded from the analysis, suggesting that the association between antisense transcription and the changed distribution of histone modifications is independent of promoter class (Figure 28).

Histone modification	correlation with antisense transcription	correlation with sense transcription
H3K79me3	-0.45	0.09
H3K4ac	0.42	0.22
H3K79me2	0.39	0.05
H3K56ac	0.37	0.31
H3K36me3	-0.35	0.10
H3K9ac	0.31	0.32
H3K14ac	0.27	0.49
HR2me2a	-0.26	-0.50
H2BK123ub	-0.26	0.36
H3R2me1	-0.24	0.04
H3K4me3	0.15	0.64
H3K4me2	0.19	0.42
H3K4me1	-0.07	-0.19

Table 4: Genome-wide correlations between sense and antisense transcription and thirteen different histone modifications.

Shown are the Spearman correlation coefficients for antisense (blue) and sense transcription (red). The darker the colour the stronger the correlation, whether positive or negative. Values were determined by binning the coding portion of the genome into 50bp windows and determining the level of sense and antisense transcription and the histone modifications within each window.

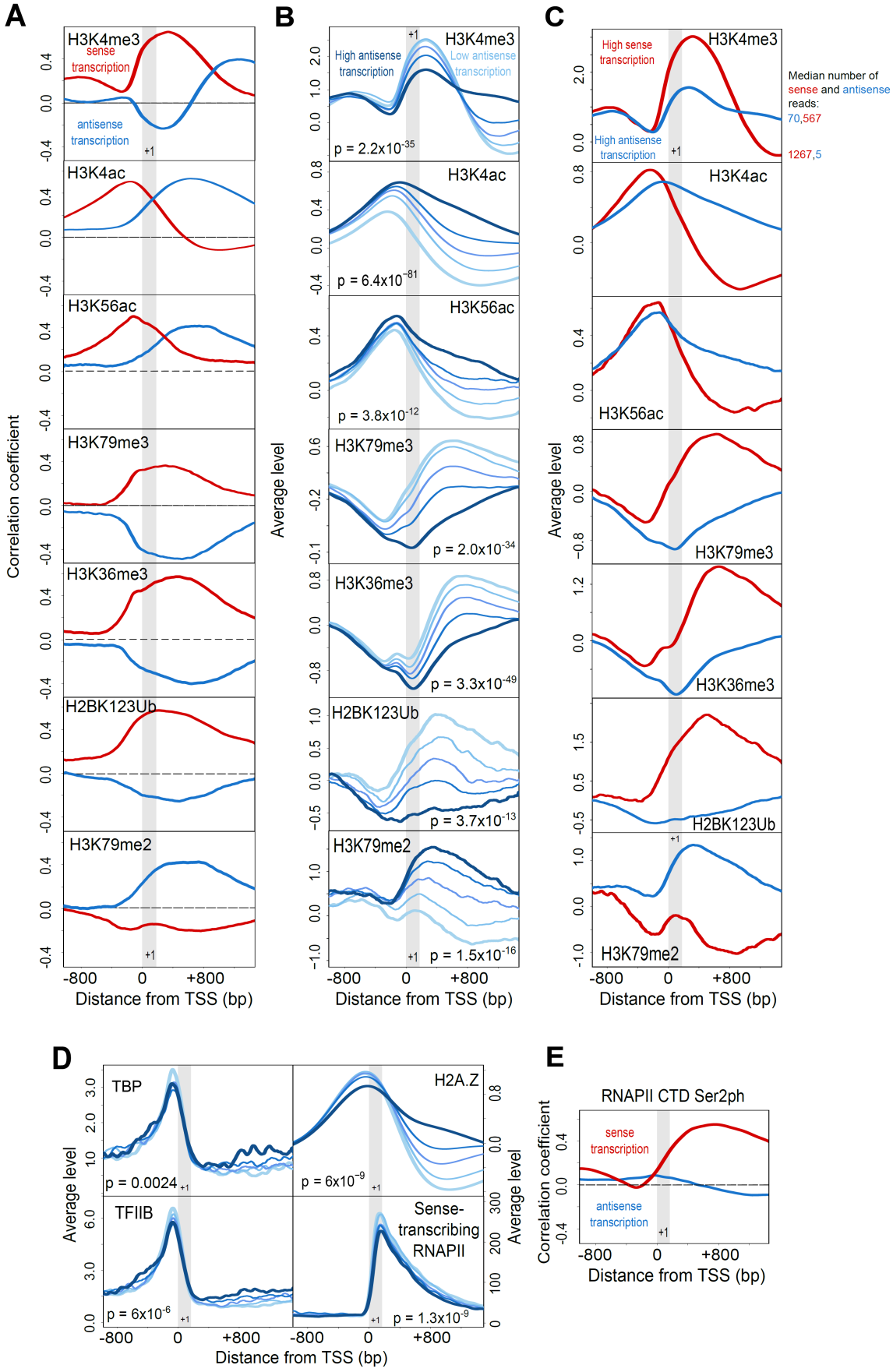

A
B
C
D
E
H3K4me3
sense transcription
antisense transcription
H3K4ac
H3K56ac
H3K79me3
H3K36me3
H2BK123Ub
H3K79me2
Correlation coefficient
Average level
Distance from TSS (bp)
-800
0
+800
+1
High antisense transcription
Low antisense transcription
p = 2.2x10-35
p = 6.4x10-81
p = 3.8x10-12
p = 2.0x10-34
p = 3.3x10-49
p = 3.7x10-13
p = 1.5x10-16
High sense transcription
Median number of sense and antisense reads:
70,567
1267,5
TBP
H2A.Z
TFIIB
Sense-transcribing RNAPII
p = 0.0024
p = 6x10-9
p = 6x10-6
p = 1.3x10-9
RNAPII CTD Ser2ph

Figure 27: Antisense transcription is associated with broad changes in histone modifications independent of sense transcription

(A) The correlation coefficient between the levels of seven different histone modifications in 10bp windows around the TSS (0) and, separately, the number of sense and antisense NET-seq reads in the 300bp window described in Figure 25. **(B-C)** The average levels of seven different histone modifications around the TSS (0) in the five gene classes described in Figure 25, subject to **(B)** varying antisense transcription and **(C)** high antisense transcription (blue), and high sense transcription (red) in the 300bp window. **(D)** The average levels of TBP, TFIIB, H2A.Z and sense-transcribing RNAPII. TBP and TFIIB do not spread into the body of the gene with increasing antisense transcription, as H2A.Z does. **(E)** Correlation coefficient between level of RNAPII CTD Ser2 phosphorylation and sense and antisense transcription. Where shown, p-values were determined using the Wilcoxon rank sum test, comparing the genes in the highest class to the lowest class near the TSS. All histone modification data in **A**-**C** was normalised first to levels of H3.

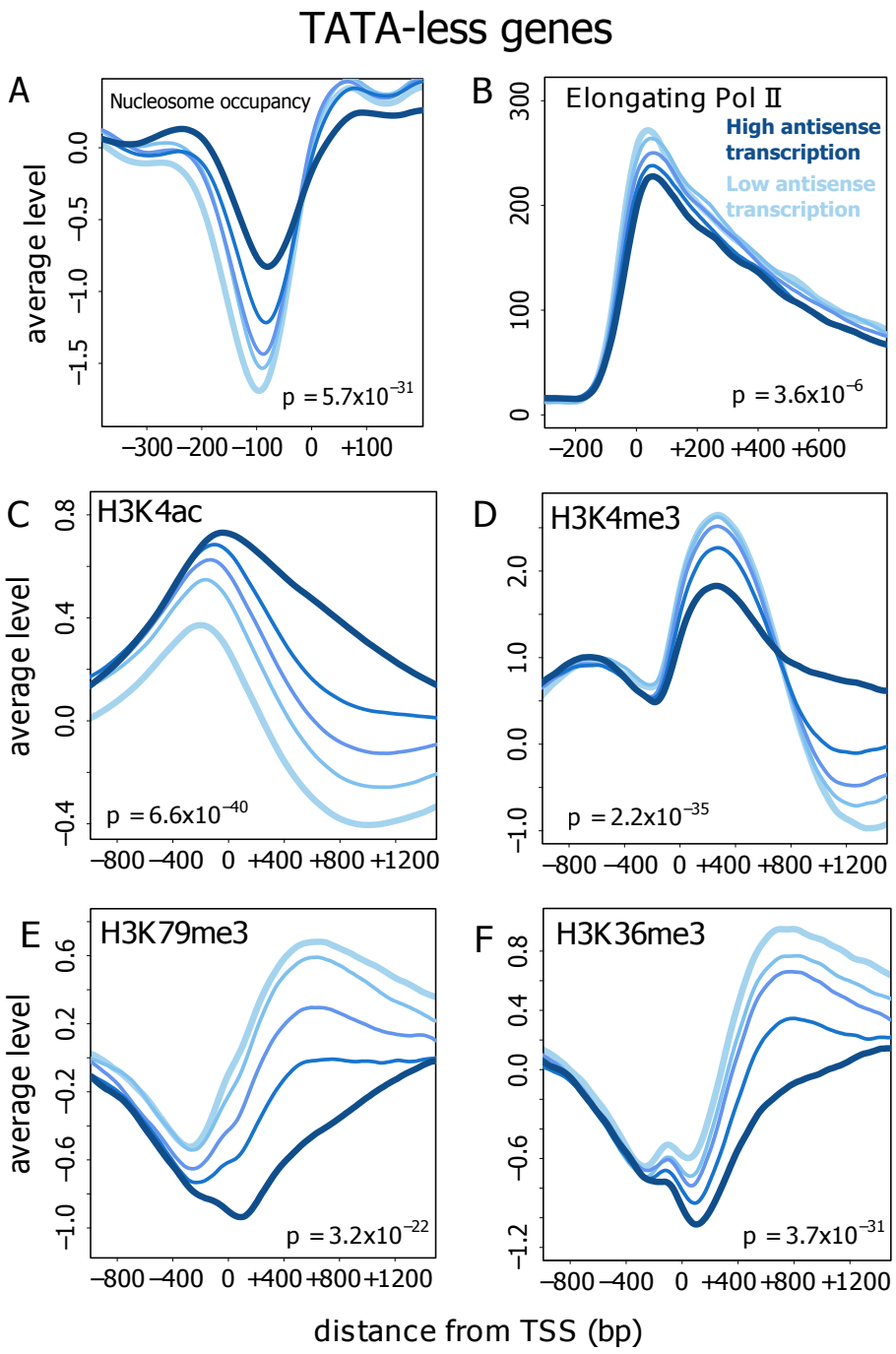


Figure 28: Trends associated with antisense transcription remain when genes with TATA-boxes are excluded from the analyses.

Shown are the average levels of **(A)** nucleosome occupancy, **(B)** sense transcribing RNAPII, **(C)** H3K4ac, **(D)** H3K4me3, **(E)** H3K79me3 and **(F)** H3K36me3 in the vicinity of the TSS of TATA-less (Basehoar et al. 2004; Huisinga and Pugh 2004) genes with varying levels of antisense, as defined in the main text and Figure 25. This analysis suggests that although genes with TATA-boxes generally have higher levels of antisense transcription than genes without TATA-boxes, the effect of antisense transcription affects TATA-less promoters in the same way as genome-wide. It is concluded that antisense transcription affects all classes of yeast gene and is not specific for SAGA-regulated, TATA-box genes.

5.3.5. H3K4 methylation and H2A.Z spread into the coding region of genes subject to antisense transcription

H3K4me3 was found to be strongly correlated with nascent sense transcription (Figure 27A), primarily within the early transcribed region (Figure 27B), as described previously (Kirmizis et al. 2007). Conversely, levels of H3K4me3 were significantly decreased at the promoter and early transcribed region of genes subject to antisense transcription (Figure 27C; $p=6.4 \times 10^{-81}$), but were increased further downstream, suggesting the mark has been redistributed by antisense transcription. This pattern is consistent with a model in which antisense-transcribed genes have two, overlapping and convergent transcription units, resulting in transcription-dependent redistribution of H3K4 methylation. To explore this further, the distributions of H3K4 mono-, di- and trimethylation across a gene were assessed, i.e. where most of a given mark is localised in a gene body (Figure 29). It was found that whereas the marks are well partitioned across those genes (H3K4me3, then H3K4me2 then H3K4me1 from 5' to 3') with low levels of antisense transcription, in genes subject to high antisense transcription the marks are more evenly spread across the gene body, suggesting they have been redistributed. In support of this, levels of the variant histone Htz1 (H2A.Z) showed a similar redistribution from the promoter into the body of genes subject to high antisense transcription (Figure 27D). However, TFIIB and TBP remain associated at the promoter, suggesting that the redistribution observed is not a result of cryptic promoters in the transcribed region (Figure 27D). Note also that TFIIB and TBP do not show even a modest reduction in genes subject to high antisense transcription, supporting a model in which antisense transcription does not simply repress sense transcription.

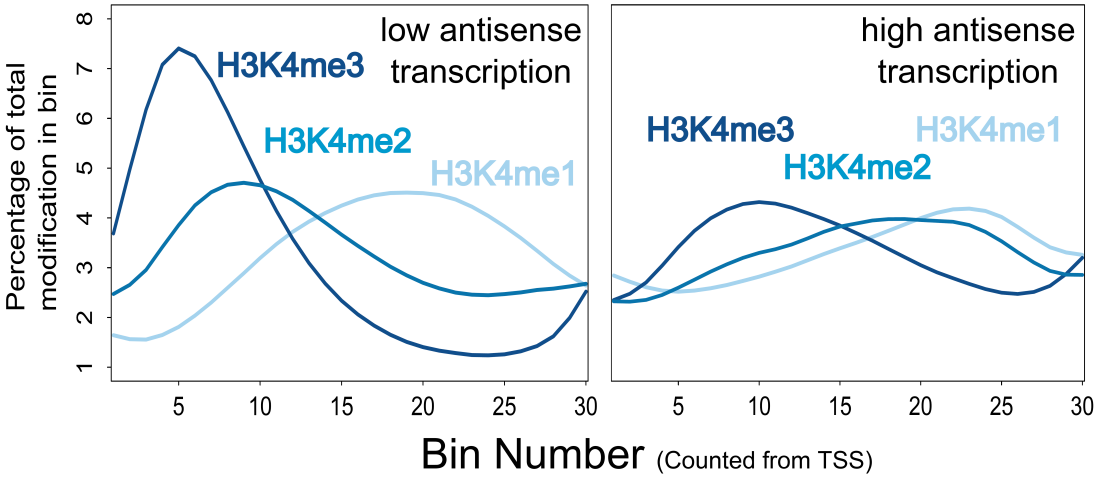


Figure 29: Genes with high antisense transcription show more evenly distributed H3K4 methylation profiles.

The distribution of H3K4me1, H3K4me2 and H3K4me3 within genes with both low and high levels of antisense transcription (left and right panels respectively). Genes were divided into the 30 bins and the level of modification determined as a percentage of the total modification across the gene. The peak of H3K4me3 normally found at the beginning of the transcription unit is much less pronounced in genes with high antisense transcription, in line with previous observations on single genes (Kirmizis et al. 2007; Pokholok et al. 2005). One explanation for the more even distribution of H3K4 methylation marks over genes with high antisense transcription is that it is the result of having two, overlapping, and opposing transcription units. This is in line with the rise in H3K4me3 seen in the gene body of genes with high antisense transcription, a possible consequence of 3' promoters directing antisense transcription.

5.3.6. H3K36me3, H3K79me3 and H2BK123ub are reduced in genes subject to antisense transcription, in contrast to sense transcription

H3K36me3, H3K79me3 and H2BK123ub are associated with the stabilization of nucleosome structures and suppression of new histone incorporation (Batta et al. 2011; De Vos et al. 2011; Fleming et al. 2008; Rufiange et al. 2007; Lee et al. 2012; Smolle et al. 2012; Venkatesh et al. 2012). Strikingly, all three modifications showed strong negative correlations with nascent antisense transcription genome-wide (Figure 27A). Sense transcription in the 300bp window showed a strong association with these modifications, primarily over the gene body (Figure 27B). By contrast, there was a clear decrease in all three modifications from the promoter and well into the transcribed region in genes subject to increasing levels of antisense transcription (Figure 27C).

H3K79me3 has previously been associated with gene length, with longer genes showing higher levels in the gene body (Schulze et al. 2011), however, varying antisense transcription altered levels of H3K79me3 and H3K36me3 not just over the gene body but at the promoter as well (Figure 30). Interestingly, H3K79me2 shows a strong positive correlation with nascent antisense transcription, unlike H3K79me3 (Figure 27D) supporting functionally distinct methylation states at H3K79 (Nguyen and Zhang, 2011).

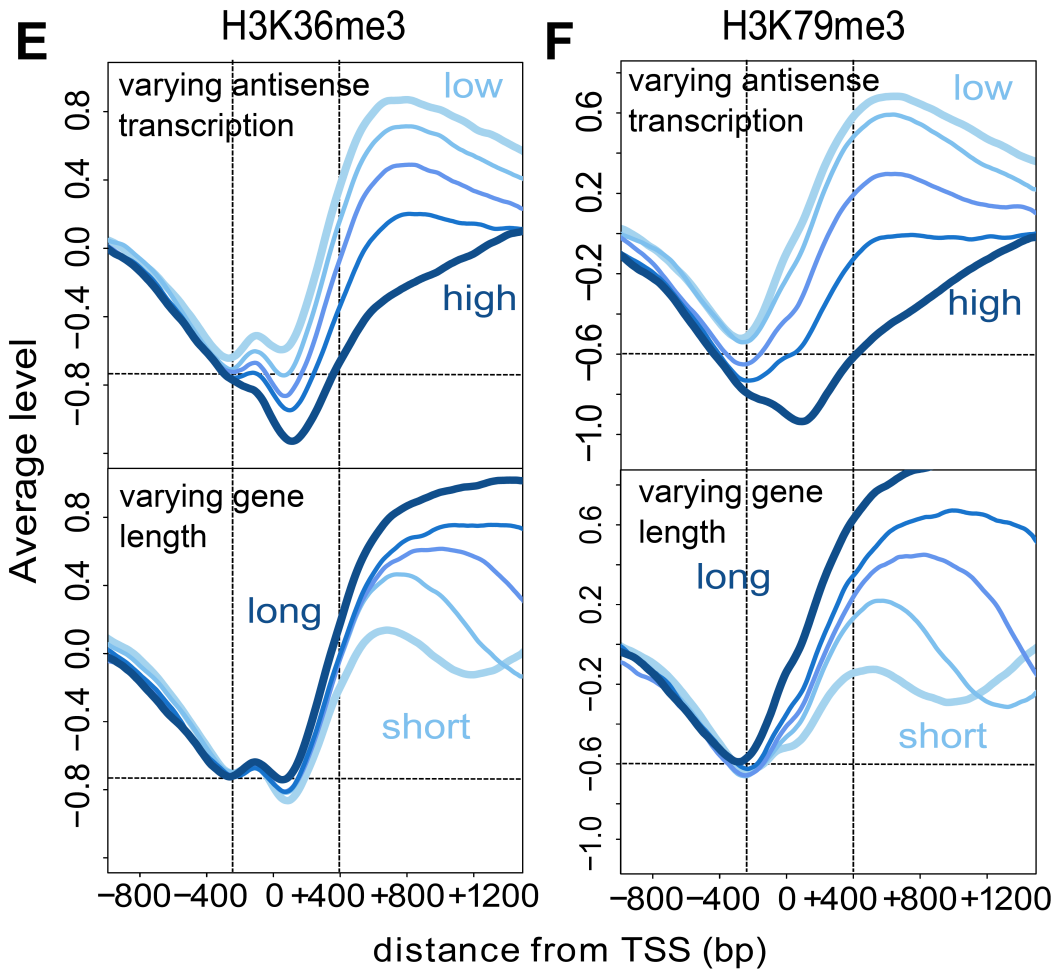


Figure 30: The association of antisense transcription with decreased H3K36me3 and H3K79me3 is not related to gene length.

The association of antisense transcription with decreased H3K36me3 **(A)** and H3K79me3 **(B)** at the promoters of genes is independent of the known association between gene length and these marks. Shown is the average level for both marks at varying gene length (divided into groups of approximately 500, 1,000, 1,500, 2,000 and 2,500bp) compared to the average levels in genes with varying levels of antisense. Horizontal dotted black lines correspond to the minimum average value found in genes of varying length. Vertical dotted black lines have been added to illustrate a region approximately encapsulating the promoter and the -1, +1 and +2 nucleosomes. Neither H3K79me3 nor H3K79me3 show substantial changes in this region when length is varied, but do when the level of antisense is varied.

These findings are contrary to expectations if genes with antisense transcription are considered to have two overlapping, convergent transcription units. If sense and antisense transcription were both associated with these marks in the same way then one would expect them both to have positive correlations across the gene body. This provides further evidence that antisense transcription may be inherently different from

sense transcription. In support of this, antisense transcription shows no correlation with phosphorylation of the RNAPII CTD at Ser2, in marked contrast to sense transcription (Figure 27D).

5.3.7. Promoters subject to antisense transcription exhibit increased histone turnover rates

Intriguingly, acetylation at K4, K9, K14 and K56 on H3 showed positive correlations with both nascent sense and antisense transcription genome-wide, but at different regions of the gene (Figure 27A-B, Figure 31A-B). In both types of transcription, acetylation is high across the promoter. Within the gene body however, things are quite different. Here, histone acetylation levels drop dramatically in genes subject to high sense transcription, but remain high in those subject to high antisense transcription (Figure 27C, Figure 31C). Given that the patterns of methylation at H3K4 are consistent with transcription-mediated re-distribution, maintaining these patterns of acetylation would require constant replenishment of acetylated histones at the promoter region, potentially by histone turnover. To probe a possible connection between histone turnover and modification, the correlation between turnover and modification genome-wide was determined, in a similar fashion to that described above for transcription and modification. The histone modifications most strongly correlated with antisense transcription were also those most strongly correlated with histone H3 turnover genome-wide; namely H3K79me3 and H3K36me3 (negatively correlated with both) and H3K4ac and K3K56ac (positively correlated with both) (Figure 32A). This suggests that the changes in histone modifications are due to this increased turnover – diluting established modifications with those common to soluble histones – as opposed to changes in the recruitment of histone modifying enzymes (supported by Table 2).

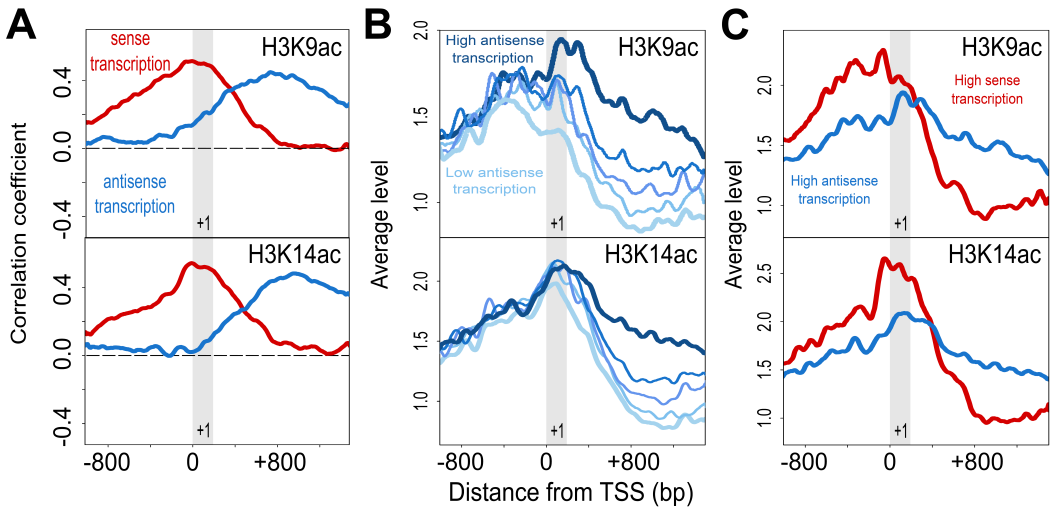


Figure 31: Antisense transcription is associated with a rise in H3K9ac and H3K14ac across the gene body.

(A) The correlation coefficient between the levels of H3K9ac and H3K14ac in 10bp windows around the TSS (0) and, separately, the number of sense and antisense NET-seq reads in the 300bp window described in Fig. 1A. **(B-C)** The average levels of H3K9ac and H3K14ac around the TSS (0) in the five gene classes described in Fig. 1A, subject to **(B)** varying antisense transcription and **(C)** high antisense transcription (blue), and high sense transcription (red) in the 300bp window. H3K9ac and H3K14ac show a similar relationship with antisense transcription compared to H3K4ac – antisense transcription is associated with an increase of H3K4ac at the promoter and, to a greater extent, the +1 nucleosome and remainder of the gene body.

5.3.8. Promoters subject to antisense transcription exhibit increased histone turnover rates

To assess histone turnover rates at promoters, a genome-wide map was utilized in which the rate of displacement of Myc-tagged H3 by Flag-tagged H3 was measured following induction of Flag-H3 expression (Dion et al. 2007). Histone turnover was significantly higher between nucleosomes -1 and +4 of genes with high antisense transcription in the 300bp window compared to those with no/low antisense transcription in the 300bp window (Figure 32B; p=9.6x10-7, 1.7x10-22, 1.9x10-24, 5.4x10-21 and 4.9x10-25 respectively). These trends remained when an alternative source of turnover data was utilised (Venkatesh et al. 2012)(data not shown), and when the H3K36 methyltransferase *SET2* was deleted – despite deletion of this enzyme resulting in a genome-wide reduction of histone turnover (Figure 33)(Venkatesh et al.

2012). To confirm that increased histone turnover was a feature of antisense transcription, and not transcription more generally, histone turnover was assessed for genes with varying levels of sense transcription (Figure 32C). Despite large changes in sense transcription between groups, histone turnover did not vary substantially at the promoter, being consistent with levels of antisense transcription, but did show the expected relationship at nucleosomes +3 and +4. This lack of association between sense transcription and histone turnover at the promoter agrees with previous findings (Dion et al. 2007). Taken together, these results suggest that although histone turnover is a feature of the canonical eukaryotic promoter, it is not correlated with the level of sense transcription, but rather with the level of antisense transcription.

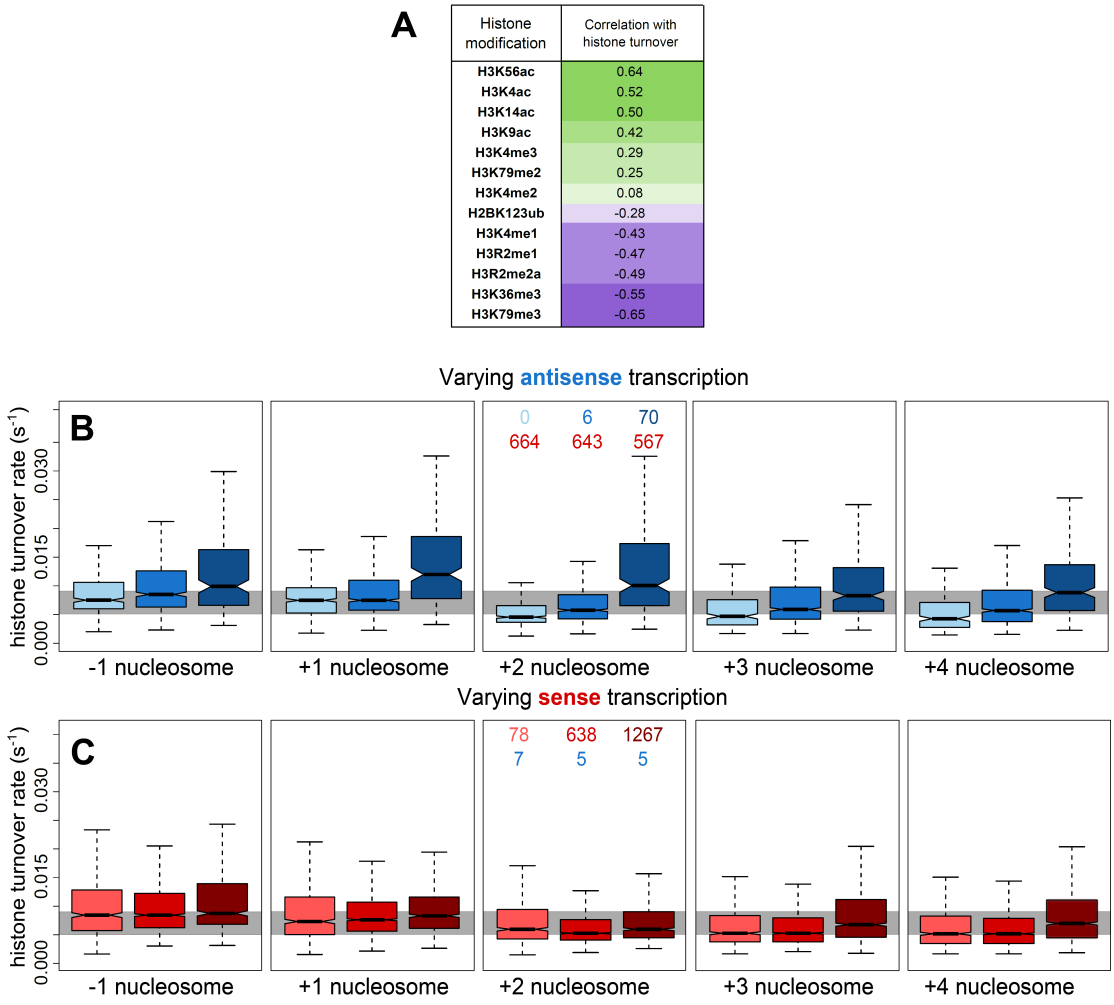


Figure 32: Antisense transcription is more strongly associated with increased histone turnover than sense transcription

(A) The Spearman's correlation coefficients obtained by correlating histone turnover genome-wide with the levels of thirteen different histone modifications within the same probed regions. **(B-C)** Boxplots showing the distributions of histone turnover rates at the -1 to +4 nucleosomes of genes with **(B)** varying levels of antisense transcription, coloured as in Figure 25, with the three intermediate classes combined into a single group, or **(C)** sense transcription in the 300bp window. Median levels of sense and antisense transcription in each group are given in red and blue respectively. The top and bottom of the grey box indicates the median value of all probes overlapping a -1 nucleosome or histone turnover genome-wide, respectively.

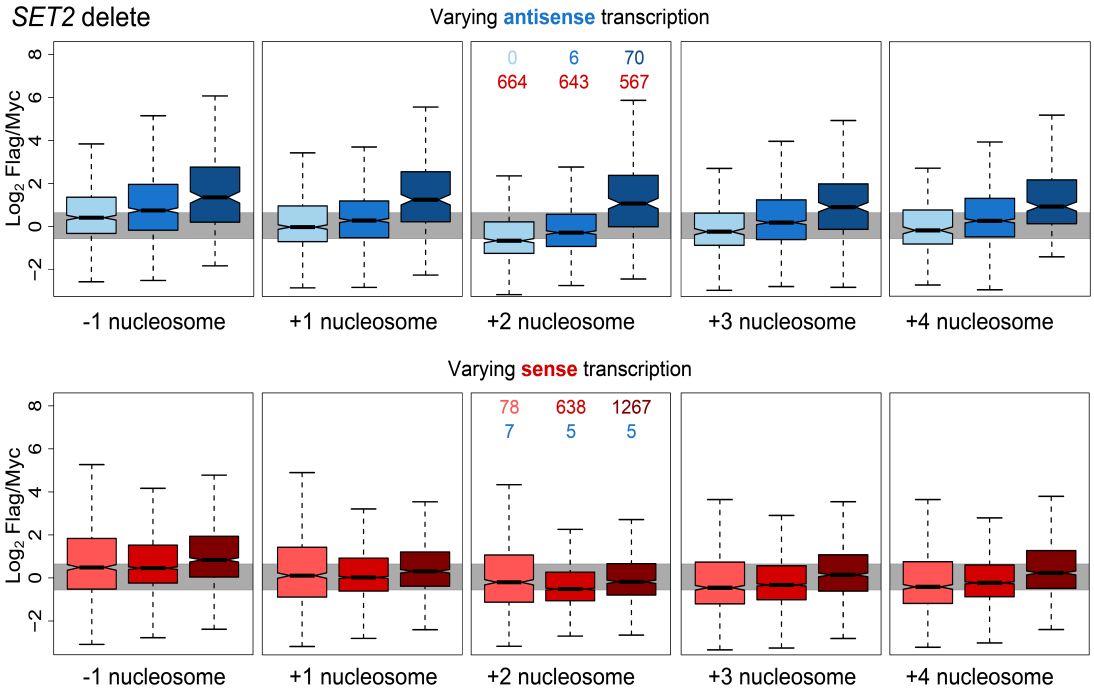


Figure 33: Antisense transcription is still associated with increased histone turnover following deletion of *SET2*, an enzyme involved in enhancing turnover.

(Top) Boxplots showing the distributions of the log2 Flag/Myc ratio (an estimate of histone turnover rate) at the -1 to +4 nucleosomes of genes with varying levels of antisense transcription, specifically the five classes defined in Figure 25, with the three intermediate classes combined into a single group. Median values of sense and antisense transcription in each group are given in red and blue respectively. The bottom of the grey box indicates the median value of histone turnover genome-wide, while the top of the box indicates the median value of all probes overlapping a -1 nucleosome. **(Bottom)** Boxplots of histone turnover at the -1 to +4 nucleosomes, with the genes divided on the basis of sense transcription in the 300bp window shown in Figure 25.

5.4. Discussion

Here it has been demonstrated that genes subject to high levels of antisense transcription also show pronounced differences in a broad range of chromatin features, both at the sense promoter and within the gene body. Generally, features associated with newly deposited dynamic chromatin – acetylation, turnover, chromatin remodelling enzymes and histone chaperones – are increased, while features associated with established chromatin – H3K79me3, H3K36me3 and H2B123ub – are reduced. Notable is the absence in changes of histone modifying enzymes between classes,

suggesting that altered modification patterns could be due instead to changes in histone turnover. That nucleosome occupancy is also increased suggests that histone deposition is the prevailing process rather than strictly turnover, although the increased occupancy could also result from closure of the NDR by chromatin remodelling. Below possible mechanisms that could explain the observed associations are explored.

5.4.1. Exploring the association between antisense transcription and chromatin modulation

Antisense transcription is associated with a rise in the level of chromatin remodelling enzymes and changes in histone modification patterns. Broadly, these associations could be explained by one of three basic mechanisms: 1) that antisense transcription is causing these chromatin changes, 2) that the chromatin changes are conducive to antisense transcription or 3) that the association is due to them both being governed by the same upstream process. An important point in considering 2) is that many of the chromatin changes described here are distinct from the assumed point of antisense transcript initiation - i.e. they occur in the promoter region or just downstream of the TSS, whereas antisense transcription initiation tends to occur just upstream of a gene's TTS (Xu et al. 2011), making it more likely that antisense transcription is causing these chromatin changes than vice versa. Of course, it is possible that changes at the promoter/early coding region are also reflected in changes at the site of antisense transcription initiation; i.e. that genes with high antisense transcription generally possess more transcriptionally permissive chromatin. Furthermore, some of the changes observed in the gene body are consistent with an increase in spurious transcription: it has been shown that Set2-dependent H3K36 methylation and Rpd3S-dependent deacetylation are required for the suppression of spurious transcription in the gene body Li et al., (2007), and indeed genes subject to high antisense transcription

show high acetylation and low H3K36me3 in the gene body. However, as discussed in chapter 3 and in Rhee and Pugh, (2012), antisense transcription is directed by PIC formation, whereas here even in genes subject to high levels of antisense transcription levels of PIC components in the gene body are not increased. Finally, if the chromatin of these genes were more transcriptionally permissive then one might expect more transcription upon the sense strand as well, and as demonstrated in chapter 4 this is not the case.

It is possible that high antisense transcription and the observed chromatin changes could both result from a distinct mechanism. For example, gene looping is able to direct the binding of Isw2 (Yadon et al. 2013), and regulate non-coding transcription at bidirectional promoters (Tan-Wong et al. 2012). Furthermore, gene looping has been shown to involve interaction with the nuclear pore (Tan-Wong et al. 2009). Nuclear pore association has been linked to histone acetylation (Mizzen et al. 1996). Furthermore, using data collected across chromosome VI (Schmid et al., 2006), it was found that antisense transcription is correlated with nuclear pore binding at the promoter (data not shown). Thus it is possible that looping, antisense transcription and chromatin dynamics are linked, though clearly further effort will be needed to gain more insight, potentially necessitating the use of available genome-wide data identifying long range DNA interactions (Naumova et al. 2012). However, in chapter 6, experimental evidence will be presented demonstrating that antisense transcription can bring about changes in chromatin modifications, supporting a model in which antisense transcription can itself modulate chromatin. Due to the absence of enrichment for histone modifying enzymes, the enrichment for chromatin remodelling enzymes and histone chaperones, and the increase in histone turnover observed at genes subject to high antisense transcription, antisense transcription may bring about changes in chromatin modification through increased histone deposition and turnover (Figure 34).

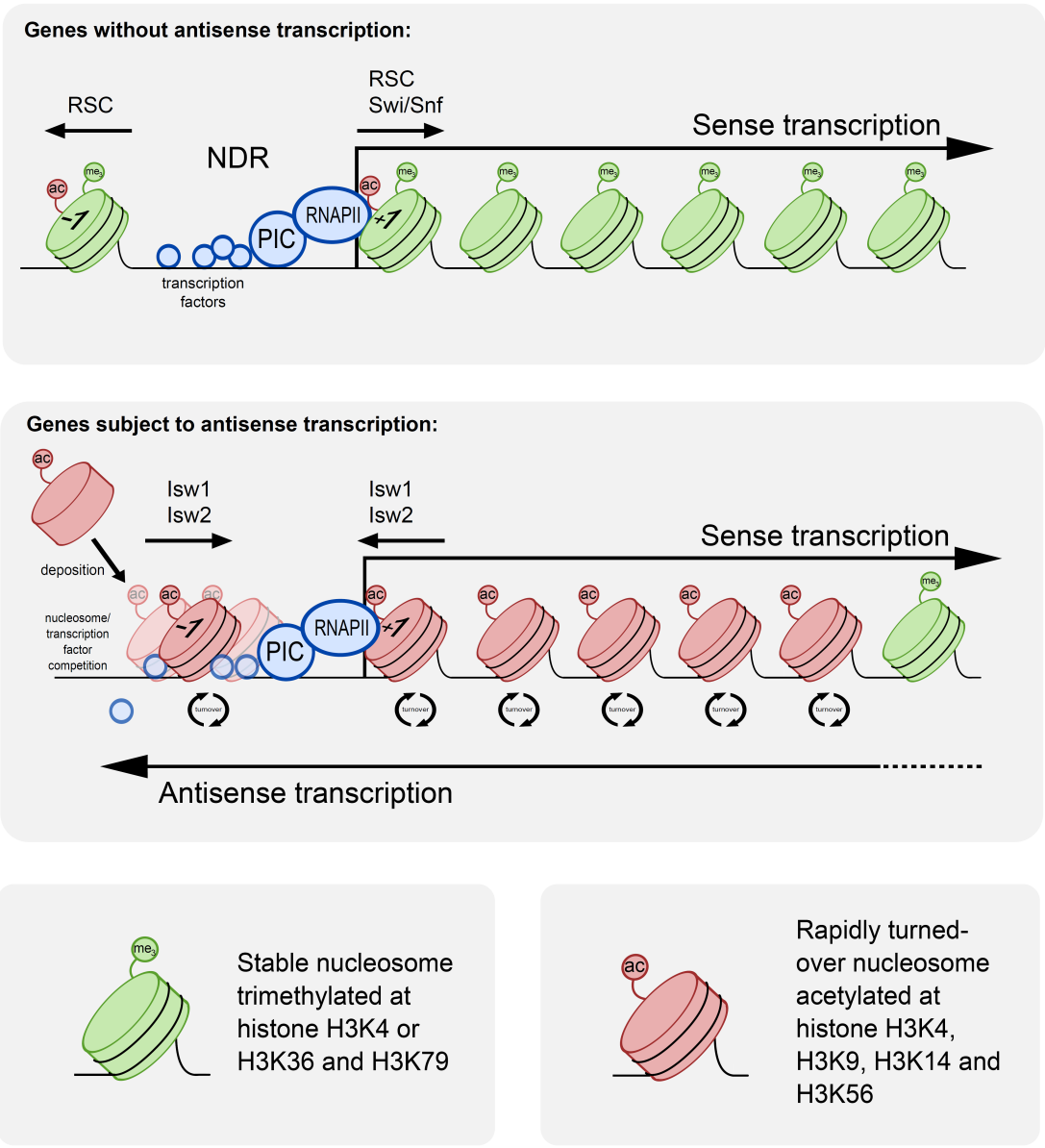


Figure 34: A model demonstrating how antisense transcription might alter gene chromatin

A depiction of how genes might behave in the presence and absence of antisense transcription. In the presence of sense transcription only, the gene has an open NDR at its promoter, maintained by RSC and Swi/Snf complexes, and histones that are hypoacetylated and trimethylated at H3K4, H3K36 and H3K79me3. In the presence of antisense transcription, closure of the NDR occurs, mediated by Isw1 and Isw2, and potentially histone deposition. Nucleosome movement might allow for an increased number of potential promoter configurations in terms of occupied TFBSs, as discussed below, which could explain the association between noise and antisense transcription established in chapter 6. Increased turnover as a result of antisense transcription leads to removal of histones and their replacement with those from the nuclear pool, resulting in increased acetylation at the promoter and gene body.

5.4.2. Antisense transcription is ubiquitous but not always abundant

The profile of nascent sequence reads across many genes is symmetrical, with high levels at the beginning of the transcribed region that tail off to lower levels on both the sense and antisense strands (Churchman and Weissman 2011). Thus in the 300bp window immediately downstream of the promoter, antisense reads tend to be low while sense reads are higher. Nevertheless, when genes are grouped on the basis of levels of antisense transcription in this window, large changes in the chromatin features described above were found. Even when comparing the first two classes, one still observes changes in modifications – i.e. there is a *graded* response to antisense transcription, rather than simply a sudden jump in levels from e.g. the fourth group to the fifth. If the assumption is made that antisense transcription is *causing* these chromatin changes then it implies that antisense transcription can modulate chromatin even at low levels relative to the amount of sense transcription. Thus, antisense transcription could be a general feature of most genes – rather than simply a novelty found in a select group – present at low but nevertheless functional levels. It should still be noted that the presence of an antisense transcript *is* associated with transcriptional plasticity (chapter 3, Xu et al., 2011), suggesting that although most genes may be capable of supporting functionally relevant levels of antisense transcription, some are capable of supporting higher levels than others across different environmental conditions. Another important consideration is that, due to the nature of deep-sequencing data, many antisense transcription events may be missed. At a higher sequencing depth it is possible that one could capture lower-frequency antisense transcription events, such that genes in the lowest class might show evidence of antisense transcription. If such data were available then it might be interesting to repeat the analyses shown here, to assess whether differences in chromatin features between classes become even larger.

5.4.3. Sense and antisense transcription have distinct chromatin associations

Antisense transcription and sense transcription do not show the same associations – not just at the sense promoter, but genome-wide as well. For example, antisense transcription is associated with a rise in acetylation within the gene body, whereas sense transcription is associated largely with a peak of acetylation across the promoter. Perhaps the most striking observations however are H3K36me3 – which is positively correlated with sense transcription in the gene body but negatively correlated with antisense transcription to the same extent – and histone turnover itself, which is more strongly associated with antisense than with sense transcription. That these changes are most prominent in the gene body rather than the 5' or 3' end suggests that they are associated with transcription elongation, rather than initiation. This raises the intriguing possibility that antisense transcription and sense transcription might in some way be fundamentally distinct processes. While it has been demonstrated that antisense transcription initiation involves components of the RNAPII PIC machinery (Rhee and Pugh, 2012), there are other ways in which the polymerase complex could differ, which might perhaps explain the unique observed association, and perhaps also relate to why antisense transcripts are often quickly degraded (van Dijk et al. 2011; Neil et al. 2009). The subunit composition of the elongating polymerase could itself differ: for example, RNAPII contains a subunit, Rpb4, which is not essential for viability, is only found in approximately 20% of cells under optimal conditions, but is found at higher percentages under conditions of stress (Choder and Young 1993; Kolodziej et al. 1990), demonstrating that RNAPII subunit composition can be regulated to functional ends. RNAPII subunit composition has been shown to be altered by the HMG-protein Nhp6a and the mediator component Med3p (Xue and Lehming, 2006), while the RING finger ubiquitin ligase Asr1 can disengage the Rpb4/Rpb7 heterodimer from the polymerase

complex via ubiquitination (Daulny et al. 2008). RNA processing factors have also been found to modulate chromatin. The cap binding complex, for example, plays important roles in transcription termination and export (Das et al. 2000), and has recently been shown to support Set2-dependent H3K36me3 (Hossain et al. 2013). Differential processing of antisense transcripts could therefore also result in the polymerase complex having unique chromatin-regulating activities.

An observation that might provide some insight into this is the finding that sense transcription is strongly correlated with phosphorylation on Ser2 of the RNAPII CTD, specifically downstream of the promoter, whereas antisense transcription is not. Modification of the CTD is known to play important roles at all stages of the transcription process (Hsin and Manley 2012). Amongst these known roles is the recruitment of both histone modifying and chromatin remodelling enzymes. Recruitment of the Set2 methyltransferase, for example, is dependent upon Ser2 phosphorylation (Krogan et al. 2003b), while recruitment of the Hos2 deacetylase and Rpd3S complex is dependent upon Ser5 phosphorylation (Govind et al. 2010). Given that these enzymes bring about H3K36 methylation and histone deacetylation respectively it is tempting to propose that a lack of CTD modification of the antisense-transcribing polymerase leads to a lack of recruitment of these enzymes, resulting in the observed differences. However, was this the case, then there would be no correlation between antisense transcription and these marks, rather than a positive correlation for acetylation and a negative correlation for H3K36me3 as is the case. Examples of chromatin remodelling enzymes that can be recruited to RNAPII are described below. That differentially modified RNAPIIs can recruit different proteins provides an example of how distinct polymerase complexes could exist, and in turn differentially affect chromatin.

Ultimately, however, though sense and antisense transcription could be intrinsically different, whether through polymerase subunit composition, CTD modification or additional proteins, it still raises the question as to what might be the initial cause of this distinction, the upstream effector, as it were, that either makes sense transcription distinct from antisense transcription or vice versa. It is possible that a sequence element might play this role, though use of a Gibbs' sampling algorithm (Neuwald et al. 1995) has so far failed to identify unique elements found at antisense transcript TSSs not found at sense transcript TSSs (data not shown). Another possibility is that bidirectional sense/antisense promoters, though capable of being independently regulated, might still possess a mechanism that defines which promoter is directing non-coding antisense transcription and which the coding sense transcription, a sort of promoter 'dominance' effect. It might be worth comparing such bidirectional promoters with those promoters situated between two divergent coding genes, to investigate whether they contain unique features that might explain possible differences in transcription.

One study that does not necessarily agree with the findings here is that by Houseley et al., (2008). In their study, they found that ablation of the antisense transcript at *GAL10* resulted in *reduced* H3K36me3, rather than increased as might be expected from our analysis. Deletion of *EAF3*, which recruits the Rpd3S deacetylase complex to H3K36me3, reduced the reported effect of antisense transcription upon *GAL10* induction. That this non-coding RNA *increases* H3K36me3 but *decreases* acetylation within its transcription unit actually suggests that it is behaving more like sense transcription than antisense transcription. If other antisense transcripts can be found that reduce H3K36me3 and increase acetylation, as suggested by the above analysis, then comparison of the mechanism of transcription in these cases could shed light on how different *modes* of transcription might arise.

5.4.4. Chromatin remodelling enzymes and histone chaperones can be recruited to RNAPII

Chromatin remodelling enzymes are generally thought to be recruited by direct interaction with transcriptional activators or by binding to specific histone modifications (Peterson and Workman 2000; Carey et al. 2006; Attikum et al. 2004; Santos-Rosa et al. 2004; Memedula and Belmont 2003; Goldmark et al. 2000). However, there is evidence that chromatin remodelling enzymes may be recruited directly by RNAPII, and that this can be achieved in a manner that is independent of histone marks. The SWI/SNF complex has been shown to travel through the coding regions of genes with elongating Pol II (Schwabish and Struhl 2007), the Rsc4 component of the RSC complex is known to bind to the Rpb5 component of RNA polymerases (Soutourina et al. 2006), Spt6 and the FACT complex interact with elongating RNAPII (Schwabish and Struhl 2006; Mason and Struhl 2003), and Isw1 has been found within the gene bodies of actively transcribed genes (Morillon et al., 2003). Moreover, Ino80, in cooperation with Spt6, has been shown to associate with the ORFs of stress-response genes in a manner that is dependent upon elongating RNAPII, but independent of histone modificatio ns such as H3K4me3 (Klopf et al. 2009), whilst the recruitment of Ino80 and Snf2 to the promoter of *INO1* is independent of the presence of histone acetylases (Konarzewska et al. 2012). Similarly, the recruitment of Swi/Snf to *GAL1* requires RNAPII but not histone acetylation (Lemieux and Gaudreau 2004). Based on these findings it is certainly plausible that an antisense-processing transcription complex might introduce chromatin remodelling enzymes and histone chaperones to the promoters and 5' ends of genes, and in doing so bring about the observed changes in modification state without necessarily recruiting histone modifying enzymes.

5.4.5. Antisense transcription is associated with a redistribution of H3K4 methylation and H2A.Z

At genes with high antisense transcription there appears to be a 'spreading' of certain chromatin features. Mono-, di, and trimethylation of H3K4 show a flattening in profile, rather than the defined partitioning seen in other genes, while the histone variant H2A.Z, like H3K4me3, shows increased levels in the body of the gene. Given that H2A.Z is typically associated with the +1 nucleosome (Zhang et al. 2005; Albert et al. 2007), this might suggest the presence of antisense promoters adjacent to the sense promoter, however this is not in keeping with the lack of observation of increased TFIIB or TBP in the same region (Figure 27D), our previous observation that peaks of TBP and TFIIB occur at the 3' end (chapter 3) of genes with antisense transcripts, or the observation that antisense transcripts tend to initiate just upstream of the 3' end (Xu et al., 2011). Rather, it appears that these features might be the result of marks spreading from the promoter, which would presumably necessitate the movement of histones themselves. How might such a movement occur? Work on the *SRG1* non-coding RNA might shed light on a possible mechanism (Thebault et al. 2011). *SRG1* is a non-coding RNA that initiates upstream of the *SER3* gene, and is transcribed in the same orientation (Martens et al. 2004). By replacing the *SRG1* promoter with the inducible promoter of *GAL1*, Thebault et al., (2011) observed a rise in H3 occupancy over the *SER3* promoter following induction. However, using a similar tagging system to that described above to distinguish old and newly bound histones, they found that old, previously bound histones were being moved into *SER3* promoter, not newly deposited histones. These results suggest that the passage of RNAPII can in some scenarios result in previously deposited nucleosomes being re-deposited in its wake, leading to the movement of histones, in the case of *SER3,* from the gene body into the promoter. However, in the case of antisense transcription, such a mechanism would result in the reverse; the

movement of histones from promoter to gene body, which could explain the observed spreading. This histone movement was found to be dependent upon Spt2, which depends on the histone chaperone Spt6 for its recruitment to *SRG1* (Thebault et al. 2011). Intriguingly, levels of Spt6 were increased at the promoters of genes subject to high antisense transcription, thus it is possible that a similar mechanism could be occurring here. Note that it is entirely plausible that the 'spooling' of nucleosomes in this way could occur simultaneously with new nucleosome deposition – i.e. histones are deposited at the promoter and early coding region, for example, and then moved further into the gene body.

5.4.6. Chromatin remodelling enzymes and the NDR

Here, an enrichment of chromatin remodelling enzymes and a closure of the NDR has been identified at those genes subject to high antisense transcription. Components of a broad range of chromatin remodelling complexes were identified: RSC, ISW1, ISW2, SWI/SNF and INO80. That this list is so large is possibly surprising; one might expect to discover a single 'agent' of antisense transcription that is substantially enriched beyond the rest. Similarly, two histone chaperones and one putative histone chaperone were identified, thus it is possible that antisense transcription is associated with enrichment of different remodellers/chaperones at different sense promoters, or else that it is simply non-specific in its recruitment, and so leads to multiple remodellers/chaperones being recruited to all promoters subject to antisense transcription. Not all reported functions of these remodelling enzymes are in keeping with the changes observed here. For example, in RSC-depleted cells the NDR was found to be smaller (Hartley and Madhani 2009), suggesting RSC plays a role in opening the NFR, not closing as our results would suggest. Isw2, on the other hand, does play a role in NDR closure (Whitehouse et al. 2007). Isw1 is thought to function within the body of genes, where it

is proposed to move nucleosomes towards the 3' end (Tirosh et al. 2010). However, if this behaviour is in some way dependent upon the directionality of transcription, then it is possible that antisense transcription could reverse this process, moving nucleosomes instead towards the 5' end. Using genome-wide ChIP combined with deletion analysis, Yen et al., (2012) concluded that chromatin remodelling enzymes are generally associated with specific nucleosomes relative to the start and end sites of genes (i.e. the -1, +1, +2 nucleosomes etc.), and also that they each drive a net movement of nucleosomes that is enzyme-specific. However, given the observations described above, in which remodelling enzymes are recruited to elongating polymerases, and the fact that antisense-transcribed genes are composed of two overlapping transcription units, it is possible that such established rules cannot be applied so simply to these genes. Another important consideration is that chromatin remodelling enzymes can also behave as histone chaperones, thus the increased promoter occupancy could be entirely or in part dependent upon increased histone deposition.

5.4.7. Repressive mechanisms and deacetylation: dealing with discrepancies

As discussed in chapter 1, there are numerous reported studies in which antisense transcription causes repression of sense transcription. This would appear to be at odds with our finding that sense and antisense transcription are not correlated genome-wide. It is worth noting that several of these studies rely on deletion of exosome components to stabilise CUTs (Berretta et al. 2008; Camblong et al. 2007, 2009), leading to resultant repression. Such an additional stabilisation might not occur physiologically, and as such the repression might never occur either. A further possibility is that, while antisense transcription may indeed repress sense transcription, those genes with extensive antisense transcription might have *compensated* for its presence. The antisense transcription-mediated repression might be an unwanted by-product of otherwise

beneficial chromatin changes, and as such the sense promoters are restructured to meet this. In support of this, genes with annotated antisense transcripts more frequently have TFBSs at their sense promoter than those without (data not shown). Another possibility is that repression occurs largely through *trans*-mediated effects, but that the majority of antisense transcripts function genome-wide in *cis*, as proposed by Camblong et al., (2009). Finally, it is possible that antisense transcripts might repress, activate or else have no effect upon sense levels, depending upon the gene or perhaps also the conditions, and hence explaining the lack of correlation genome-wide. However, this would seem to be at odds with the larger number of studies that report repressive effects than they do activatory. Based on the much broader changes in chromatin that are observed, some consistent with activation (e.g. acetylation) others with repression (e.g. NDR closure), it is possible that the effects of antisense transcription may be more nuanced than simple activation or repression, and that any small reductions in sense transcription (the marginal reduction in sense-transcribing RNAPII, for example) might be incidental to antisense transcription's main function.

The second major discrepancy identified is that several studies in yeast have reported recruitment of deacetylase enzymes to antisense-transcribed genes, or the requirement of deacetylases for antisense transcription-mediated repression. Antisense transcript stabilisation led to Hda1 recruitment to *PHO84* (Camblong et al. 2007), the Rpd3 complex was recruited to the *GAL1-10* promoter (Houseley et al. 2008; Pinskaya et al. 2009a), and histone deacetylase inhibitors prevented suppression of the Ty1 retrotransposon by its CUT (Berretta and Morillon 2009). However, work in this chapter demonstrates that genes subject to antisense transcription have high levels of H3 acetylation across their promoters and gene bodies. This would appear to be a discrepancy: surely if antisense transcription was directing deacetylation then one would expect to find low acetylation levels? It is possible, however, that acetylation

levels could still be high in the presence of elevated levels of deacetylase enzymes, so long as there is also a competing process that replenishes the acetylation marks. Histone turnover could fulfil this role by providing acetylated histones to the chromatin. What might be the purpose of this apparently futile process? As discussed below, continual acetylation and subsequent deacetylation is a component of transcription cycling, thought to be necessary for genes to respond to changing conditions (Métivier et al. 2004; Metivier et al. 2003), thus antisense transcription might provide an improved platform for deacetylation and transcription cycling via histone turnover. Note that there *were* significantly enriched levels of Rpd3 at the promoters of genes subject to high antisense transcription (Table 2), thus it is possible that these genes are more subject to high rates of histone deacetylation, but that this is overcompensated for by the high rates of histone turnover, resulting in higher steady-state levels of acetylation overall. As mentioned above, Rpd3 has also been shown to behave as a histone chaperone (Chen et al. 2012), and it is possible that it functions in this capacity and/or as a deacetylase at these genes.

5.4.8. Possible functions of antisense-mediated chromatin modulation

This chapter has so far described novel associations between antisense transcription and chromatin, and described possible mechanisms by which antisense transcription might affect chromatin. However, how might these chromatin changes in turn affect gene behaviour?

The promoters of genes with antisense transcription also share common dynamic chromatin features with the promoters of genes that are highly plastic. Transcriptionally plastic genes possess occupied promoters, have higher histone turnover rates and are more extensively regulated by chromatin remodelling enzymes (Tirosh and Barkai, 2008). That antisense transcription is associated with

transcriptionally plastic promoter features is in agreement with our findings in chapter 3, supporting the idea that antisense transcription enhances a gene's plasticity and makes it more amenable to regulation. Antisense transcripts have already been implicated in transcriptional plasticity (Xu et al., 2011) but, until now, the association between antisense transcription and these chromatin features has not been evident. How might antisense transcription enhance plasticity in this way? In chapter 4 it was established that there is no correlation between sense and antisense transcription within the 300bp window downstream of the TSS, suggesting that antisense transcription does not unanimously up- or downregulate transcription. However, one possible mechanism is that antisense transcription could increase the potential number of chromatin configurations permissible at a promoter, by, for example, recruiting chromatin remodelling enzymes. More possible configurations would mean more potential combinations of bound transcription factors. This could in turn increase the variety of transcription complexes that can form across different environments, and so enhance plasticity. In support of this, genes with more TFBS – i.e. those with inherently more possible permutations of bound and unbound sites – are more transcriptionally plastic (Lin et al. 2010). The association between antisense transcription and NFR could provide an explanation as to why antisense transcription can be both activatory and repressive under different genomic contexts – closing of the NDR could either occlude or unveil TFBSs involved in stabilisation of the transcription complex (Figure 35).

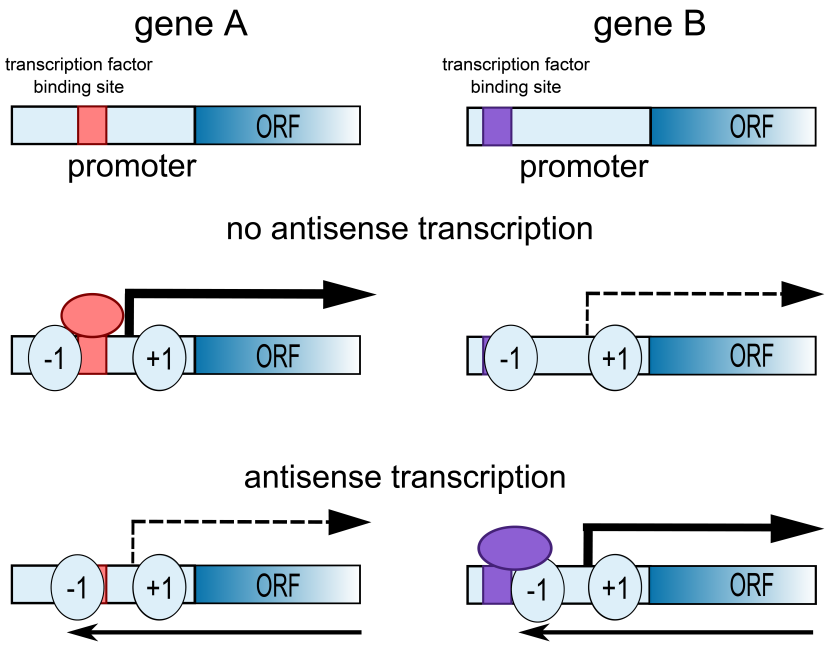


Figure 35: A model of how antisense transcription could be either activatory or repressive under different genomic contexts.

A schematic showing two genes, A and B, with distinct promoter arrangements. In A, a single TFBS is more proximal to the TSS than in B. In the absence of antisense transcription, positioning of the -1 and +1 nucleosomes results in the TFBS in A being amenable to transcription factor binding (red oblong) and thus high levels of sense transcription can be supported. In B however, the TFBS is occluded by the -1 nucleosome, leading to low levels of sense transcription. In the presence of antisense transcription, closure of the NDR (potentially brought about by increased levels of chromatin remodelling enzymes, as discussed in the text) results in the TFBS in A being occluded by the -1 nucleosome, leading to a decrease in sense transcription, while in B it results in the TFBS being made available for binding by a transcription factor (purple oblong), leading to an increase in the level of sense transcription.

By continually promoting histone turnover, antisense transcription could be 'resetting' the modification state of a gene, taking established marks and replacing them with those associated with newly bound histones. The constant deposition, movement of histones and replacement of marks could have the effect of disrupting bound transcription complexes. Studies of the estrogen receptor in human cells demonstrated that the formation of a transcriptionally productive complex at promoters occurred in cycles – assembly of transcription factors, permissive modifications and chromatin remodelling followed by non-permissive modifications and disassembly (Métivier et al. 2004; Metivier et al. 2003). These cycles were proposed to restrict the extent of activation, and

in doing so allow a cell to respond to sudden changes in condition – if a promoter is tightly bound by a particular productive complex directing a particular 'level' of transcription, then it will continue to direct such a level in spite of changing external signals, and so fail to respond. Continual removal of such complexes will mean that with each new assembly available signals will be able to set the new level of transcription. It is possible in a sense/antisense system the sense transcription complex is established by available transcription factors but then removed by subsequent antisense transcription. This would fit with the observation that genes with antisense transcripts are more plastic. In the next chapter a similar idea is developed – that antisense transcription confines sense transcription to 'bursts', as opposed to continual, low level transcription.

Strikingly, many of the chromatin features identified here as being associated with antisense transcription are also associated with genes that show high expression variability – otherwise known as expression *noise*. A previous study identified distinct promoter classes at yeast genes, using available genome-wide data, that were classified largely upon their NDR size, and which were associated with different levels of gene expression noise (Zaugg and Luscombe 2012). The analysis presented above suggests that antisense transcription is associated with changes in promoter state that are highly reminiscent of these classes, as shown in Figure 36. In the next chapter, this possible association with gene expression noise shall be explored, building towards the hypothesis that antisense transcription might, in fact, be behaving as a noise generating mechanism.

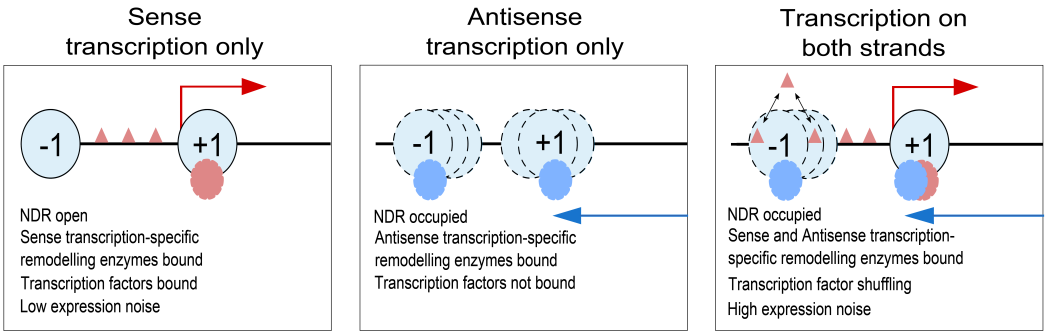


Figure 36: Sense and antisense transcription are associated with distinct promoter classes

Schematics detailing a model in which sense and antisense transcription are associated with distinct promoter classes. In the presence of sense transcription, the NFR is open, TFBSs are occupied, and expression is noisy. In the presence of antisense transcription, but no sense transcription, the NDR is closed and the -1 and +1 nucleosomes potentially poorly positioned. In the presence of both types of transcription, The NDR is closed, however the +1 nucleosome is well positioned to allow sense transcription. However, the promoter configuration results in high gene expression noise, the reasons for which shall be discussed in chapter 6.

Chapter Six

Changing gene expression modes: a role for antisense transcription as a noise-generating mechanism

6. Changing gene expression modes: a role for antisense transcription as a noise-generating mechanism

6.1. Summary

Antisense transcription is associated with changes in gene chromatin. These changes are similar to those associated with gene 'noise' – variability in gene expression within a genetically identical population of cells. Intrinsically noisy genes adopt a distinct mechanism of gene expression, termed transcription *bursting*, distinguished by infrequent bouts of rapid re-initiation events. Bursting gives rise to multiple, nascent transcripts at the site of transcription, which can be detected as bright, nuclear dots using RNA-FISH. Here, the possibility that antisense transcription can itself enforce a bursting mode of sense transcription - and so enhance expression noise –is explored. Using genes in which the level of antisense transcription can be experimentally varied, it is shown that reducing antisense transcription causes a gene to revert to a less bursty, constitutive expression mode. Three genes, *GAL1, GAL10* and *HMS2* all behave in this manner. The levels of histone acetylation change in these genes, in keeping with the previous genome-wide analysis (Chapter 5), and supporting the hypothesis that histone acetylation contributes to bursting. However, deletion analysis suggests that the origin of this acetylated chromatin is not via promoter-recruited modifying enzymes. Most strikingly, deletion of the noise-suppressing complex Set3 only enhances noise in the *presence* of antisense transcription, suggesting that antisense transcription provides the crucial chromatin substrate upon which noise-suppressing processes can act, and explaining why not all antisense-transcribed genes need be noisy.

6.2. Introduction

Antisense transcription is associated with a closed nucleosome architecture at the 5' sense promoter, increased nucleosome occupancy and H3 acetylation, and decreased H3K79me3, H3K36me3 and H2BK123 ubiquitination. Intriguingly, these chromatin features have been previously associated with gene noise (Weinberger et al. 2012; Tirosh and Barkai 2008), and so the question was posed as to whether there was a link between noise and antisense transcription. To begin, it is necessary to review our understanding of gene noise and its association with chromatin and transcription.

6.2.1. Expression noise is an inevitable consequence of cellular systems

In statistics, noise refers to recognised amounts of unexplained variation within a sample. In cell biology, noise has come to refer to variation within a clonal (genetically identical) population that arises from fluctuations in molecular operations – i.e. variation in phenotype in the absence of any variation in genotype. *Expression* noise refers to variation in gene expression within a clonal population. Noise arises due to stochastic biological processes, and can be classed as either intrinsic or extrinsic noise (Thattai and van Oudenaarden 2001). Intrinsic noise refers to the stochastic variation that is associated with the molecular processes directly driving gene expression, or else the functions of the expressed gene product itself. Both transcription and translation can contribute to intrinsic noise (Ozbudak et al. 2002). Extrinsic noise represents the variation resulting from environmental differences between cells i.e. in cell-to-cell contacts or external signalling molecules that have the potential to modify gene expression. To measure noise, it has been necessary to use single cell techniques such as flow cytometry and time lapse microscopy. For example, Elowitz et al., (2002) used an *E. coli* system in which expression of the fluorescent proteins CFP and YFP were controlled by identical regulatory sequences (Figure 37). By comparing the fluorescence of the two

reporter genes both within individual cells and across the population they were able to interrogate intrinsic and extrinsic noise separately, finding that both acted substantially upon the overall variation.

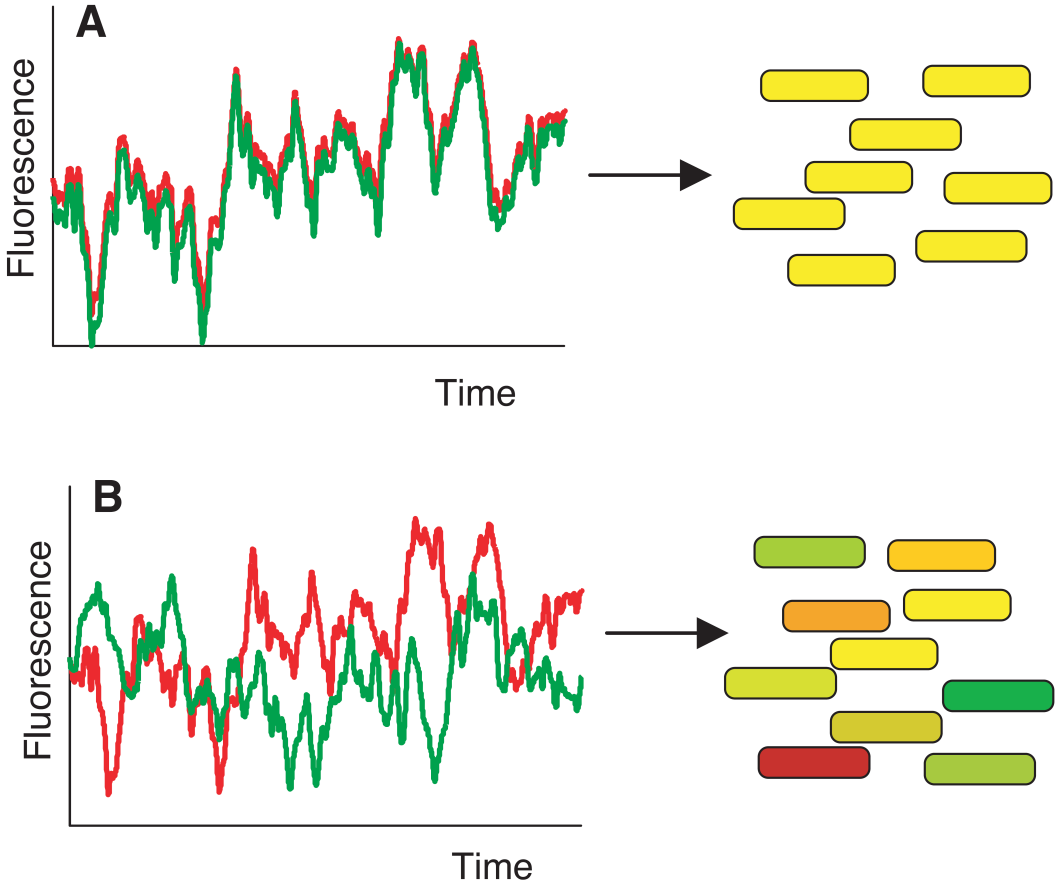


Figure 37: Measuring intrinsic noise in a population of cells.

Intrinsic noise can be measured by assessing fluorescence from cells with copies of the *CFP* and *YFP* genes under the control of identical promoters. Cells expressing equivalent amounts of fluorescent protein appear yellow, while cells expressing more of one than the other appear red or green. Shown on the left is the level of fluorescence of both fluorescent proteins within a single cell over time. (**A)** If the genes are under the control of 'quiet' promoters (i.e. those not directing intrinsically noisy expression) then cells show coordinated expression of both proteins over time within a single cell (left), though the precise amount will vary because of extrinsic noise. **(B)** If genes are under the control of promoters directing intrinsically noisy expression then cells show uncorrelated expression, such that individual cells can have different amounts of protein at any given time. Figure taken from Elowitz et al., (2002).

6.2.2. Noise can be regulated to the benefit of the organism

Noise can arise simply because cellular reactions often comprise only a small number of molecules per cell (e.g. DNA, RNA and proteins). This leads to large concentration

fluctuations which can spread through genetic networks (Thattai and van Oudenaarden 2001), and which could lead to an impairment of functionality or fidelity in biological processes, disrupting the orchestrated physiology of the cell (Rao et al. 2002). More specifically, noise could result in haploinsufficiency in a subset of cells, something which has been proposed to contribute to prostate neoplasia in mice (Magee et al. 2003), or alternatively lead to a loss of cellular homeostasis (metabolic or developmental) or disruption of the stoichiometries of protein complexes (Lehner 2008; Fraser et al. 2004; Batada et al. 2006). With such possible consequences, it might seem prudent to limit noise within a biological system. Indeed, Elowitz et al., (2002) found that deletion of the *recA* gene drastically increased noise in their experimental system, suggesting cells actively suppress noise (in the case of *recA* via the rescuing of stalled replication forks), while in development of the *Drosophila* embryo, morphogen gradients with high embryo-embryo variability are able to give rise to well-defined patterns of downstream gene expression with low embryo-embryo variability, following a phenomenon described as noise 'filtration' (Houchmandzadeh et al. 2002).

While this might implicate noise as simply an unwanted but inevitable property of cellular systems, there is also evidence to suggest that noise can be exploited by the organism. There are well studied instances in which noise underlies biological processes, particularly those in which some sort of binary decision must be made. For example, stochastic molecular events are thought to govern the decision between entering a lytic or lysogenic state in bacteriophage lambda following infection of its *E coli* host. Simply, this decision is governed by two independently produced regulatory proteins operating at low concentrations (Arkin et al., 1998; Johnson et al., 1979). Stochastic variations in the concentrations of these proteins result in a genetically identical population adopting different cell fates (Arkin et al., 1998) – i.e. the decision making mechanism is dependent upon the stochastic processes underlying expression

of both proteins. Noise-dependent mechanisms also underlie latency in human immumodeficiency virus (HIV) - i.e. the decision as to whether the virus will undergo lysis or enter into a reservoir – again governed by stochastic fluctuations in the Tat protein (Weinberger et al. 2005). Noise is also essential for decision making in bacteria (Lopez et al. 2009), determining which of a number of different developmental choices a bacterium will make during conditions of starvation, while in higher eukaryotes it is thought to be a fundamental component of cell differentiation during organismal development (Arias and Hayward 2006).

6.2.3. Noise can enhance environmental responsiveness

In yeast, gene expression noise has been found to be associated with responses to environmental fluctuations. When presented with a medium containing both glucose and galactose (at high and low levels respectively) the *GAL* genes show a bimodal expression within the population (Biggar and Crabtree 2001), indicating that cells decide stochastically between growing on glucose or galactose, something which presumably increases the fitness of the population. This bimodal expression was not observed in yeast grown solely on galactose. A similar observation has been observed on a genome-wide scale (Newman et al. 2006). Here, high-throughput flow cytometry was used in conjunction with a library of GFP-tagged yeast strains (in which genes were tagged at their C-terminus), in order to assign a noise value to roughly half of the genes in the yeast genome. Consequently, questions could begin to be asked about what sorts of genes are noisy and what sorts are not. Strikingly, environmentally responsive genes were found to be noisy while genes involved in protein synthesis (a class of so-called 'housekeeping' genes) were found not to be. In support of this, the stress response in yeast has been shown to rely on noisy signalling mechanisms (Petrenko et al. 2013). Here, noise in signalling pathways were shown to result in genetically identical cells

responding differently to stress conditions. This suggests that it is generally advantageous to express environmentally responsive genes at variable levels – i.e. for some cells to respond more than others to the environmental cue. One can envisage that the noisiness of a particular gene might be evolutionarily tuned to reflect the specific cost/benefit trade-off of noise at that gene.

Why do environmentally responsive genes tend to be noisy? Under fixed environmental conditions, one would predict that all cells would adopt the fastest growing phenotype, not adopt a variety of different phenotypes as is the case. However, outside of controlled laboratory conditions cells tend to encounter fluctuating environments. Stochastic fluctuation models have been used to demonstrate that in such changing conditions, a phenotypically heterogeneous population of cells has the capacity to achieve higher net growth rates than a homogenous one (Thattai and Oudenaarden 2004). Simply, a heterogeneous population samples many different phenotypes – hedging its bets, as it were – and so can anticipate sudden unforeseen changes in the environment.

6.2.4. Transcription bursting

As stated above, expression noise can result from fluctuations in mRNA levels between cells in a population. Indeed, Newman et al., (2006) have proposed that this is the *major* source of expression noise, based on the observation that mRNA and protein levels in single cells tended to be correlated. But by what mechanism does noisy transcription arise, and is it different from 'quiet' transcription?

It has been proposed that transcription occurs in a digital fashion – that is, the transcription apparatus at a gene exists in either an "on" or "off" state at any given time (Hume 2000). In such a model, a population of cells would, upon activation of a gene, possess a subpopulation in which that gene was being transcribed (i.e. in which it was

"on"), while in the remaining cells the gene would not be being transcribed (and therefore be "off"). The higher the level of activation, the higher the fraction of cells at any given time in which the gene is "on". This is opposed to a model in which genes have a graded response, whereby in a population of cells every gene is being transcribed at the same rate, reflecting the activation level. Support for the digital model came from a study in which activation of the PAI-2 gene in murine macrophages was shown to result in an increase in the fraction of cells with high PAI-2 levels (Costelloe et al. 1999). Expression of a *lac*Z reporter gene in macrophages in response to varying concentrations of external stimuli followed a similar pattern – the higher the signal, the higher the fraction of expressing cells (Ross et al. 1994). This led to the hypothesis that transcriptional regulation involves changing the probability of transcription initiation, not the rate, and that transcription initiation is a stochastic process. The *transcription noise* resulting from such a stochastic process has been shown to contribute substantially to the level of heterogeneity within a clonal population (Blake et al. 2003). The mechanism of stochastic transcription is thought to be dependent upon rapid rounds of transcription re-initiation whilst the gene is in the "on" state. Mutation of the TATA-box – an element which supports re-initiation (Yean and Gralla 1997) – was shown to decrease expression noise (Blake et al. 2003).

If all genes adopt this stochastic mechanism of transcription initiation, then why are some genes noisy and others quiet (Newman et al. 2006)? Zenklusen et al., (2008) shed light on this by utilising RNA-FISH to define two extreme modes of gene transcription – the two-state model described above, and a one-state model (Figure 38). They demonstrated that the rapid bouts of re-initiation associated with digital, two-state transcription could be visualised as bright, nuclear dots – termed transcription *bursts,* which arise from multiple elongating polymerases, and therefore nascent mRNAs, being present at the site of transcription. Conversely, low noise levels were achieved through a

mode of transcription in which single transcription events are clearly separated in time – i.e. though initiation occurs, the probability of *re*-initiation is low, which they refer to as non-correlated or constitutive transcription. For a given gene they were able to class it as adopting either a *bursting* mode or a *constitutive* mode of transcription, by determining the number of nascent transcripts (visible as nuclear dots), and by estimating the number of nuclear transcripts (by dividing the intensity of the nuclear dots by the average intensity of the cytoplasmic dots)(Figure 39). For example, the gene *DOA1* was characterised by a narrow distribution of cytoplasmic mRNAs (i.e. lower variation) with generally 0 or 1 nuclear, nascent transcripts at any given time. This gene is therefore adopting a constitutive more of transcription. *PDR5*, however, showed a much broader distribution of cytoplasmic mRNAs (i.e. high variation), along with many more nascent transcripts, and was therefore classed as *bursty.* One can see simply by comparing the cytoplasmic mRNA distributions of these two genes that *PDR5* has more cell-cell variation, and Zenklusen et al., (2008) used computational modelling to demonstrate that this is a result of the differences in nuclear transcript distributions.

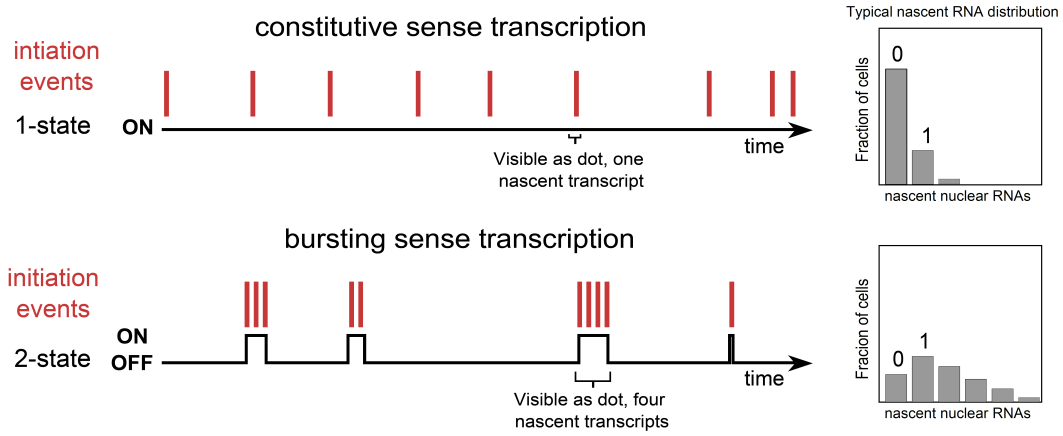


Figure 38: A two-state and one-state model of transcription.

(**Top)** A one-state model of sense transcription, in which a gene is constantly in an active state, giving rise to frequent, single, initiation events (red dashes), and a profile of nascent transcription in which single transcripts are most frequently observed. Genes whose behaviour is consistent with this model are described as constitutive. **(Bottom)** A two state model, in which genes switch between an on and an off state. In the on state the gene undergoes rapid rounds of re-initiation, while in the off state no transcription occurs. This gives rise to a nascent transcription profile in which multiple nascent transcripts are frequently observed. Adapted from Hebenstreit (2013).

Intriguingly, the *KAP104* gene showed a behaviour that was halfway between these two other genes, both in terms of its nascent and cytoplasmic transcription profiles (Figure 40)(Zenklusen et al., 2008). This suggests that bursting and constitutive transcription represent extreme types of behaviour rather than distinct modes. Hornung et al., (2012) formalised this continuum by defining the terms burst *size* and burst *frequency*. Burst size refers to the duration of a burst, i.e. the number of successive rounds of re-initiation within a single "on" step (Figure 38), while burst frequency represents the number of bursts. In the examples above, *PDR5* would have a high burst duration but a low burst frequency, *DOA1* would have a low burst duration but a high burst frequency, and *KAP104* would have values that were intermediate. Note that in this way it is possible for two genes to have the same overall expression level within a population whilst

having different noise values. This suggests that noise levels can be set at the site of transcription, but how is this achieved?

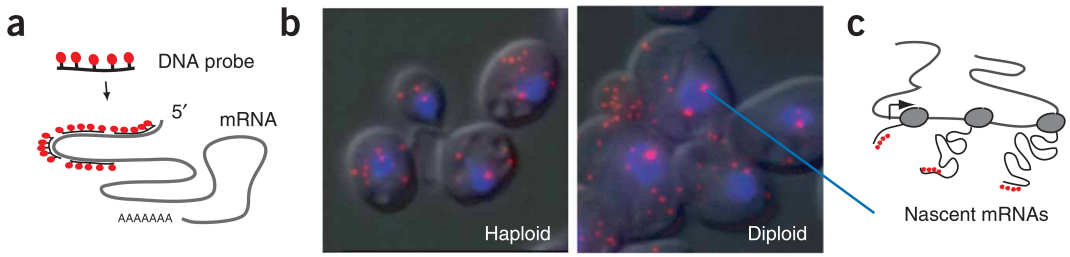


Figure 39: Multiple nascent mRNAs are detectable by RNA-FISH as bright dots in the nucleus

(**A)** Multiple probes designed to bind to the 5' end of mRNAs allows for detection of the nascent transcript while it is still bound to the polymerase. Multiple probes are used to improve signal-to-noise ratio and increase fluorescence. **(B)** RNA-FISH images showing bright nuclear dots (indicated by the blue line), and less intense cytoplasmic dots. **(C)** Bright nuclear dots are thought to represent sites where multiple polymerases are transcribing the same gene, hence why the intensities of such dots are greater than those in the cytoplasm, which are thought to represent single mRNAs. Taken from Zenklusen et al., (2008).

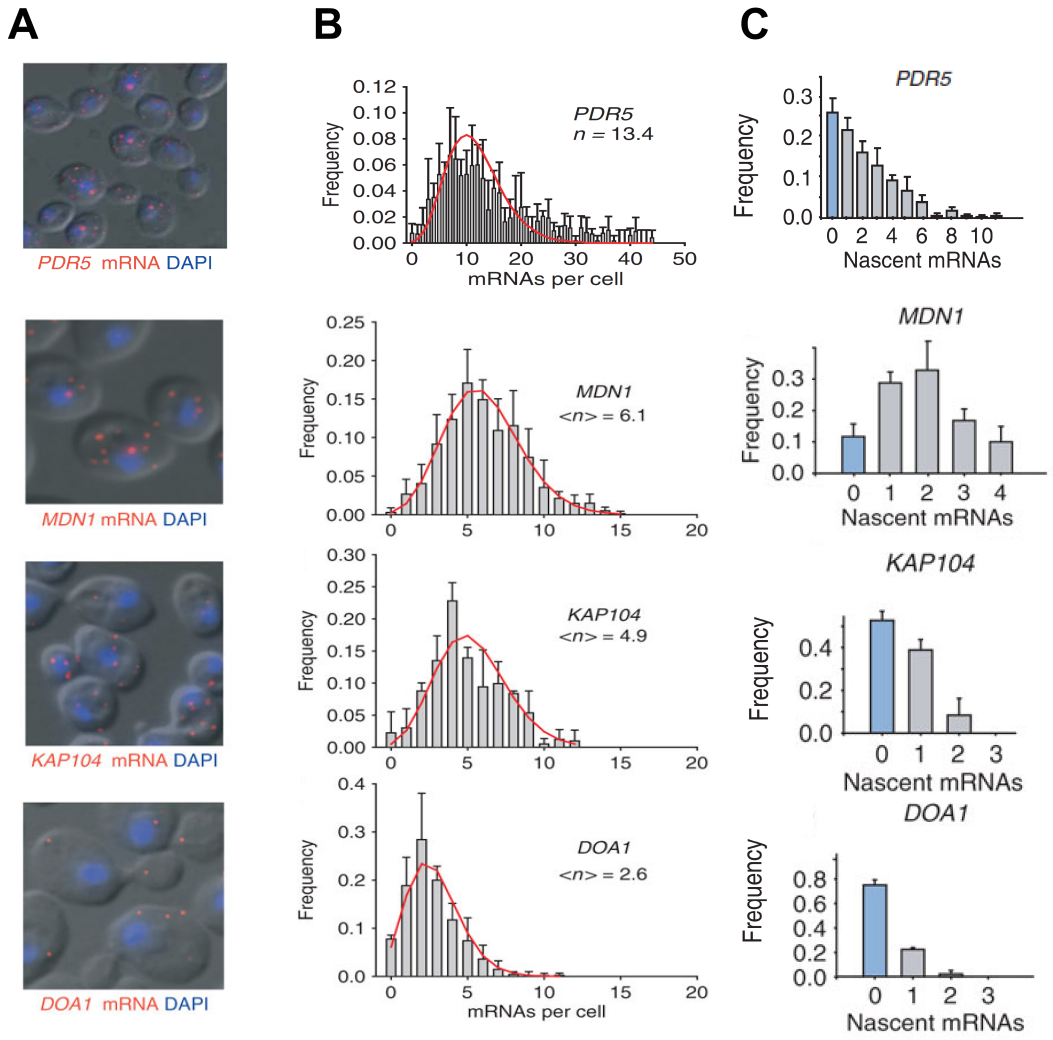


Figure 40: Different genes have distinct distributions of cytoplasmic and nuclear transcripts, reflecting their different transcription modes

(A) RNA-FISH images showing cells probed for the mRNA transcripts of four different genes. Visible are bright dots within the DAPI stained nucleus, thought to represent multiple nascent transcripts at the site of transcription. Also visible are dots within the cytoplasm, thought to represent single transcripts. *PDR5* and *MDN1* show a higher incidence of nuclear dots than *KAP104* and *DOA1*. **(B)** Histograms in which the number of cytoplasmic dots have been counted per cell. *PDR5* and *MDN1* have a larger number of cytoplasmic mRNAs than *KAP104* but also a broader distribution, indicative of a higher standard deviation in cytoplasmic transcript count i.e. noisier transcription. **(C)** Histograms in which the number of *nuclear* dots have been counted per cell. *PDR5* and *MDN1* have a higher incidence of multiple nascent RNAs than *DOA1* and *KAP104*. In *MDN1*, this is thought to arise from the gene being much longer than the other three (15kb), meaning that the chance of multiple polymerases being on the gene at the same time is much higher. In *PDR5*, however, which is of a similar size to *KAP104* and *DOA1*, this is thought to arise from a difference in the transcription initiation mechanism, in which transcription occurs in bouts of rapid re-initiation, a more inherently stochastic mechanism resulting in the wider observed distributions of cytoplasmic transcripts. Adapted from Zenklusen et al., (2008).

6.2.5. Setting the transcription mode: mechanisms of noise regulation

Noise has been found to be associated with a number of promoter features. Noisy genes often possess a TATA-box and have high nucleosome occupancy over their promoter (Zaugg and Luscombe, 2012; Choi and Kim, 2009; Field et al., 2008; Tirosh and Barkai, 2008). Given the TATA-box's ability to support re-initiation (Yean and Gralla 1997), it is clear how it could promote noise by increasing burst size, but how might nucleosome positioning affect noise? It is conceivable that having more nucleosomes over a promoter occludes more transcription factor binding sites. If the nucleosomes are mobile (or 'fuzzy') as has been shown to be the case at noisy genes (Zaugg and Luscombe 2012) then different binding sites might be occluded in different cells, leading to higher resultant variability. This would also suggest that chromatin remodelling could play important roles in noise regulation by structuring the promoter, and indeed studies at the *PHO5* promoter found that transcription noise was highly sensitive to mutations of chromatin remodelling enzymes and mutations of upstream activator sequences involved in the recruitment of such enzymes (Raser and Shea 2004).

A recent study utilised a genetic screen to identify factors involved in noise (Weinberger et al. 2012). Briefly, 137 different chromatin factors were selected and mutated to assess their effects on the expression noise of a reporter gene driven by 11 different promoters. They found that burst size (and therefore noise) could be increased by prevention of H2B ubiquitination and Set3-dependent deacetylation. Specifically, deletion of *LGE1* and *RAD6* (required for H2BK123 ubiquitination) and deletion of *SET3* and *HOS2* (both components of the *SET3* deacetylase) increased burst size and therefore noise, while preventing H2BK123 via the H2B-K123R mutation had a similar effect. Deletion of *SET3* resulted in an increase in levels of H3K9ac within gene bodies genome-wide, while H2B ubiquitination was found to promote *SET3* dependent deacetylation, suggesting that the two modifications are linked. Weinberger et al., (2012) proposed

that both modifications altered burst size (and therefore noise) by affecting RNAPII processivity – ubiquitination positively and acetylation negatively. This is based on the observations that blocking H2B ubiquitination increases sensitivity to drugs interfering with transcription elongation (Kim and Buratowski 2009), and that *SET3* was found to work preferentially at those genes with low RNAPII efficiency (which is linked to processivity) (Weinberger et al. 2012). Taking the above observations together, it seems that noise can be tuned at a gene by operations at the promoter – controlling access via the nucleosomes and setting the re-initiation rate via the TATA-box – and in the gene body, through modulation of Pol II processivity.

6.2.6. A hypothesis: antisense transcription as a noise-generating mechanism.

In chapter 5 a number of chromatin features were identified as being associated with high levels of antisense transcription. Strikingly the three features described above as being associated with noise – high promoter occupancy, high H3 acetylation in the gene body and low H2B ubiquitination in the gene body – are also associated with antisense transcription. Furthermore, it is easy to identify links between these and many of the other chromatin features associated with antisense transcription. For example, H3K79me3 and H3K36me3 have been linked to H2BK123ub, are also found in the gene body, and have all been linked to nucleosome stability (Batta et al. 2011; De Vos et al. 2011; Fleming et al. 2008; Lee et al. 2012; Rufiange et al. 2007; Venkatesh et al. 2012; Smolle et al. 2012). H3 acetylation on the other hand, specifically H3K56ac, has been linked to histone turnover (Kaplan et al. 2008; Li et al. 2008), which is high in genes subject to high antisense transcription. It is easy to envisage that increased histone turnover and occupancy would result in decreased RNAPII processivity, which would in turn result in an increased burst size, and hence increased noise. The focus of this chapter is to assess how these features relate to noise through a bionformatical analysis,

and then to assess experimentally whether there is a link between antisense transcription and transcription bursting.

6.3. Results

6.3.1. Antisense transcription is associated with increased noise

On the basis of the observations in chapter 5 it was hypothesised that genes subject to high antisense transcription would have higher transcription noise. The previous study described above (Newman et al. 2006) was used in order to assign a measure of noise to the majority of genes in the yeast genome in order to test this hypothesis. Strikingly, the class of genes subject to the highest levels of antisense transcription were significantly noisier than those without (Figure 41, $p = 6 \times 10^{-8}$, Wilcoxon rank sum test). Given that noise is associated with TATA-boxes (Tirosh and Barkai 2008), TATA-box containing-genes were removed from the analysis to assess whether the association remained, and indeed it did (Figure 41). This is in accordance with the finding in chapter 5 that antisense transcription-associated changes in chromatin were independent of TATA-boxes. Intriguingly, amongst TATA-box containing genes, antisense transcription was also associated with changes in noise, suggesting that antisense transcription can confer additional noise on top of that provided by the TATA-box.

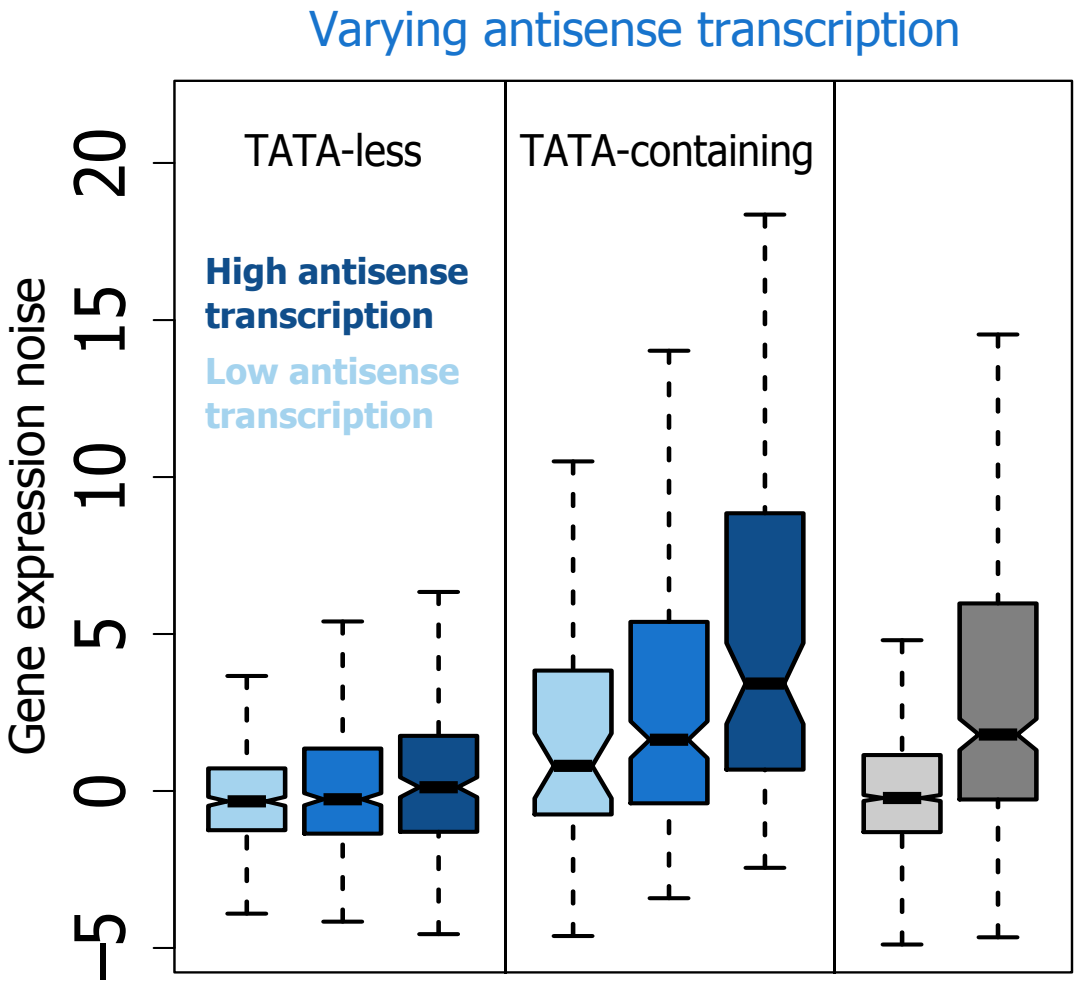


Figure 41: Antisense transcription is associated with gene expression noise.

Boxplots showing the levels of antisense transcription in genes with varying levels of antisense. Groups are as those in chapter 5, with darker blue indicating higher antisense transcription, with the middle three classes combined into one group. Because of the known association between TATA-boxes and noise, (Tirosh and Barkai 2008), genes were subdivided TATA-containing and TATA-less. An association was still observed in both classes. For comparative purposes, the noise levels of all TATA-less genes (light grey) and TATA-containing genes (dark grey) are also shown.

There are two important and related considerations to be made with regards to this analysis. Firstly, noise was measured through GFP levels – i.e. it is gene expression noise that is being measured, not transcriptional noise, and these are clearly related but not necessarily identical parameters. Secondly, GFP was inserted at the 3' ends of genes. This is likely to interfere with antisense transcription initiation – deleting the 3' end of

genes, for example, has been shown to ablate antisense transcription (Uhler et al. 2007). Nevertheless, an association is still observed. It is possible that the GFP insertion is removing or reducing antisense transcription at a subset of genes within this analysis, i.e. those in which the antisense promoter is not wholly or partially upstream of the insertion sight. If this is the case, then it might be expected that the true association – between antisense transcription levels and sense transcription noise at wild type genes – would be higher.

6.3.2. *GAL1* has an antisense transcript and adopts a bursting mode of transcription

To determine whether noise was a feature of genes with antisense transcription *GAL1,* which has a XUT initiating from its 3' end, was selected as a candidate gene for experimental analysis (chapter 3). If antisense transcription confers a bursting mode of transcription to *GAL1* then one would expect to see nuclear dots indicative of such bursting, as reported previously (Zenklusen et al. 2008). RNA FISH was used to visualise sense transcription in *GAL1*, with four probes complementary to the early coding region of *GAL1,* each containing four fluorophores. The aim was to calculate the number of nascent transcripts present within the nucleus – transcription bursts would result in some cells having multiple RNAPII molecules present on *GAL1* at the same time, and multiple nascent transcripts. To estimate the number of nascent transcripts, the intensity of a given nuclear dot was compared to the intensity of cytoplasmic dots. Cytoplasmic dots were assumed to correspond to single transcripts, and as such an average cytoplasmic intensity was determined for each image and the intensity of each nuclear dot divided by this to obtain the number of nuclear, nascent sense transcripts. Cells were only included in the analysis if they showed evidence of sense transcription, i.e. if they had cytoplasmic signal above a background signal (calculated using a *GAL1*

knockout strain). Thus, a cell with cytoplasmic signal above background and a nuclear dot with twice the intensity of a cytoplasmic dot was considered to have two nascent transcripts, whereas if the nuclear dot had an intensity equivalent to that of a cytoplasmic dot then it was considered to have one. If no nuclear dot was present then the cell was considered to have no nascent transcripts, unless the cytoplasmic signal was below background, in which the case the cell was not considered at all for the purposes of the histograms shown below. Thus this analysis only considered cells in which the gene of interest is in a 'primed state' – i.e. in one of the patterns of transcription shown in Figure 38, rather than in a completely 'off' state in which no transcription initiation occurs at any time point. It also removes from consideration any cells that might not have been spheroplasted effectively, and which might therefore be transcriptionally active cells in which probing did not occur, rather than being transcriptionally inactive.

Following *GAL1* induction, bright nuclear dots were identified in 85% of induced cells, up to five times the intensity of a single cytoplasmic dot – i.e. up to five nascent transcripts (Figure 42A). This is comparable to *PRD5,* for example, which also displays a bursting expression profile (Figure 40; Zenklusen et al., 2008). Additionally, *GAL10*, which shares its promoter with *GAL1* and has closely coordinated expression (Gandhi et al. 2011), was studied using the same RNA-FISH approach. *GAL10* showed a similar distribution, again indicating a bursting mode of transcription (Figure 42B). If antisense transcription does enforce bursting, then it is possible that antisense transcription into the *GAL1-10* bidirectional promoter is enforcing bursting of both genes, perhaps by modulating the promoter chromatin environment. A separate RNA-FISH study assessing sense and antisense transcription at *GAL10* found that both transcripts could be found separately within the nucleus, and that they were only very rarely present together (1.2% of cells; Figure 43). This finding is similar to that of a previous study in *PHO84*

(Castelnuovo et al. 2013). If antisense transcription does modulate sense transcription then this finding suggests that the effects of antisense transcription are retained after the round of antisense transcription has itself finished, i.e. antisense transcription modulates sense promoter chromatin which in turn modulates subsequent rounds of sense transcription.

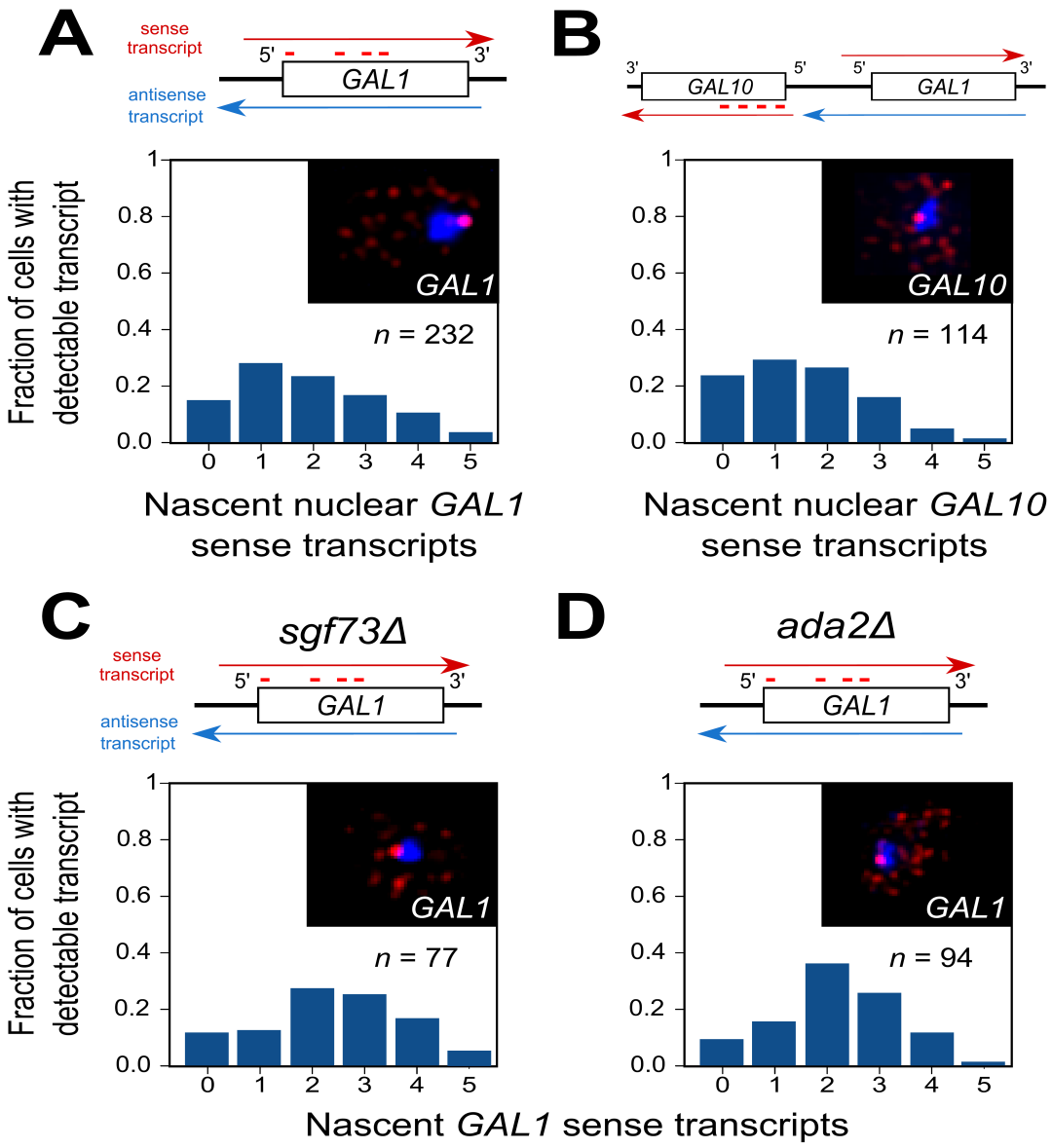


Figure 42: Genes subject to antisense transcription adopt a bursting mode of sense transcription.

(A-B) Sense transcription at *GAL1* and *GAL10*, which share a promoter that is subject to *GAL1* antisense transcription. Shown are the distributions of nascent, nuclear sense transcripts of both genes following induction in galactose for 2 hours in the presence of antisense transcription, determined by RNA-FISH. Positions of the FISH probes are shown on the schematic. *n* describes the number of cells with detectable transcript from which the data were generated. The inset shows representative cells containing bright nuclear dots. **(C-D)** Distributions of nascent, nuclear *GAL1* sense transcripts in the *sgf73Δ* and *ada2Δ* strains. Image acquisition was performed by Françoise Howe and Simon Hänni. Image analysis was performed by the author.

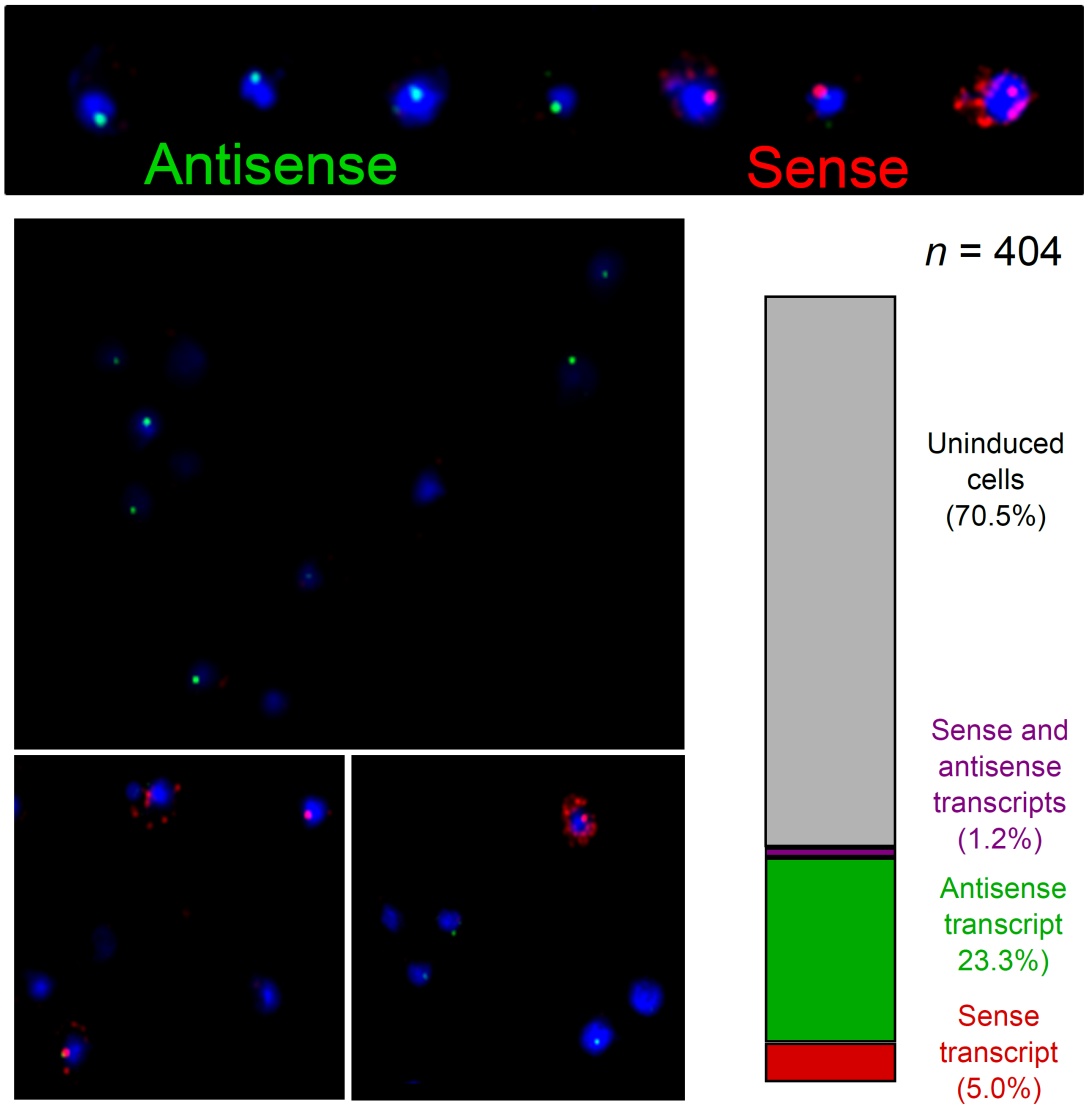


Figure 43: Sense and antisense transcription are generally not present in the same cell.

GAL10 has a relatively stable antisense (SUT013) suitable for RNA-FISH studies (Houseley et al. 2008). Cells were grown in YPD and then induced for 90 minutes in YPG. RNA-FISH in which the sense (red) and antisense (green) transcripts of *GAL10* were probed for simultaneously. FISH probes do not overlap and hybridize to different regions of *GAL10*. Images show selected individual cells illustrating typical images for the sense and antisense transcripts. The antisense transcript is predominately associated with the DAPI stained nucleus. Shown are three separate fields of cells double labelled for sense and antisense transcripts. n refers to the number of cells analysed. The bar shows percentages of cells demonstrating either sense transcription, antisense transcription or both. The numbers of cells expressing both is approximately what one would expect if sense and antisense transcription are uncorrelated processes (5% of 23.3% = 1.2%). Experiments were performed by Simon Hänni.

6.3.3. Varying antisense transcription at *GAL1* changes the histone modification pattern without changing the level of sense transcription

As discussed in chapter 3, and shown in (Figure 44), inserting the terminator sequence of *ADH1* into the ORF of *GAL1* resulted in redefinition of the transcription unit whilst maintaining levels of the redefined sense and antisense transcripts. Mapping of these transcripts revealed that the *GAL1* antisense transcript TSS was 17bp downstream of the TTS of the *GAL1* sense transcript, while the antisense transcript end site was 128bp upstream of the sense TSS (Figure 45), well into the *GAL1-10* promoter, suggesting that it might be able to modulate sense promoter structure.

Intriguingly, the inserted *ADH1* terminator contains a consensus TATA-box in the antisense orientation, which was hypothesised to grant the terminator its promoter-like activity. Mutation of this TATA-like sequence did indeed reduce antisense transcript levels, even in an *XRN1 (KEM1)* deletion strain, suggesting that antisense transcription itself has been reduced (Figure 44 and Figure 46). Furthermore, sense transcription was also not affected by this deletion, indicating that antisense transcription is acting neither to enhance nor repress sense transcription at this gene, in keeping with the lack of correlation between sense and antisense transcription observed in chapter 4. The ability to ablate antisense transcription at *GAL1* provided an experimental system in which the effects of antisense transcription on both transcription mode and chromatin could be observed.

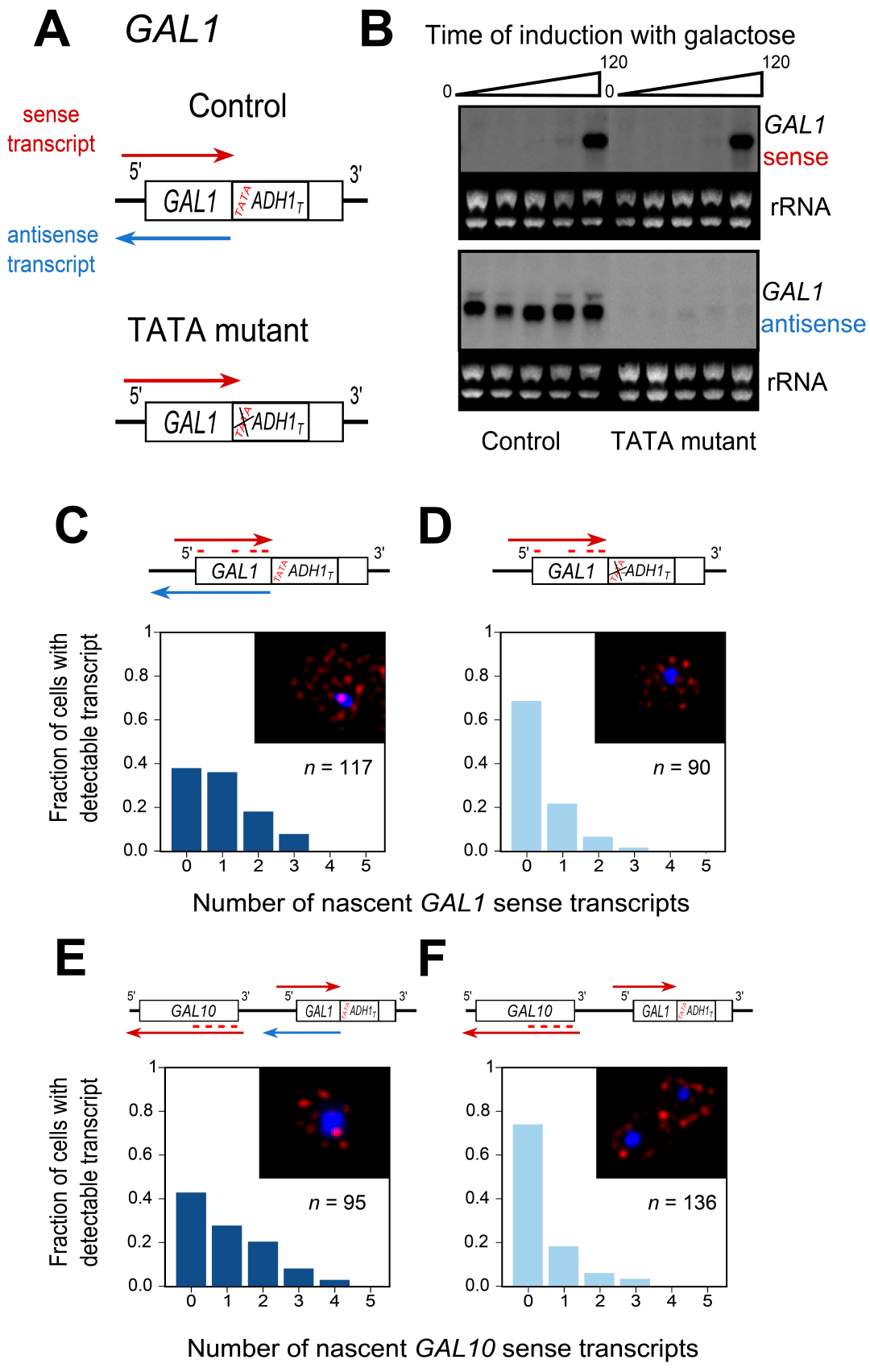

A
GAL1
Control
sense transcript
5'
GAL1
TATA
ADH1T
3'
antisense transcript
TATA mutant
B
Time of induction with galactose
0
120
0
120
GAL1 sense
rRNA
GAL1 antisense
rRNA
Control
TATA mutant
C
D
Fraction of cells with detectable transcript
n = 117
n = 90
Number of nascent GAL1 sense transcripts
E
F
GAL10
Fraction of cells with detectable transcript
n = 95
n = 136
Number of nascent GAL10 sense transcripts

Figure 44: Experimentally varying antisense transcription changes the mode of sense transcription.

(**A)** Antisense transcription into the *GAL1* promoter from the inserted *ADH1* terminator (*ADH1*$_T$) can be disrupted by mutation of a TATA-like sequence within *ADH1*$_T$. **(B)** Northern blot probing for the *GAL1* sense and antisense transcripts in both strains shown in A during an induction time-course (times in min). rRNA in the ethidium-stained gel is shown as a loading control. **(C-D)** Distribution of nascent nuclear *GAL1* sense transcripts in the control and TATA mutant *GAL1* strains after a 2 hour induction in galactose, as determined by RNA-FISH. Positions of the FISH probes are indicated on the schematic. The inset shows representative cells. *n* describes the number of cells with detectable transcript from which the data was generated. **(E-F)** Distribution of nascent, nuclear *GAL10* sense transcripts in the control and TATA mutant *GAL1* strains after a 2 hour induction in galactose. Strain design and construction and northern blotting were performed by Simon Hänni. Image acquisition was performed by Françoise Howe and Simon Hänni. Image analysis was carried out by the author.

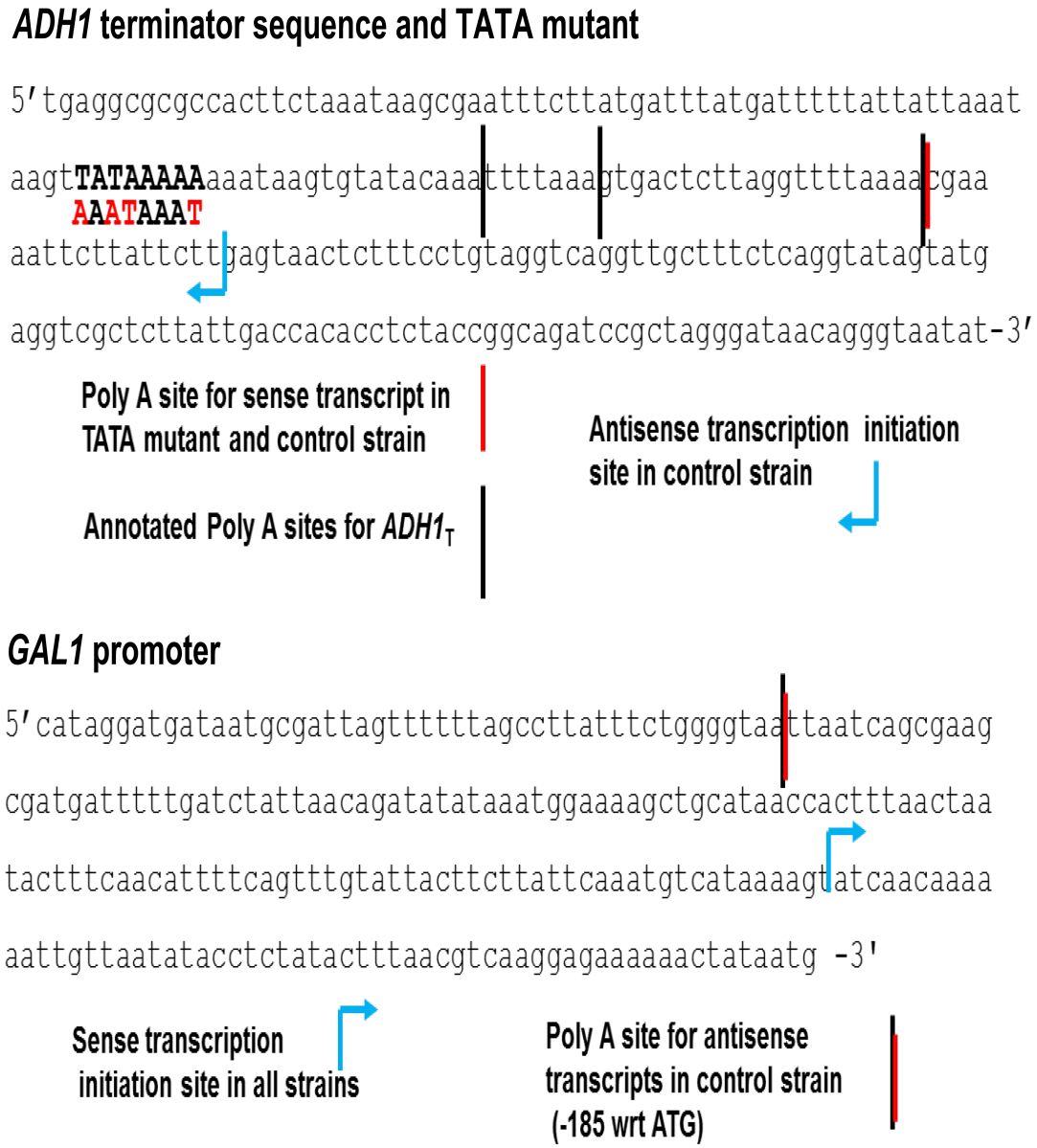


Figure 45: Mapping of sense and antisense transcripts in the *GAL1:ADH1*$_T$ reveals that antisense transcription enters the sense promoter.

(Top) *GAL1:ADH1* sense/antisense RNA mapping by Reverse Ligation mediated-PCR (Grange 2008)(with decapping and dephosphosphorylation) at the 3‘ region of the construct with *ADH1*$_T$ inserted at +757bp in the *GAL1* ORF. Also shown is the position of the TATA-like sequence, as well as its sequence following mutation. **(Bottom)** *GAL1* sense/antisense RNA mapping by RL-PCR at the 5‘ region in the *GAL1* promoter. 3‘ end sites were confirmed using 3‘ RACE. Experiments were performed by Simon Hänni.

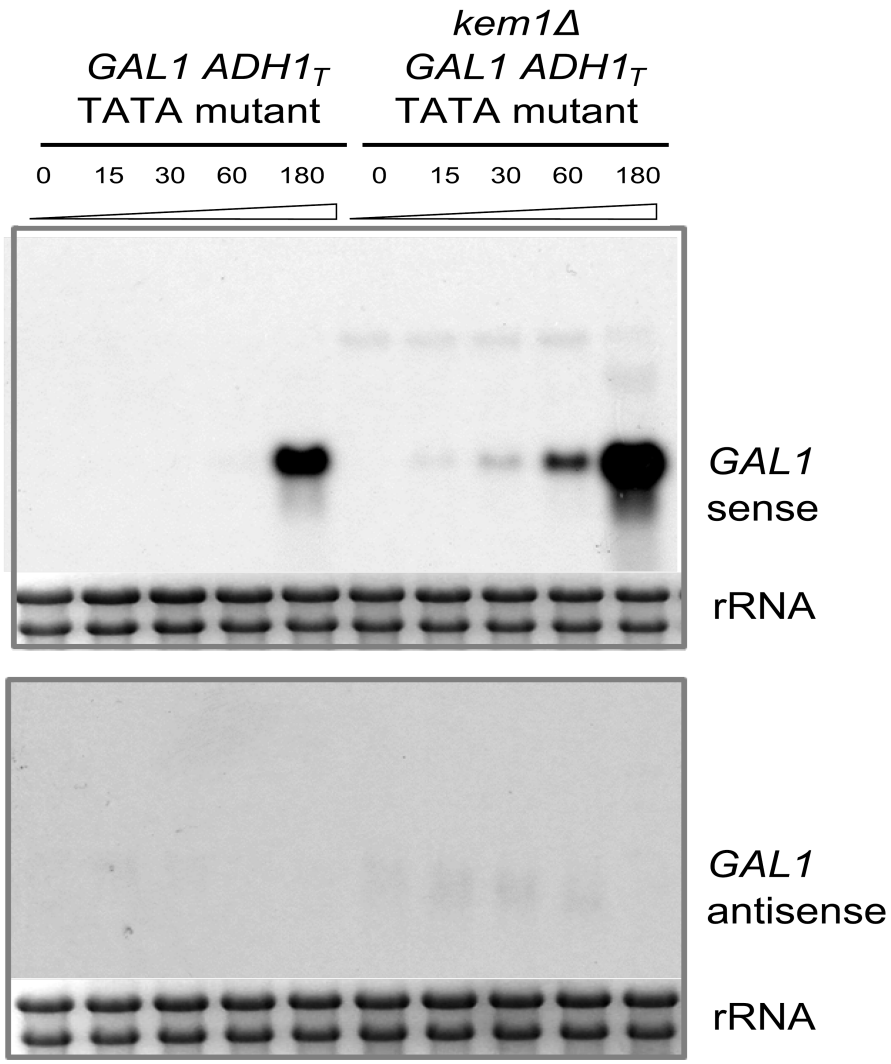


Figure 46: Mutation of the TATA-like sequence in ADH1$_T$ reduces antisense transcript levels by reducing antisense transcription, *not* via changes in stability.

Northern blots showing levels of sense and antisense transcription in wild type cells and a *XRN1(KEM1)* delete. Cell were grown in YPD and induced in YPG for 0, 15, 30, 60 or 120 minutes, using a probe corresponding to +356 to +651 relative to the ATG of *GAL1*. Levels of the antisense transcript do not increase substantially following deletion of *KEM1*, confirming that loss of the TATA-like sequence does not result in an unstable transcript but in an actual reduction of transcription.

The genome-wide analysis from chapter 5 shows that antisense transcription is associated with changes in chromatin at the promoter and early coding region. If

antisense transcription were causing these changes then there should also be differences in chromatin within these constructs. The levels of histone modifications were measured, specifically H3, a series of H3 acetylation marks (K9,14,18,23ac), H3K4me3 and H3K36me3 across both *GAL1* constructs. Based on the genome-wide analysis, it was predicted that H3 and H3 acetylation levels would fall following ablation of antisense transcription, that H3K36me3 would rise, and that H3K4me3 would become less evenly spread across the gene. ChIP experiments were performed in glucose-containing media, in the absence of the sense transcript (which is induced by galactose), in order to strictly assess changes associated with antisense transcription rather than sense.

In agreement with the predictions, levels of H3 and H3 acetylation (all four marks) did fall following mutation of the TATA-like sequence and reduction of antisense transcription (Figure 47A-B). This suggests that antisense transcription is itself causing these chromatin changes. The difference in H3 levels was most pronounced at the site adjacent to the initiation of the antisense transcript. H3K36me3 was lower in the construct with higher antisense transcription (Figure 47C), again as predicted. However, a large difference was only observed within the *GAL1-10* promoter, and not also over the gene body as predicted. Nevertheless, a loss of H3K36me3 and increased H3 and H3 acetylation is consistent with a rise in histone deposition within the promoter. H3K4me3 did not strictly behave as predicted - levels of the modification were ubiquitously higher across the gene, and were highest immediately upstream of the site of antisense initiation (Figure 47D). Given that these experiments were conducted in glucose-containing media, and given that the construct with antisense transcription has an additional transcription unit, this rise might be due to the association between transcription and H3K4me3 in the 5' prime region of a transcription unit (Liu et al., 2005; Pokholok et al., 2005). Generally, however, the observed changes are consistent

with the predictions of the genome-wide analysis in chapter 5. Furthermore, these changes are not consistent with what would be expected if the antisense transcript were behaving more like a canonical sense transcript, in which case there would be a gain of H3 and a loss of H3K36me3 upon loss of transcription. It can be concluded, therefore, that the antisense transcript in the *GAL1* construct behaves in a manner similar to antisense transcripts genome-wide, and that it is causing chromatin changes in the promoter and gene body of *GAL1.*

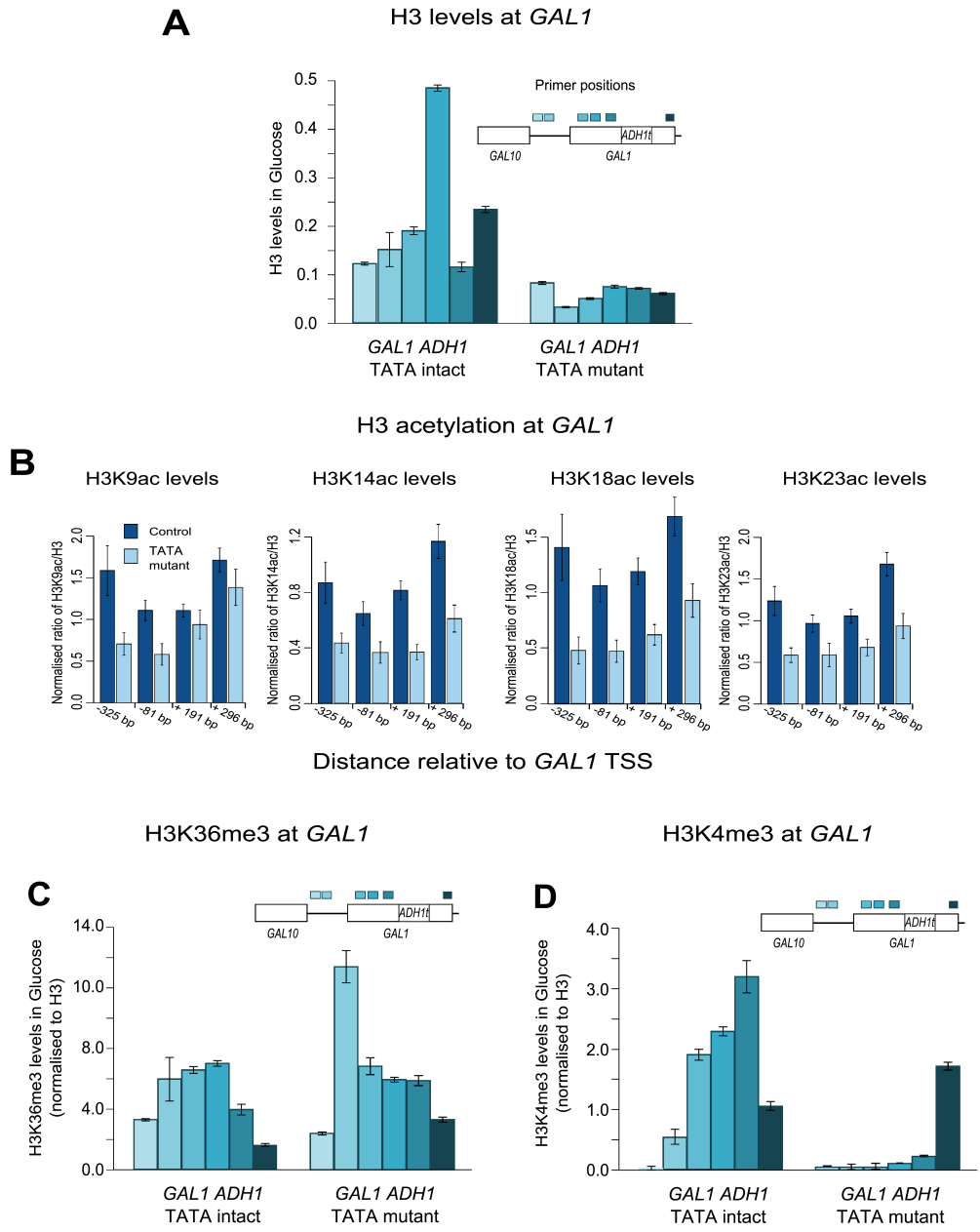


Figure 47: Reduction in antisense transcription at *GAL1* results in changes in histone levels in histone modification, largely agreeing with genome-wide analyses.

ChIP experiments showing the levels of H3, H3K4me3 and H3K36me3 at *GAL1* in the *GAL1-ADH1*$_T$ strain, both before and after mutation of the TATA-like sequence. Chromatin was prepared from strains grown in glucose-containing media as described previously (Pinskaya et al. 2009b). **(A)** H3 levels across the gene are reduced following reduction of antisense transcription, as predicted by the genome-wide analysis. **(B)** Levels of H3 acetylation are reduced following reduction of antisense transcription. **(C)** H3K36me3 levels are increased at the promoter following reduction of antisense transcription. **(D)** Levels of H3K4me3 are decreased following reduction of antisense transcription, as opposed to the spreading predicted from the genome-wide analysis. ChIP experiments were performed by Françoise Howe and Karolina Chocian.

6.3.4. Reducing antisense transcription at *GAL1* changes the gene expression mode from bursting to constitutive

That reduction of antisense transcription should lead to changes in chromatin modification is consistent with the genome-wide analysis presented in chapter 5. Reduction of antisense transcription should also result in a loss of bursting, given that these chromatin features (high occupancy and H3 acetylation) are associated with noisy genes, and so the *GAL1-ADH1*$_T$ strains before and after deletion of the TATA-box were compared.

Following insertion of the *ADH1* terminator and redefinition of the *GAL1* transcription unit, a bursting pattern was still observed by RNA-FISH, with up to 3 nascent transcripts per nuclear dot in 62% of induced cells (Figure 44C). Strikingly, however when antisense transcription was reduced by mutation of the TATA-like sequence only 30% of induced cells possessed nuclear dots, most of which represented only a single nascent transcript (Figure 44D). This profile was similar to the constitutive gene *DOA1* (Zenklusen et al. 2008), suggesting a switch to a constitutive mode of transcription. Thus the reduction of antisense transcription has indeed resulted in a loss of bursting at *GAL1*. *GAL10* also showed a similar profile, with a loss of bursting upon reduction of *GAL1* antisense transcription (Figure 44E-F), demonstrating that *GAL1* antisense transcription can modulate sense transcription at both genes. That *GAL10* follows a similar pattern of behaviour to *GAL1* suggests that the changes at *GAL1* are not a result of changes in nuclear retention of transcripts within the nucleus, which could result from defects in 3' end processing or termination caused by modification of the 3' end.

It is worth noting that the *GAL1-ADH1*$_T$ construct shows a lower incidence of dots (with a lower intensity) than the wild type gene. A possible explanation for this is that there is perhaps less antisense transcription at these genes. A more likely explanation,

however, is that the sense transcript in this strain is shorter. Thus RNAPII spends less time upon the gene, and so the probability of their being multiple RNAPII complexes on the gene simultaneously is lower.

6.3.5. *SET3* deletion has no effect upon bursting at *GAL1*

As discussed in the introduction, the *SET3* deacetylase has been shown to play a role in noise suppression at genes (Weinberger et al. 2012). *SET3* might therefore be dampening the effects of antisense transcription upon bursting, and so the effect of its deletion upon *GAL1* transcription in the WT and both *GAL1-ADH1*$_T$ constructs strains was assessed, anticipating an increase in bursting, potentially in all strains. However, deletion of *SET3* did not have notable effects upon the distribution of nascent transcripts, suggesting that it does not act to repress transcription bursting at *GAL1* (Figure 48). To assess whether histone acetylation might be necessary for bursting at *GAL1,* components of the SAGA complex were also deleted, specifically *SGF73* and *ADA2*, which enhance the Gcn5-dependent acetyltransferase activity (Cai et al. 2011), anticipating that a loss of histone acetylation would lead to a loss of bursting. However, bursting was not increased in these deletion strains. On the contrary, bursting appeared to be marginally increased, consistent with previous reports suggesting that Gcn5 and SAGA act to repress noise (Figure 42C-D) (Raser and Shea 2004; Weinberger et al. 2012). What remains to be seen is whether deletion of any of these activities does, in fact, alter levels of acetylation at these genes, to begin to assess whether acetylation itself is a component of noise.

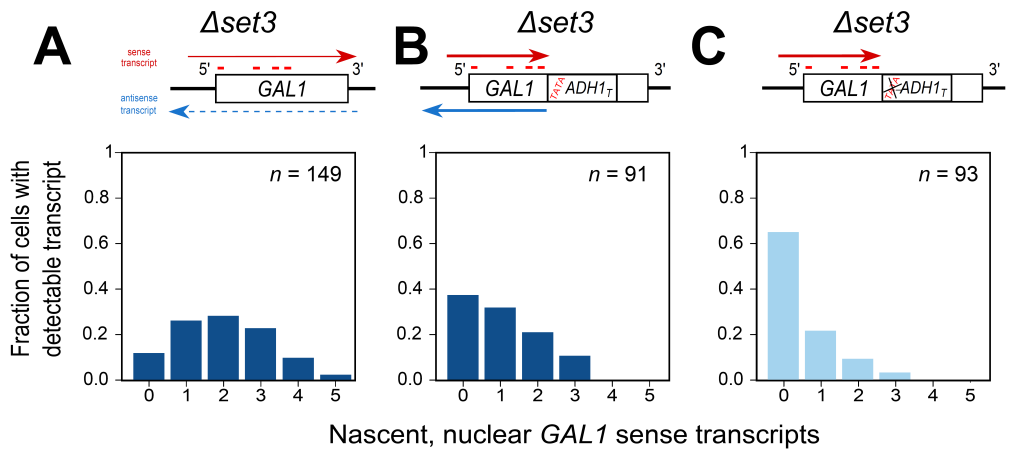


Figure 48: Deletion of *SET3* does not alter transcription mode of *GAL1* or the *GAL1* constructs.

Nascent, nuclear sense transcript profile at **(A)** *GAL1*, **(B)** *GAL1*:*ADH1*$_T$ insert and **(C)** *GAL1*:*ADH1*$_T$ insert in which the TATA-like sequence is mutated, all following deletion of *SET3*. Shown are the distributions of nascent sense transcripts of all three genes following induction in galactose for 2 hours, determined by RNA-FISH. Positions of the FISH probes are shown on the schematic. *n* describes the number of cells with detectable transcript from which the data was generated. Strains were constructed by Anitha Nair. Image acquisition was performed by Françoise Howe. Image analysis was carried out by the author.

6.3.6. Reduction of antisense transcription at *HMS2* changes its transcription mode

To expand the analysis, antisense transcription was reduced at another gene, to assess whether this would also result in a loss of noise. *HMS2* is a gene involved in the suppression of pseudohyphal growth in yeast, and is similar to the heat shock transcription factors. NET-seq data demonstrated that it has very high levels of antisense transcription immediately downstream of its TSS, and so fell within the highest antisense-transcribed class of genes (see chapter 5, Figure 25), as well as high levels of sense transcription. RNA-FISH was used to assess whether *HMS2* sense transcription showed a bursting transcription profile (Figure 49). Wild-type *HMS2* had both a sense and stable antisense transcript under glucose conditions (Figure 49A-B), and up to 3 nascent sense transcripts per nuclear dot in 51% of cells with detectable cytoplasmic transcript (Figure 49C), suggesting that bursting does occur at this gene.

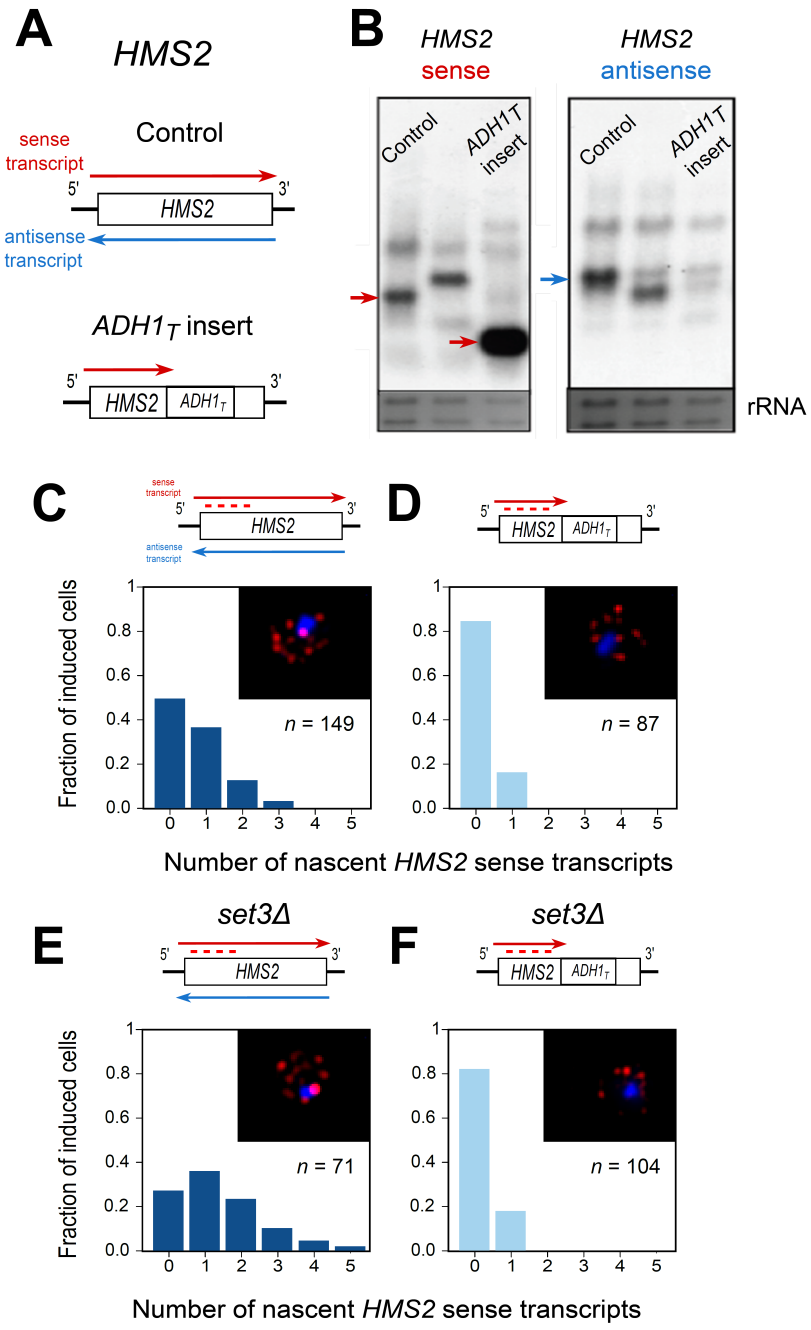


Figure 49: Set3 suppresses noise at *HMS2* only in the presence of antisense transcription.

(**A**) Antisense transcription into the *HMS2* promoter can be disrupted by insertion of *ADH1*$_T$. (**B**) Northern blot probing for the *HMS2* sense and antisense transcripts in both strains shown in **A.** Cells were grown in YPD. Red and blue arrows show bands representing sense and antisense transcripts respectively. rRNA in the ethidium-stained gel is shown as a loading control. Note that the middle lanes are not relevant to this study. (**C-D**) Distribution of nascent, nuclear *HMS2* sense transcripts in the control and *ADH1*$_T$-insert *HMS2* strains, as determined by RNA-FISH. (**E-F**) Distribution of nascent, nuclear *HMS2* sense transcripts in both strains following deletion of *SET3*. Strains design and construction and image acquisition was performed by Tania Nguyen and Françoise Howe. Image analysis was performed by the author.

The next aim was to assess whether it was possible to reduce antisense transcription at *HMS2*, and whether this could change the bursting profile. Unlike in *GAL1*, insertion of an *ADH1* terminator into the ORF of *HMS2* did not redefine the transcription unit but rather abolished antisense transcription, while increasing levels of transcription of a short sense transcript (Figure 49A-B). RNA-FISH showed that this increase in sense transcription was due, at least in part, to more cells within the population expressing *HMS2* sense transcript above background following $ADH1_T$ insertion (63% compared to 34%). Strikingly, RNA-FISH revealed that this strain did indeed show a less bursty expression profile – only single nuclear transcripts were observed in 16% of cells with detectable cytoplasmic transcript (Figure 49D). Thus antisense transcription appears to be enforcing a noisier mode of expression at *GAL1*, *GAL10* and *HMS2*.

6.3.7. Deletion of the histone deacetylase *SET3* increases bursting only in the presence of antisense transcription

Despite having higher levels of antisense transcription than *GAL1* (Figure 50) (690 reads in the 300bp window compared to 53 reads, class 5 and 4 respectively as in chapter 5, Figure 25), wild-type *HMS2* showed a lower incidence of bright nuclear dots, suggesting it is less bursty. *HMS2* might therefore be subject to additional regulation from factors acting to suppress noise. *SET3* has been shown to play a role in noise suppression, and so this gene was deleted to find out whether this might enhance bursting at *HMS2* (Weinberger et al. 2012).

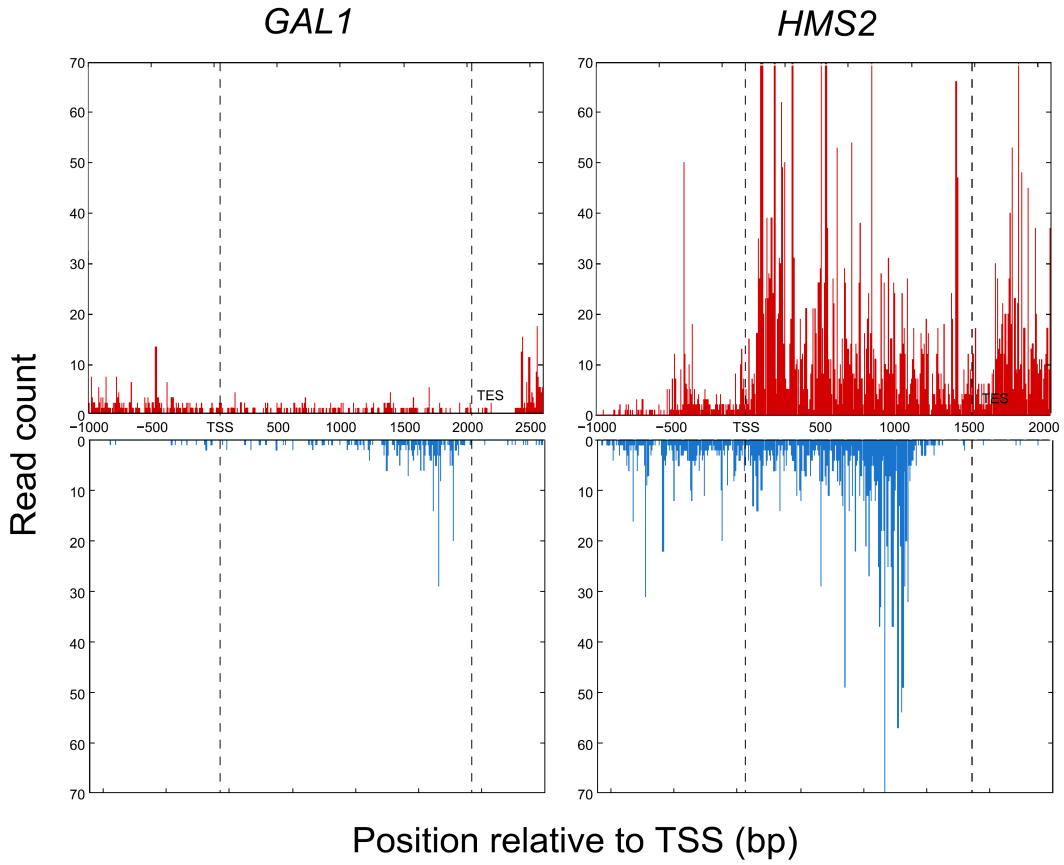


Figure 50: *HMS2* is subject to higher levels of antisense transcription than *GAL1*.

Maps of nascent sense (red) and antisense (blue) transcription at *GAL1* and *HMS2* as determined by NET-seq (Churchman and Weissman 2011). *HMS2* has higher levels of antisense transcription than *GAL1*, despite appearing less bursty. Note that NET-seq was performed under glucose conditions, whereas RNA-FISH for *GAL1* was conducted in galactose conditions, thus antisense transcription levels could differ.

Unlike at *GAL1,* deletion of *SET3* did indeed increase bursting at the intact *HMS2* gene, with up to five nascent, nuclear transcripts present in 77% of cells – a profile more similar to that seen in wild-type *GAL1* (Figure 49E). Strikingly, however, the *HMS2-ADH1*$_T$ construct with reduced antisense transcription did *not* show a similar increase in bursting upon *SET3* deletion, with exclusively single nascent, nuclear transcripts present in only 19% of cells (Figure 49F). Thus *SET3* deletion suppresses bursting transcription only in the presence of antisense transcription, suggesting that it is antisense

transcription itself that provides the substrate upon which the Set3 complex acts, and that this substrate is responsible for enhancing a gene's bursting characteristics. That *SET3* deletion should affect *HMS2* but not *GAL1* also suggests that noise can be negatively regulated in a gene-specific manner, perhaps explaining why not all genes with antisense transcription might be noisy.

6.4. Discussion

The work in this chapter supports a model in which antisense transcription modulates promoter chromatin, and in doing so, changes the mode of sense transcription from constitutive to bursting. Reduction of antisense transcription resulted in a reduction of bursting at *GAL1*, *GAL10* and *HMS2*. Furthermore, at *GAL1* antisense transcription caused increased H3, H3 acetylation and decreased H3K36me3 levels at the promoter – in line with the genome-wide analysis (chapter 5). Perhaps most strikingly, deletion of the noise suppressor *SET3* only increased noise at *HMS2* in the *presence* of antisense transcription. As Set3 is a chromatin regulator, this suggests that antisense transcription provides the proper chromatin substrate on which Set3 can operate to supress noise. Antisense transcription may therefore have a general role as a noise generating mechanism – a previously unreported consequence of this phenomenon – and suggests that this may be its crucial biological function.

6.4.1. Antisense transcription as a noise generating mechanism

Transcriptional noise is likely a consequence of enhanced nucleosome dynamics (Choi and Kim 2009; Hornung et al. 2012; Weinberger et al. 2012), and such dynamics may be conferred by antisense transcription; namely reduced nucleosome stabilizing modifications such as H2BK123ub, H3K36me3 and H3K79me3 coupled with increased nucleosome turnover and occupancy at the promoter and extending into the gene body.

NDR closure makes the promoter more bistable and therefore stochastic in nature, forcing it to switch between rapid bursts of transcription and chromatin repression (Cairns 2009). The interaction between promoter elements such as the TATA-box and nucleosomes in the vicinity could contribute to transcription re-initiation, leading to bursts and thus noise (Hornung et al. 2012). An additional contributory factor to bursting are permissive histone modifications in the early coding region that increase the chance of transition into the elongation phase (Weinberger et al. 2012). Evidence in this chapter supports the hypothesis that antisense transcription brings about such changes – in occupancy, modification, and dynamics – through histone turnover downstream and in the vicinity of the promoter. This chromatin template could then be tuned by enzymatic activity, such as via the Set3 complex, the deletion of which enhanced bursting at *HMS2* (Figure 51).

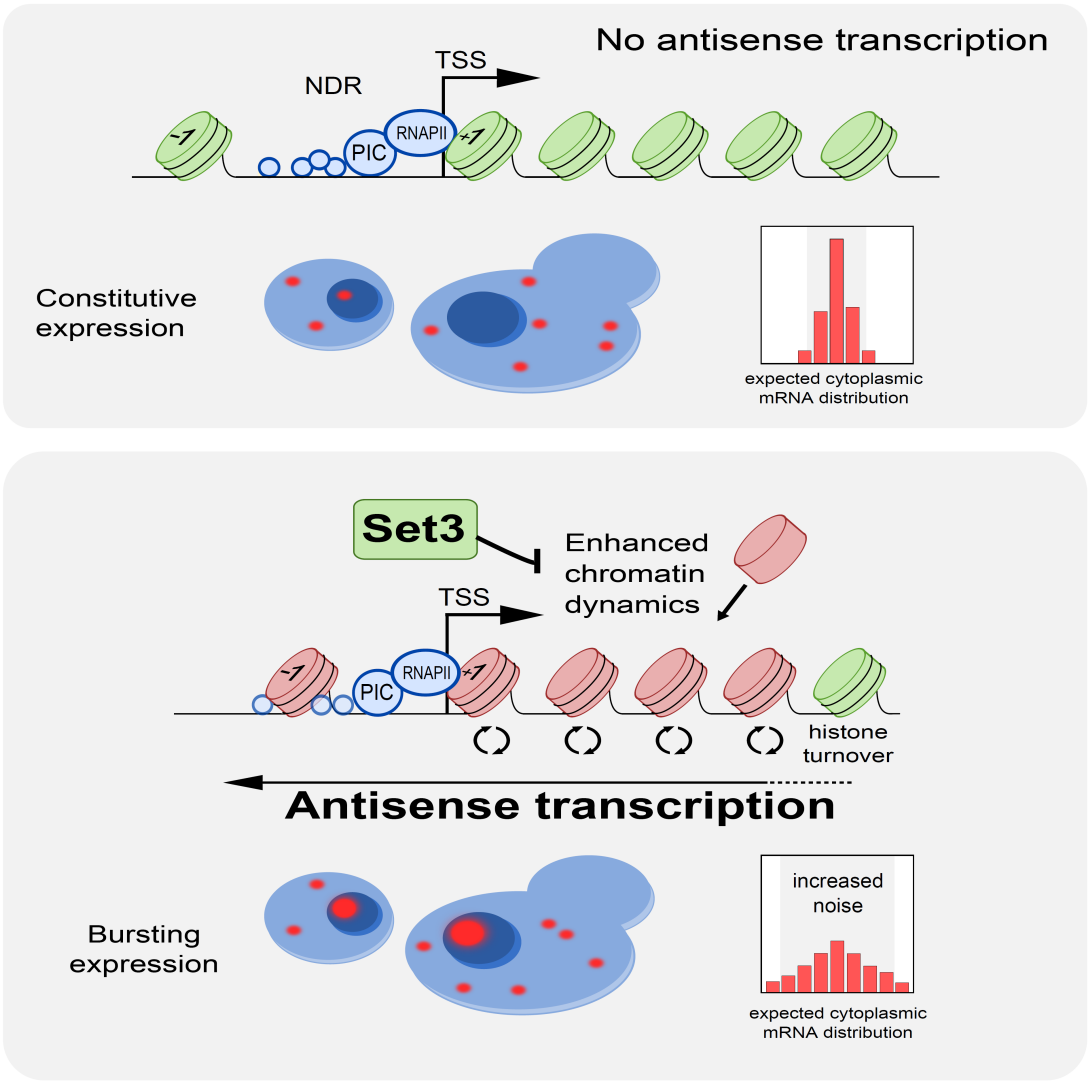


Figure 51: Antisense transcription establishes a chromatin environment supporting noisy transcription, which other factors can then suppress.

In the absence of antisense transcription, the open NDR and chromatin environment support a less noisy mode of transcription, resulting in low variability in the cytoplasmic mRNA count. Antisense transcription leads to NDR closure and the establishment of a chromatin environment that supports transcription bursting, leading to high variation in the cytoplasmic mRNA count. Factors such as Set3 that might be capable of reversing the changes to chromatin could then suppress this noise by acting upon the new chromatin state established by antisense transcription. Cycling arrows represent histone turnover. Red nucleosomes are hyperacetylated, recently turned-over nucleosomes. Green nucleosomes are hypoacetylated nucleosomes with H3K4me3, H3K36me3 and H3K79me3, which have not been recently turned over. Histograms show the distribution of cytoplasmic mRNAs expected from cells with and without antisense transcription. The grey box shows the standard deviation of these distributions.

The role of antisense transcription as an important component of noise is evidenced by the observation that the noise-enhancing phenotype of the *SET3* deletion only occurs in

the *presence* of antisense transcription. Set3 is proposed to reduce burst size by deacetylating nucleosomes in the early transcribed region (Weinberger et al. 2012), and therefore the action of HDACs, recruited perhaps via H2BK123ub, on acetylated nucleosomes could explain why not all genes with antisense transcription show noise. Antisense transcription could therefore represent a crucial, missing piece in our understanding of gene regulation, namely how the baseline level of intrinsic noise is set, and explains how cells might tune the noisiness of a given gene without having to modulate the signals received at the sense promoter. That antisense transcription shows evidence of being independently regulated by transcription factors (chapter 3) also suggests that noise could be different for a given gene under given environmental conditions, something which could not be achieved through noise generation via a TATA-box, which will obviously be present under all conditions.

Though antisense transcription may itself be enforcing bursting, it is unclear precisely which downstream changes are themselves causing bursting. For example, deletion of *ADA2* and *SGF73,* which contribute to the HAT activity of SAGA, did not decrease bursting, as one might expect if histone acetylation is an important component of this transcription mode. It is of course possible that deletion of these activities does not result in reduced acetylation at these genes. However, another possibility is that acetylation is not required for bursting, but rather that under normal conditions it is in an indicator of histone turnover, if histone deposition itself enforces bursting, either directly or through other changes in modification – such as reduction of H3K36me3. Clearly, more extensive deletion analysis will need to be performed using these experimental systems, combined with ChIP of H3 levels and histone modifications, RNA-FISH to assess bursting, and northern blotting to assess whether sense and/or antisense transcript levels have themselves changed.

Chapter Seven

Discussion

7. Discussion

Within this work the nature of non-coding antisense transcription within the budding yeast *Saccharomyces cerevisiae* has been explored, addressing questions regarding its origins (where it is produced, how it is produced, and by how much), its regulation, and also its possible function. These questions were addressed on a genome-wide scale, making use of publically available datasets to assess, broadly, how genes and chromatin vary in the presence of different levels of antisense transcription. In this way, hypotheses were developed that could then be tested experimentally, identifying a hitherto unknown role for antisense transcription in switching genes from a constitutive mode of sense transcription to a noisy one.

7.1. Antisense transcription is a general feature of genes

This work demonstrates that antisense transcription is abundant genome-wide – roughly a third of genes in the yeast genome possess previously defined antisense transcripts, while the average level of antisense transcription at these genes is almost half that of sense transcription. Furthermore, approximately 75% of genes possess a level of antisense transcription that is associated with changes in chromatin, suggesting it is functional even at low levels. Antisense transcription is directed by a promoter complex that binds in the vicinity of the terminator, and which is again present at genes with defined antisense transcripts, at average levels roughly half of those found at the sense promoter. These '3' promoters' show evidence of regulation, being enriched for TFBSs even in the absence of downstream sense promoters, suggesting that antisense transcription can be regulated. Antisense transcription is therefore a pervasive, abundant and regulated genome-wide process, and should be considered a general feature of budding yeast genes, rather than a sophistication of certain genes, while genes

themselves should be considered as having two opposing and regulatable promoter activities.

7.2. Antisense transcription enforces changes in promoter chromatin and sense transcription mode

Antisense transcription is associated broadly with changes in chromatin. Significant changes were observed for *all* histone modifications assessed, while, strikingly, no correlation between sense and antisense transcription was observed. The changes observed are consistent with changes associated with transcription bursting, and indeed, removal of antisense transcription at *GAL1* and *HMS2* resulted in a loss of bursting. Removal of antisense transcription also resulted in changes in histone H3 levels, H3K36me3 and H3 acetylation that were consistent with the observed associations. Furthermore, antisense transcription was associated with increased plasticity – the capacity to change in response to changing environmental conditions, something that is also linked with noise (Tirosh and Barkai 2008). Taken together with the finding that antisense promoters frequently possess TFBSs, this suggests that antisense transcription could be tuned at a given gene to define a gene's noise level and capacity to respond to changing environmental conditions.

7.3. The operation of sense and antisense transcription in real-time: a model

It is evident that both sense and antisense transcription can occur across the same gene under given conditions (Neil et al. 2009; Xu et al. 2009; Churchman and Weissman 2011). This does not necessarily mean that they occur contemporaneously, but rather that in a given population of cells a gene can be transcribed in the antisense direction in

one cell while being transcribed in the sense direction in another. But are these two processes coordinated? In chapter 6 it is shown that the percentage of cells with both sense and antisense transcripts is what would be expected given uncoordinated transcription, while in chapter 4 it is shown that sense and antisense transcription levels at the same gene are not correlated genome-wide. Thus sense and antisense transcription would appear to occur in a manner independent of one another, presumably regulated by their own associated promoters. Of course, work in chapter 6 suggests that antisense transcription *can* change the sense promoter, although in ways that change the sense transcription *mode* without necessarily altering the level – i.e. limiting transcription to infrequent but large bursts rather than frequent but small bursts – and so making the gene noisier.

It is proposed, based on the genome-wide and experimental analysis described in this work, that antisense transcription establishes these changes through modulation of the chromatin environment – but how long will such changes last? Does a single round of antisense transcription simply modify the next round of sense transcription – delaying the arrival of the next burst but ensuring that, when it does arrive, it is of a higher frequency? Or alternatively, do antisense transcripts provide an altered chromatin architecture that persists for many rounds of sense transcription? Based on genome-wide data in chapter 5, antisense transcription is associated with chromatin changes in a stepwise manner. If one round of antisense transcription was sufficient to bring about long-lasting changes then one might not expect such a gradual increase. Thus it is suggested that antisense transcription-induced changes in chromatin persist only transiently, sufficient only to alter the transcription mode of immediately subsequent sense transcription rounds.

If one considers a gene that is subject to a tenth as much antisense transcription as sense transcription, then one would expect one round of antisense transcription for

every ten rounds of sense transcription. This single antisense transcription round might bring about changes in chromatin that persist for the next round (or perhaps burst) of sense transcription. This round of sense transcription might then revert the chromatin back, such that subsequent rounds of sense transcription behave in a non-bursting fashion. Thus sense and antisense transcription might be competing to define the chromatin itself. One hypothesis might be that the chromatin at a gene exists in one of two states at any given time: a sense transcription state (low acetylation of H3 across the gene body, low promoter occupancy, high H3K36me3) or an antisense transcription state (the opposite). At any given time, the chromatin state is dependent upon which strand of the gene was most recently transcribed, whereas at a population level the chromatin profile is a composite that favours the most frequent type of transcription (Figure 52).

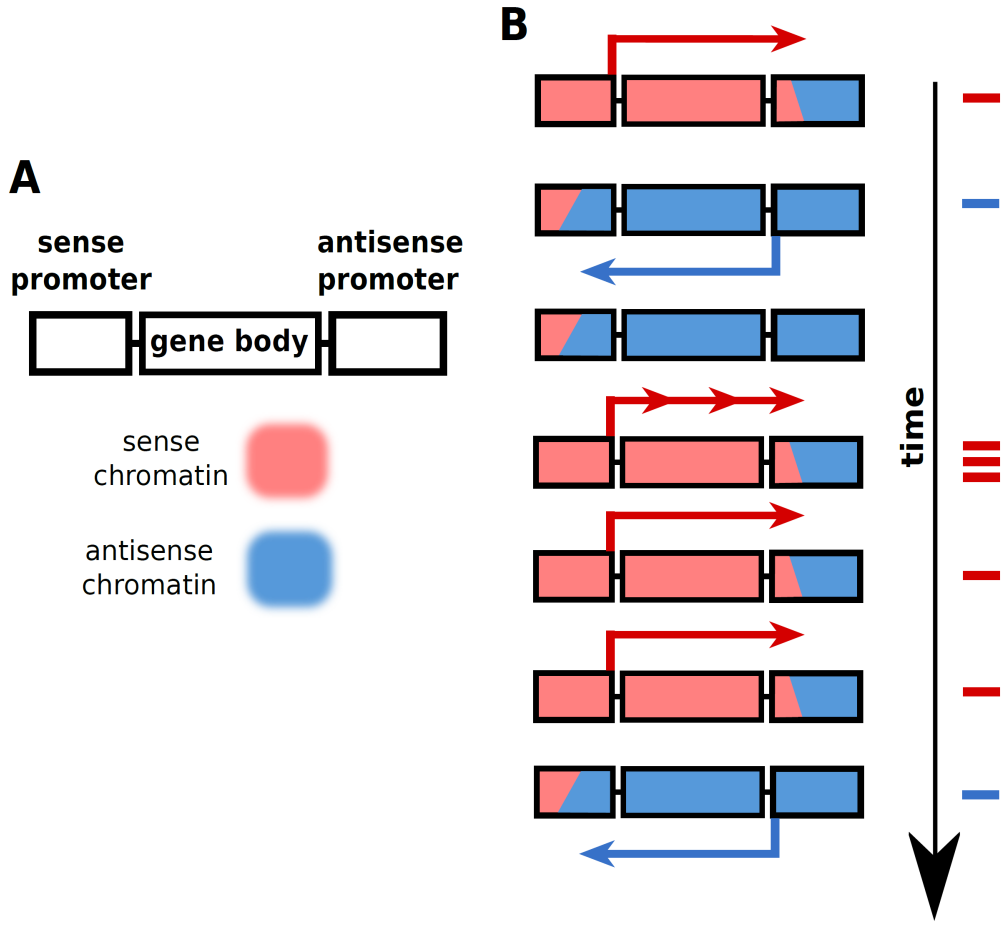


Figure 53. A model of sense and antisense transcription in real-time

A model of how sense and antisense transcription might occur at a single gene, suggesting how they relate to chromatin and gene behaviour **(A)** A typical gene is composed of a sense and antisense promoter, with the sense promoter typically possessing a higher activity than the antisense promoter (chapter 3). Genes can adopt one of two chromatin states – a sense (red) or an antisense one (blue), representing their distinct chromatin associations as described in chapter 5. **(B)** Genes undergo stochastic transcription of either the sense or antisense strand. At any given time either the sense or antisense strand can be transcribed, giving rise to the appropriate chromatin state described in A. Shown is an example of how a gene might change in time. Initially, a single round of sense transcription gives rise to a sense chromatin state. Then, a round of antisense transcription changes this to an antisense chromatin state. In an antisense chromatin state, sense transcription initiation has a lower probability of occurring than in a sense chromatin state, however when it does it occurs in a rapid burst, in this case of three rounds. This reverts the chromatin back to a sense chromatin state, such that sense initiation occurs with higher frequency but at lower burst size, until a round of antisense transcription changes it back again. The black arrow represents time progression; red and blue rectangles represent sense and antisense initiation events respectively.

To test this hypothesis, it would be worth assessing how genes vary depending on the *ratio* of antisense transcription to sense transcription, rather than simply antisense

transcription itself. Genes with a high ratio would be expected to adopt an antisense chromatin state, while those with a low ratio would favour a sense state, irrespective of the absolute levels of sense and antisense transcription. Of course, there are likely to be other mechanisms at play, independent of sense and antisense transcription, that might be changing chromatin to an entirely different state between rounds of sense and antisense transcription, which would make the absolute level important as well as the ratio – i.e. the longer the stretches of time between rounds of transcription the higher the probability of other changes occurring.

7.4. Insights into mammalian antisense transcription: a new perspective?

Here antisense transcription is identified as a modulator of chromatin dynamics and an enhancer of transcription noise in the budding yeast *Saccharomyces cerevisiae*. But might antisense transcription be performing a similar role in mammals?

A recent study by Conley & Jordan, (2012), published in the wake of a study by this lab that forms the basis of chapter 3 (Murray et al., 2012), provides a comprehensive analysis of antisense transcription in humans which in many ways mirrors this work. Here, they identify many thousands of antisense promoters using CAGE analysis, which characterises the 5' ends of transcripts, and find that such promoters are generally present towards the 3' ends of genes, as shown in yeast by Xu et al., (2011). Antisense transcripts varied between different cell types, and showed evidence of activatory histone marks at their promoters that correlated with the level of RNAPII occupancy – providing evidence that antisense transcription can be regulated, and analogous to our finding that antisense transcription shows evidence of transcription factor-directed regulation (chapter 3). Antisense transcription from these promoters was abundant, though not as high as from sense promoters, in agreement

with the findings of chapter 4. Perhaps most importantly, they found that the activity of sense promoters did not correlate with the levels of activity of their associated antisense promoters, agreeing with findings in chapter 4. This study therefore corroborates several of the key features identified here with regards to antisense transcription, but what about its effects on the sense promoter?

Xu et al., (2011) found evidence that, as in yeast, antisense transcription is associated with increased transcriptional plasticity in mammals. Using the genome-wide, strand-specific RNA-seq conducted by He et al., (2008), they found that genes with evidence of antisense transcription displayed greater variation in sense transcription across five different human cell lines. This mirrors what is shown here in yeast (chapter 3), and also suggests that antisense transcription might be important in establishing the changes in transcriptional programme associated with differentiation. Tirosh and Barkai, (2008) have linked plasticity with transcriptional *noise*, something that also plays important roles in differentiation (Arias and Hayward 2006), so it is possible that, as in yeast, both these gene features are associated (or even caused by) antisense transcription. What remains to be seen is whether sense promoter chromatin shows the same associations with histone modification in mammals as it does in yeast. Given the availability of genome-wide ChIP data for numerous histone modifications in different cell types, an exciting future analysis would be to assess whether changes in antisense promoter activity at genes (Conley and Jordan, 2012) are associated with the expected changes in promoter chromatin (increased nucleosome occupancy, H3 acetylation etc.). The work described here identifies a number of key features in yeast that are associated with, and in some cases possibly caused by, antisense transcription, which could now be looked for in higher eukaryotes, to assess whether antisense transcription performs an evolutionarily conserved function.

7.5. Future perspectives

7.5.1. What makes sense and antisense transcription different?

It has been shown here that antisense transcription has distinct chromatin associations from sense transcription – for example, it is more strongly correlated with histone turnover, acetylation over the gene body, and is negatively correlated with H3K36me3 whereas sense transcription is positively correlated. What is more, it has been shown experimentally that antisense transcription is necessary for high H3 levels, H3 acetylation and lower H3K36me3 at *GAL1,* suggesting that antisense transcription is related to these features by causation, rather than association. What is unclear is how transcription conducted by the same polymerase (RNAPII) could have such diverse relationships with chromatin. One possible clue lies in the observation that antisense transcription was not correlated with Ser2 phosphorylation of the RNAPII CTD (chapter 5). Another possibility is that the differential processing of transcripts might give rise to different downstream effects, and to this end it might be fruitful to subdivide antisense-transcribed genes on the basis of whether they possess SUTs, CUTs or XUTs, and see whether these differ in terms of promoter chromatin. For example, it might be that the transcription of SUTs, being free from extensive degradation (whether by Rrp6 or Xrn1), behaves in a manner similar to sense transcription, while the recruitment of factors like Rrp6 by CUTs might be instrumental in bringing about the changes in promoter chromatin associated with antisense transcription. This still raises the question of what initially defines whether a given transcription unit behaves like canonical sense transcription or the antisense transcription described here, with regards to chromatin – i.e. what is the initial 'spark' that tips RNAPII in favour of one sort of transcription or the other. Perhaps when two transcription units overlap, one exerts dominance over the other, defining which will be the sense and which the antisense transcript. Examining promoters genome-wide might help establish which factors play a role in 'choosing'

between sense or antisense transcription. One way to achieve this might be to utilise the large genome-wide study by Venters et al., (2011), in which the levels of 202 different factors were determined at the majority of sense promoters. One could compare two classes of bidirectional promoter – one in which coding sense transcription is being directed in both directions, and the other in which coding sense transcription is being directed in one direction and non-coding antisense transcription is being directed in the other – i.e. the two types of bidirectional promoter considered in chapter 4. Such an analysis would reveal if any of these factors are enriched or depleted at promoters directing antisense transcription.

7.5.2. The relationship between antisense transcription and the sense promoter TATA-box

It is curious that many of the features associated with antisense transcription – noise, plasticity, high promoter occupancy etc. – are also associated with TATA-box containing genes (Blake et al. 2003; Tirosh and Barkai 2008). Despite their similar associations, they are not themselves strongly linked – genes with high antisense transcription are only slightly enriched for TATA-boxes, while removing TATA-box containing genes from the analysis did not alter the observed changes in histone modifications and nucleosome occupancy (chapter 5). That they are both associated with similar features raises the question of whether their associations (or indeed effects) are cumulative – i.e. whether the presence of both features results in more pronounced promoter changes than just one or the other. The genome-wide noise analysis described in chapter 6 would suggest that this is the case – genes with high antisense and a TATA-box are noisier than those with low antisense and a TATA-box, and vice versa (chapter 6). Nevertheless, it would be fruitful to analyse TATA-containing genes and assess how genes subject to high antisense differ compared to those subject to low in terms of chromatin features.

Experimentally, one might also assess how deletion or mutation of a TATA-box within the sense promoter affects transcription bursting in the presence and absence of antisense transcription. The *GAL1* promoter possesses a TATA-box, which has been shown to influence transcription initiation (Giardina and Lis 1993), thus the *GAL1-ADH1*$_T$ system might represent a means of studying whether the increased bursting provided by antisense transcription is dependent in turn upon the increased burst size provide by the TATA-box, or vice versa. The TATA-box is thought to increase burst size by directing rapid re-initiation (Yean and Gralla 1997; Hornung et al. 2012). Antisense transcription, meanwhile, might increase burst size by creating a permissive environment for passage into elongation, perhaps by increasing acetylation at the early coding region (Carey et al. 2006; Dobson and Ingram 1980). Alternatively, it could decrease burst frequency by increasing nucleosome occupancy at the promoter. Of course, to maintain a gene's overall sense transcription level while still increasing noise, burst frequency would have to decrease while burst size would have to increase (Hornung et al. 2012) – i.e. promote the PIC's ability to re-initiate while inhibiting its initial formation. Nevertheless, it is possible to see how both the TATA-box and antisense transcription could achieve increased noise via alternative mechanisms, such that they are not dependent upon one another (Figure 54), and bioinformatics and experimental approaches would help to validate this.

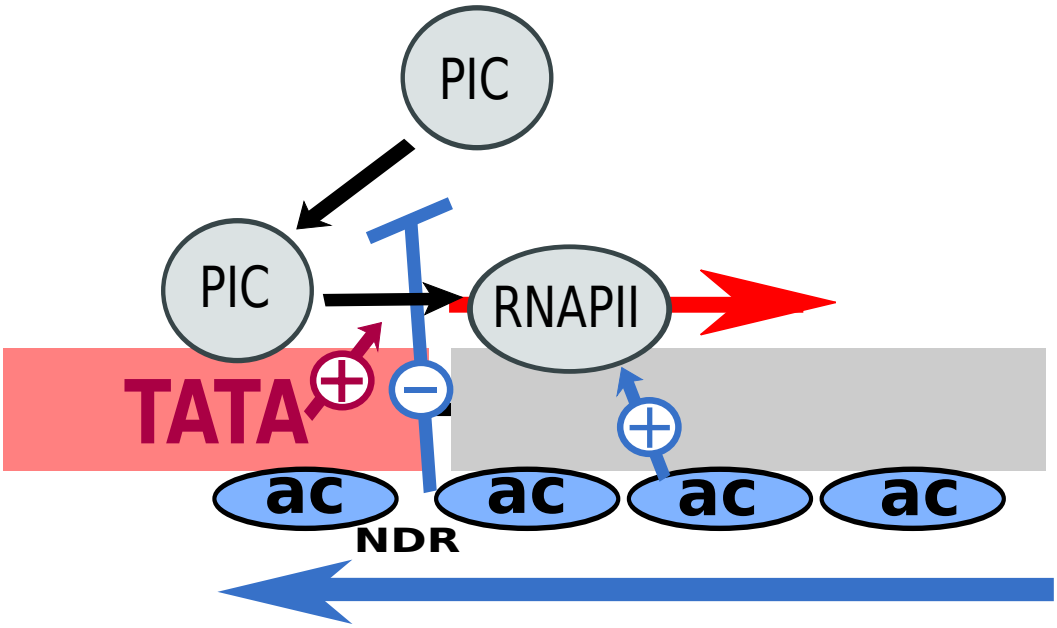


Figure 54. Possible relationship between bursting, the TATA-box and antisense transcription

A model of how the TATA-box and antisense transcription might confer a bursting transcription mode to genes. Formation of the PIC involves its recruitment to the promoter (red box), whereupon it recruits RNAPII and directs initiation of RNAPII and entry into elongation across the gene body (grey box). The PIC can remain bound and direct subsequent rounds of re-initiation. Noisy transcription is the consequence of increased burst size (the rate of re-initiation and entry into elongation) and decreased burst frequency (formation of an active PIC at the promoter). The TATA-box (maroon) is proposed to increase burst size by increasing the rate of re-initiation. Antisense transcription (blue arrow) is associated with increased acetylated nucleosomes (blue oblongs) and closure of the NDR. Increased acetylation could increase burst size by increasing rate of re-entry into elongation, while the closure of the NDR could decrease burst frequency by preventing formation of an active PIC. Blue arrows indicate positive or negative downstream effects of antisense transcription respectively. The maroon arrow represents a positive effect of the TATA-box. The red arrow represents sense transcription.

7.5.3. Extending the genome-wide analysis

In validating the hypothesis that antisense transcription can generally alter transcription at a gene, it would be necessary to assess how changes in antisense transcription as a result of changing conditions are associated with changes in chromatin genome-wide. Ideally, this would require NET-seq experiments – previous attempts to study antisense transcription across changing conditions are limited by their use of tiling arrays measuring steady state transcript levels (Neil et al. 2009; Xu et

al. 2009). By determining the level of antisense transcription across different conditions, as well as levels of histone modifications such as H3 acetylation and H3K36me3 in the same conditions, one could assess whether these features change as expected alongside changing levels of antisense transcription within the 300bp window of genes. This analysis, however, would not demonstrate a causative effect of antisense transcription upon chromatin. Certain studies have identified mutations that might be causing global changes in pervasive transcription. Carrozza et al., (2005) found that deleting the *SET2* methyltransferase resulted in an increase in spurious transcription within the ORFs of certain genes, while Kim et al., (2012) showed on a much wider scale that deletion of the *SET3* histone deacetylase had a similar effect. It remains unclear to what extent this 'spurious' transcription is equivalent to antisense transcription – i.e. whether cryptic internal promoters are synonymous with 3' antisense promoters. A potentially suitable candidate was identified by Churchman and Weissman (2011), who found that deleting components of the Rpd3S histone deacetylase complex resulted in an increase in the level of antisense transcription from bidirectional promoters. Strikingly, the effect of this mutation was specific to antisense transcription – sense-transcribing RNAPII density over the 5' end of sense transcripts was generally unaffected. Antisense transcripts in these mutants still had the same start and end sites as in the wild-type strain, suggesting that the increase is not a result of what might be considered spurious internal transcription. This finding agrees with one of the main proposals resulting from this work – that antisense transcription does not generally repress sense transcription – given that increasing antisense transcription globally did not alter sense transcription levels. Combining these data with genome-wide ChIP data in the same mutant strains would allow one to assess whether increased antisense transcription results in the expected changes in chromatin, though of course, this would likely be complicated by

effects of Rpd3S on chromatin that are independent of its effects of antisense transcription.

It is observed here that antisense transcription is not well correlated with Ser2 phosphorylation. However, this approach is limited by the fact that the Ser2 phosphorylation data used are not strand-specific. To overcome this, NET-seq could be utilised; comparing NET-seq in which RNAPII is immunoprecipitated via a Flag tag as performed previously (Churchman and Weissman 2011) with a NET-seq experiment in which it is immunoprecipated via Ser2 phosphorylation, or else by a protein that binds to Ser2 phosphorylation, such as Set2 (Li et al. 2003). The prediction would be that the ratio of Ser2 phosphorylation to RNAPII levels on the sense strand of genes would be higher than on the antisense strand, and would allow insight into a possible explanation for why antisense transcription has different associations with (and possibly effects on) chromatin compared to sense transcription.

7.6. Closing remarks

The work described here aims to encourage a more quantitative consideration of antisense transcription. That is, rather than considering whether or not a gene has an antisense transcript, one should consider how much antisense transcription a gene is subject to. Genes appear generally to possess the means to support antisense transcription, irrespective of whether antisense transcription is observed under a given condition. It is well documented that 5' NDRs behave as promoters of bidirectional transcription (Neil et al. 2009; Xu et al. 2009), and here it is shown that the 3' NDR can also support antisense transcription, and that terminators can behave as promoters of antisense transcription when inserted into ORFs. Such transcriptionally permissive platforms possess TFBSs that allow them to be regulated. Thus, while antisense transcription at a given gene might be so low as to evade detection, under different

conditions the same gene might support high levels. This is supported by the existence of both TATA-boxes and (albeit less pronounced) NDRs at genes without annotated antisense transcripts. Perhaps most striking is the finding that, as one considers groups of genes with steadily increasing antisense transcription, chromatin features change in a stepwise fashion – across the entire complement of yeast genes considered. Thus these chromatin changes are not simply specific to those genes with the most antisense transcription, suggesting that antisense transcription is present at functional levels at the vast majority of genes. That antisense transcription should be so abundant, and that it can both modulate chromatin and enhance the noise of gene expression, suggests that it should be considered as an important and canonical feature of yeast genes.

8. Bibliography

Adkins MW, Howar SR, Tyler JK. 2004. Chromatin Disassembly Mediated by the Histone Chaperone Asf1 Is Essential for Transcriptional Activation of the Yeast PHO5 and PHO8 *Mol Cell* **14**: 657–666.

Adkins MW, Tyler JK. 2004. The histone chaperone Asf1p mediates global chromatin disassembly in vivo. *J Biol Chem* **279**: 52069–74.

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**: 387–93.

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF. 2007. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature* **446**: 572–6.

Albert I, Wachi S, Jiang C, Pugh BF. 2008. GeneTrack - a genomic data processing and visualization framework. *Bioinformatics* **24**: 1305–6.

Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA. 2013. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**: 360–3.

Almer A, Rudolph H, Hinnen A. 1986. Removal of positioned nucleosomes from the yeast PH05 promoter upon PH05 induction releases additional upstream activating DNA elements. *EMBO J* **5**: 2689–2696.

Anderson JD, Lowary PT, Widom J. 2001. Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol* **307**: 977–85.

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* **100**: 837–841.

Arias AM, Hayward P. 2006. Filtering transcriptional noise during development: concepts and mechanisms. *Nat Rev Genet* **7**: 34–44.

Arigo JT, Eyler DE, Carroll KL, Corden JL. 2006. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell* **23**: 841–51.

Arkin A, Ross J, Mcadams HH. 1998. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in phage lambda-infected Escherichia coli cells. *Genetics* **149**:1633-48

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 A Resolution. *Science (80-)* **289**: 905–920.

Basehoar AD, Zanton SJ, Pugh BF. 2004. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**: 699–709.

Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz B-J, Hurst LD, Tyers M. 2006. Stratus not altocumulus: a new view of the yeast protein interaction network. *PLoS Biol* **4**: e317.

Bataille AR, Jeronimo C, Jacques P-É, Laramée L, Fortin M-È, Forest A, Bergeron M, Hanes SD, 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**: 158–70.

Batista PJ, Chang HY. 2013. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**: 1298–307.

Batta K, Zhang Z, Yen K, Goffman DB, Pugh BF. 2011. Genome-wide function of H2B ubiquitylation in promoter and genic regions. *Genes Dev* **25**: 2254–65.

Berger SL. 2007. The complex language of chromatin regulation during transcription. *Nature* **447**: 407–12.

Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL. 2004. Global nucleosome occupancy in yeast. *Genome Biol* **5**: R62.

Berretta J, Morillon A. 2009. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep* **10**: 973–82.

Berretta J, Pinskaya M, Morillon A. 2008. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. *Genes Dev* **22**: 615–26.

Bian C, Xu C, Lee KK, Burke TL, Tempel W, Barsyte D, Li J, Wu M, Zhou BO, Fleharty BE, et al. 2011. Sgf29 binds histone H3K4me2 / 3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J* **30**: 2829–2842.

Biddick R, Young ET. 2005. Yeast mediator and its role in transcriptional regulation. *C R Biol* **328**: 773–82.

Biggar SR, Crabtree GR. 2001. Cell signaling can direct either binary or graded transcriptional responses. **20**: 3167–3176.

Blake WJ, KÆrn M, Cantor CR, Collins JJ. 2003. Noise in eukaryotic gene expression. *Nature* **422**: 633–7.

Boeger H, Griesenbeck J, Strattan JS, Kornberg RD. 2004. Removal of Promoter Nucleosomes by Short Article Disassembly Rather Than Sliding In Vivo. **14**: 667–673.

Breathnach R, Chambon P. 1981. Expression of eukaryotic split genes encoding for proteins. *Annu Rev Biochem* **50**: 349–383.

Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, Dent SY, Winston F, Allis CD. 2001. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev* **15**: 3286–95.

Brown CR, Mao C, Falkovskaia E, Law JK, Boeger H. 2011. In vivo role for the chromatin-remodeling enzyme SWI/SNF in the removal of promoter nucleosomes by disassembly rather than sliding. *J Biol Chem* **286**: 40556–65.

Bruno M, Flaus A, Stockdale C, Rencurel C, Ferreira H, Owen-Hughes T. 2003. Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. *Mol Cell* **12**: 1599–606.

Bryne JC, Valen E, Tang M-HE, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A. 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res* **36**: D102–6.

Burgess RJ, Zhou H, Han J, Zhang Z. 2010. A role for Gcn5 in replication-coupled nucleosome assembly. *Mol Cell* **37**: 469–80.

Bushnell DA, Westover KD, Davis RE, Kornberg RD. 2004. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* **303**: 983–8.

Butler JEF, Kadonaga JT. 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* **16**: 2583–92.

Cai L, Sutter BM, Li B, Tu BP. 2011. Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol Cell* **42**: 426–37.

Cairns BR. 2009. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**: 193–8.

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**: 1534–45.

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

Carey M, Li B, Workman JL. 2006. RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Mol Cell* **24**: 481–7.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia W-J, Anderson S, Yates J, Washburn MP, et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**: 581–92.

Carvin CD, Kladde MP. 2004. Effectors of lysine 4 methylation of histone H3 in Saccharomyces cerevisiae are negative regulators of PHO5 and GAL1-10. *J Biol Chem* **279**: 33057–62.

Castelnuovo M, Rahman S, Guffanti E, Infantino V, Stutz F, Zenklusen D. 2013. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* **20**: 851–8.

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, et al. 2004. Unbiased Mapping of Transcription Factor Binding Sites along Human Chromosomes 21 and 22 Points to Widespread Regulation of Noncoding RNAs 3380 Central Expressway. *Cell* **116**: 499–509.

Chaban Y, Ezeokonkwo C, Chung W-H, Zhang F, Kornberg RD, Maier-Davis B, Lorch Y, Asturias FJ. 2008. Structure of a RSC-nucleosome complex and insights into chromatin remodeling. *Nat Struct Mol Biol* **15**: 1272–7.

Chakravarthy S, Park Y-J, Chodaparambil J, Edayathumangalam RS, Luger K. 2005. Structure and dynamic properties of nucleosome core particles. *FEBS Lett* **579**: 895–8.

Chatterjee S, Struhl K. 1995. Connecting a promoter-bound protein to TBP bypasses the need for a transcriptional activation domain. *Nature* **374**: 820–822.

Chen W, Struhl K. 1988. Saturation mutagenesis of a yeast his3 "TATA element": genetic evidence for a specific TATA-binding protein. *Proc Natl Acad Sci U S A* **85**: 2691–5.

Chen X-F, Kuryan B, Kitada T, Tran N, Li J-Y, Kurdistani S, Grunstein M, Li B, Carey M. 2012. The Rpd3 core complex is a chromatin stabilization module. *Curr Biol* **22**: 56–63.

Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, et al. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**: 1149–54.

Cheng W, Zhang Z, Wang J. 2013. Long noncoding RNAs: New players in prostate cancer. *Cancer Lett*.

Cho EJ, Kobor MS, Kim M, Greenblatt J, Buratowski S. 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev* **15**: 3319–29.

Choder M, Young RA. 1993. A Portion of RNA Polymerase II Molecules Has a Component Essential for Stress Responses and Stress Survival. *Mol Cell Biol* **13**: DOI: 10.1128/MCB.13.11.6984.

Choi JK, Kim Y-J. 2009. Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nat Genet* **41**: 498–503.

Chung H-R, Dunkel I, Heise F, Linke C, Krobitsch S, Ehrenhofer-Murray AE, Sperling SR, Vingron M. 2010. The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS One* **5**: e15754.

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

Conley AB, Jordan IK. 2012. Epigenetic regulation of human cis-natural antisense transcripts. *Nucleic Acids Res* **40**: 1438–45.

Conte I, Banfi S, Bovolenta P. 2013. Non-coding RNAs in the development of sensory organs and related diseases. *Cell Mol Life Sci*.

Conway L, Wickens M. 1985. A sequence downstream of A-A-U-A-A-A is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes. *Proc Natl Acad Sci U S A* **82**: 3949–53.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845–48.

Cormack BP, Struhl K. 1992. The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell* **69**: 685–96.

Costelloe EO, Stacey KJ, Antalis TM, Hume DA. 1999. Regulation of the plasminogen activator inhibitor-2 (PAI-2) gene in murine macrophages. Demonstration of a novel pattern of responsiveness to bacterial endotoxin. *J Leukoc Biol*. **66**: 172-82.

Cui P, Liu W, Zhao Y, Lin Q, Ding F, Xin C, Geng J, Song S, Sun F, Hu S, et al. 2012. The association between H3K4me3 and antisense transcription. *Genomics Proteomics Bioinformatics* **10**: 74–81.

Danckwardt S, Kaufmann I, Gentzel M, Foerstner KU, Gantzert AS, Gehring NH, Neu-Yilik G, Bork P, Keller W, Wilm M, et al. 2007. Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *EMBO J* **26**: 2658–69.

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 The Role of Nuclear Cap Binding Protein Cbc1p of Yeast in mRNA Termination and Degradation. *Mol Cell Biol* **20**: 2872–2838.

Daulny A, Geng F, Muratani M, Geisinger JM, Salghetti SE, Tansey WP. 2008. Modulation of RNA polymerase II subunit composition by ubiquitylation. *Proc Natl Acad Sci U S A* **105**: 19649-54.

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**: 5320–5.

Dimmeler S, Nicotera P. 2013. MicroRNAs in age-related diseases. *EMBO Mol Med* **5**: 180–90.

Dion MF, Kaplan T, Kim M, Buratowski S, Friedman N, Rando OJ. 2007. Dynamics of replication-independent histone turnover in budding yeast. *Science* **315**: 1405–8.

Dobson ME, Ingram VM. 1980. In vitro transcription of chromatin containing histories hyperacetylated in vivo. *Nucleic Acids Res* **8**: 4201–4220.

Dvir A, Conaway JW, Conaway RC. 2001. Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Curr Opin Genet Dev* **11**: 209–14.

Dvir A. 2002. Promoter escape by RNA polymerase II. *Biochim Biophys Acta* **1577**: 208–223.

Edwalds-gilbert G, Prescott J, Falck-pedersen E. 1993. 3 ' RNA processing efficiency plays a primary role in generating termination-competent RNA polymerase II elongation complexes. *Mol Cell Biol* **13**: 3472–3480.

Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. *Science* **297**: 1183–6.

Erkina TY, Zou Y, Freeling S, Vorobyev VI, Erkine a M. 2010. Functional interplay between chromatin remodeling complexes RSC, SWI/SNF and ISWI in regulation of yeast heat shock genes. *Nucleic Acids Res* **38**: 1441–9.

Fan X, Moqtaderi Z, Jin Y, Zhang Y, Liu XS, Struhl K. 2010. Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. *Proc Natl Acad Sci U S A* **107**: 17945–50.

Feldmann H, Aigle M, Aljinovic G, André B, Baclet MC, Barthe C, Baur A, Bécam AM, Biteau N, Boles E, et al. 1994. Complete DNA sequence of yeast chromosome II. *EMBO J* **13**: 5795–809.

Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol* **4**: e1000216.

Fillingham J, Recht J, Silva AC, Suter B, Emili A, Stagljar I, Krogan NJ, Allis CD, Keogh M-C, Greenblatt JF. 2008. Chaperone control of the activity and specificity of the histone H3 acetyltransferase Rtt109. *Mol Cell Biol* **28**: 4342–53.

Fishburn J, Mohibullah N, Hahn S. 2005. Function of a eukaryotic transcription activator during the transcription cycle. *Mol Cell* **18**: 369–78.

Fitzgerald M, Shenkt T. 1981. The Sequence 5 ' -AAUAAA-3 ' Forms Part of the Recognition Site for Polyadenylation of Late SV40 mRNAs. *Cell* **24**: 251–260.

Fleming AB, Kao C-F, Hillyer C, Pikaart M, Osley MA. 2008. H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. *Mol Cell* **31**: 57–66.

Floer M, Wang X, Prabhu V, Berrozpe G, Narayan S, Spagna D, Alvarez D, Kendall J, Krasnitz A, Stepansky A, et al. 2010. A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. *Cell* **141**: 407–18.

Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. 2004. Noise minimization in eukaryotic gene expression. *PLoS Biol* **2**: e137.

Gandhi SJ, Zenklusen D, Lionnet T, Singer RH. 2011. Transcription of functionally related constitutive genes is not coordinated. *Nat Struct Mol Biol* **18**: 27–34.

Gardner KE, Allis CD, Strahl BD. 2011. Operating on chromatin, a colorful language where context matters. *J Mol Biol* **409**: 36–46.

Giardina C, Lis JT. 1993. DNA melting on yeast RNA polymerase II promoters. *Science* **261**: 759–762.

Goldmark JP, Fazzio TG, Estep PW, Church GM, Tsukiyama T. 2000. The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell* **103**: 423–33.

Goodman AJ, Daugharthy ER, Kim J. 2013. Pervasive antisense transcription is evolutionarily conserved in budding yeast. *Mol Biol Evol* **30**: 409–21.

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**: 234–46.

Grange T. 2008. Sensitive detection of mRNA decay products by use of reverse-ligation-mediated PCR (RL-PCR). *Methods Enzymol* **448**:445-66.

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

Green TL, Hart RP. 1988. Mutations in Poly(A) Site Downstream Elements Affect In Vitro Cleavage Activity. *Mol Cell Biol* **8**: 1839–1841.

Greenblatt J. 1991. Roles of TFIID in transcriptional initiation by RNA polymerase II. *Cell* **66**: 1067–70.

Gromak N, West S, Proudfoot NJ. 2006. Pause Sites Promote Transcriptional Termination of Mammalian RNA Polymerase II. *Mol Cell Biol* **26**: 3986–3996.

Guillemette B, Bataille AR, Gévry N, Adam M, Blanchette M, Robert F, Gaudreau L. 2005. Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol* **3**: e384.

Guillemette B, Drogaris P, Lin H-HS, Armstrong H, Hiragami-Hamada K, Imhof A, Bonneil E, Thibault P, Verreault A, Festenstein RJ. 2011. H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. *PLoS Genet* **7**: e1001354.

Güldener U, Heck S, Fielder T, Beinhauer J, Hegemann JH. 1996. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24**: 2519–24.

Guo Z, Sherman F, Sherman F. 1996. Signals sufficient for 3 ' -end formation of yeast mRNA. *Mol Cell Biol* **16**: 2772–2776.

Haimovich G, Medina DA, Causse SZ, Garber M, Millán-Zambrano G, Barkai O, Chávez S, Pérez-Ortín JE, Darzacq X, Choder M. 2013. Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* **153**: 1000–11.

Hampshire AJ, Rusling DA, Broughton-Head VJ, Fox KR. 2007. Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods* **42**: 128–40.

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne J, Reynolds DB, Yoo J, et al. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99–104.

Harrison BR, Yazgan O, Krebs JE. 2009. Life without RNAi: noncoding RNAs and their functions in *Saccharomyces cerevisiae*. *Biochem* **87**: 767-79.

Hartley PD, Madhani HD. 2009. Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**: 445–58.

He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. 2008. The antisense transcriptomes of human cells. *Science* **322**: 1855–7.

Hebenstreit D. 2013. Are gene loops the cause of transcriptional noise? *Trends Genet* **29**: 333–338.

Henikoff S, Shilatifard A. 2011. Histone modification: cause or cog? *Trends Genet* **27**: 389–96.

Henry KW, Wyce A, Lo W-S, Duggan LJ, Emre NCT, Kao C-F, Pillus L, Shilatifard A, Osley MA, Berger SL. 2003. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**: 2648–63.

Hieronymus H, Yu MC, Silver PA. 2004. Genome-wide mRNA surveillance is coupled to mRNA export. *Genes Dev* **18**: 2652–62.

Hinnebusch AG. 2005. Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* **59**: 407–50.

Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ. 2012. RNA polymerase II collision interrupts convergent transcription. *Mol Cell* **48**: 365–74.

Holstege FC, Fiedler U, Timmers HT. 1997. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J* **16**: 7468–80.

Hongay CF, Grisafi PL, Galitski T, Fink GR. 2006. Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell* **127**: 735–45.

Hopper AK, Phizicky EM. 2003. tRNA transfers to the limelight. *Genes Dev* **17**: 162–80.

Hornung G, Bar-Ziv R, Rosin D, Tokuriki N, Tawfik DS, Oren M, Barkai N. 2012. Noise-mean relationship in mutated promoters. *Genome Res* **22**: 2409–17.

Hossain MA, Chung C, Pradhan SK, Johnson TL. 2013. The yeast cap binding complex modulates transcription factor recruitment and establishes proper histone H3K36 trimethylation during active transcription. *Mol Cell Biol* **33**: 785–99.

Houchmandzadeh B, Wieschaus E, Leibler S. 2002. Establishment of developmental precision and proportions in the early Drosophila embryo. *Nature* **415**: 798–802.

Hough BR, Smith MJ, Britten RJ, Davidson EH. 1975. Sequence Complexity of Heterogeneous Nuclear RNA in Sea Urchin Embryos. *Cell* **5**: 291–299.

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.

Hsin J-P, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**: 2119–37.

Hughes AL, Jin Y, Rando OJ, Struhl K. 2012. A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol Cell* **48**: 5–15.

Hughes JD, Estep PW, Tavazoie S, Church GM. 2000. Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. *J Mol Biol* **296**: 1205–14.

Huisinga KL, Pugh BF. 2004. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Mol Cell* **13**: 573–85.

Hume DA. 2000. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. *Blood* **96**:2323–8.

Humphrey T, Birse CE, Proudfoot NJ. 1994. RNA 3' end signals of the S.pombe ura4 gene comprise a site determining and efficiency element. *EMBO J* **13**: 2441–51.

Humphrey T, Sadhale P, Platt T, Proudfoot N. 1991. Homologous mRNA 3' end formation in fission and budding yeast. *EMBO J* **10**: 3503–11.

Hwang WW, Venkatasubrahmanyam S, Ianculescu AG, Tong A, Boone C, Madhani HD, Francisco S. 2003. A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Mol Cell* **11**: 261–266.

Hyman LE, Moore CL. 1993. downstream of yeast polyadenylation sites . Termination and pausing of RNA polymerase II downstream of yeast polyadenylation sites. *Mol Cell Biol* **13**: 5159–5167.

Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N. 2002. Revealing modular organization in the yeast transcriptional network. *Nat Genet* **31**: 370–7.

Ivanovska I, Jacques P-É, Rando OJ, Robert F, Winston F. 2011. Control of chromatin structure by spt6: different consequences in coding and regulatory regions. *Mol Cell Biol* **31**: 531–41.

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**: 318–356.

Jacques Cote CLP and JW. 1998. Perturbation of nucleosome core structure by the SWI/SNF complex persists after its detachment, enhancing subsequent transcription factor binding. *Proc Natl Acad Sci U S A* **95**: 4947–4952.

Jamai A, Imoberdorf RM, Strubin M. 2007. Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. *Mol Cell* **25**: 345–55.

Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* **293**: 1074–80.

Jiang L, Smith JN, Anderson SL, Ma P, Mizzen C a, Kelleher NL. 2007. Global assessment of combinatorial post-translational modification of core histones in yeast using contemporary mass spectrometry. LYS4 trimethylation correlates with degree of acetylation on the same H3 tail. *J Biol Chem* **282**: 27923–34.

Johnson AD, Meyer BJ, Ptashne M. 1979. Interactions between DNA-bound repressors govern regulation by the lambda phage repressor. *Proc Natl Acad Sci U S A* **76**: 5061–5065.

Juven-Gershon T, Kadonaga JT. 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol* **339**: 225–9.

Kao C-F, Hillyer C, Tsukuda T, Henry K, Berger S, Osley MA. 2004. Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. *Genes Dev* **18**: 184–95.

Kaplan CD, Laprade L, Winston F. 2003. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301**: 1096–9.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**: 362–6.

Kaplan T, Liu CL, Erkmann JA, Holik J, Grunstein M, Kaufman PD, Friedman N, Rando OJ. 2008. Cell cycle- and chaperone-mediated regulation of H3K56ac incorporation in yeast. *PLoS Genet* **4**: e1000270.

Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SPA, Gingeras TR. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**: 916–9.

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, et al. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**: 1484–8.

Kasten MM, Szerlong H, Erdjument-Bromage H, Tempst P, Werner M, Cairns BR. 2004. Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. *EMBO J* **23**: 1348–59.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–54.

Kim J, Iyer VR. 2004. Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles. *Mol Cell Biol* **24**: 8104–8112.

Kim T, Buratowski S. 2009. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* **137**: 259–72.

Kim T, Buratowski S. 2007. Two Saccharomyces cerevisiae JmjC domain proteins demethylate histone H3 Lys36 in transcribed regions to promote elongation. *J Biol Chem* **282**: 20827–35.

Kim T, Xu Z, Clauder-Münster S, Steinmetz LM, Buratowski S. 2012. Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. *Cell* **150**: 1158–69.

Kim TH, Barrera LO, Zheng M, Qu C, Singer M a, Richmond T a, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436**: 876–80.

Kim TS, Liu CL, Yassour M, Holik J, Friedman N, Buratowski S, Rando OJ. 2010. RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. *Genome Biol* **11**: R75.

Kirmizis A, Santos-Rosa H, Penkett CJ, Singer MA, Green RD, Kouzarides T. 2009. Distinct transcriptional outputs associated with mono- and dimethylated histone H3 arginine 2. *Nat Struct Mol Biol* **16**: 449–51.

Kirmizis A, Santos-Rosa H, Penkett CJ, Singer MA, Vermeulen M, Mann M, Bähler J, Green RD, Kouzarides T. 2007. Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **449**: 928–32.

Kirov N, Tsaneva I, Einbinder E, Tsanev R. 1992. In vitro transcription through nucleosomes by T7 RNA polymerase. *EMBO J* **11**: 1941–7.

Klopf E, Paskova L, Solé C, Mas G, Petryshyn A, Posas F, Wintersberger U, Ammerer G, Schüller C. 2009. Cooperation between the INO80 complex and histone chaperones determines adaptation of stress gene transcription in the yeast Saccharomyces cerevisiae. *Mol Cell Biol* **29**: 4994–5007.

Knee R, Murphy PR. 1997. Regulation of gene expression by natural antisense RNA transcripts. *Neurochem Int* **31**: 379–392.

Kolodziej PA, Woychik N, Liao S, Young RA. 1990. RNA polymerase II subunit composition, stoichiometry and phosphorylation. *Mol Cell Biol* **10**: 1915–1920.

Komarnitsky P. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**: 2452–2460.

Konarzewska P, Esposito M, Shen C-H. 2012. INO1 induction requires chromatin remodelers Ino80p and Snf2p but not the histone acetylases. *Biochem Biophys Res Commun* **418**: 483–8.

Kornberg RD. 1996. RSC, an essential, abundant chromatin-remodeling complex. *Cell* **87**: 1249–60.

Krishnamurthy S, He X, Reyes-Reyes M, Moore CL, Hampsey M. 2004. Ssu72 is an RNA Polymerase II CTD phosphatase. *Mol Cell* **14**: 387–394.

Kristjuhan A, Svejstrup JQ. 2004. Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo. *EMBO J* **23**: 4243–52.

Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M, et al. 2003a. The Paf1 Complex Is Required for Histone H3 Methylation by COMPASS and Dot1p : Linking Transcriptional Elongation to Histone Methylation. *Mol Cell* **11**: 721–729.

Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, Richards DP, Beattie BK, Emili A, Boone C, et al. 2003b. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol* **23**: 4207–4218.

Kuehner JN, Brow D a. 2008. Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. *Mol Cell* **31**: 201–11.

Kulaeva OI, Gaykalova DA, Pestov NA, Golovastov V V, Vassylyev DG, Artsimovitch I, Studitsky VM. 2009. Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. *Nat Struct Mol Biol* **16**: 1272–8.

Kulaeva OI, Gaykalova DA, Studitsky VM. 2007. Transcription through chromatin by RNA polymerase II: histone displacement and exchange. *Mutat Res* **618**: 116–29.

La Cava 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**: 713–24.

Lapidot M, Pilpel Y. 2006. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep* **7**: 1216–22.

Lee J-S, Garrett AS, Yen K, Takahashi Y-H, Hu D, Jackson J, Seidel C, Pugh BF, Shilatifard A. 2012. Codependency of H2B monoubiquitination and nucleosome reassembly on Chd1. *Genes Dev* **26**: 914–9.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39**: 1235–44.

Lehner B. 2010. Conflict between noise and plasticity in yeast. *PLoS Genet* **6**: e1001185.

Lehner B. 2008. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol Syst Biol* **4**: 170.

Lemieux K, Gaudreau L. 2004. Targeting of Swi/Snf to the yeast GAL1 UAS G requires the Mediator, TAF IIs, and RNA polymerase II. *EMBO J* **23**: 4040–50.

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, Margaritis T, van de Pasch L a L, van Heesch S a a C, Brok MO, Groot Koerkamp MJ, et al. 2011. The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell* **42**: 536–49.

Li B, Gogol M, Carey M, Pattenden SG, Seidel C, Workman JL. 2007. Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. *Genes Dev* **21**: 1422–30.

Li B, Howe L, Anderson S, Yates JR, Workman JL. 2003. The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **278**: 8897–903.

Li B, Pattenden SG, Lee D, Gutiérrez J, Chen J, Seidel C, Gerton J, Workman JL. 2005. Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc Natl Acad Sci U S A* **102**: 18385–90.

Li Q, Zhou H, Wurtele H, Davies B, Horazdovsky B, Verreault A, Zhang Z. 2008. Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. *Cell* **134**: 244–55.

Lin Z, Wu WS, Liang H, Woo Y, Li W-H. 2010. The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation. *BMC Genomics* **11**: 581.

Ling MHT, Ban Y, Wen H, Wang SM, Ge SX. 2013. Conserved expression of natural antisense transcripts in mammals. *BMC Genomics* **14**: 243.

Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ. 2005. Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biol* **3**: e328.

Liu X, Bushnell D a, Wang D, Calero G, Kornberg RD. 2010. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327**: 206–9.

Liu Y, Warfield L, Zhang C, Luo J, Allen J, Lang WH, Ranish J, Shokat KM, Hahn S. 2009. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol Cell Biol* **29**: 4852–63.

Longtine MS, McKenzie A, 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**: 953–61.

Lopez D, Vlamakis H, Kolter R. 2009. Generation of multiple cell types in Bacillus subtilis. *FEMS Microbiol Rev* **33**: 152–63.

Lorch Y, LaPointe JW, Kornberg RD. 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**: 203–10.

Lorch Y, Zhang M, Kornberg RD. 1999. Histone octamer transfer by a chromatin-remodeling complex. *Cell* **96**: 389–392.

Lubliner S, Keren L, Segal E. 2013. Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res* **41**: 5569–5581.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997a. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251–260.

Luger K, Rechsteiner TJ, Flaus AJ, Waye MMY, Richmond TJ. 1997b. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol* **272**: 301–11.

Luger K, Richmond TJ. 1998. The histone tails of the nucleosome. *Curr Opin Genet Dev* **8**: 140–146.

Luse DS, Jacob GA. 1987. Abortive Initiation by RNA Polymerase II in Vitro at the Adenovirus 2 Major Late Promoter. *J Biol Chem* **262**: 14990–14997.

MacIsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E. 2006. An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics* **7**: 113.

Magee JA, Abdulkadir SA, Milbrandt J. 2003. Haploinsufficiency at the Nkx3.1 locus. A paradigm for stochastic, dosage-sensitive gene regulation during tumor initiation. *Cancer Cell* **3**: 273–283.

Maicas E, Friesen JD. 1990. A sequence pattern that occurs at the transcription initiation region of yeast RNA polymerase II promoters. *Nucleic Acids Res* **18**: 3387–3393.

Martens J a, Laprade L, Winston F. 2004. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature* **429**: 571–4.

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**: 8323–33.

Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF. 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* **18**: 1073–83.

Medler S, Al Husini N, Raghunayakula S, Mukundan B, Aldea A, Ansari A. 2011. Evidence for a complex of transcription factor IIB with poly(A) polymerase and cleavage factor 1 subunits required for gene looping. *J Biol Chem* **286**: 33709–18.

Memedula S, Belmont AS. 2003. Sequential recruitment of HAT and SWI/SNF components to condensed chromatin by VP16. *Curr Biol* **13**: 241–6.

Meneghini MD, Wu M, Madhani HD. 2003. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112**: 725–36.

Métivier R, Penot G, Carmouche RP, Hübner MR, Reid G, Denger S, Manu D, Brand H, Kos M, Benes V, et al. 2004. Transcriptional complexes engaged by apo-estrogen receptor-alpha isoforms have divergent outcomes. *EMBO J* **23**: 3653–66.

Metivier R, Penot G, Hubner MR, Reid G, Brand H, Kos M, Gannon F. 2003. Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**: 751–763.

Milcarek C, Price R, Penman S. 1974. The metabolism of a poly(A) minus mRNA fraction in HeLa cells. *Cell* **3**: 1–10.

Millar CB, Xu F, Zhang K, Grunstein M. 2006. Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. *Genes Dev* **20**: 711–22.

Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dölken L, et al. 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7**: 458.

Mischo HE, Proudfoot NJ. 2013. Disengaging polymerase : Terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta* **1829**: 174–185.

Mizuguchi G, Shen X, Landry J, Wu W-H, Sen S, Wu C. 2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**: 343–8.

Mizzen CA, Yang XJ, Kokubo T, Brownell JE, Bannister AJ, Owen-Hughes T, Workman J, Wang L, Berger SL, Kouzarides T, et al. 1996. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**: 1261–70.

Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L, Workman JL, Washburn MP. 2009. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol Cell* **34**: 168–78.

Murray SC, Serra Barros A, Brown D a, 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**: 2432–44.

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**: 1344–9.

Narlikar GJ, Fan H, Kingston RE. 2002. Cooperation between Complexes that Regulate Chromatin Structure and Transcription. *Cell* **108**: 475–487.

Naumova N, Smith EM, Zhan Y, Dekker J. 2012. Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods* **58**: 192–203.

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**: 33000–10.

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**: 1038–42.

Neuwald AF, Liu JS, Lawrence CE. 1995. Gibbs motif sampling: detection of bacterial outer membrane protein repeats. *Protein Sci* **4**: 1618–32.

Newburger DE, Bulyk ML. 2009. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **37**: D77–82.

Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. 2006. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**: 840–6.

Ng HH, Feng Q, Wang H, Erdjument-bromage H, Tempst P. 2002. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev* **16**: 1518–27.

Ng HH, Robert F, Young R a, Struhl K. 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11**: 709–19.

Noma K, Allis CD, Grewal SI. 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**: 1150–5.

Novina CD, Sharp P. 2004. The RNAi revolution. *Nature* **430**: 161–4.

O'Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, Mellor J, Proudfoot NJ. 2004. Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* **36**: 1014–8.

Okoniewski MJ, Miller CJ. 2006. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics* **7**: 276.

Osborne BI, Guarente L. 1989. Mutational analysis of a yeast transcriptional terminator. *Proc Natl Acad Sci U S A* **86**: 4097–101.

Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. 2002. Regulation of noise in the expression of a single gene. *Nat Genet* **31**: 69–73.

Papamichos-Chronakis M, Watanabe S, Rando OJ, Peterson CL. 2011. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**: 200–13.

Perocchi F, Xu Z, Clauder-Münster S, Steinmetz LM. 2007. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res* **35**: e128.

Peterson CL, Workman JL. 2000. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr Opin Genet Dev* **10**: 187–92.

Petrenko N, Chereji R V, McClean MN, Morozov A V, Broach JR. 2013. Noise and interlocking signaling pathways promote distinct transcription factor dynamics in response to different stresses. *Mol Biol Cell* **24**: 2045–57.

Phillips DM, Johns EW. 1965. a Fractionation of the Histones of Group F2a From Calf Thymus. *Biochem J* **94**: 127–30.

Pinskaya M, Gourvennec S, Morillon A. 2009a. H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J* **28**: 1697–707.

Pinskaya M, Nair A, Clynes D, Morillon A, Mellor J. 2009b. Nucleosome remodeling and transcriptional repression are distinct functions of Isw1 in Saccharomyces cerevisiae. *Mol Cell Biol* **29**: 2419–30.

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**: 517–27.

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Heick J. 2008. RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science* **322**: 1851–1854.

Proudfoot N. 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* **3**: 272–278.

Proudfoot NJ. 2011. Ending the message: poly(A) signals then and now. *Genes Dev* **25**: 1770–82.

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**: 752–62.

Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD. 2005. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**: 233–48.

Ransom M, Williams SK, Dechassa ML, Das C, Linger J, Adkins M, Liu C, Bartholomew B, Tyler JK. 2009. FACT and the proteasome promote promoter chromatin disassembly and transcriptional initiation. *J Biol Chem* **284**: 23461–71.

Rao C V, Wolf DM, Arkin AP. 2002. Control, exploitation and tolerance of intracellular noise. *Nature* **420:** 231-7.

Raser JM, Shea EKO. 2004. Control of stochasticity in eukaryotic gene expression. *Science* **1811**: 1811-4.

Rhee HS, Pugh BF. 2011. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**: 1408–19.

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

Richmond RK, Sargent DF, Richmond TJ, Luger K, Mader AW. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **18**: 251-60.

Rodnina M V, Beringer M, Wintermeyer W. 2007. How ribosomes make peptide bonds. *Trends Biochem Sci* **32**: 20–6.

Rodriguez CR, Cho E-J, Keogh M-C, Moore CL, Greenleaf AL, Buratowski S. 2000. Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA Polymerase II. *Mol Cell Biol* **20**: 104–112.

Rodrıguez-Gil A, Garcia-Martinez J, Pelechano V, de la Cruz M, 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.

Roeder RG. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**: 327–35.

Ross IANL, Browne CM, Hume DA. 1994. Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol Cell Biol* **72**: 177–185.

Royce TE, Rozowsky JS, Gerstein MB. 2007. Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. *Nucleic Acids Res* **35**: e99.

Rufiange A, Jacques P-E, Bhat W, Robert F, Nourani A. 2007. Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Mol Cell* **27**: 393–405.

Santos-Rosa H, Bannister AJ, Dehe PM, Géli V, Kouzarides T. 2004. Methylation of H3 lysine 4 at euchromatin promotes Sir3p association with heterochromatin. *J Biol Chem* **279**: 47506–12.

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, Schreiber SL, Mellor J, Kouzarides T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**: 407–411.

Santos-Rosa H, Schneider R, Bernstein BE, Karabetsou N, Morillon A, Weise C, Schreiber SL, Mellor J, Kouzarides T. 2003. Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol Cell* **12**: 1325–32.

Schmid M, Arib G, Laemmli C, Nishikawa J, Durussel T, Laemmli UK. 2006. Nup-PI: the nucleopore-promoter interaction of genes in yeast. *Mol Cell* **21**: 379–91.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671–675.

Schroeder SC. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev* **14**: 2435–2440.

Schulze JM, Hentrich T, Nakanishi S, Gupta A, Emberly E, Shilatifard A, Kobor MS. 2011. Splitting the task: Ubp8 and Ubp10 deubiquitinate different cellular pools of H2BK123. *Genes Dev* **25**: 2242–7.

Schwabish M a, Struhl K. 2006. Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol Cell* **22**: 415–22.

Schwabish M a, Struhl K. 2007. The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo. *Mol Cell Biol* **27**: 6987–95.

Schwabish MA, Struhl K. 2004. Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol Cell Biol* **24**: 10111–10117.

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn R a, Young R a, Sharp P a. 2008. Divergent transcription from active promoters. *Science* **322**: 1849–51.

Shandilya J, Roberts SGE. 2012. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta* **1819**: 391–400.

Shendure J. 2008. The beginning of the end for microarrays? *Nat Methods* **5**: 585–7.

Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR. 2008. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol* **6**: e65.

Singer VL, Wobbe CR, Struhl K. 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev* **4**: 636–645.

Singh BN, Hampsey M. 2007. A transcription-independent role for TFIIB in gene looping. *Mol Cell* **27**: 806–16.

Smale ST, Kadonaga JT. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**: 449–79.

Smith-Vikos T, Slack FJ. 2013. MicroRNAs circulate around Alzheimer's disease. *Genome Biol* **14**:

Smolle M, Venkatesh S, Gogol MM, Li H, Zhang Y, Florens L, Washburn MP, Workman JL. 2012. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol* **19**: 884–92.

Sneppen K, Dodd IB, Shearwin KE, Palmer AC, Schubert RA, Callen BP, Egan JB. 2005. A mathematical model for transcriptional interference by RNA polymerase traffic in Escherichia coli. *J Mol Biol* **346**: 399–409.

Soutourina J, Bordas-Le Floch V, Gendrel G, Flores A, Ducrot C, Dumay-Odelot H, Soularue P, Navarro F, Cairns BR, Lefebvre O, et al. 2006. Rsc4 connects the chromatin remodeler RSC to RNA polymerases. *Mol Cell Biol* **26**: 4920–33.

Stagoj MN, Comino A, Komel R. 2005. Fluorescence based assay of GAL system in yeast Saccharomyces cerevisiae. *FEMS Microbiol Lett* **244**: 105–10.

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**: 327–31.

Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA. 2006. Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell* **24**: 735–46.

Stormo GD. 2000. DNA binding sites: representation and discovery. *Bioinformatics* **16**: 16–23.

Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* **403**: 41–45.

Strahl BD, Grant PA, Briggs SD, Sun Z, Bone JR, Caldwell JA, Mollah S, Cook RG, Shabanowitz J, Hunt DF, et al. 2002. Set2 Is a Nucleosomal Histone H3-Selective methyltransferase that mediates transcriptional repression. *Mol Cel Biol* **22**: 1298–1306.

Strahl BD, Ohba R, Cook RG, Allis CD. 1999. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proc Natl Acad Sci U S A* **96**: 14967–72.

Struhl K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* **14**: 103–5.

Stünkel W, Kober I, Seifar KH. 1997. A nucleosome positioned in the distal promoter region activates transcription of the human U6 A Nucleosome Positioned in the Distal Promoter Region Activates Transcription of the Human U6 Gene. *Mol Cell Biol* **17**: 4397–4405.

Suk K, Choi J, Suzuki Y, Ozturk SB, Mellor JC, Wong KH, MacKay JL, Gregory RI, Roth FP. 2011. Reconstitution of human RNA interference in budding yeast. *Nucleic Acids Res* **39**: e45

Svejstrup JQ, Feaver WJ, LaPointe J, Kornberg RD. 1994. RNA polymerase transcription factor IIH holoenzyme from yeast. *J Biol Chem* **269**: 28044–8.

Tan-Wong SM, Wijayatilake HD, Proudfoot NJ. 2009. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev* **23**: 2610–24.

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**: 671–5.

Taverna SD, Ilin S, Rogers RS, Tanny JC, Lavender H, Li H, Baker L, Boyle J, Blair LP, Chait BT, et al. 2006. Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol Cell* **24**: 785–96.

Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sá-Correia I. 2006. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. *Nucleic Acids Res* **34**: D446–51.

Thattai M, van Oudenaarden A. 2001. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A* **98**: 8614–9.

Thattai M, van Oudenaarden A. 2004. Stochastic gene expression in fluctuating environments. *Genetics* **530**: 523–530.

Thebault P, Boutin G, Bhat W, Rufiange A, Martens J, Nourani A. 2011. Transcription regulation by the noncoding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. *Mol Cell Biol* **31**: 1288–300.

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**: 853–64.

Thiriet C, Hayes JJ. 2005. Replication-independent core histone dynamics at transcriptionally active loci in vivo. *Genes Dev* **19**: 677–82.

Thompson DM, Parker R. 2007. Cytoplasmic decay of intergenic transcripts in Saccharomyces cerevisiae. *Mol Cell Biol* **27**: 92–101.

Tirosh I, Barkai N. 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome Res* **18**: 1084–91.

Tirosh I, Sigal N, Barkai N. 2010. Widespread remodeling of mid-coding sequence nucleosomes by Isw1. *Genome Biol* **11**: R49.

Tse C, Sera T, Wolffe AP, Hansen JC. 1998. Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol Cell Biol* **18**: 4629–4638.

Tsubota T, Berndsen CE, Erkmann JA, Smith CL, Yang L, Freitas MA, Denu JM, Kaufman PD. 2007. Histone H3-K56 acetylation is catalyzed by histone chaperone-dependent complexes. *Mol Cell* **25**: 703–12.

Tu S, Bulloch EMM, Yang L, Ren C, Huang WC, Hsu P-H, Chen CH, Liao C-L, Yu HM, Lo WS, et al. 2007. Identification of histone demethylases in Saccharomyces cerevisiae. *J Biol Chem* **282**: 14262–71.

Turner BM. 2000. Histone acetylation and an epigenetic code. *BioEssays* **22**: 836–45.

Uhler JP, Hertel C, Svejstrup JQ. 2007. A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc Natl Acad Sci U S A* **104**: 8011–6.

Van Attikum H, Fritsch O, Hohn B, Gasser SM. 2004. Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA Double-Strand Break Repair NCCR Frontiers in Genetics Program. *Cell* **119**: 777–788.

Van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Né P, Loeillet S, et al. 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475**: 114–7.

Van Dijk K, Ding Y, Malkaram S, Riethoven J-JM, Liu R, Yang J, Laczko P, Chen H, Xia Y, Ladunga I, et al. 2010. Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in Arabidopsis thaliana. *BMC Plant Biol* **10**: 238.

Van Leeuwen F, Gafken PR, Gottschling DE. 2002. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109**: 745–56.

Vanácová S, 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.

Venkataraman K, Brown KM, Gilmartin GM. 2005. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* **19**: 1315–27.

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* **489**: 452–5.

Venkatesh S, Workman JL. 2012. Non-coding transcription SETs up regulation. *Cell Res* 1–3.

Venters BJ, Pugh BF. 2009. A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome Res* **19**: 360–371.

Venters BJ, Wachi S, Mavrich TN, Andersen BE, Jena P, Sinnamon AJ, Jain P, Rolleri NS, Jiang C, Hemeryck-Walsh C, et al. 2011. A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. *Mol Cell* **41**: 480–92.

Verreault a, Kaufman PD, Kobayashi R, Stillman B. 1996. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**: 95–104.

De Vos D, Frederiks F, Terweij M, van Welsem T, Verzijlbergen KF, Iachina E, de Graaf EL, Altelaar a FM, Oudgenoeg G, Heck AJR, et al. 2011. Progressive methylation of ageing histones by Dot1 functions as a timer. *EMBO Rep* **12**: 956–62.

Wang A, Kurdistani SK, Grunstein M. 2002. Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science* **298**: 1412–4.

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

Warner JR. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24**: 437–40. Weake VM, Workman JL. 2010. Inducible gene expression: diverse regulatory mechanisms. *Nat Rev Genet* **11**: 426–37.

Weinberger L, Voichek Y, Tirosh I, Hornung G, Amit I, Barkai N. 2012. Expression noise and acetylation profiles distinguish HDAC functions. *Mol Cell* **47**: 193–202.

Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, Schaffer D V. 2005. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **122**: 169-82.

Weiner A, Chen H V, Liu CL, Rahat A, Klien A, Soares L, Gudipati M, Pfeffner J, Regev A, Buratowski S, et al. 2012. Systematic dissection of roles for chromatin regulators in a yeast stress response. *PLoS Biol* **10**: e1001369.

White SA, Allshire RC. 2008. RNA-mediated chromatin silencing in fission yeast. *Curr Top Microbiol Immunol* **320**:157-83

Whitehouse I, Rando OJ, Delrow J, Tsukiyama T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**: 1031–5.

Widlund HR, Vitolo JM, Thiriet C, Hayes JJ. 2000. DNA sequence-dependent contributions of core histone tails to nucleosome stability: differential effects of acetylation and proteolytic tail removal. *Biochemistry* **39**: 3835–41.

Williams SK, Truong D, Tyler JK. 2008. Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc Natl Acad Sci U S A* **105**: 9000–5.

Wittenberg C, Reed SI. 2005. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. *Oncogene* **24**: 2746–55.

Wyers F, Rougemaille M, Badis G, Rousselle JC, 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**: 725–37.

Xiao T, Kao C, Krogan NJ, Sun Z, Greenblatt JF, Osley MA, Strahl BD. 2005. Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Mol Cell Biol* **25**: 637–651.

Xu F, Zhang Q, Zhang K, Xie W, Grunstein M. 2007. Sir2 deacetylates histone H3 lysine 56 to regulate telomeric heterochromatin structure in yeast. *Mol Cell* **27**: 890–900.

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

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

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**: 5110–22.

Yadon AN, Singh BN, Hampsey M, Tsukiyama T. 2013. Article DNA Looping Facilitates Targeting of a Chromatin Remodeling Enzyme. *Mol Cell* **50**: 93–103.

Yang X, Zaurin R, Beato M, Peterson CL. 2007. Swi3p controls SWI/SNF assembly and ATP-dependent H2A-H2B displacement. *Nat Struct Mol Biol* **14**: 540–7.

Yassour M, Kaplan T, Fraser HB, Levin JZ, Pfiffner J, Adiconis X, Schroth G, Luo S, Khrebtukova I, Gnirke A, et al. 2009. Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *Proc Natl Acad Sci U S A* **106**: 3264–9.

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

Yean D, Gralla J. 1997. Transcription reinitiation rate: a special role for the TATA box **17**: 3809-16.

Yen K, Vinayachandran V, Batta K, Koerber RT, Pugh BF. 2012. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* **149**: 1461–73.

Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science* **309**: 626–30.

Yudkovsky N, Ranish JA, Hahn S. 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**: 225–9.

Zaugg JB, Luscombe NM. 2012. A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. *Genome Res* **22**: 84–94.

Zeng L, Zhou MM. 2002. Bromodomain: an acetyl-lysine binding domain. *FEBS Lett* **513**: 124–8.

Zenklusen D, Larson DR, Singer RH. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* **15**: 1263–71.

Zenklusen D, Singer RH. 2010. Analyzing mRNA expression using single mRNA resolution fluorescent in situ hybridization. *Methods Enzymol* **470:** 641-59

Zhang H, Roberts DN, Cairns BR. 2005. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123**: 219–31.