University of Oxford

Transcriptional regulation by non-coding RNAs in *Saccharomyces cerevisiae*

Doctorate of Philosophy in Biochemistry

Ana Serra Barros

The Queen's College

Abstract

Genome-wide studies in *Saccharomyces cerevisiae* have revealed that the majority of the genome is transcribed on both strands, producing both coding and non-coding RNAs (ncRNAs). Initially, these ncRNAs were regarded as spurious transcripts but some have since been shown to have important roles as transcriptional regulators. Very little is understood about how ncRNAs are initiated, terminated and processed or how this influences their function. To address these questions, the expression, stability, and subcellular localization of the ncRNAs at the endogenous *GAL* locus was analysed. This revealed a complex interleaved transcript map, challenging the conventional view of a transcription unit (TU) flanked by 5' sequences or promoters (P) that initiate transcription and 3' regions, known as terminators (T), which control events such as transcript cleavage, polyadenylation, export and transcription termination. By creating conventional (P_{GAL}-T) or unconventional (P_{GAL}-P) hybrid TUs at the *GAL* locus, in which a promoter or terminator is positioned downstream of a galactose-inducible promoter, this work shows that both promoters and terminators are able to initiate antisense transcription to yield stable antisense transcripts. The data suggest that terminators contribute to efficient but variable expression from the promoter. An unconventional P-P TU, lacking a terminator, is transcribed on both strands but the sense transcript remains at low levels, through the repressive action of antisense transcription, and is retained in the nucleus. In contrast, the conventional P-T bi-directional TUs are plastic, with the Rrp6 component of the nuclear exosome and TATA-like sequences in the 3' UTR determining whether the predominant transcript is antisense or sense. By relieving the repressive action of

antisense transcription, this allows high levels of sense transcript to accumulate in the cytoplasm, contributing to gene expression, supporting a novel mode of gene regulation involving components of RNA quality control pathways acting through the 3' region of genes.

Ana Serra Barros

The Queen's College

Submitted for the degree of Doctor of Philosophy

Trinity Term 2012

Acknowledgments

I would like to, first of all, thank Manuel Santos, Alexandre Akoulitchev and Jane Mellor, for all the support and guidance throughout these years.

A big thanks to everyone at the Mellor lab, past and present, Mike, Dave, Pete, Anitha, Jon, Françoise, Tania, Simon, Struan and Ivan for making the working environment the best anyone could have. A special thanks to Françoise and Tania for the immense task of proofreading my thesis over and over again.

To all the amazing people I have met during these 'short' 7 years that I have lived in Oxford and that have made my living abroad a very entertaining and maturing experience, a big thank you.

And last but certainly not least I would have not been here if it was not for my mother.

Publications

Murray, S.C., Serra Barros, A., Brown, D.A., Dudek, P., Ayling, J., and Mellor, J. (2011). A pre-initiation complex at the 3′ -end of genes drives antisense transcription independent of divergent sense transcription. Nucleic Acids Research 40(6):2432-2444.

Work performed by others was used in this thesis with their approval.

Below is a list:

RNA FISH – Simon Haenni

Figure 22, Figure 44 and Figure 63

5’ RACE mapping and construction of the *GAL1* frameshift mutation - Peter Dudek

Reb1 Northern blot and ChIP – Michael Youdell

Figure 14

Design and construction of the *GAL10 ADH1*^{T} deletion strains - Struan Murray

Design and construction of the *GAL10 TEF*^{T} LoxP and *GAL10 ADH1*^{T} LoxP strains – David Brown

Table of Contents

Chapter 1

Introduction

1.1 Transcription

The genetic information stored in the DNA is transcribed into a pre-mRNA through the process of transcription by RNA Polymerase II (RNAPII), which is then processed into mRNA before export from the nucleus to the cytoplasm and translation into a protein. The act of transcribing a sequence of DNA into an RNA molecule is a process of successive steps including, i) initiation, ii) elongation and iii) termination. Transcriptional initiation refers to the recruitment of the basal transcriptional machinery to the transcription start site (Sikorski and Buratowski, 2009), while the subsequent action of the RNAPII travelling along the DNA template, synthesising an RNA molecule, is called transcriptional elongation (Lavelle, 2007). The final step of dissociation of RNAPII from the DNA and consequent cleavage of the RNA molecule is designated as transcriptional termination (Moore and Proudfoot, 2009). There are three eukaryotic RNA polymerases, with RNAPII being the one responsible for transcription of almost all mRNAs in a cell, whereas the other two, RNAPI and RNAPIII, are responsible for the transcription of functionally well-described non-coding RNAs including rRNAs, tRNAs and other small RNAs (Woychik et al., 1990). The core RNAPII is composed of 12 subunits (Rpb1-Rpb12) and this complex, in association with numerous other proteins and complexes, including the general transcription factors, the mediator complex and chromatin remodelling complexes, is referred to as the RNAPII holoenzyme. Rpb1 is the largest subunit and it includes a C-terminal domain (CTD) that has functions in practically all steps of transcription. The CTD is comprised of a repeated sequence, YSPTSPS, which can be phosphorylated mainly at serines 2 and 5. It is thought that the phosphorylation pattern of the CTD is responsible for the regulation of many processes (Thomas and Chiang, 2006).

1.1.1 Promoter elements and transcription machinery

The DNA region that efficiently recruits the transcriptional machinery that directs initiation of transcription is called the promoter (Butler and Kadonaga, 2002; Juven-Gershon and Kadonaga, 2010). Eukaryotic promoters contain a number of *cis*-regulatory elements (CREs) that are responsible for the recruitment of *trans*-acting sequence specific factors. CREs can be classified as enhancers in metazoans or upstream activating/repressive sequences (UAS/URS) in yeast, depending on their effect on transcription. A core promoter region can be defined as the minimal sequence of DNA required for efficient assembly of the machinery and initiation of transcription (Butler and Kadonaga, 2002). In metazoans, several elements are part of the core promoter including the Initiator (Inr), the downstream promote element (DPE), motif ten element (MTE) and TFIIB recognition element (BRE) (Venters and Pugh, 2010). In yeast, the elements associated with the core promoter are the TATA element and Initiator. The TATA element, with a 5'-TATA(T/A)A(A/T)(A/G)-3' consensus sequence, is usually located around 40-100 bp (25-30 bp in metazoans) upstream the transcription start site (TSS) while the Initiator element is situated adjacent to the TSS, and lacks a strong consensus sequence (Sikorski and Buratowski, 2009; Struhl, 1995; Venters and Pugh, 2010). However, it was demonstrated that variations of the TATA consensus sequence can still function as a TATA motif (Singer et al., 1990).

Transcription initiation is achieved by the recruitment of trans-acting factors usually denominated as general transcription factors (GTFs) to form the pre-initiation complex (PIC) (Thomas and Chiang, 2006). These include, TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. A stepwise association of the GTFs to a promoter to form the PIC begins with TFIID, which

is a large protein complex containing TBP (TATA-binding protein) and several TAFs (TBP-associated factors) that recognizes and binds to the TATA motif, although not exclusively (Juven-Gershon and Kadonaga, 2010; Venters and Pugh, 2010). Then recruitment of TFIIA and TFIIB contributing to the stabilization of the DNA-bound TFIID brings in RNAPII with TFIIF. Followed by TFIIE and finally TFIIH which contributes for promoter clearance and transcriptional initiation (Thomas and Chiang, 2006). Transcriptional activators are usually sequence-specific proteins that bind to the promoters to regulate gene expression. These work cooperatively with the GTFs in what is called “activated” as opposed to “basal” transcription.

Additional factors regulate the process of transcription initiation that can either be in a positive, transcriptional activators, or negative, transcriptional repressors, manner. The transcriptional activators contain a DNA-binding domain (DBD) as well as an activation domain (AD) which triggers transcription. The mechanisms of action include i) increased PIC assembly; ii) involvement in later steps after PIC assembly, for *e.g.*, initiation, elongation, reinitiation; iii) recruitment of chromatin remodelling complexes. Examples of well-defined transcription activators are Gal4p and Gcn4p (Green, 2005). Transcriptional repressors (for *e.g.* Gal80p, Mig1p) on the other hand function by i) competing with activators for the DNA element, ii) direct binding to the AD of the activator, iii) binding to DNA and disturbing the communication between the activator and the transcriptional machinery (Struhl, 1995).

1.1.2 Nucleosomes and gene expression

The accessibility of the DNA elements by trans-acting factors is constrained by the existence of a chromatin structure (Kornberg and Lorch, 1992). The DNA is packaged into the nucleus by the association with highly basic proteins, histones, which neutralizes the negative charge of DNA. Two copies of each of the four canonical histones, H3, H4, H2A, H2B, are arranged in an octamer, designated nucleosome, the basic unit of the chromatin structure, which contains 147 bp of DNA wrapped around 1.65 times (Workman and Buchman, 1993). Nucleosome positioning is still a very active research topic but with recent development of genome-wide technology it is emerging that various determinants are involved, including the underlying DNA sequence and chromatin factors (Jansen and Verstrepen, 2011; Jiang and Pugh, 2009; Mavrich et al., 2008). Histones can be subjected to various modifications, including methylation, acetylation, phosphorylation, ubiquitination SUMOylation, at numerous amino acids residues. These modifications regulate chromatin organization and DNA-associated processes, such as transcription (Lennartsson and Ekwall, 2009).

The positioning of nucleosomes at a promoter influences gene expression. It is becoming more evident that there is a pattern of organization of nucleosomes within genes. A -1 nucleosome, located -300 to -150 from the TSS, that can regulate the accessibility of the transcription machinery to some of the promoter elements, and a well-positioned +1 nucleosome, located after the TSS, delimit a well-defined 5' UTR nucleosome free region (NFR). The +1 nucleosome usually contains histone variants (such as H2A.Z and H3.3) and histone tail modifications (Jiang and Pugh, 2009). This organization of -1, NFR and +1 nucleosomes delineates the region where the transcription machinery assembles. Further

into a gene the nucleosomes are found to be less organized and the distribution looks as if it they are randomly positioned. At the 3' end of a gene the nucleosome occupancy becomes once more well-organized into a 3' NFR (Fan et al., 2010; Jiang and Pugh, 2009).

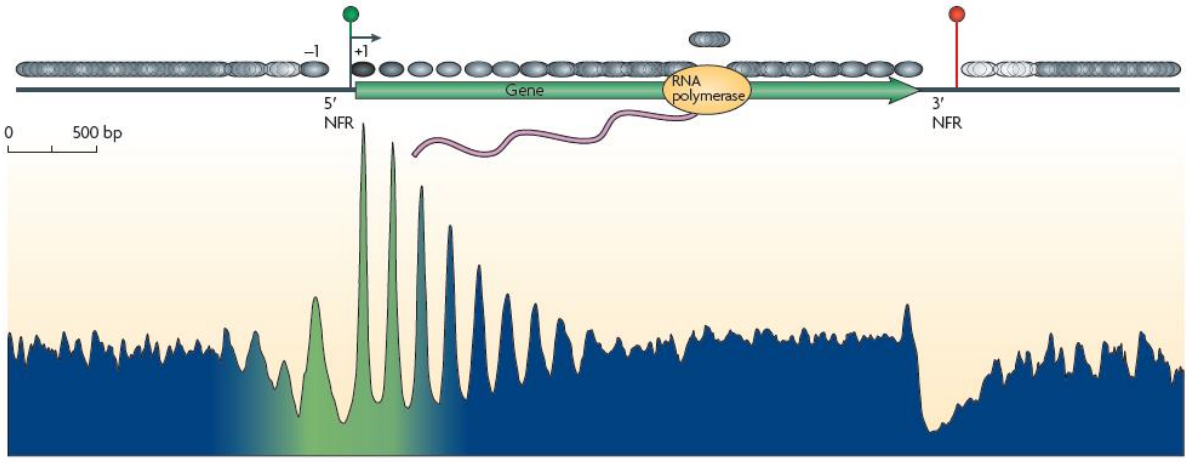


Figure 1. Nucleosome positioning at yeast genes. Nucleosomes (grey) organization within a gene, with the 5' and 3' nucleosome free regions (NFRs). Transcription start site (TSS) represented with a green dot and the termination at the 3'end with a red dot. The peaks in the graph below represent the positioning of the nucleosomes and the green colour the presence of histone variants and modifications, such as H2A.Z, acetylation and H3 K4 methylation opposed to low levels of these modifications represented in blue. Reproduced from (Jiang and Pugh, 2009).

1.1.3 3' end processing events

There are a number of processing events that an mRNA needs to be subjected to in order to transport the information encoded in the sequence to the cytoplasm of a cell. These include the addition of a 5' end capping structure, by the Cap-binding complex (CBC), splicing of introns if they are present, cleavage and polyadenylation of the 3'end. All of these events occur co-transcriptionally and are regulated by the RNAPII CTD, as mentioned (Lykke-Andersen and Jensen, 2007).

Termination of RNAPII transcription of mRNAs includes the transcription through the 3' end region, recognition of the poly(A)/cleavage site by a protein complex, cleavage and processing of the RNA molecule and consequent dissociation of the RNAPII from the DNA (Proudfoot, 1989). The actual site of termination is stochastic and it occurs downstream of the cleavage/polyadenylation site. Transcription termination is interlinked with 3' end processing of the RNA established by observations that both processes are dependent on the same DNA sequences and that certain factors involved in the cleavage and polyadenylation are required for efficient termination (Buratowski, 2005). This co-transcriptional process is thought to be mediated by the phosphorylated status of the RNAPII CTD. The RNAPII CTD is phosphorylated at serine 5 by TFIIH at the promoter and an additional kinase, Ctk1, phosphorylates serine 2 during elongation. This differential pattern or code of CTD phosphorylation can recruit different factors. The code of CTD phosphorylation is thought to highly regulate initiation, elongation, termination and transcription-coupled mRNA processing processes (Thomas and Chiang, 2006). At least, two different mechanisms of 3'end-processing coupled transcription termination have been discovered (Figure 2) (Venters and Pugh, 2010).

Polyadenylation-dependent termination of mRNAs requires the recognition of the poly(A) site by a large poly(A) complex. In yeast the 3' end signal is more complex and less conserved than in metazoans. There are at least 3 elements that constitute the minimal 3' end region: i) the UA-rich efficiency element; ii) the A-rich positioning element; and iii) the site of polyadenylation, PyAn (Zhao et al., 1999). Also, some polyadenylation sites have been found to work in both directions (Aranda et al., 1998). For cleavage of the mRNA different factors are required including: Cleavage/polyadenylation factor IA (CF IA)

complex (comprised of Rna14, Rna15, Pcf11 and a polypeptide); Cleavage/polyadenylation factor IB (CF IB); Cleavage factor II (CF II). For polyadenylation, the necessary factors are: CF IA; CF IB; polyadenylation factor I (PF I); Poly(A)-binding protein I; poly(A) nuclease (PAN); and Poly(A) polymerase (Pap1) (Millevoi and Vagner, 2010; Moore and Proudfoot, 2009). Not much is known about the exact mechanism of cleavage and polyadenylation in yeast but it is thought that all factors associate into a large poly(A) complex, which recognizes the elements on the nascent mRNA, cleaves it at the PyAn site and polyadenylates the newly created 3'end (Zhao et al., 1999). Furthermore, it is has been found that cleavage of the mRNA is not required for termination but the polyadenylation is essential (Zhang et al., 2005). The addition of the poly(A) tail has also been found to be required for mRNA export (Brodsky and Silver, 2000).

Recently, another mechanism of termination was found. Ndr1-complex-mediated termination of small RNAs (such as, snRNAs and snoRNAs) is regulated by the recognition of sequences at the nascent transcript and subsequent binding of Ndr1 and Ndr3 and cleavage. Processing of the pre sn- and snoRNAs comprises of the trimming of an extended 3' end by the exosome until it reaches a specific RNA-bound protein. The TRAMP (Trf4/5-Air1/2-Mtr4) a polyadenylation complex, stimulates the exosome activity. The newly uncovered cryptic unstable transcripts (CUTs) are also terminated by this mechanism but the lack of specific RNA-bound factors means that the transcript gets quickly degraded (Lykke-Andersen and Jensen, 2007; Mapendano et al., 2010; Moore and Proudfoot, 2009).

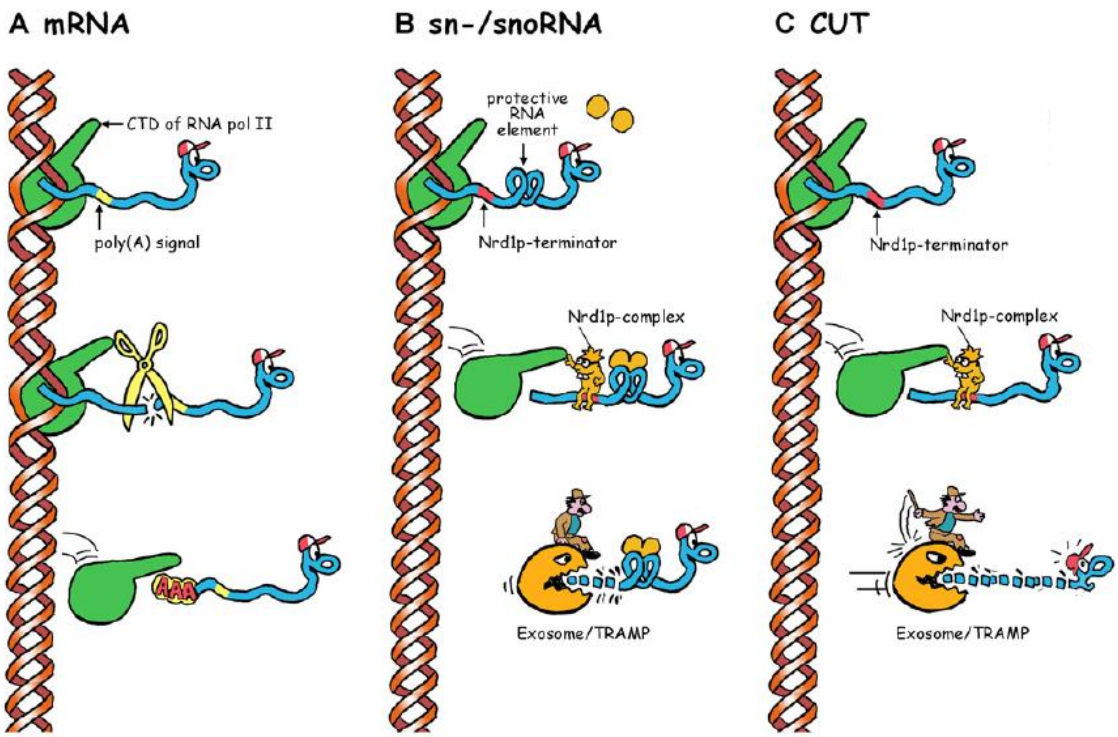


Figure 2. Different mechanisms of transcription termination. (A) Recognition of the poly(A) site by the poly(A) complex directs cleavage and polyadenylation of the transcript. **(B)** Recognition of a sequence-specific element by the Ndr1 complex causing transcription termination and the recruitment of the exosome and TRAMP complexes. Trimming of the 3'end occurs up to the location of sequence-specific RNA-bound proteins which block further degradation. **(C)** Transcription termination of CUTs is also dependent of the Ndr1 complex but the lack of protective RNA elements causes the entire degradation of the transcript by the exosome and TRAMP complexes. Reproduced from (Lykke-Andersen and Jensen, 2007).

1.1.4 Nuclear mRNA quality control

All the steps required for production of a matured mRNP are susceptible to errors in processing and these can be very detrimental for the cell (Culbertson, 1999; Frischmeyer and Dietz, 1999; Hentze and Kulozik, 1999). To prevent these errors from accumulating in the cell, several nuclear mRNA quality control (mQC) mechanisms identify and degrade any improperly processed/packaged mRNAs (Fasken and Corbett, 2005). Additionally,

mechanisms in the cytoplasm also exert mQC (Maquat, 2002, 2004; Vasudevan et al., 2002).

A key component in the mRNA biogenesis is the exosome, a multicomponent complex that includes exonucleases and accessory factors (Mitchell et al., 1997). The exosome is required for the maturation of ribosomal (rRNAs), small nuclear RNAs (snoRNAs) and pre-mRNAs, as well as for RNA degradation. Additionally, it targets the improperly processed/packaged mRNAs for degration (Fasken and Corbett, 2009). In yeast, the exosome core is constituted by 9 subunits arranged in a 6-subunit ring (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46 and Mtr3) and trimeric cap (Rrp4, Rrp40 and Csl4). Additionally, two 3'-5'-exoribonucleases associate with the complex; the essential subunit, Dis3p (also known as Rrp44p), present both in the nucleus and cytoplasm and the nuclear Rrp6p subunit (Lykke-Andersen et al., 2011). Co-factors of the exosome can recognize and target aberrant RNAs and subsequently recruit the complex. One example of a co-factor is the TRAMP complex. Mechanisms of mQC rapidly degrade aberrant transcripts with abnormal 3'ends or poly(A) tails (Burkard and Butler, 2000; Hilleren et al., 2001). The exosome has also been found to retain aberrant mRNAs near the transcription site and near the nuclear pore (Abruzzi et al., 2006; Galy et al., 2004; Schmid and Jensen, 2008).

In addition to the 3'-5' exonucleolitic activity in the nucleus, there is also a 5'-3'-exonuclease, Rat1, that is involved in the maturation of snRNAs and pre-RNA processing (Geerlings et al., 2000). In the cytoplasm the major component involved in mRNA decay is Kem1p (also known as Xrn1p) (Long and McNally, 2003).

1.1.5 mRNA export

One of the known mechanisms by which mRNAs are exported from the nucleus to the cyroplasm includes the association of an mRNA export adaptor, Yrap1p/REF and the ATPase/RNA helicase Sub2p to the mRNA during transcription. Upon cleavage/polyadenylation they recruit the Mex67p-Mtr2p nuclear receptor. This results in the release of Sub2p and interaction with the nucleoporins resulting in the translocation of the mRNAP and consequent export to the cytoplasm (Vinciguerra and Stutz, 2004). Recently, it has been shown that only a small proportion of mRNAPs are found to be associated with Yrap1p/REF (Longman et al., 2003), and that the Mex67p-Mtr2p interaction with Nlp3p (an SR-like protein) contributes to the association with the mRNP (Gilbert and Guthrie, 2004). This indicates that there must be other factors involved in the mRNP export, besides Yrap1p/REF and Sub2p. The direct relationship between mRNA export and transcription has been proposed by the discovery that Yrap1p/REF and Sub2p associates with THO-associated TREX (Transcription-Export) complex that has been associated to transcription elongation (Chavez et al., 2000; Libri et al., 2002; Straszer et al., 2002; Zenklusen et al., 2002). Additionally, a protein has been identified, Sus1p, required for mRNP export, which interacts with both SAGA (histone acetylase) and the Sac3-Thp1p complex (mRNA export), and is recruited to the *GAL1* promoter upon induction, further linking the mRNA export to transcription (Rodŕiguez-Navarro et al., 2004).

Recent studies have found that TREX associates with the nuclear exosome (a 3'-5' exonucleolity complex). The THO complex and mRNA export adaptors, Yra1p and Sub2p when mutated result in low levels and retention of 3' end defective RNAs in nuclear foci.

Upon Rrrp6p mutation the expression levels were rescued and the mRNAs were exported. The results in this study suggest that the release of improperly 3' end formed RNAs is under the control of Rrp6p and/or exosome (Libri et al., 2002; Zenklusen et al., 2002). In general terms, this study suggests a connection between mRNA export and RNA quality control mechanisms.

1.2 Pervasive transcription

1.2.1 Transcription of non-coding RNAs

The genome-wide tiling array experiments revealed the fact that the genome is profusely transcribed giving rise to numerous previously unknown non-coding RNAs (ncRNAs) (Churchman and Weissman, 2011; David et al., 2006; Katayama et al., 2005; Perocchi et al., 2007; Xu et al., 2009; Yassour et al., 2010). This lead to the perception that transcription is not restricted to functional regions, like genes, but it is on the contrary pervasive. Different classes of non-coding transcripts, usually based on their length, have been described in metazoans and are summarised in Table 1. In yeast, three classes of ncRNAs have been described based on their RNA stability, including stable unannotated transcripts (SUTs (David et al., 2006)), cryptic unstable transcripts (CUTs (Wyers et al., 2005)) and Xrn1-sensitive unstable transcripts (XUTs (van Dijk et al., 2011)). Antisense transcripts, which are transcribed from the opposite strand to a coding transcript, are thought to be originated from the inherent bi-directionality of the promoters (Beck and Warren, 1988; Wang et al., 2011; Xu et al., 2009). These conclusions were drawn not only from the expression profiling of transcripts but also from the direct measure of transcription using, for *e.g.*, the NET-Seq (nascent elongating transcript sequencing)

methodology which maps elongating RNAPII (Churchman and Weissman, 2011). However, recent studies have identified a PIC or partial PIC (without RNAPII) in the promoters known to drive antisense transcription (Venters and Pugh, 2009). Furthermore, it was found that this occupancy of the transcriptional machinery in the antisense promoters was not just due to neighbouring promoters or transcription (Murray et al., 2011). The architecture of the *GAL10* internal promoter was found to contain the -1, NFR, +1 nucleosome configuration, arguing against a spurious initiation of transcription due to the availability of regions of the genome (Venters and Pugh, 2009).

Abbreviation	**Full name**	**Length (nt)**	**Transcript stability**	**Organism**	**Technology**
PARS	Promoter-associated short RNA	22 - 200	Stable	Human	Tiling array
TASR	Terminator-associated short RNA	22-200	Stable	Human	Tiling array
PALRs	Termination-associated short RNA	>1000	Stable	Human	Tiling array
sRNA	Short RNA	<200	Stable	Human	Tiling array
		<200			RNA Seq
		50 - 200			Tiling array

lRNA	Long RNA	>200	Stable	Human	Tiling array
tiRNA	Transcription initiation RNA	~18	Stable	Human, Chicken, *Drosophila*	RNA Seq
		12-27		Human, Chicken	RNA Seq
TSSa-RNA	Transcription start site-associated RNA	20-90	Stable	Mouse	RNA Seq
SUT	Stable unannotated transcript	~760	Stable	Yeast	Tiling array
PROMPT	Promoter upstream transcript	-	Unstable	Human	RNA Seq
CUT	Cryptic unstable transcript	~200-600	Unstable	Yeast	Tiling array, RNA Seq
XUT	Xrn1-sensitive transcript	-	Unstable	Yeast	RNA Seq

Table 1. Description of the non-coding transcripts annotated to date in various species and the methodology that led to their finding. Adapted from (Wei et al., 2011).

1.2.2 Functionality of ncRNAs

Even though, in general terms, the function of the ncRNAs is unknown there have been a few studies in yeast, which suggest that ncRNAs are involved in regulatory mechanisms that influence gene expression (Wei et al., 2011). The mechanisms known to date vary depending on the gene and are described below.

The gene *SER3* encodes for phosphoglycerate dehydrogenase and is repressed under rich media conditions. A non-coding transcript denominated as *SRG1* is produced upstream of the *SER3* promoter and represses transcription of the coding gene. The mechanism is suggested to be the *in cis* transcription through the promoter causing the Spt2-dependent deposition of nucleosomes which are maintained by the Paf1 complex, resulting in occlusion of the activators binding sites. A simplistic schematic of the mechanism is shown in Figure 3 (A) (Hainer et al., 2011; Martens et al., 2004; Martens et al., 2005; Pruneski et al., 2011; Thebault et al., 2011).

IME4 is a meiotic gene that becomes activated in starvation conditions, leading the cells into meiosis and sporulation. This gene is repressed in haploid cells by an antisense transcript, *RME2* (regulator of meiosis 2). In diploid cells *RME2* is repressed by the a1-α2 complex allowing activation of *IME4* expression (Hongay et al., 2006). It has been recently found that another meiotic gene, *ZIP2,* also produces a repressive antisense transcript (*RME3*) and whose mechanism seems similar to that of *IME4*. Both antisense transcripts work exclusively *in cis*, however it was found that it is not through an interference mechanism at the promoter. The exact mechanism is still unknown (Figure 3 B) (Gelfand et al., 2011).

PHO5, a phosphate metabolic gene, is repressed in high phosphate conditions by four well-positioned nucleosomes. Upon induction in low phosphate conditions, the four nucleosomes are evicted from the promoter by the binding of the activators, Pho4p and Pho2p. This remodelling of the promoter is achieved by the action of the remodelling complexes, SAGA, Swi/Snf, INO80 and Asf1 chaperone. Interestingly, it was found that an antisense transcript, produced in repressed conditions and degraded by Rrp6, facilitated the promoter remodelling achieved during induction. Furthermore, mutations repressing elongation have been shown to affect the kinetics of the promoter remodelling due to non-production of this ncRNA. This is therefore one example of an activating ncRNA, whose mechanism is simplified in Figure 3 (C) (Uhler et al., 2007).

PHO84, another phosphate metabolic gene, has the best characterized mechanism of ncRNA function. Two antisense transcripts, a long one reading through to the next ORF and a short one, the size of the *PHO84* ORF, are stabilized upon deletion of *RRP6*, resulting in repression of the promoter. It was found that in the absence of Rrp6p the histone deacetylase, Hda1, was recruited to the promoter. Interestingly, the *PHO84* expression was also reduced in aged cells, and further investigation revealed that Rrp6p is down-regulated, implicating that the antisense transcripts are stabilized in aged cells. Further investigation, also found that besides the *in cis* recruitment of Hda1/2/3 complex to the *PHO84* promoter, the antisense transcript could also act *in trans* by an Hda1-independent mechanism (Figure 3 D) (Camblong et al., 2009; Camblong et al., 2007a).

The *GAL10* gene contains an internal promoter, which has been shown to produce long antisense transcript in a Reb1-dependent way, in repressed conditions. Upon induction the expression of these transcripts is repressed. The antisense transcripts are suggested

to play a repressive role in the induction of both *GAL10* and *GAL1*, which share a bidirectional promoter. This repression has been suggested to be achieved by two independent mechanisms, although sharing the view that it is due to co-transcriptional chromatin remodelling. Firstly, a study by Houseley et al. has suggested that the expression of a long ncRNA deposited H3K36me3 by Set2, which then leads to the recruitment of the Rpd3(S) HDAC complex and the deacetylation of the promoter (Houseley et al., 2008). Secondly, Pinskaya et al. suggested that the ncRNAs co-transcriptionally deposition of H3K4me2/3 by Set1 recruited the Rpd3(S) to the promoter (Figure 3 E) (Pinskaya et al., 2009).

Even though a few mechanisms are known, the general function of ncRNAs is still largely unknown. There are several theories about the function of promoter-associated ncRNAs, either in a positively or negatively matter, including: i) Displacement of positioned nucleosomes (Taft et al., 2009); ii) Contributing towards a larger pool of RNAPII and transcriptional machinery near the TSS (Preker, 2008; Seila et al., 2009)); iii) Stimulating upstream initiation due to negative supercoiling of the DNA (Seila, 2008); iv) Preventing spreading of repressive chromatin marks (Seila et al., 2009); v) Maintaining an open conformation at the promoter to prevent stochastic variation of transcription (Tirosh and Barkai, 2008); vi) Removing promoter-bound transcription factors (Neil et al., 2009); vii) Competition for the same pool of transcription machinery (Neil et al., 2009).

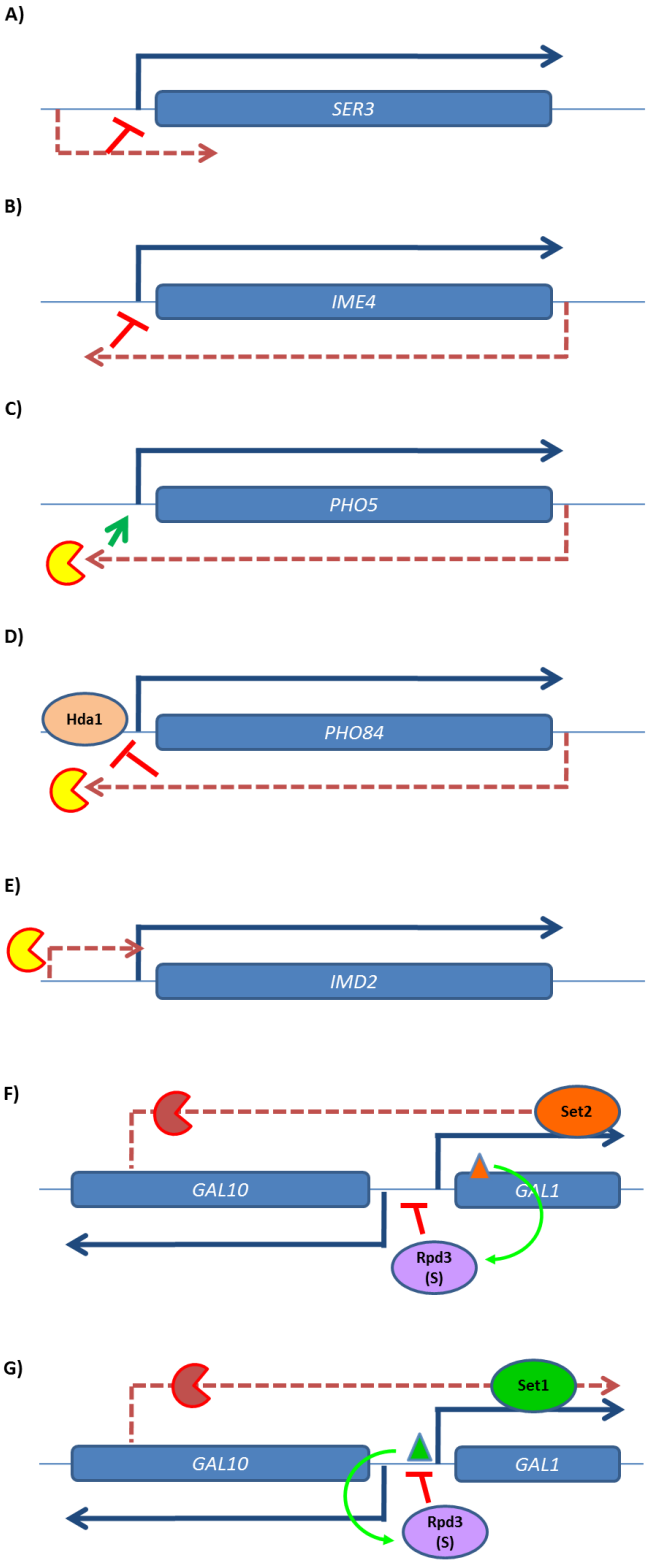


Figure 3. Mechanisms of action of ncRNAs in yeast. Schematics showing the mechanisms through each ncRNA represses or activates transcription. The genes are *SER3* (A), *IME4* (B), *PHO5* (C), *PHO84* (D), *IMD2* (E), *GAL10 –GAL1* (F). For description of each different mechanism see text. Blue arrows – sense transcript; Red dotted lines – ncRNAs; Green arrows – activation; Red T – repression; Yellow figure – Rrp6p; Dark red figure – Trf4p/Xrn1p; Orange triangle – H3K36me3; Green triangle – H4K3m2/3

1.3 The *Saccharomyces cerevisiae GAL* regulon

1.3.1 Overview of the *GAL* locus

One of the best characterized eukaryotic systems of transcriptional regulation is the *Saccharomyces cerevisiae GAL/MEL* regulon. The *GAL* genes encode the enzymes required for use of galactose as a carbon source, including those necessary for its transport into the cell and galactose catabolism *via* glycolysis. The *GAL* locus is tightly regulated at the transcriptional level by carbon source, being highly activated when galactose is present in the medium, inactive-repressed in glucose and inactive-poised in glycerol and raffinose (Bhat and Murthy, 2001; Lohr et al., 1995).

The *GAL/MEL* regulon comprises the structural genes *GAL2* (Permease), *GAL7* (galactose-1-phosphate uridyl transferase), *GAL10* (uridine diphosphoglucose epimerase), *GAL1* (galactokinase), *MEL1* (α-galactokinase) and the regulatory genes, *GAL3*, *GAL4* and *GAL80* (Bhat and Murthy, 2001; Lohr et al., 1995). *GAL7-10-1* are located as a cluster in chromosome II, with *GAL10-1* sharing a bidirectional promoter.

1.3.2 Regulation of the *GAL* locus

Galactose rapidly activates transcription of the structural *GAL* genes from an undetectable or basal level up to a high level of expression. On the other hand, in glucose, expression of these genes is not detectable and thus the mechanisms of *GAL* locus regulation must be very efficient and precise. Tight regulation of the locus depends on the close relationship between an activator (Gal4p), a repressor (Gal80p) and an inducer (Gal3p), all of which are encoded by the regulatory genes mentioned above (Platt and Reece, 1998; Traven et al., 2006).

1.3.3 Transcription activator: Gal4p

Gal4p is a transcriptional activator that binds to an upstream activating sequence (UAS_{GAL}) present in the promoter regions of both the structural and regulatory genes of the *GAL* locus. The UAS_{GAL} is a 17 bp consensus sequence (5'CGG-N11-CCG3') with a variable affinity for Gal4p. The number of UAS present in a promoter depends on the gene, for example, the *GAL10-1* promoter, has four UAS_{GAL} and *GAL7* has two while *GAL80* and *GAL3* have one each (Lohr et al., 1995). *GAL4* and *GAL80* (despite the presence of an UAS_{GAL}) are not regulated through GAL-specific mechanisms, but contain other DNA sequence elements at their promoters. The variable affinity of a UAS for Gal4p and their differing abundance at promoters, affects the fold induction of transcription, although it is not yet clear how this is regulated. In the case of *GAL10-1*, *GAL7* and *GAL80* promoters, chromatin structure might play a role since the UAS exist in a nucleosome-free region, thereby regulating the accessibility for Gal4p binding (Lohr et al., 1995; Traven et al., 2006). Other explanations might include the cooperative binding of Gal4p to the multiple sites or the presence of other regulatory elements in the vicinity (Traven et al., 2006).

Gal4 protein is a long polypeptide with 881 aminoacids and contains several functional domains including a $Zn_{(II)}Cys_6$ binuclear cluster DNA-binding domain, a linker domain, a dimerization domain and two acidic transcription activation domains (ARI and ARII). Gal4p binds to the UAS_{GAL} as a dimer, in which each $Zn_{(II)}Cys_6$ cluster interacts with the CGG triplets present at the UAS_{GAL} (Traven et al., 2006).

1.3.4 The relation between Gal4p, Gal80p and Gal3p

Expression of *GAL* genes is repressed in the presence of glucose, a phenomenon known as glucose repression, due to the binding of Gal80p to one of the transcription activation domains (ARII) of Gal4p, preventing its interaction with the transcription machinery and subsequent activation of transcription (Traven et al., 2006). However, studies with different mutants have shown that transcription activation and the binding of Gal80p are mutationally separable functions (Lohr et al., 1995). Gal80p inhibition is overcome through an interaction with the inducer Gal3p in response to galactose and ATP, allowing Gal4p to consequently activate transcription. Cells with *gal3* mutations display a severe phenotype known as long term adaptation (LTA) in which galactose induction takes 2-5 days in contrast to several minutes in wild type cells. In addition, cells defective in *GAL3* and any of the *GAL* loci (*GAL1-10-7*) genes cannot be induced (Bhat et al., 1990). The mechanism by which Gal3p interacts with Gal80p has not yet been clarified and two fairly different models have been proposed (Figure 4). One model suggests that Gal3p binds to the Gal4p-Gal80p complex, causing a conformational change in the Gal4p-Gal80p complex that allows transcription activation. Several observations favour this model, for example, the recent FRET (fluorescence resonance energy transfer) data showing that Gal4p and Gal80p stay associated in the presence of galactose (Bhaumik et al., 2004; Leuther and Johnston, 1992; Parthun and Jaehning, 1992; Platt and Reece, 1998; Sellick and Reece, 2005). On the other hand, other observations propose that Gal3p sequesters Gal80p in the cytoplasm, releasing Gal4p from the inhibitory effects of Gal80p. The data that supports this model shows that Gal3p is predominantly, if not exclusively, cytoplasmic (Peng and Hopper, 2000, 2002; Sil et al., 1999).

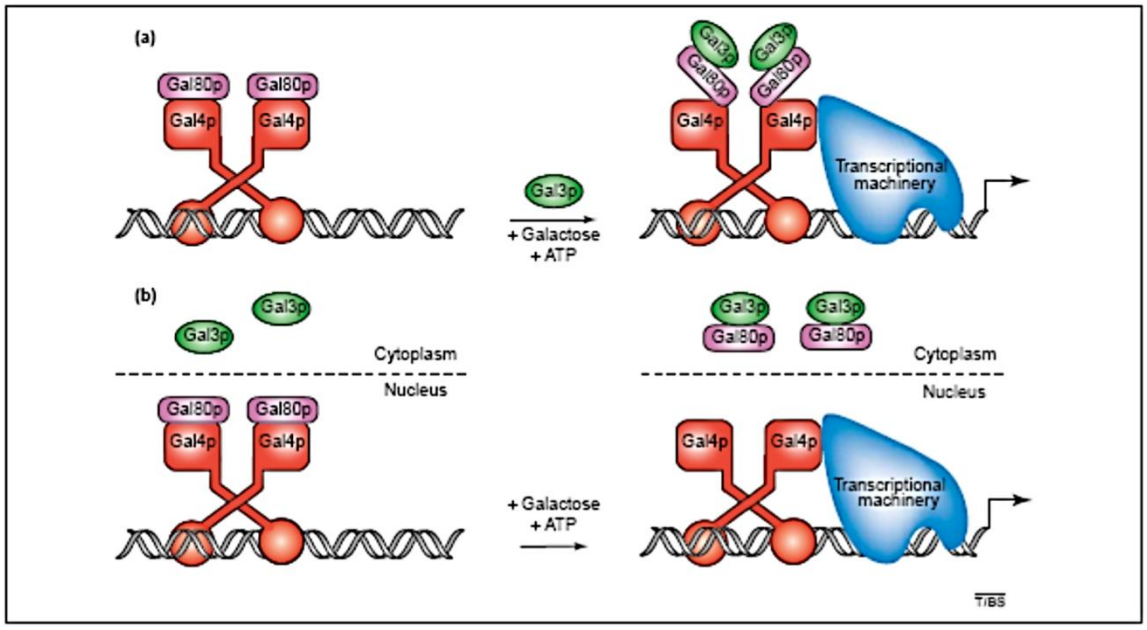


Figure 4. Models for the mechanism of Gal3p. (A) Non-dissociative - Gal3p in the presence of galactose and ATP migrates to the nucleus and binds to Gal80p changing its conformation with Gal4p which allows the binding of the transcriptional machinery. **(B)** Dissociative - Gal3 sequesters Gal80p in the cytoplasm in the presence of galactose and ATP causing a decrease in the levels of Gal80p in the nucleus therefore releasing Gal4p repression allowing its interaction with the transcriptional machinery. Reproduced from (Sellick and Reece, 2005).

1.3.5 Transcription activation by Gal4p

Gal4p activates transcription through its activation domain by recruiting the general transcription machinery and coactivators to the promoters (Traven et al., 2006). Although the target(s) of Gal4p are not yet clear, several proteins have been identified that interact with it including TBP, TFIIB, Gal11p (component of Mediator), Cdk8p (also known as Srb10), SWI/SNF, SAGA, Srb4p (component of Mediator), Sug1p and Sug2p (proteasome components) (Traven et al., 2006). Recent studies using FRET have shown that SAGA and Mediator complexes are recruited to the UAS_{GAL} while the pre-initiation complex (PIC) is loaded into the core promoter. The main target of Gal4p *in vivo* was identified as being Tra1, a subunit of the SAGA (Spt/Ada/Gcn5/acetyltransferase) complex, that recruits the complex to the UAS_{GAL} and subsequently recruiting Mediator, that depends on SAGA

presence and integrity (Bhaumik et al., 2004). However, some studies contradict the FRET results by showing that Mediator can interact with Gal4p in the absence of SAGA (Bryant and Ptashne, 2003; Larschan and Winston, 2005).

In the presence of galactose, Gal4p is released from Gal80p inhibition and recruits SAGA and Mediator that in turn recruits PIC to the core promoter and activates transcription.

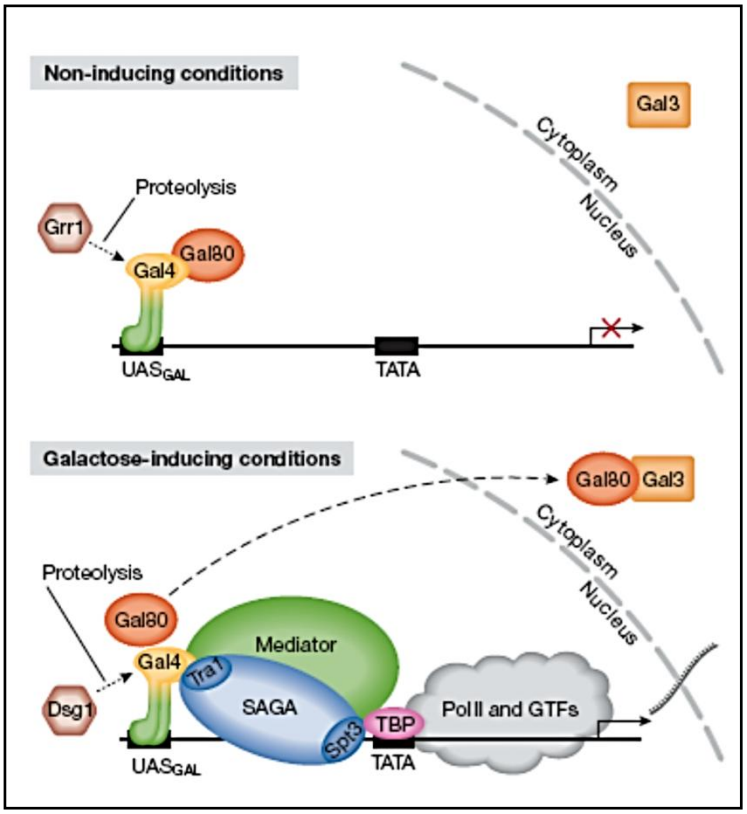


Figure 5. Transcription activation by Gal4p. In the absence of galactose, Gal80p is bond to Gal4p suppressing its interaction with the transcriptional machinery. Grr1 is responsible for regulating Gal4p levels. In the presence of galactose, Gal3p releases Gal80p repression (based on the dissociative model) allowing Gal4p to interact with SAGA and the Mediator which in turn will recruit the general transcription factors (GTFs) and Polymerase II (Pol II) initiates transcription. Dsg1 is responsible for regulating the turnover of transcriptionally active Gal4p. SAGA, Spt-Ada-Gcn5-acetyltransferase; TBP, TATA-binding protein. Reproduced from (Traven et al., 2006).

1.3.6 Glucose repression of the *GAL* locus

In the presence of glucose, the *GAL* locus is strongly repressed. The repression mechanisms are dependent on a global glucose repressor, Mig1p that binds to GC-rich upstream repression sequence (URS) present in the *GAL1*, *GAL3* and *GAL4* promoter. In addition, it can also bind to Gal4p, repressing the structural and regulatory *GAL* genes indirectly. The glucose repression mechanisms involve three described pathways. *GAL4* and *GAL3* expression are decreased by the binding of Mig1p to their URS. A decrease in Gal4p levels leads to reduced *GAL* gene transcription while less Gal3p leads to more Gal80p available to repress Gal4p. Furthermore, the binding of Mig1p to URS on *GAL1* promoter as well as to the Gal4p bound to the UAS_{GAL} increases the repression levels on the *GAL* locus.

1.4 Aims

The aims of this thesis are to identify and characterised the non-coding transcripts at the *GAL* locus in order to study how they influence galactose metabolism, particularly induction and repression of the structural genes. The role of non-coding RNAs, and other regulatory factors, in processes such as transcriptional memory and inducibility will be studied. RNA fluorescence *in situ* hybridization (FISH) will be used to determine the subcellular localization of non-coding and coding transcripts, particularly antisense transcripts. Finally, hybrid transcription units will be engineered at the *GAL* locus in order to study how antisense transcripts, initiated by either a bidirectional promoter or a terminator, influence the *GAL* promoters.

Chapter 2

Materials and Methods

2.1 Strains

All strains used in this study were derived from BY4741 (Euroscarf) unless otherwise stated (Table 2). Deletion strains were constructed as described previously using PCR-mediated deletion and modification of genomic Open Reading Frames with *KanMX*, *HisMX* or *TRP* selectable markers (Longtine et al., 1998). In short, DNA for yeast transformation was generated by PCR using primers with sequences homologous to the regions flanking the desired genomic insertion site. The following plasmids were used as template plasmid DNA: pFA6a-*KanMX6* (confers Kan resistance to transformants), pFA6a-*TRP1* and pFA6-*HisMX6* (confers tryptophan and histidine prototrophy to transformants respectively).

Marker-less disruptions were produced by Cre-lox recombination. Disruption cassettes consisting of a *LEU2* marker flanked by loxP recombination sites were amplified from plasmid pUG72 by PCR using primers with sequences homologous to the regions flanking the desired genomic insertion site (Gueldener et al., 2002). These DNA fragments were transformed into yeast. Positive transformants were selected and transformed with pSH62 (*his*$^{+}$), which contains the gene encoding Cre recombinase under the control of the *GAL1* promoter. Basal expression of the Cre recombinase from the leaky *GAL1* promoter drove excision of the *LEU2*marker. Kanamycin-sensitive, leucine auxotrophs were identified by replica plating and were sequenced. The residual 106 bp *loxP* signature sequence is shown below with the *loxP* element highlighted in grey.

5’-CAGCTGAAGCTTCGTACGCTGCAGGTCGACAACCCTTAATATAACTTCGTATAATGTATGCTAT ACGAAGTTATTAGGTGATATCAGATCCACTAGTGGCCTATGC-‘3

Strain	Parent	Genotype	Origin
BY4741		*MATa; his3Δ; leu2Δ; met15Δ; ura3*	Euroscarf
gal1Δ	BY4741	*gal1Δ::kanMX*	Euroscarf
gal10Δ	BY4741	*gal10Δ::kanMX*	Euroscarf
gal7Δ	BY4741	*gal7Δ::kanMX*	Euroscarf
gal3Δ	BY4741	*gal3Δ::kanMX*	Euroscarf
gal4Δ	BY4741	*gal4Δ::kanMX*	Euroscarf
gal1Δgal3Δ	BY4741	*gal1Δ::kanMX; gal3Δ::His5MX*	This study
pRS316	BY4741	pRS316	This study
pRS316 *ADH1*^{p}*-GAL1*	BY4741	pRS316 *ADH1*^{p}*-GAL1*	This study
pRS316 *ADH1*^{p}*-GAL1*_FS	BY4741	pRS316 *ADH1*^{p}*-GAL1*_FS	This study
gal10Δ pRS316	BY4741	*gal10Δ::kanMX* pRS316	This study
gal10Δ pRS316 *ADH1*^{p}*-GAL1*	BY4741	*gal10Δ::kanMX* pRS316 *ADH1*^{p}*-GAL1*	This study
gal1Δ pRS316	BY4741	*gal1Δ::kanMX* pRS316	This study

gal1Δ pRS316 *ADH1*^{p}*-GAL1*	BY4741	*gal1Δ::kanMX* pRS316 *ADH1*^{p}*-GAL1*	This study
gal1Δ pRS316 *ADH1*^{p}*-GAL1*_FS	BY4741	*gal1Δ::kanMX* pRS316 *ADH1*^{p}*-GAL1*_FS	This study
gal3Δ pRS316	BY4741	*gal3Δ::kanMX* pRS316	This study
gal3Δ pRS316 *ADH1*^{p}*-GAL1*	BY4741	*gal3Δ::kanMX* pRS316 *ADH1*^{p}*-GAL1*	This study
gal3Δ pRS316 *ADH1*^{p}*-GAL1*_FS	BY4741	*gal3Δ::kanMX* pRS316 *ADH1*^{p}*-GAL1*_FS	This study
gal1Δgal3Δ pRS316	BY4741	*gal1Δ::kanMX*; *gal3Δ::His5MX* pRS316	This study
gal1Δgal3Δ pRS316 *ADH1*^{p}-GAL1	BY4741	*gal1Δ::kanMX*; *gal3Δ::His5MX* pRS316 *ADH1*^{p}*-GAL1*	This study
gal1Δgal3Δ pRS316 *ADH1*^{p}*-GAL1*_FS	BY4741	*gal1Δ::kanMX; gal3Δ::His5MX* pRS316 *ADH1*^{p}*-GAL1*_FS	This study
rrp6Δ	BY4741	*rrp6Δ::kanMX*	Euroscarf
kem1Δ	BY4741	*kem1Δ::kanMX*	Euroscarf
G1 *TEF*^{P}	BY4741	*GAL1*$_{(+757\ bp)}$ *AgTEF*^{P}*:His3MX:AgTEF*^{T}	This study
G1 *ADH1*^{T}	BY4741	*GAL1*$_{(+757\ bp)}$ *ScADH1*^{T}*:AgTEF*^{P}*:His3MX:AgTEF*^{T}	This study
G1 *3' END TEF*^{P}	BY4741	*GAL1*$_{(3'END)}$ *AgTEF*^{P}*:His3MX:AgTEF*^{T}	This study
G1 *3' END ADH1*^{T}	BY4741	*GAL1*$_{(3'END)}$ *ScADH1*^{T}*:AgTEF*^{P}*:His3MX:AgTEF*^{T}	This study

G1 *IG TEF^{P}*	BY4741	*GAL1$_{(IG)}$ AgTEF^{P}:His3MX:AgTEF^{T}*	This study
G1 *IG ADH1^{T}*	BY4741	*GAL1$_{(IG)}$ ScADH1^{T}:AgTEF^{P}:His3MX:AgTEF^{T}*	This study
G1 *TEF^{P} rrp6Δ*	BY4741	*GAL1$_{(+757\ bp)}$ AgTEF^{P}:His3MX:AgTEF^{T}; rrp6Δ::KanMX*	This study
G1 *ADH1^{T}rrp6Δ*	BY4741	*GAL1$_{(+757\ bp)}$ ScADH1^{T}:AgTEF^{P}:His3MX:AgTEF^{T};*	This study
G1 *3' END TEF^{P} rrp6Δ*	BY4741	*GAL1$_{(3'END)}$ AgTEF^{P}:His3MX:AgTEF^{T}; rrp6Δ::KanMX*	This study
G1 *3' END ADH1^{T} rrp6Δ*	BY4741	*GAL1$_{(3'END)}$ ScADH1^{T}:AgTEF^{P}:His3MX:AgTEF^{T};*	This study
G1 *IG TEF^{P} rrp6Δ*	BY4741	*GAL1$_{(IG)}$ AgTEF^{P}:His3MX:AgTEF^{T}; rrp6Δ::KanMX*	This study
G1 *IG ADH1^{T} rrp6Δ*	BY4741	*GAL1$_{(IG)}$ ScADH1^{T}:AgTEF^{P}:His3MX:AgTEF^{T}; rrp6Δ::KanMX*	This study
G10 *TEF^{P}*	BY4741	*GAL10$_{(+1453\text{-}2007\ bp)}$ AgTEF^{P}:KanMX:AgTEF^{T}*	Youdell, M.
G10 *ADH1^{T}*	BY4741	*GAL10$_{(+1453\text{-}2007\ bp)}$ ScADH1^{T}:AgTEF^{P}:KanMX:AgTEF^{T}*	Youdell, M.
G10 *ADH1^{T}* #1	BY4741	*GAL10$_{(+1453\text{-}2007\ bp)}$ ScADH1^{T}#1:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.
G10 *ADH1^{T}* #2	BY4741	*GAL10$_{(+1453\text{-}2007\ bp)}$ ScADH1^{T}#2:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.
G10 *ADH1^{T}* #3	BY4741	*GAL10$_{(+1453\text{-}2007\ bp)}$ ScADH1^{T}#3:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.
G10 *ADH1^{T}* #4	BY4741	*GAL10$_{(+1453\text{-}2007\ bp)}$ ScADH1^{T}#4:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.

G10 *ADH1^{T}* #5	BY4741	*GAL10*$_{(+1453\text{-}2007\ bp)}$ *ScADH1^{T}#5:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.
G10 *ADH1^{T}* #6	BY4741	*GAL10*$_{(+1453\text{-}2007\ bp)}$ *ScADH1^{T}#6:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.
G10 *ADH1^{T}* #7	BY4741	*GAL10*$_{(+1453\text{-}2007\ bp)}$ *ScADH1^{T}#7:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.
G10 *ADH1^{T}* #8	BY4741	*GAL10*$_{(+1453\text{-}2007\ bp)}$ *ScADH1^{T}#8:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.
G10 *ADH1^{T}* #9	BY4741	*GAL10*$_{(+1453\text{-}2007\ bp)}$ *ScADH1^{T}#9:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.
G10 *ADH1^{T}* #10	BY4741	*GAL10*$_{(+1453\text{-}2007\ bp)}$ *ScADH1^{T}#10:AgTEF^{P}:KanMX:AgTEF^{T}*	Murray, S.
gal10-3'Δ40bpADH1$_{T}$	BY4741	gal10-3'Δ::40bpADH1$_{T}$KanMX	Murray, S.
gal10-3'Δ40XADH1$_{T}$	BY4741	gal10-3'Δ::40XADH1$_{T}$KanMX	Murray, S.
G1 *ADH1^{T}* #4	BY4741	*GAL1*$_{(+757\ bp)}$ *ScADH1^{T}#4:AgTEF^{P}:His3MX:AgTEF^{T}*	This study
G1 *ADH1^{T}* #5	BY4741	*GAL1*$_{(+757\ bp)}$ *ScADH1^{T}#5:AgTEF^{P}:His3MX:AgTEF^{T}*	This study
FT4		*MATa, ura3-52, trp1-Δ63, his3-Δ200, leu2::PET56*	Vogelauer, M.
Reb1_BSΔ	FT4	*gal10-3'Δ::pMV12* (EcoRI/Xho1 - Reb1 BSΔ)	Vogelauer, M.
gal3Δ	FT4	*gal3Δ::kanMX*	This study
Reb1_BSΔ*gal3Δ*	FT4	*gal10-3'Δ::pMV12* (EcoRI/Xho1 - Reb1 BSΔ) ; *gal3Δ::KanMX*	This study

gal4Δ	FT4	*gal4Δ::kanMX*	This study
Reb1_BSΔ*gal4Δ*	FT4	*gal10-3'Δ::pMV12* (EcoRI/Xho1 - Reb1 BSΔ) ; *gal4Δ::KanMX*	This study

Table 2. Strain genotypes.

2.2 Media Composition

Minimal media was based on complete supplement mixture (Qbiogene) plus 2 % glucose, plus 2 % Difco yeast nitrogen base (BD and Co.). Rich medium (YP) was 1% bactopeptone, 1% Difco yeast extract (BD and Co.) supplemented with 2% glucose (for YPD) or 2% galactose (for YPG).

2.3 Yeast culture

All strains were inoculated overnight in 3-5 ml YPD unless otherwise stated. Cells were diluted to 0.2 OD_{600} in 2% YPD or 2% YPG in the desired final volume and allowed to grow to an OD_{600} of 0.4-0.6. For galactose induction, the cells were then washed with water, added to fresh 2% YPG, and harvested at the time points indicated. Induction to steady-state indicates a 3 hour growth in galactose-containing medium.

2.4 Transformation of *S. cerevisiae*

Transformation was performed using the standard lithium acetate protocol (Ito et al., 1983). Cells were grown to log phase in YPD media and harvested by centrifugation. Cells were washed in H_20 and competence was induced by resuspension in 0.1 M LiAc/TE. Transformation was performed by mixing 100 µl of the competent cell suspension with 700 µl 40% PEG/0.1 M LiAc/TE, 10 µl calf thymus DNA and 10 µl of the PCR product to be

transformed. Cells were incubated at 30^{o}C for 30 minutes and subjected to heat shock at 42^{o}C for 20 minutes. For higher transformation rates, 40 μl of DMSO was added prior to heat shock. Cells were pelleted gently, resuspended in H_20 and plated onto appropriate media to allow for selection. Transformants were confirmed by PCR analysis of genomic DNA preparations from several colonies.

2.5 Preparation of genomic DNA from *S. cerevisiae*

Cells were allowed to grow in YPD media for a minimum of 12 hours, pelleted by centrifugation at 13,000 rpm and resuspended in 200 μl of extraction buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-Cl pH 8.0, 1 mM EDTA pH 8.0), 200 μl acid washed glass beads and 200 μl of 4:1 phenol:chloroform. Following vigorous agitation by vortexing for 20 minutes, and centrifugation at 13,000 rpm at 4°C for 15 minutes, the upper aqueous phase was transferred into a new eppendorf. The DNA was then precipitated with 500 μl ethanol:ammonium acetate solution (6 volumes ethanol : 1 volume 7.5 mM ammonium acetate) and centrifugation at 13,000 rpm at 4°C for 20 minutes followed by resuspension of the pellet in 100 μl H_20.

2.6 Chromatin Immunoprecipitation (ChIP)

Chromatin Immunoprecipitation was performed based on the method previously described by Morillon *et al*. (Morillon et al., 2003). Cells were grown in 50 ml YPD to a density of ~1.5 x 10^7 cells/ml followed by cross linking in 1X PBS for 30 minutes by the addition of formaldehyde (1 % final concentration) (Fisher chemicals). Cross linking was stopped by the addition of 2.5 ml 2.5 M glycine and cells pelleted and washed twice with ice-cold 1X PBS. Cells were then resuspended in 500 μl FA buffer (0.1 % SDS, 1% Triton-X-

100, 10 mM Hepes, 0.1 % sodium deoxycholate, 150 mM NaCl, 1 mM AEBSF, 1 EDTA-free proteinase inhibitor tablet per 10 ml). Lysis of the yeast cells was performed by grinding with acid washed glass beads in a MagnaLyser (Roche) twice for 1 minute. Fixed chromatin was sonicated using a bioruptor (Cosmo Bio) for 30 minutes at 4^{O}C (1 minute pulses with 20 second intervals) to a fragment size of less than 500 bp. Immunoprecipitations were performed by incubation of a one in ten dilution of 100 µl of fixed chromatin sample with the relevant antibody (Table 2). Purification of the immunoprecipitated protein/DNA complexes was achieved by adding 40 µl 50 % slurry protein A-sepharose beads (CL-4B Amersham) for a minimum of 90 minutes at room temperature. Following binding, beads were washed with TSE-150/500 (1 % Triton-X-100, 0.1% SDS, 2 mM EDTA, 20 mM Tris-Cl pH 8.0, 150/500 mM NaCl) for 3 minutes each, LiCl wash (0.25 M LiCl, 1 % NP-40, 1 % sodium dioxycholate, 1 mM EDTA, 10Mm Tris-Cl pH 8.0) and TE for 15 minutes each. Elution of bound protein/DNA was achieved by 30 minute incubation at 65°C with 500 µl elution buffer (1 % SDS, 0.1 M $NaHCO_3$) pre-warmed to that temperature. Crosslinking was reversed by incubation of a minimum of 3 hours at 65°C after the addition of 7 µl of 5 M NaCL. This was followed by treatment with 1 µl of RNaseA and 1µl proteinase K (Roche) prior to purification with QIAquick PCR purification columns (Qiagen).

Epitope	Volume used (µl)	Supplier
Rpb1 (y80)	50	Santa Cruz

Table 3. Antibody used for ChIP

DNA was quantified using Real Time PCR (RT-PCR) with the Rotorgene-3000 system (Corbett Robotics). PCR was performed using Sensimix (Quantance SYBR green). Primer positions are shown in Table 4. Quantitation of the DNA was calculated based on a standard curve generated from serially diluted sonicated yeast genomic DNA. % Input was calculated form three replicates of IP, control IP (no antibody) and input DNA using the following calculation;

[(Σ IP signals / n IP samples) – (Σ No Ab signals / n No Ab samples)] /(Σ input signals / n input samples)

Gene	PCR product	Sequence of forward and reverse
GAL1 P -139 (from *GAL1* ATG)	1	5'-TGGAAAAGCTGCATAACCACT-3'
		5'-ACCTGAGTTCAATTCTAGCGC-3'
GAL1 +133 (from *GAL1* ATG)	2	5'-CGCTTATGATGCTAAACCGG-3'
		5'-TGATTTTGATATGCTTTGCG-3'
GAL1 +238 (from *GAL1* ATG)	3	5'-CCGTCAAAGTTTTGAACGAGA-3'
		5'-CTTATGTCACAATTGATCCTTCTG-3'
GAL1 +596 (from *GAL1* ATG)	4	5'-GCGTATTACGGTCGTTGCAG-3'
		5'-GAGTTCAAACCGCAGTTGAA-3'
GAL10 +176 (from *GAL10* ATG)	5	5'-CCGAATCAATTTTATATTCTT-3'
		5'-CCGAATCAATTTTATATTCTT-3'

Table 4. Primer pairs used for qPCR.

2.7 Extraction of RNA from yeast cells

Extraction of total RNA was performed using hot phenol extraction. Cells were resuspended in 400 µl TES (10 mM Tris-Cl pH 7.5, 5 mM EDTA pH 8.0, 1 % SDS) and an equivalent volume of acid phenol pH 4.5 (Qbiogene) and incubated for 20 minutes at 65^{o}C with shaking. After an incubation at -80°C for 30 minutes and centrifugation at 13,000 rpm for 15 minutes at room temperature, the aqueous phase was recovered. The RNA was then precipitated by the addition of ethanol containing 10 mM sodium acetate and incubated for 10 minutes at -80^{o}C. The RNA was pelleted by centrifugation at 13,000 rpm at 4°C and resuspended in RNase free water. Poly(A)+ isolation from total RNA performed following the batch protocol of the Oligotex mRNA Kit (Qiagen). The Poly(A)- fraction was obtained by concentrating the washes of the poly(A)+ isolation with the RNeasy Kit (Qiagen).

2.8 Northern blotting

15 µg of RNA was separated on 1.1% formaldehyde FA gels for about 2 hours and transferred to Magma nylon membranes by wet blotting overnight in 20X SSC (Maniatis and Fristish, 1982). After fixing the RNA to the membrane by baking for 2 hours at 80^{o}C, the membranes were blocked for 2 hours in PerfectHyb Plus (Sigma) at 65^{o}C. Radio-labelled probe was then added and hybridised overnight at 65^{o}C, after which non-specifically bound probe was removed by washing the membranes twice in 1X SSC/0.1 % SDS and once in 0.2X SSC/0.1% SDS and 0.05X SSC/0.1% SDS for 20 minutes each at 65^{o}C. Membranes were typically exposed to photographic film for between 1 hour and 1 week. Levels of total RNA loaded were monitored by the levels of rRNA species by ethidium

bromide staining. To note that all Northern blotting experiments in this thesis were biologically repeated, at least, three separate times.

2.9 Generation of strand-specific radiolabelled RNA probes by *in vitro* transcription

The product of interest was amplified by PCR with strand-specific primers containing the unidirectional T7 promoter sequence at 5' end of the strand of interest. The probes were then produced *in vitro* with recombinant T7 polymerase and radioactive UTP for 1 hour at 37^{O}C. 1 µl of DNase was added for 15 min at 37^{O}C. Probes were then purified using Sephadex G-50. Primers used for generation of probes are shown in Table 5.

Primers	Position	Sequence
GAL7 IVT F	+115 (from *GAL7* ATG)	5'-TTCAGCAGCTTGTTCCGAAG-3'
GAL7 IVT R	+1053 (from *GAL7* ATG)	5'-AGGAGGCTGCTTACAAGCCC-3'
GAL10/7 IG IVT F	+2250 (from *GAL10* ATG)	5'-CATGTGAGGAAATTCGCTGT-3'
GAL10/7 IG IVT R	+2769 (from *GAL10* ATG)	5'-GCCCATAAATATTTCCGACC-3'
GAL10/1 IG IVT F	-131 (from *GAL10* ATG)	5'-AGAGCCCCATTATCTTAGCC-3'
GAL10/1 IG IVT R	-474 (from *GAL10* ATG)	5'-CTGGGGTAATTAATCAGCGA-3'
GAL10 F	+44 (from *GAL10* ATG)	5'-ATCCAGCACCACCTGTAACC-3'
GAL10 R	+2058 (from *GAL10* ATG)	5'-TTGGACCCGTAAGTTTCACC-3'

Table 5. Primers used for *in vitro* transcription generation of strand-specific probes.

2.10 Generation of strand-specific radiolabelled DNA probes by asymmetric PCR

The product of interest was amplified by PCR. A second round of PCR with only a forward or reverse primer, depending on the strand of interest, was performed with radioactive dCTP. The probe was then purified using Sephadex G-50.

Asymmetric Primers	Position	Sequence
GAL10 F	+171	5'-CTTGACCAAGCATCACATTCC-3'
GAL10 R	+1145	5'-CCAGATTTCAAGCCACGTTTG-3'
GAL1 5' F	+356	5'-TGTGTCGGACTGGTCTAATT-3'
GAL1 5' R	+651	5'-TAACAATGGCGGTATGGATC-3'
GAL1 3' F	+945	5'-CGTTTATTATGCCAGATATC-3'
GAL1 3' R	+1198	5'-GAATCTTTAAGAGTCTTGAA-3'

Table 6. Primers used for asymmetric PCR generation of strand-specific probes.

2.11 Tiling array

The procedure for sample preparation for tiling array hybridization was performed as described in Perocchi et al. (Perocchi et al., 2007). In summary, wild-type cells were grown to an OD_{600} of 0.6 in 2% YPD, washed and then inoculated in 2% YPG for 3 hours before harvesting. Total RNA was extracted and poly(A) RNA isolated using the method described in 2.7. The poly(A) sample was treated with RNase-free DNaseI using the Turbo DNA-free kit (Ambion). For first-strand cDNA synthesis, 20 µg of total RNA or 9 µg of poly(A) RNA was mixed with 4.5 µg of random hexamers and incubated at 70°C for 10

min, then transferred to ice. The synthesis included 2000 U of SuperScript II Reverse Transcriptase, 50 mM Tris–Cl, 75 mM KCl, 3 mM MgCl2, 0.01 M DTT, 0.25 mM dNTP mix (Invitrogen), 20 µg/ml ActinomycinD in a total volume of 200 µl at 42°C for 1 h. Samples were then subjected to RNase H treatment for 20 min at 37°C (30 U RNase H, Epicentre, 60 U of RNase Cocktail, Ambion). First-strand cDNA was purified with Affymetrix cDNA cleanup columns and eluted with 30 µl DEPC water. 5.5 µg of cDNA was digested with 10 U/µl of Uracil-DNA Glycosylase (UDG) and 1000 U/µl of apurinic/apyrimidinic endonuclease (APE) in a total volume of 50 µl at 37°C for 1 h to yield fragments of 50–100 bp in size. Each sample was 3′ end-labelled with 0.07 mM Biotin-N6-ddATP (Enzo Life Sciences) using 400 U of Terminal Transferase (Roche) for 1 h at 37°C. For hybridization to the arrays, 4.5 µg of the cDNA was denatured in a solution containing 100 mM MES, 1 M NaCl, 20 mM EDTA pH 8.0, 0.01% Tween-20, 50 pM control oligonucleotide B2 (Affymetrix), 0.1 mg/ml herring sperm DNA, and 0.5 mg/ml BSA in a total volume of 300 µl, from which 220 µl was hybridized per array. Hybridizations were carried out at 45°C for 16 h with 60 r.p.m. rotation.

2.12 RNA Fluorescence *In Situ* Hybridization (FISH)

The RNA FISH protocol is adapted from Zenklusen et al. (Zenklusen and Singer, 2010). In short, DNA probes of ~50nt and ~50% GC content were designed and five modified bases (amino-allyl dT) spaced by about ~10nt, for the incorporation of the fluorophore, were added. For the labelling of the probes, a total of 5 µg was purified using the QIAquick Nucleotide Removal Kit (Qiagen) and eluted with 40 µl of H_2O. The probes were then lyophilized in a SpeedVac, resuspended in 10 µl of 0.1 M sodium bicarbonate pH 9.0 and added to the dye-containing tube (CyDye™ GE Healthcare, Cy3, Cy5 or Cy3.5). The tube

was vortexed vigorously followed by a quick spin. The reaction was incubated overnight at room temperature with low speed shaking. The probes were purified using the QIAquick Nucleotide Removal Kit (Qiagen) and eluted with 100 µl of elution buffer. The concentration and efficiency of the labelling was measured using a spectrophotometer. Probes were stored in the dark at -20°C. The labelling efficiency was calculated as described in Zenklusen et al. (Zenklusen and Singer, 2010). The cells were grown and induced as described in 2.3, cross-linked with 4% (v/v) of paraformaldehyde for 45 minutes at room temperature and spheroplasted in 25 U/ml of lyticase enzyme, 1.2 M sorbitol, 100 mM $KHPO_4$ pH 7.5, 20 mM ribonucleoside-vanadyl complex (VRC) and 20 mM β-mercaptoethanol until >80% of digestion was achieved. Cells were then washed with 1 ml 1.2 M sorbitol and 100 mM $KHPO_4$ pH 7.5, before resuspension in 1.5 ml 1.2 M sorbitol, 100 mM $KHPO_4$ pH 7.5, 20 mM VRC and 150 µl of cells attached to poly-L-lysine-treated coverslips for 30 minutes at 4°C. The coverslips were then washed in 2 ml of 1.2 M sorbitol, 100 mM $KHPO_4$ pH 7.5, 20 mM VRC, 2 ml of 70% ethanol for at least 3 h at -20°C. Samples were rehydrated twice with 2 ml of 2X SSC for 5 minutes at room temperature and washed with 40% formamide in 2X SSC. For the hybridization, 0.5 ng of each probe, 10 µg *E. coli* tRNA, 10 µg salmon sperm DNA were mixed and lyophilized in a SpeedVac. 12 µl of 40% formamide, 2X SSC, $NaHPO_4$ pH 7.5 was added and the probes were denatured at 95°C for 3 minutes followed by the addition of 12 µl of 2X SSC, 2 mg/ml BSA, 10 mM VRC. Hybridization was done overnight at 37°C in a parafilm-sealed chamber, where the coverslips with the cells facing down were placed onto 22 µl of the hybridization mixture. The coverslips were then subjected to a series of washes, twice at 37°C for 15 minutes with 40% formamide in 2X SSC, once at room temperature for 15 minutes with 2X SSC, 0.1% Triton X-100, once at room temperature for 15 minutes with

1X SSC, once for 2 minutes with 0.5 µg/ml DAPI in 1X PBS and once for 2 minutes with 1X PBS. The coverslips were then dipped into 100% ethanol and left to dry followed by the mounting onto a microscope slide (cells facing down), polymerized overnight, sealed with nail polish and imaged. The imaging was performed in collaboration with Ilan Davis' laboratory, using a quantitative microscope (DeltaVision CORE: Wide-field fluorescence deconvolution imaging). The quantitation of the RNA FISH raw data was achieved using the ImageJ software. For each intact cell, the plane with the maximum signal in the channels used was selected and maximum, mean and integrated signals measured and recorded (alternatively, a selection of slices (planes 10-25) which included the nucleus and most of the cytoplasm was maximally projected or signals summed up to retrieve the respective parameters (max signals from maximal projections, means and integrated values from summed up slices)). As negative controls, non-induced or knockout cells and/or no-probe controls were used. The error bars were calculated from standard deviations using the T-test and a confidence interval of 0.01 or 0.05 as indicated.

Chapter 3

Non-coding transcripts at the GAL locus

3.1 Introduction

The recent development of several high-throughput techniques such as high resolution tilling arrays, RNA-sequencing (RNA-Seq), cDNA library sequencing and SAGE (serial analysis of gene expression) has provided researchers with the tools, for the first time, to have a better understanding of cellular transcription and transcripts (Churchman and Weissman, 2011; David et al., 2006; Katayama et al., 2005; Perocchi et al., 2007; Xu et al., 2009; Yassour et al., 2010). This has led to the unexpected finding that there is abundant transcription genome-wide and the total number of transcripts identified exceeds, by a large fold, the number of protein-coding genes (David et al., 2006; Ito et al., 2008; Katayama et al., 2005). The transcripts were mapped to all sites of the genome, including intergenic regions, heterochromatin domains and previously unannotated regions. These studies have also shown that non-coding RNAs (ncRNAs) can be transcribed on both strands, in relation to a coding gene, and can range from only a few base pairs to several kilobases long. A transcript that is transcribed in the same orientation as an open reading frame (ORF) is referred to as a sense RNA (S) whereas if transcribed in the opposite orientation is referred to as an antisense RNA (aRNA or AS). When a gene has both sense and antisense transcripts of similar sizes, starting and/or ending in its promoter and terminator regions, it is referred to as a SAP (sense-antisense pair) (Xu et al., 2009). Several classes of ncRNAs have been identified in various organisms including snRNA (small nuclear RNA), snoRNA (small nucleolar RNA), miRNA (micro RNA), siRNA (small-interfering RNA), piRNA (piwi-interacting RNA) and lncRNA (long non-coding RNA). They can also be classified in relation to their stability, for example, SUTs (stable unannotated transcripts), CUTs (cryptic unstable transcripts) and XUTs (Xrn1 unstable transcripts)

(Berretta and Morillon, 2009; Tisseur et al., 2011). There is still controversy as to whether this pervasive transcription is just noise, a by-product of transcription itself or whether it is a regulated process with a functional outcome. There have been a few studies so far, in *S. cerevisiae*, which suggest that ncRNAs are involved in regulatory mechanisms that influence gene expression. These are described in detail in chapter 1.

The genes encoding for the enzymes responsible for the catabolism of galactose into glucose-6-phosphate are *GAL7* (UDP-glucose-4-epimerase), *GAL10* (galactose mutarose) and *GAL1* (galactokinase). These genes are located in a cluster on chromosome II that is referred to, in this study, as the *GAL* locus (Figure 6). Galactose activates transcription of these genes to very high levels, while glucose quickly represses expression. Tight regulation of the locus depends on the close relationship between an activator (Gal4p), a repressor (Gal80p) and an inducer (Gal3p). In the absence of galactose, Gal80p is bound to the activator domain in Gal4p, consequently repressing the main activator. When galactose is added to the medium, it is detected and possibly bound to by Gal3p, which in turn binds Gal80p. This Gal80p-Gal3p association relieves the repressive effect that Gal80p has on the Gal4p activator domain, which can now bind to the UAS_{GAL} (Upstream Activating Sequence) present in the promoters of the GAL genes and recruit the PIC (Pre-initiation complex), initiating transcription. However, there are still many uncertainties regarding this model, including whether Gal4p remains bound to the UAS_{GAL} in the presence of glucose, whether Gal3p is exclusively cytoplasmic or if it is also nuclear, possibly shuttling to the nucleus to bind Gal80p, and whether Gal3p binds galactose directly or a galactose by-product.

Figure 6. Schematic of the *GAL* locus organization. The three genes *GAL7*, *10* and *1* are localized from position 274427 to 280607 on chromosome II. The orientation of the genes is given by the arrows. The green bars identify the Gal4 binding sites or UAS_{GAL} (Upstream Activating Sequence). The light blue box refers to the *GAL10* internal promoter and the yellow lines represent the Reb1p binding sites.

3.2 Several non-coding RNAs are present at the *GAL* locus in the induced and repressed states

Xu et al. published a *S. cerevisiae* data set using their recently developed/updated high-resolution tiling array methodology that discriminates transcription on the two DNA strands (Xu et al., 2009). This study, using poly (A)+ enriched RNA, identified a large number of stable (SUTs) and unstable (CUTs) non-coding transcripts genome-wide. Examining the data set for the *GAL* locus, where the colour blue indicates hybridization and light yellow indicates no hybridization, it is observable that a SUT (SUT013), antisense to *GAL10*, and a CUT (CUT445), antisense to *GAL1*, were detectable in repressive conditions, such as YPD with 2% glucose (Figure 7). Curiously, the *GAL1* antisense identified here as a CUT, not only can be observed in the other growth conditions tested, but it also has been identified in a separate study as a XUT (van Dijk et al., 2011).

As expected, the coding transcripts for *GAL7*, *10* and *1* are only observed in the induced YPGal condition, to very high levels, identifiable by the very dark blue colour, which means hybridization of a high amount of transcripts. However, low levels of the *GAL7* sense transcript can also be observed in non-induced/repressive conditions. Interestingly,

a strong hybridization signal is clearly visible in the intergenic regions between *GAL1* - *FUR4* and *GAL10* – *GAL7*. This is not entirely unexpected, since a *GAL10-GAL7* read-through transcript has been described (Greger and Proudfoot, 1998), and therefore, it is possible that the *GAL1-FUR4* intergenic signal is also due to read-through transcription, and indeed it will be confirmed later in this work. Furthermore, there are more regions of hybridization, implicating the presence of additional transcripts, although these have not been annotated.

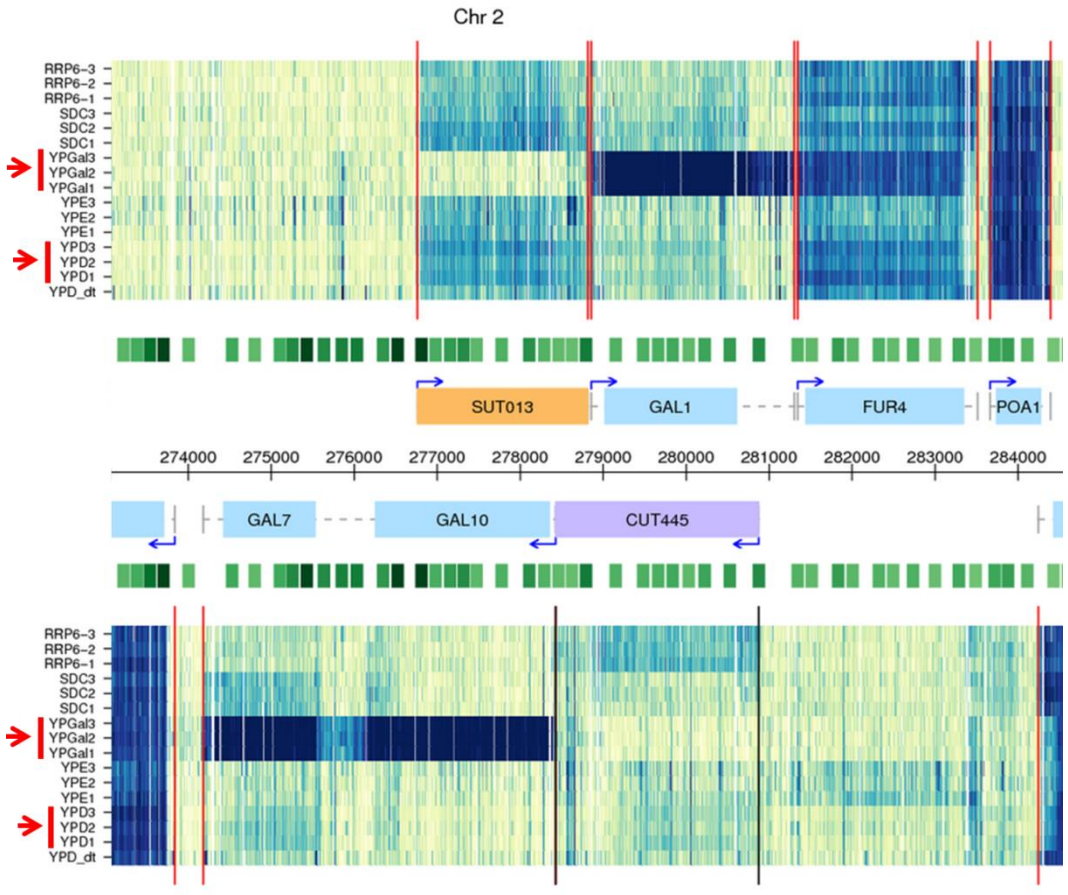


Figure 7. Tiling array of poly(A)+ RNA performed by the Steinmetz laboratory (Xu et al., 2009). Data acquired from http://steinmetzlab.embl.de/cgi-bin/viewNFRsharing.pl?showSamples=NFRsharing&gene=GAL10. The figure shows the annotated ORFs (blue boxes) in the middle either above or below the scale depending on the orientation of the gene. A CUT (blue/purple) and a SUT (orange) are also represented. The rows of data at the top and bottom of the figure represent each strand, *e.g.* for *GAL1* the top data set is the sense strand and the bottom the antisense strand. The colour code is green/yellow for low or no hybridization and dark for high. The green boxes represent the nucleosome positioning (Mavrich et al., 2008). The red arrows indicate the significant rows for comparison (YPGal and YPD).

However, careful examination of the growth conditions used by the Steinmetz group, revealed the fact that the cells were being grown to much higher densities (0.8 - 1 OD_{600}) than the ones used in this study (0.4 - 0.6 OD_{600}) (section 2.3), raising the concern that in the Steinmetz study the cells could be entering a different growth phase known as the diauxic shift. When cells are grown in batch culture, they will preferentially metabolize the sugar in which they can grow faster (usually glucose), by glycolysis, releasing ethanol into the medium. Once they exhaust the primary carbon source they will enter a different growth phase, the diauxic shift, in which they start to metabolize ethanol. The diauxic shift is characterized by slower growth and significant differences in the genome-wide transcription profile (Radonjic et al., 2005). For this reason, it was important to repeat the transcription profiling experiments in the growth conditions and with the wild type strain used in this study. A total (Figure 8 A) and poly (A)+ (Figure 8 B) RNA tiling array was performed in collaboration with the Steinmetz laboratory. The WT cells were grown and prepared as described in section 2.3 and the data is available at:

http://steinmetzlab.embl.de/cgi-bin/viewMellorLabArray.pl?showSamples=502_Glu1Vs503_Gal1&type=heatmap&gene=gal1.

Not surprisingly, the array hybridized with total RNA (Figure 8 A) has a high background compared to the poly (A)+ RNA published data set, probably due to the extremely high levels of ribosomal RNA and the pervasive nature of transcription. It was then decided to use poly (A)+ enriched RNA instead. Figure 8 B shows that much of the background is no longer detected and the putative transcripts are now very distinct.

In repressed conditions (WT_YPD sample), on the plus strand, the *GAL10* SUT described by Xu et al. is clearly observable, as a strong hybridization signal from the 3' end of *GAL10* to the promoter. However, a further signal is detected which extends to the *GAL1* terminator. The starting location for this putative transcript is not discernible. Also detectable on this strand, in glucose, is the *FUR4* sense transcript, whose hybridization signal is as strong as that just described present in glucose that runs sense to *GAL1*.

On the minus strand, in glucose, low-level hybridization might indicate expression of *GAL7*. This signal is also visible in the Xu et al. data set, however the reason for this expression is unknown. Perhaps Gal7p has a separate function from its role galactose metabolism. Also clearly visible is a small putative transcript, which starts at the 3' end sense to *GAL10*, and whose levels are comparable to those of the *GAL7* and the *GAL10* SUT transcripts. It could also be argued that a very low level transcript is observed in the *GAL1* antisense region, which compared with the Steinmetz array data, could be the transcript they described as the *GAL1* antisense CUT.

Under induced conditions (WT_YPG sample), on the plus strand, the strongest hybridization signal corresponds to the *GAL1* transcript. However, this is not the only transcript present. From approximately the middle of *GAL10,* a hybridization signal is detectable, probably due to the presence of a transcript whose levels seem slightly lower than those of the *GAL10* SUT, although the signal increases in the intergenic region between *GAL10* and *GAL1*. Likewise, a signal of similar levels is observable in the intergenic region between *GAL1* and *FUR4.* Also of note is that, in these conditions, the *FUR4* transcript levels are similar in both YPD and YPG. So, from the middle of *GAL10* to

the end of *FUR4* in galactose and on the plus strand, there is not a region which is not being transcribed although, it is not possible to say how many transcripts are present.

On the minus strand, regions of hybridization corresponding to the *GAL7* and *GAL10* transcripts and the *GAL10-7* fusion transcript are clearly identifiable (Greger and Proudfoot, 1998; Kaplan et al., 2005). Also, unmistakably there is a signal antisense to *FUR4* that seems to run through the *GAL1–FUR4* intergenic region, albeit at lower levels, perhaps indicating the presence of more than one transcript. Again, this putative transcript is also visible in Xu et al. data set, however it was not annotated.

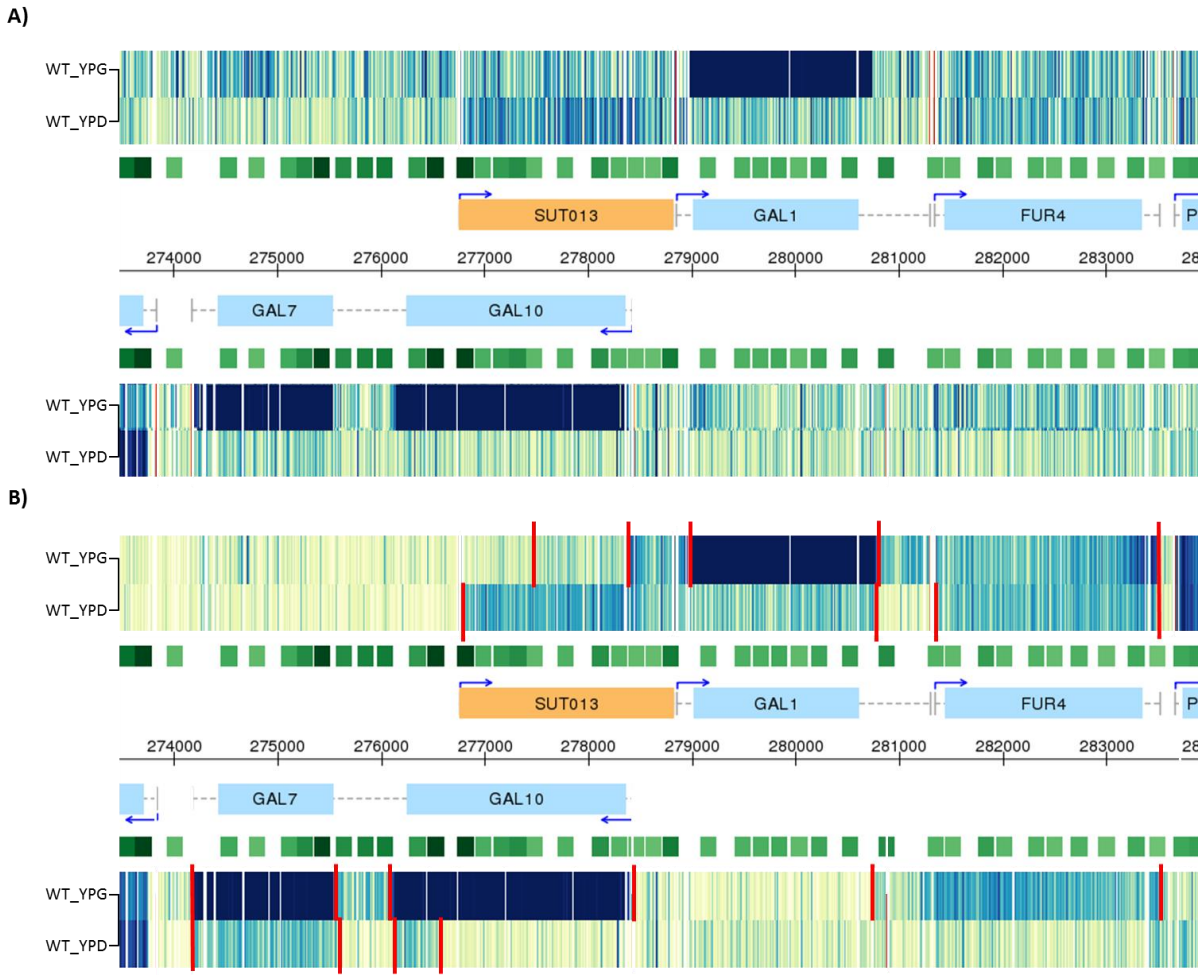


Figure 8. Tiling array of total and poly (A)+ RNA of the *GAL* locus in the induced and repressed states. The tiling array experiments were performed in collaboration with Lars Steinmetz at EMBL. Two separate samples of a WT strain were grown as in section 2.3 and then allowed to grow for a further 3 hours in fresh YPD (WT_YPD) or YPG (WT_YPG). The figure shows the annotated ORFs (light blue boxes) in the middle either above or below the scale depending on the orientation of the gene. The two rows of data at the top (plus strand) and bottom (minus strand) represent each strand, *e.g.* for *GAL1* the top data set is the sense strand and the bottom the antisense strand. The colour code is light blue for low hybridization and dark for high. The green boxes represent the nucleosome positioning (Mavrich et al., 2008). **(A)** Tiling array of total RNA. **(B)** Tiling array of poly (A)+ enriched RNA. Red vertical lines indicate the boundaries of discrete transcripts visually observed in the data set.

It is evident from this data that it is very likely that far more transcripts are present in the *GAL* locus than the ones already described in published data (Houseley et al., 2008; Pinskaya et al., 2009; van Dijk et al., 2011; Xu et al., 2009). In order to fully understand the regulation of this locus and the role that the non-coding RNAs play in it, it is necessary

to first identify all the transcripts present. However, it also evident from this data that there are regions which contain potentially overlapping transcripts and that it is impossible to discriminate between them using this methodology.

One technique that is commonly used to further define transcripts is RACE (Rapid Amplification of cDNA Ends). Indeed, some of the transcripts identified by the tiling arrays were then mapped by RACE in this laboratory (Peter Dudek - unpublished data) (Table 7). However, this technique also cannot discriminate between transcripts and its analysis becomes technically challenging when there are several overlapping transcripts.

Transcript	Strand	TSS position (relative to *GAL10* ATG)
Transcript **d**	+	+788 to +791
Transcript **f**	+	+1611
Transcript **h**	-	+1819 to +1824

Table 7. Non-coding transcripts at *GAL10* mapped by RACE. Performed by Peter Dudek. Transcripts mapped by 5' RACE, in WT cells grown in glucose. Schematic of the position of the transcripts annotated in Figure 9 (B). TSS – Transcription start site

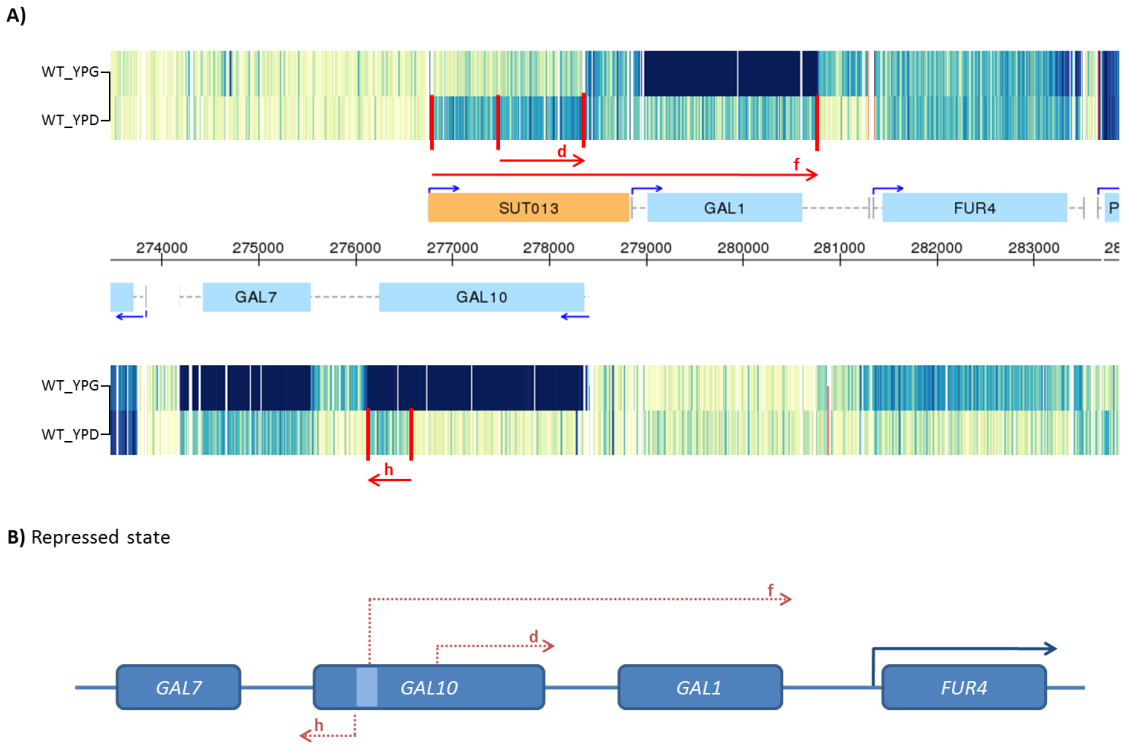


Figure 9. Comparison of the transcripts identified by RACE and Tiling Arrays. (A) Poly (A)+ RNA tiling arrays data described in Figure 8 (B). **(B)** Schematic of the starting location of the transcripts mapped by 5' RACE. End locations have not yet been mapped.

3.2.1 Control for Northern blot

Currently, the only technique that can discriminate between different overlapping transcripts is Northern blotting. With this methodology it is possible to determine how many transcripts there are in a given region, from which strand they are being produced and their approximate lengths. It is also possible to know approximately where the transcripts start and end by designing differently positioned probes across the region of interest. Given the advantages of this technique it was decided to further analyse the *GAL* locus using this method. Several probes were designed across the *GAL* locus. For optimal detection of low level transcripts a long probe is preferable, however, the longer the probe, the higher the probability of cross-hybridization and the reduced resolution

obtained on start and end locations. A combination of long and short probes was therefore used.

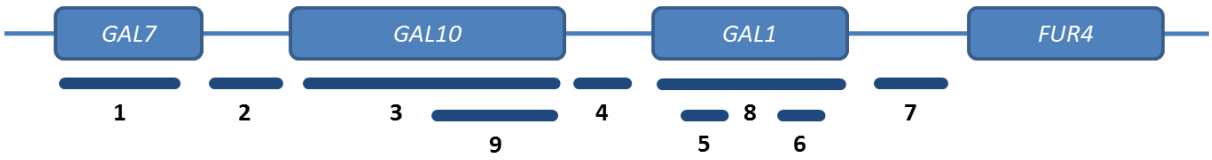


Figure 10. Position of the probes across the *GAL* locus used for Northern blotting. A schematic of the position and length of the probes. See Table 8 for name and accurate size of the probes. Both sense and antisense probes have the same length and position.

Probe	Name	Length
1	*GAL7*	900 bp
2	*GAL10/7 IG*	424 bp
3	*GAL10* Long	2034 bp
4	*GAL10/1 IG*	363 bp
5	*GAL1*	295 bp
6	*GAL1 3'*	253 bp
7	*GAL1/FUR4 IG*	464 bp
8	*GAL1* Long	1515 bp
9	*GAL10*	1030 bp

Table 8. Length of the probes used for Northern blotting. The number of the probe refers to Figure 10. Both sense and antisense probes have the same length and position (primers used for production of the probes indicated in Table 5 and Table 6).

In this study, the key advantage of the technique is the strand-specificity of the probe. Until recently, before the finding that genomes are heavily transcribed on both strands,

most studies were done with non-strand-specific probes. The best approach to obtain radiolabelled strand-specific probes was investigated and two different methods were tested. One of the techniques tested, in which the probes are labelled by an *in vitro* transcription reaction, was found not to be completely strand-specific in all cases. The reason for this was not investigated, but the sequence of the probe and/or RNA secondary structure may play a role. Another technique, which labels the probes by asymmetric PCR, was then tested and proved to be a reliable method. This method uses two rounds of PCR where, firstly, a DNA template stock is made by amplifying the sequence of interest followed by gel extraction for maximum purity and specificity. The radiolabelled probe is then obtained by another round of PCR where the template DNA is incubated with either a forward (for an antisense-detecting probe) or reverse primer (for a sense-detecting probe) and a radioactive nucleotide. All the probes were tested for specificity by the dot-blot method, where unlabelled sense and antisense probe were dotted onto a membrane, baked and then probed with the same radiolabelled sense and antisense probe. An example of the control for the *GAL1* probe is shown in Figure 11. A low level cross-hybridization is always detected probably due to the sequestration of some of the labelled probe in the high amounts of RNA dotted in the membrane, as it commonly happens with the ribosomal RNAs.

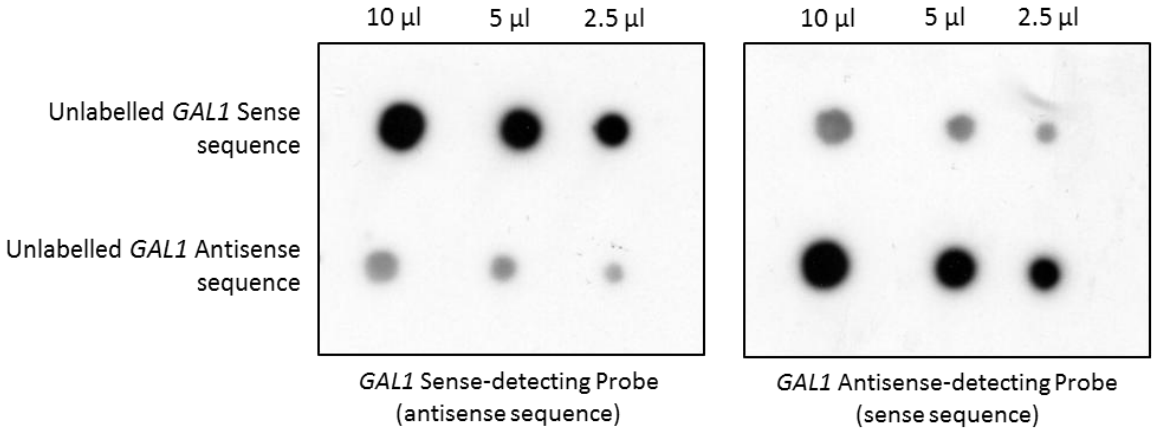


Figure 11. Strand-specificity of the *GAL1* sense and antisense probes. Dot blot analysis of the *GAL1* sense and antisense probes. Unlabelled sense and antisense probes were obtained by the method described in section 2.10 and were diluted to a concentration of 1 µg/µl before being dotted onto a membrane. The membranes were then probed with the labelled sense and antisense probes as described in section 2.8.

Another concern was raised about *GAL3* which is 90% similar to *GAL1* in its DNA sequence and thus could show cross-hybridization. Initially, a long probe that covered the entire *GAL1* ORF was used (*GAL1* Long Probe, Figure 10 and Table 8), however, due to the high similarity between the genes it was decided to design a smaller probe of about 300 bp (*GAL1* Probe, Figure 10). The new probe was tested for *GAL1* specificity and it was shown that in a *gal1Δ* strain, no *GAL3* transcripts were detected confirming that the probe is specific for *GAL1* (Figure 12). Of note is that, in a *gal3Δ* strain, all the *GAL* genes are induced slowly therefore explaining why no *GAL1* RNA is detected.

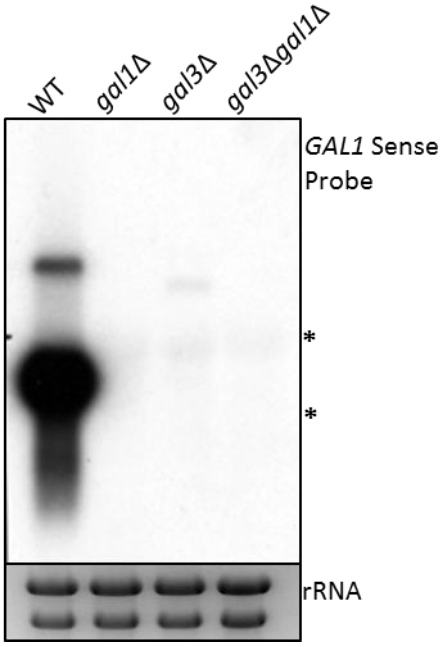


Figure 12. *GAL1* sense probe is specific to *GAL1* only. Northern blot showing the control for the sequence specificity of the probe against the strains indicated. Cells were induced for 3 h to steady state in galactose as described in 2.3. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gel (rRNA) is shown for loading control.

3.2.2 Identification of ncRNAs by Northern blot in repressed conditions

So far, RACE and tiling arrays have indicated the presence of at least two new transcripts, in addition to the already described *GAL10* SUT (Xu et al., 2009), in repressed conditions. The transcripts present in the *GAL* locus in repressed conditions were further assessed by Northern blotting and as shown in Figure 13, several were detected. Transcript **g** is detected with the *GAL10* antisense and the *GAL10/1* intergenic (sense to *GAL1*) probes and its size is around ~2.4 kb in length, which is consistent with it starting in the *GAL10* ORF and reading through to the beginning of *GAL1*. Although transcript **g** has not yet been mapped to base pair resolution, it is likely to initiate at the *GAL10* internal bi-directional promoter as it is Reb1-dependent (Figure 14). Transcript **f** is highly likely to correspond to the *GAL10* SUT described by separate studies (Houseley et al., 2008;

Pinskaya et al., 2009; Xu et al., 2009), transcribed from *GAL10* to the end of *GAL1*. This transcript was found to be produced from the *GAL10* internal bi-directional promoter and is dependent on Reb1 binding. Upon removal of either Reb1p itself or the Reb1 binding sites located at the internal promoter, the transcript is no longer produced (Houseley et al., 2008; Pinskaya et al., 2009). Analysing the strain that was used in the study of Houseley et al., in which all the Reb1 binding sites were mutated, it was observed that both transcripts (**f** and **g**) were absent (Figure 32). Transcripts **d** and **β** are not detectable with the *GAL10/1* intergenic probe (Figure 13 B), meaning that they probably do not extend far into the promoter. Transcript **d** (Figure 13 A) is around ~1 kb in length, and, its start location has been mapped by 5' RACE as starting at around +790 bp downstream of the *GAL10* ATG (Table 7), consistent with it terminating early in the promoter. Of note is that transcript **β** (of about 1.2 kb long) in this Northern blot is detected at extremely low levels, however, its levels vary between experiments and usually when transcript **β** is present, transcript **d** is not, and vice-versa, so it is possible that transcript **β** is a longer variant of transcript **d**. Transcript **h**, detected by the tiling array experiment (Figure 8), is not detectable with the *GAL10* sense probe, which is homologous to the first half of *GAL10*.

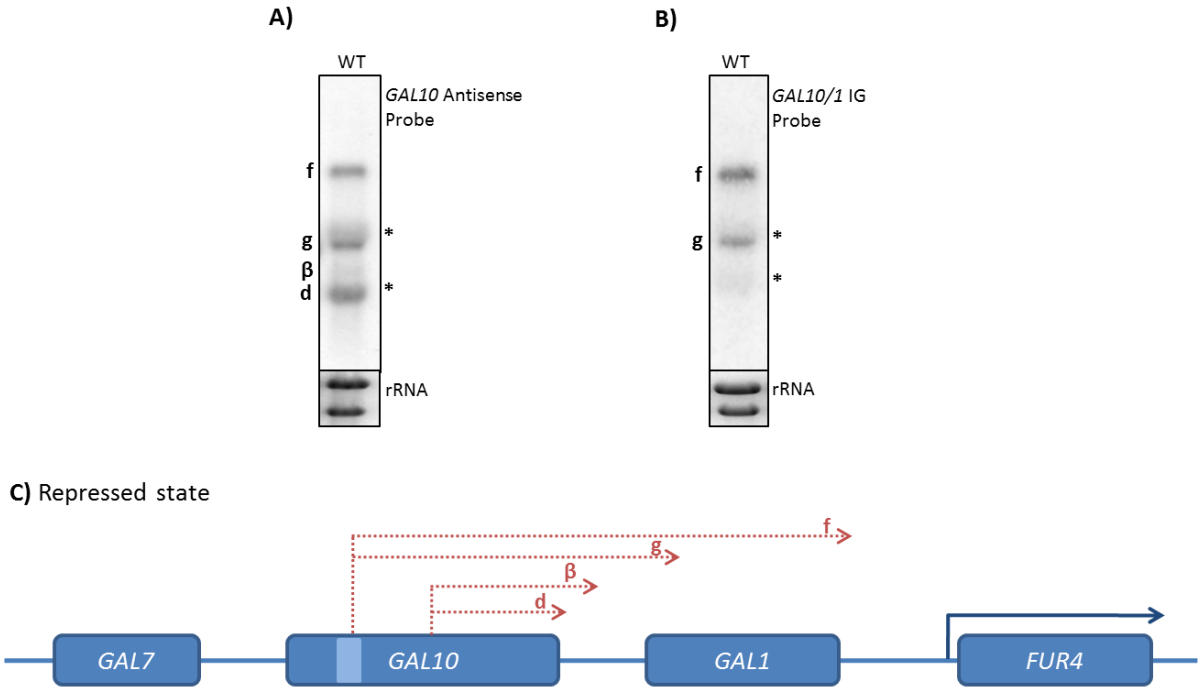


Figure 13. Transcripts present at the *GAL* locus in repressed conditions. Northern blots probed for the *GAL10* antisense strand **(A)** and *GAL10/1* intergenic region, sense to *GAL1* **(B)**. A WT strain was grown in YPD in the conditions mentioned in section 2.3. The ribosomal bands are marked with * for size guidance. **(C)** Schematic of the profile of transcripts identified in the Northern blots above. Transcript start and end sites are only approximate.

Houseley et al. have shown that transcript **f** is produced from the *GAL10* internal bi-directional promoter, which contains four Reb1-binding sites (Houseley et al., 2008). Upon mutation of the Reb1-binding sites, transcription of the ncRNA was compromised indicating that transcript **f** is dependent on Reb1p. It was therefore important not only to reproduce his data but also to investigate whether the remaining transcripts are also Reb1-dependent. Data obtained in the laboratory by Michael Youdell showed that Reb1p recruitment to the *GAL10* internal bi-directional promoter is decreased in minimal media (MM) (Figure 14 B), due to overall reduced levels of Reb1p in this medium (data not shown). Therefore, a *reb1Δ* phenotype can be simulated by growing cells in minimal media.

Thus, it is expected that transcript **f** will be absent in cells grown in minimal media and in fact this is confirmed in Figure 14, where the transcript is no longer observed. Interestingly, transcript **g** is also absent in this experiment. Therefore, transcript **g** is also Reb1-dependent indicating that it is also produced from the *GAL10* internal bi-directional promoter. Figure 14 also shows that transcript **d** is still present in the minimal media conditions indicating that it is Reb1-independent, and therefore not likely to start at the same location as transcripts **f** and **g**.

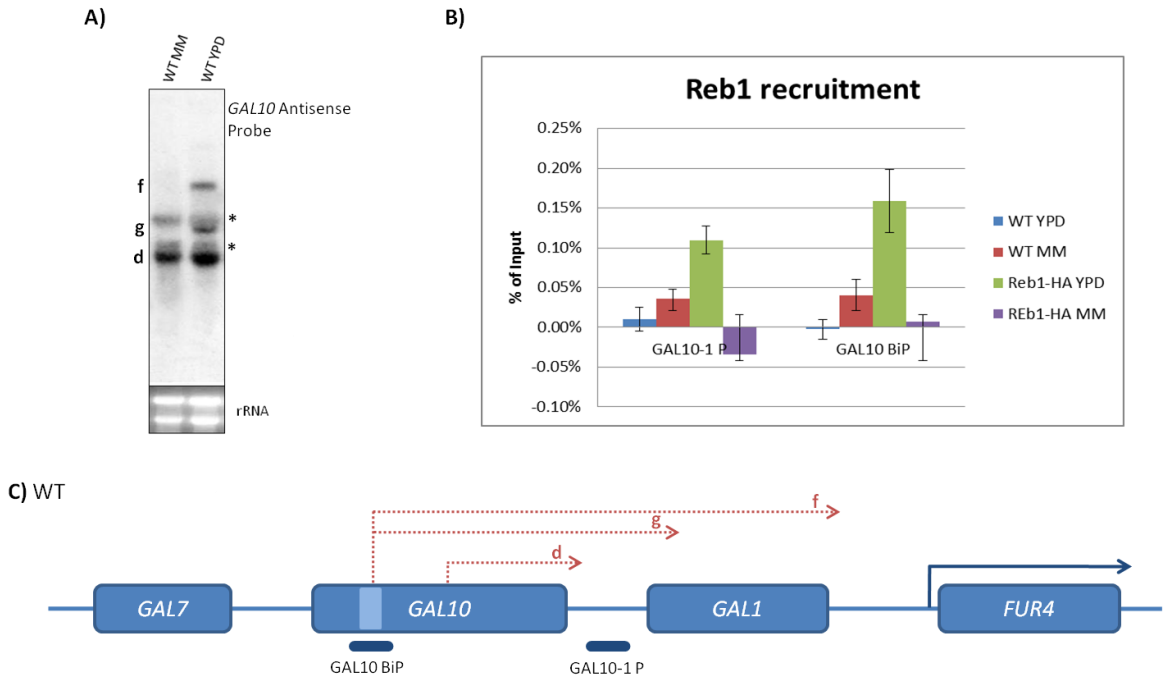


Figure 14. Transcripts f and g are Reb1-dependent. (A) Northern blot analysis of total RNA extracted from a WT strain and probed for *GAL10* antisense. Cells were grown in minimal media with 2% glucose (MM) or in YPD. Transcripts f, g and d are indicated. **(B)** Quantitative PCR (qPCR) analysis of a ChIP experiment against α-HA of a Reb1-HA tagged strain at the *GAL10-1* promoter (*GAL10-1* P) and *GAL10* internal bi-directional promoter (*GAL10* BiP). Cells were grown in minimal media with 2% glucose (MM) or in YPD. **(C)** Schematic of the profile of transcripts identified in the Northern blots above and the location of the PCR probes. Transcript start and end sites are only approximate. Northern blot, ChIP and qPCR performed by Michael Youdell.

3.2.3 Identification of ncRNAs by Northern blotting in induced conditions

In the previous section novel non-coding transcripts present in the *GAL* locus were identified in repressed conditions. Next, the transcripts present in induced conditions were assessed. Firstly, the probes that are homologous to the coding regions of *GAL7,* 10 and 1 were used (Figure 15). As expected, the coding transcripts of *GAL7* (Figure 15 A), *GAL10* (Figure 15 B) and *GAL1* (Figure 15 C) were highly transcribed after 3 hours of growth in galactose and easily detected by Northern blotting. The *GAL7* transcript, with an expected size of ~1.1 kb without the poly(A) tail of approximately 200 nucleotides, was detected just above the first ribosomal band (18S rRNA), which is known to run at between 1.5 to 2 kb. This is not surprising as it was observed that the ribosomal RNAs slow the progression of the other RNAs in an agarose gel. The *GAL10* transcript has an expected size of ~2.2 kb and was detected below the top ribosomal band (26S rRNA) which should run at ~3 kb-3.5 kb. The *GAL1* transcript migrates in the region between the two rRNA bands and has an expected size of ~1.6 kb. With these three probes, high molecular weight transcripts were also noticeable, hereby designated **e** and **c**. Taking into account the strand from which they are produced, the two intergenic probes also used (*GAL10* antisense (Figure 15 D) and *GAL1/FUR4* IG, sense to *GAL1* (Figure 15 E)) and the tiling array data (Figure 15 G), it can be concluded that these long transcripts read through *GAL10* (transcript **e**) or *GAL1* (transcript **c**) into the next gene. Transcript **e** (Figure 15 A and B) is very likely to start at the *GAL10* promoter and reads-through to the end of *GAL7*, as it is present in the Northern blots probed for *GAL10* and *GAL7*. It is unlikely that the transcript comes from the *GAL1* gene since it is not detectable with a *GAL1* antisense probe (Figure 20) and is not seen in the tiling array data (Figure 15 G). The size of the long *GAL10-GAL7* transcript seems to be around 4 kb, however it is only an

estimate as the agarose gel has very poor resolution for transcripts in this size range, but it is consistent with the observed transcript in the tiling array data (Figure 15 G). Transcript **c** (Figure 15 C) is very likely to start at the *GAL1* promoter and reads through to the end of *FUR4*, with an approximate size of 4.5 kb. Consistently, a probe used in the *GAL1–FUR4* intergenic region detects a transcript of the same size (Figure 13 C), also evident in the tiling array data (Figure 15 G).

When probing for the *GAL10* antisense strand (Figure 15 D), three transcripts were observed, whose sizes are consistent with the transcripts detected in repressed conditions (transcripts **f**, **g** and **β**/**d**) (Figure 13). It has been shown that the transcription of the *GAL10* SUT (transcript **f**) is decreased in induced conditions in the tiling array data (Figure 8 B) and in published work (Houseley et al., 2008; Pinskaya et al., 2009). However, the conditions used throughout this study include the growth of the cells to logarithmic phase in glucose-containing medium and then an induction in galactose-containing medium for usually a maximum of three hours. It is, therefore, possible that the transcripts identified with the *GAL10* antisense probe under induced conditions represent the transcripts produced in glucose, which might not yet have been degraded. Also, we do not know how quickly the repressed transcription profile switches to the induced transcription profile. However, the possibility that these transcripts differ from the ones in repressive conditions cannot be excluded. Nevertheless, in this study, the transcripts identified with the *GAL10* antisense probe in repressed and induced conditions, if of the same size, will be presumed to be equivalent. Finally, in Figure 15 (E), besides transcript **c**, a second transcript is detected (annotated with a question mark) with the *GAL1/FUR4* IG probe. Although this might represent RNA trapped or occluded in the rRNA, the tiling

array data indicates transcription in the *GAL1/FUR4* intergenic region in YPG that could explain the result observed in the Northern blot. However, because it is not directly related to the *GAL* genes it was not investigated further. The probing of *GAL10-GAL1* intergenic region is not shown here due to the fact that the probe does not detect any transcript from the induced locus.

Figure 15 (F) compiles the data obtained by Northern blotting for the induced state of the *GAL* genes. It is evident that there are several novel transcripts throughout the *GAL* locus on both strands besides those from the strongly inducible *GAL7*, *10* and *1* genes.

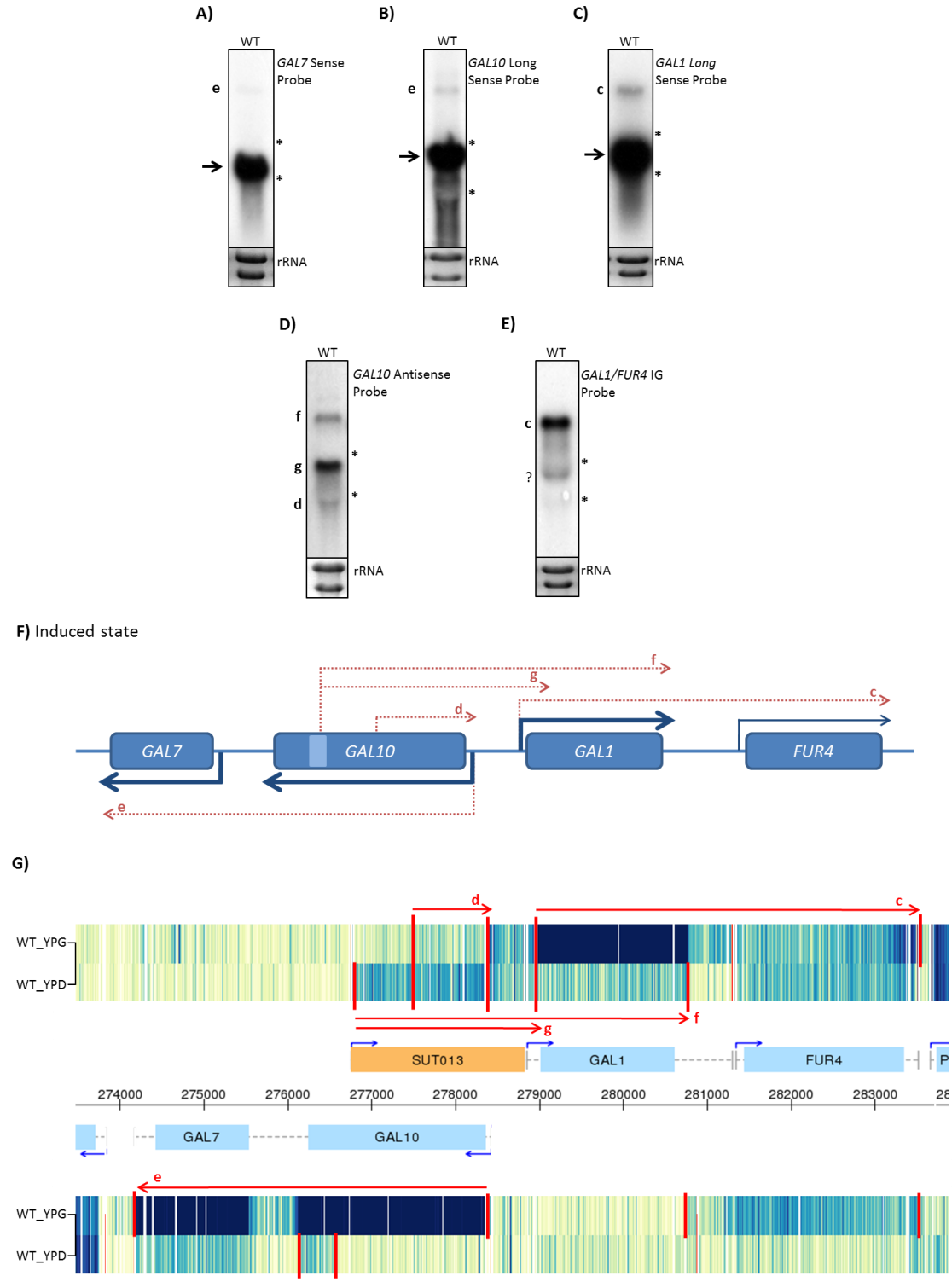


Figure 15. Transcripts present at the *GAL* locus under induced conditions. Northern blots probed for the *GAL7* **(A)**, *GAL10* **(B)** and *GAL1* **(C)** coding regions, as well as *GAL10* antisense **(D)**, *GAL1/FUR4* IG, sense to *GAL1* **(E)**. The exposure of the intergenic Northern blots (about 3 days) is higher than those of the coding regions (overnight). The WT strain was grown and induced for 3 hours in the conditions mentioned in section 2.3. The coding transcript is identified by the black arrow. The ribosomal bands are marked with * for size guidance. **F)** Schematic of the profile of transcripts identified in the Northern blots above. Transcript start and end sites are only approximate. **G)** Poly(A)+ RNA tiling array described in Figure 8 (B).

3.2.4 Novel transcripts identified in the GAL locus

An extensive investigation into the transcript profile of the *GAL* locus has unveiled several novel transcripts. Figure 16 outlines all the transcripts detected by Northern blotting and RACE and compares them to the tiling array data. It is clear that many of the transcripts would not have been detected if a technique like Northern blotting had not been used. The necessity of setting high thresholds to avoid background noise during the analysis of tiling array data also resulted in low-level transcripts not being identified.

The presence of so many stable non-coding transcripts raises the questions of how and why they are being produced.

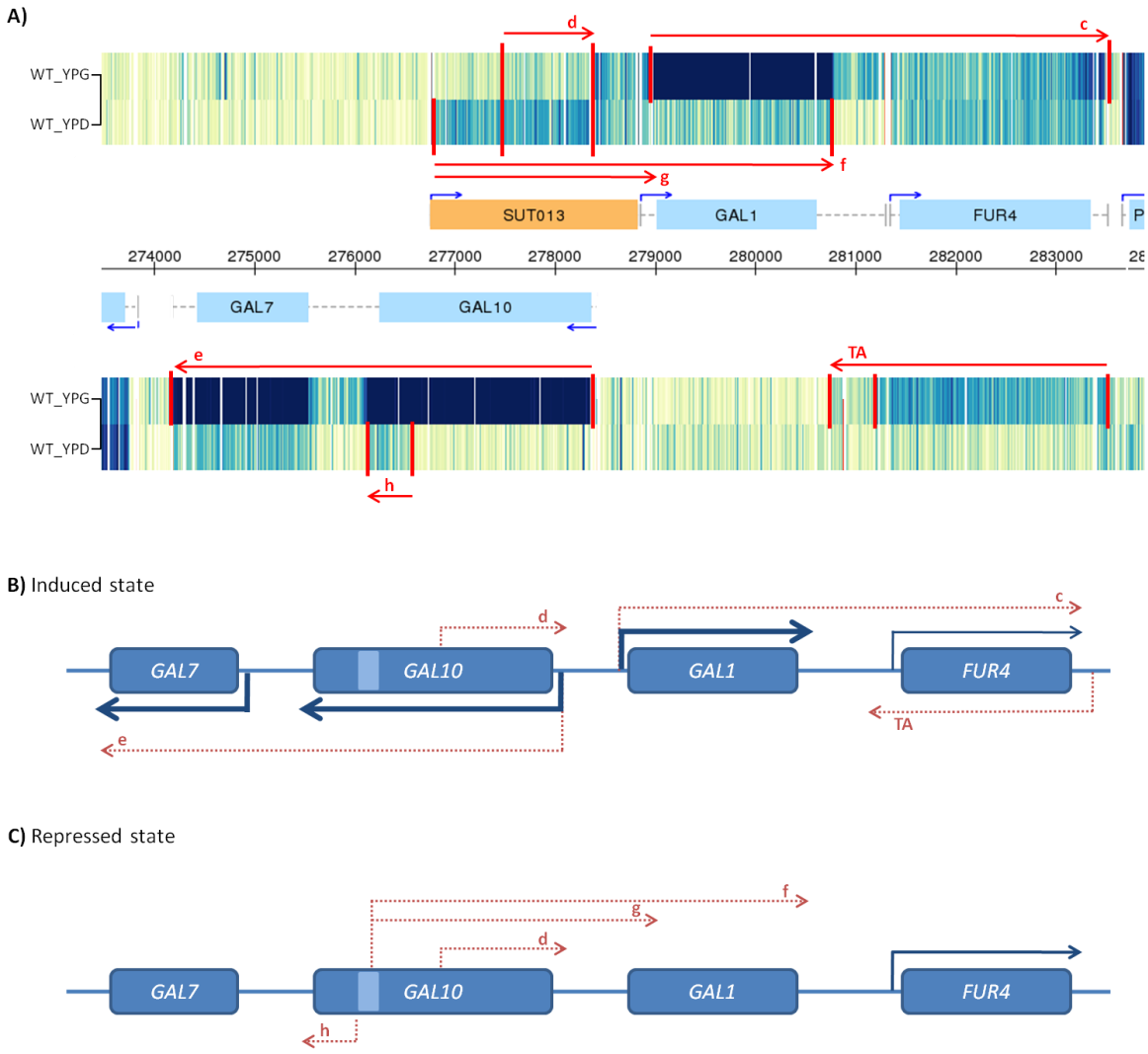


Figure 16. Comparison of the transcription profiles between the tiling array data and the data obtained by Northern blotting. (A) Poly (A)+ RNA tiling array with the high hybridization regions marked by the red vertical lines and the corresponding transcripts identified. TA refers to a transcript identified by the tilling array experiment only. Schematic of all the transcripts unveiled by Northern blotting, tilling array and RACE in induced **(B)** and repressed **(C)** states. Note that transcripts **g** and **f** are present and are annotated in the repressed state; however they may also persist for 3 hours or more during induction.

3.3 Presence of unstable transcripts at the *GAL* locus

One of the questions frequently asked about the non-coding transcripts relates to their stability. High-throughput techniques, either RNA tiling arrays performed on mutant strains or by genomic mapping of elongating polymerases, have elucidated the fact that a large proportion of the pervasive transcripts are unstable (Churchman and Weissman, 2011; van Dijk et al., 2011; Xu et al., 2009). CUTs (Cryptic Unstable Transcript) are quickly degraded by the nuclear exosome complex, which is responsible for RNA quality control, hence not visible in a WT strain. If the main enzymatic component, Rrp6p, is deleted the nuclear exosome activity is impaired and the previously unstable transcripts are stabilized. This is consequently the main approach used to study unstable transcripts. Moreover, the NET-Seq technique can be used to identify regions where the high frequency of RNA polymerase does not seem to correlate with transcript levels indicating that they have been degraded (Churchman and Weissman, 2011). Profiling of unstable transcripts performed in an *rrp6* deletion strain by the Steinmetz laboratory shows that in the *GAL* locus only one CUT, antisense to *GAL1*, is observable (Figure 17) (Xu et al., 2009). However, one could argue that this CUT is visible in other carbon sources and only slightly stabilized in the absence of Rrp6p. A transcript antisense to *GAL1* has not been found in a WT strain either by tiling array or Northern blot in this study. Since many more transcripts were unveiled under the growth conditions used in this study, it was important to investigate whether this would be the case for unstable transcripts.

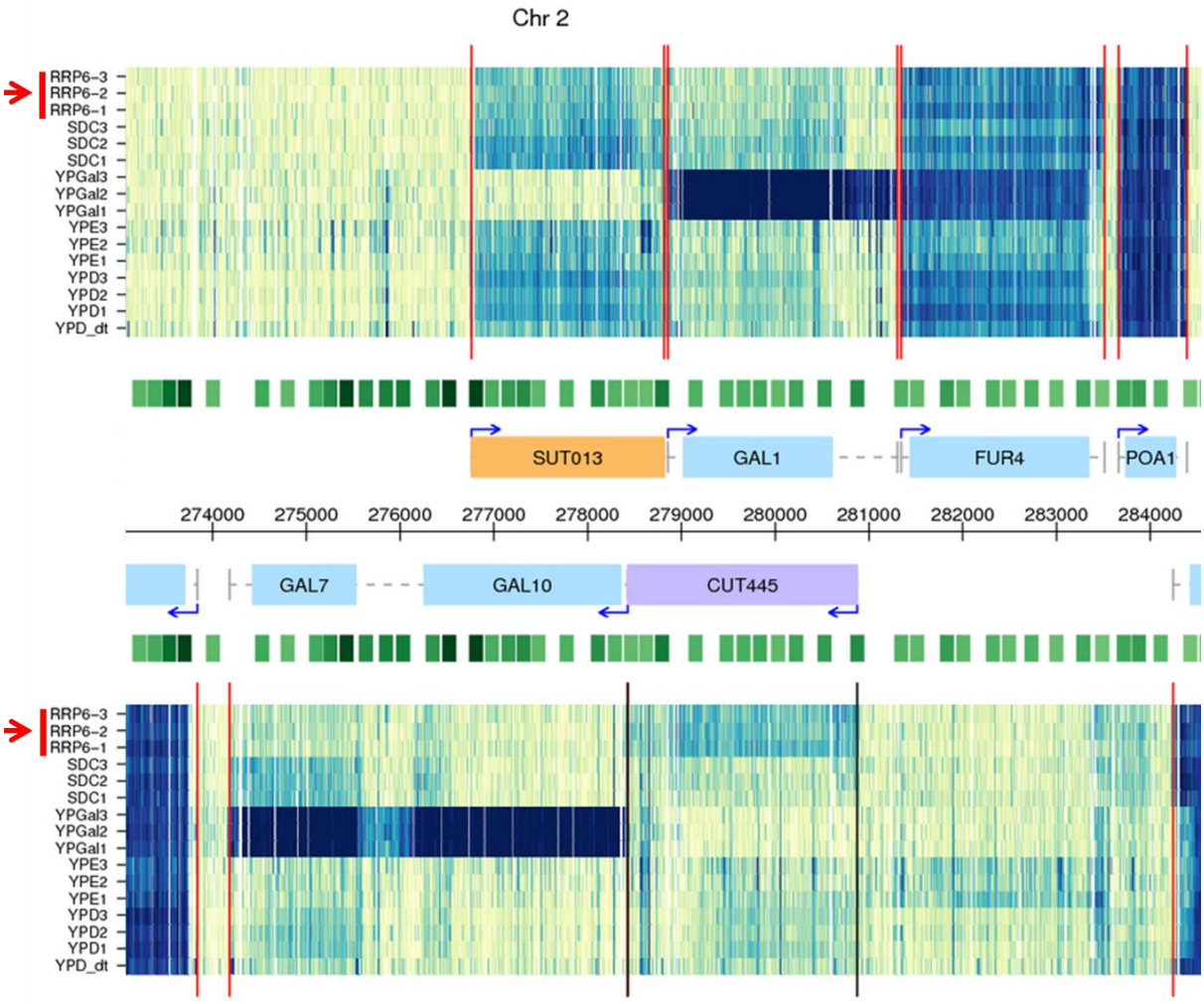


Figure 17. Tiling array of an *rrp6Δ* strain shows a *GAL1* antisense CUT. Data acquired from http://steinmetzlab.embl.de/cgi-bin/viewNFRsharing.pl?showSamples=NFRsharing&gene=GAL10. The figure shows the annotated ORFs (light blue boxes) in the middle either above or below the scale depending on the orientation of the gene. A CUT (purple) and a SUT (orange) are also represented. The rows of data at the top and bottom of the figure represent each strand, *e.g.* for *GAL1* the top data set is the sense strand and the bottom the antisense strand. The colour code is light blue for low hybridization and dark for high. The green boxes represent the nucleosome positioning (Mavrich et al., 2008). The red arrows indicate the significant rows for comparison.

The first major observation when analysing Figure 18 is the difference in the levels of the *GAL10* and *GAL1* coding transcripts between the WT and the *rrp6Δ* strains. This is probably due to a phenotype, identified later in this study, in which the *GAL* genes induce faster in the *rrp6Δ* strain than in the WT. This is not solely due to stability of the transcripts. Apart from the major increase in the levels of the coding transcripts, it was

found that the absence of Rrp6p has minimal consequences on the transcripts present at *GAL10* and *GAL1*, in induced conditions. The transcripts present at *GAL10* are not highly stabilized; however, two high molecular weight transcripts detected using the *GAL10* sense probe are observed in the absence of Rrp6p (Figure 18. A). One probably corresponds to transcript **e** given its size, whilst the other is unknown. It is possible that this unknown transcript starts at the *GAL10* promoter, like transcript **e**, but terminates in the beginning of *GAL7*, however further investigation would be necessary to confirm this. The transcripts detected with a *GAL10* antisense probe are not affected by the impairment of the nuclear exosome, with the exception of transcript **f**, which is only slightly stabilized. In this Northern blot a much higher molecular weight transcript is also visible. It is thought that this is a locus-wide transcript, also present in the other strand, which can sometimes be detected in other circumstances. At *GAL1*, transcript **c** is increased by the same proportions as the *GAL1* coding transcript, very likely also due to a faster induction in the *rrp6Δ* strain and not to increased stability (Figure 18 B). This conclusion arises from the fact that transcript **c** induces at the same rate as the coding transcript. With the *GAL1* antisense probe there is a suspected transcript that could be the *GAL1* CUT annotated by the Steinmetz group in their tiling array (Xu et al., 2009). However, if indeed present, that transcript is at extremely low levels and barely detectable. It is not possible, though, to completely discard the possibility that it is just background. In the Steinmetz tiling array data (Figure 17) the CUT has similar levels to that of the *GAL10* SUT, and therefore should be easily detected by Northern blotting. Interestingly, examination of the *rrp6Δ* strain growth conditions used for that tiling array experiment revealed the fact that minimal media was used, probably explaining the

differences observed. It is, then, possible that the *GAL1* antisense CUT is dependent on the growth conditions even in the absence of Rrp6p.

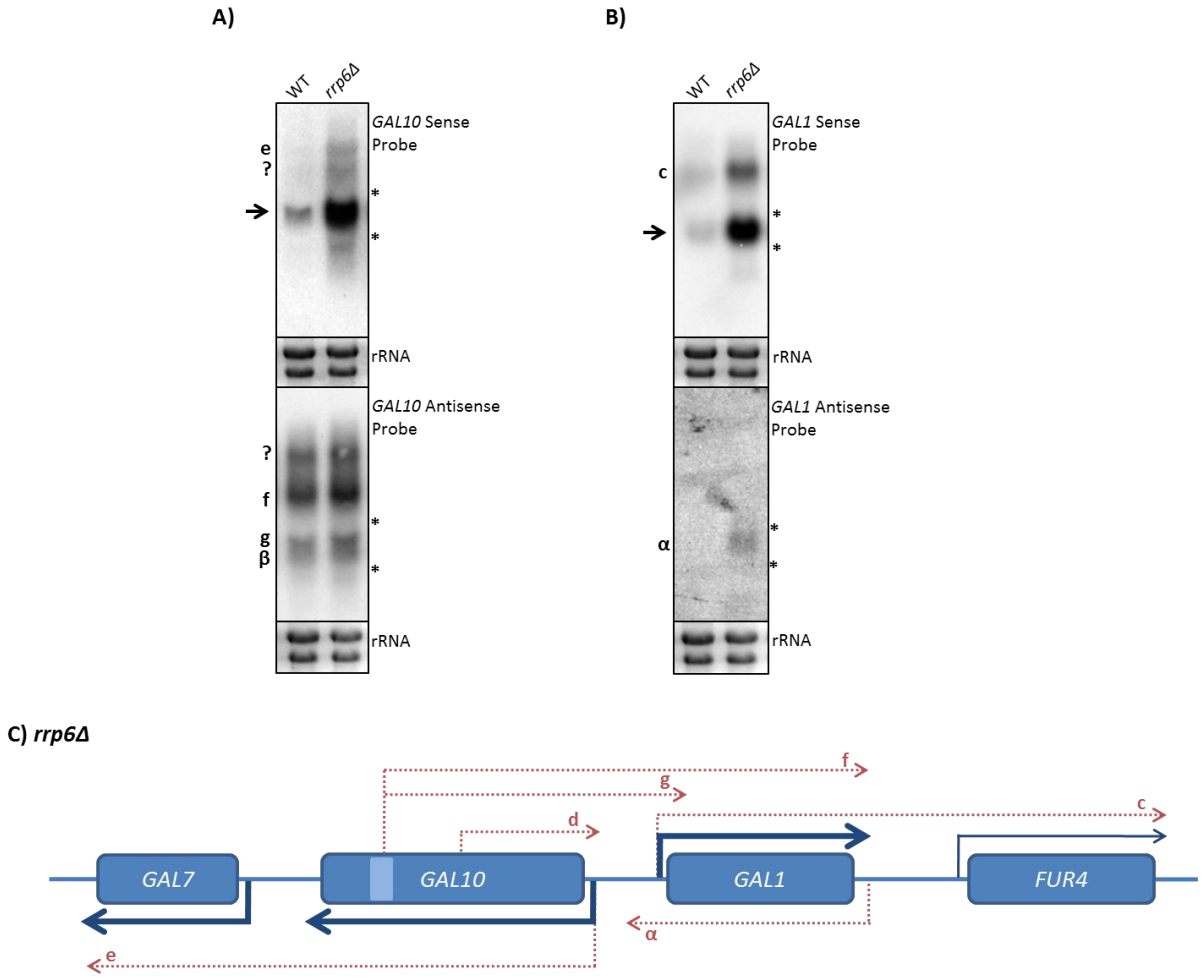


Figure 18. The exosome has minimal effects on transcripts present at *GAL10* and *GAL1*. Northern blot of WT and *rrp6Δ* strains, probed for *GAL10* and *GAL1* sense and antisense. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance. Suspected transcripts are marked with α. **(A)** WT and *rrp6Δ* cells were induced in YPG for 60 minutes. Unknown transcripts are marked with a question mark. **(B)** WT and *rrp6Δ* cells were induced in YPG for 60 minutes. **(C)** Schematic of the profile of transcripts identified in the Northern blots above in induced conditions. Transcript start and end sites are only approximate.

In a separate study, Morillon's laboratory analysed the genome-wide transcription profile, by RNA-Seq, of a *KEM1* (also known as *XRN1*) deletion strain in YPD (van Dijk et al., 2011). Kem1p is a cytoplasmatic 5'-3' exoribonuclease involved in RNA decay. This study unveiled a new class of ncRNAs called XUTs (Xrn1-sensitive unstable transcripts)

and interestingly some of the transcripts annotated in that study seem to correlate with the ones described here (Figure 19). As observed in Figure 19, four XUTs are annotated in the Morillon study. These include a small sense transcript at the 3' end of *GAL10,* which correlates with the transcript here identified as **h**; a long transcript starting at the 3' coding region of *GAL10* and terminating at the end of *GAL1*, which correlates with transcript **f;** and a *FUR4* antisense transcript, which correlates with a transcript identified by the tiling array data shown in Figure 8 (Figure 19 B, transcript **TA**). Transcripts **d** and **g** cannot be differentiated from transcript **f** by the RNA-seq method, for the same reason as with the tiling arrays.

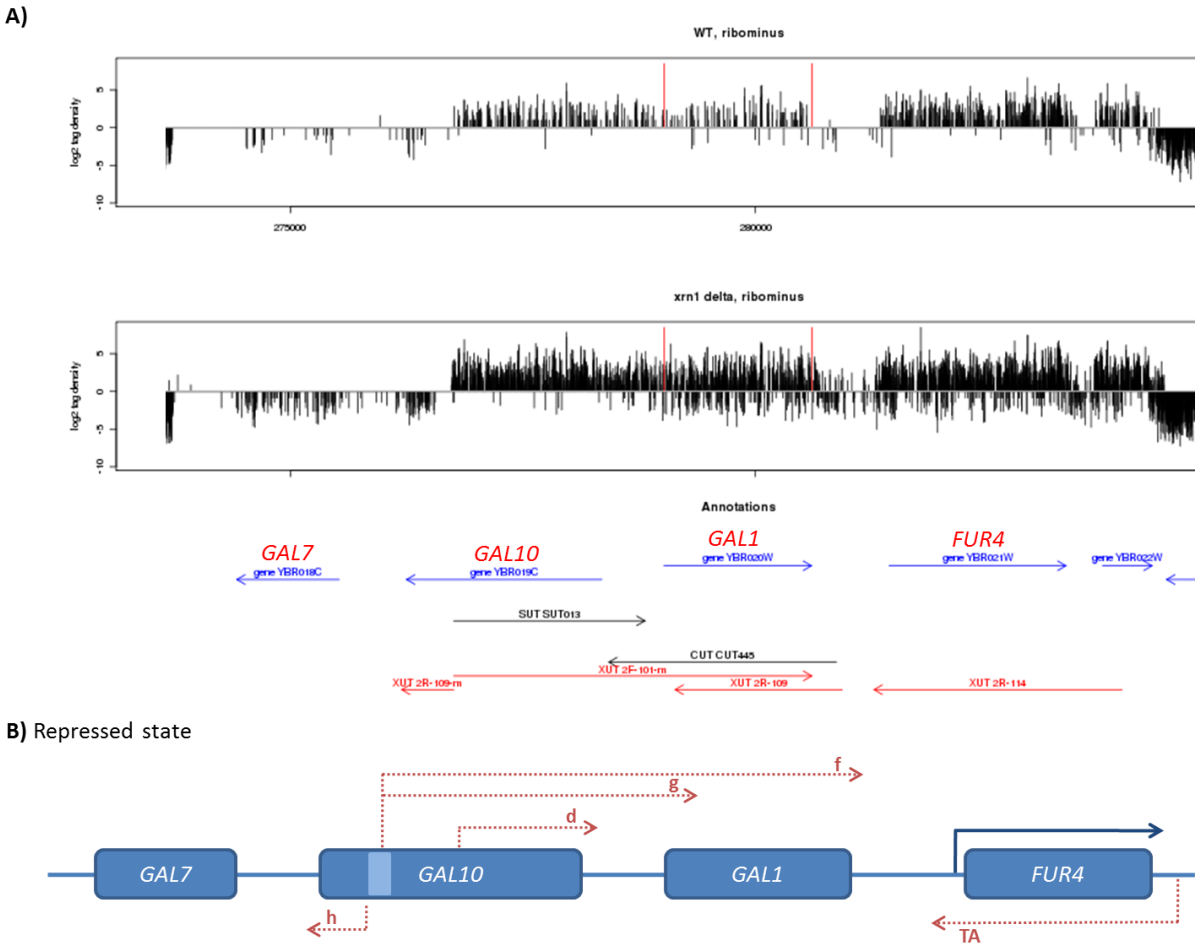


Figure 19. RNA-Seq data identify XUTs at the *GAL* locus in the repressed state. (A) Data acquired from http://vm-gb.curie.fr/XUT/images/YBR020W_1.htm. The top rectangle is the WT profile and the bottom the *xrn1Δ* profile. Inside each rectangle, the top black lines represent hits for the strand sense to *GAL1* and the bottom the other strand. The genes annotations are the blue arrows with the name in red for clearer identification. The black arrows show the SUT and CUT annotated by the Steinmetz group (Xu et al., 2009) and the red arrows the XUTs. **(B)** Schematic of the transcripts identified in this study in repressed conditions. Transcript TA (tiling array) refers to the one only identified by the tiling array experiment.

Since the Morillon group study (van Dijk et al., 2011) was only performed in YPD, it was important to check if the transcripts present in induced conditions are also susceptible to degradation by Kem1p. A *kem1Δ* strain was used and the effects were checked by Northern blot.

The absence of this RNA decay mechanism did not reveal new transcripts in induced conditions (Figure 20), apart from the appearance of a lowly expressed long transcript antisense to *GAL10*, which could be the locus-wide transcripts mentioned before. Also, the coding transcripts were slightly stabilized compared to WT. However, in repressed conditions, two new transcripts (**i** and **j**) were identified with a *GAL1* antisense probe (Figure 20. B). Transcript **i** is very likely to correspond to the XUT identified by the Morillon laboratory whereas transcript **j** is a new transcript probably starting at the *FUR4* terminator and reading through to the *GAL1* promoter.

In the case of the *GAL10* transcripts (Figure 19 A), in repressed conditions a very long transcript (with a different size from the one in YPG) is stabilized while transcripts **g** and **f** are also Kem1-sensitive, in accordance with the results shown by the Morillon group (van Dijk et al., 2011). Interestingly, the stabilization of the transcripts appears to be more prominent in repressed than in induced steady state. There can be at least two explanations for this observation. Firstly, the transcripts can be subjected to tight regulation, i.e. are only produced in repressed conditions, which is known to be true for the ones that are produced from the *GAL10* internal bidirectional promoter. Another explanation could be if the Kem1p itself is down-regulated in galactose. Indeed, in a similar way Reb1p is down-regulated in minimal media and is required for the antisense transcripts **f** and **g**.

The start and end sites of these newly found transcripts **i** and **j** have not yet been mapped but the sizes of the transcripts and the nucleosome positioning data suggest the 3'end of *GAL1* and *FUR4* as a probable region from which they are starting.

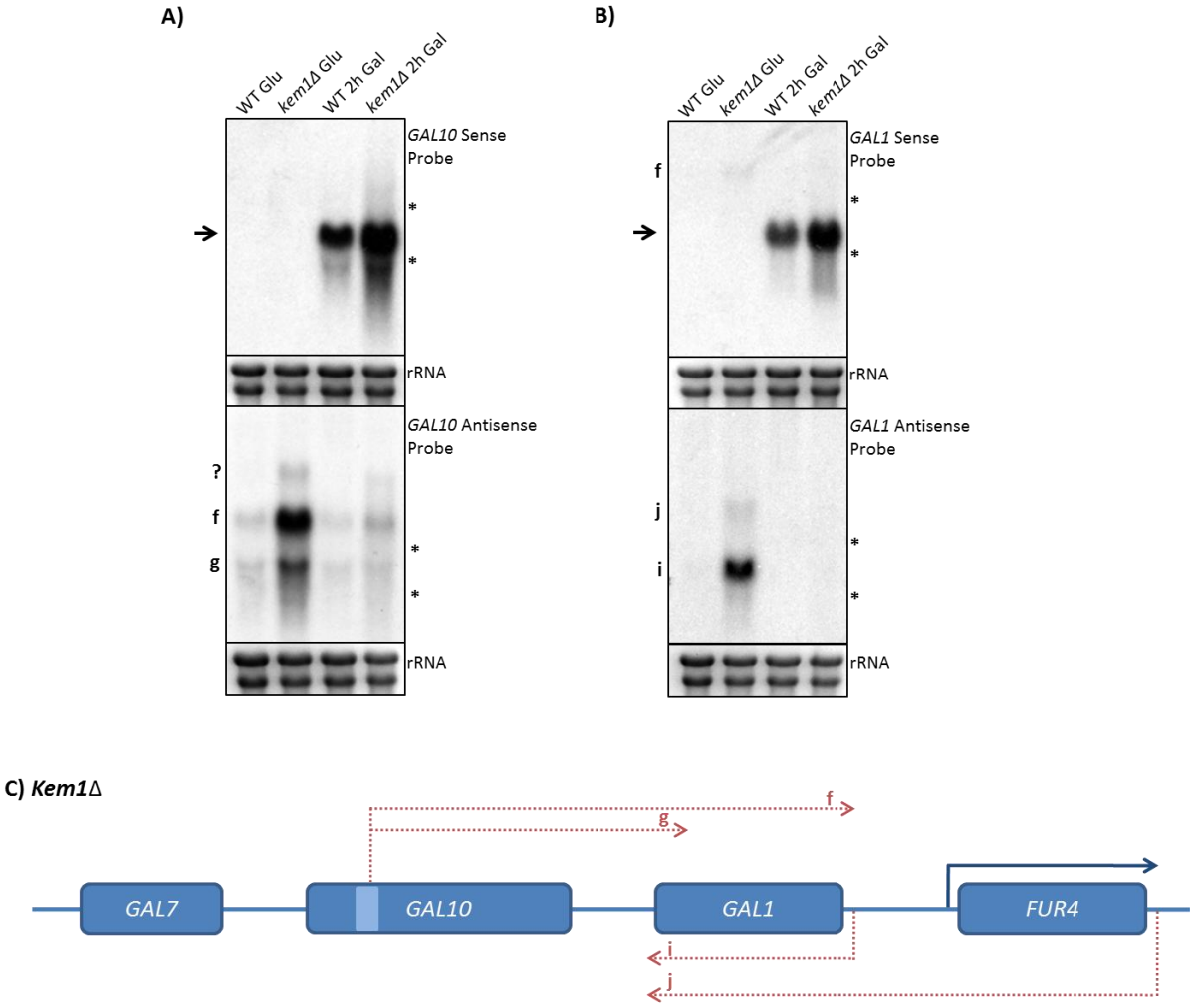


Figure 20. Effects of RNA degradation pathways on the *GAL* locus transcripts. Northern blot of WT and *kem1Δ* strains in repressed and induced conditions. The cells were grown as described in section 2.3 in YPD (lanes 1 and 2) and induced in YPG for 2 hours (lanes 2 and 3). The coding transcripts are identified by arrows. The ribosomal bands are marked with * for size guidance. **(A)** Northern blot probed for *GAL10* sense and antisense. Transcripts marked with a question mark are unknown. **(B)** Northern blot probed for *GAL1* sense and antisense. **(C)** Schematic of the profile of transcripts identified in the Northern blots above. Transcript start and end sites are only approximate.

3.4 Discussion

There are many sceptics who do not accept the concept of pervasive transcription. Amongst those, van Bakel and Timothy R. Hughes compared data sets of RNA-seq and tiling arrays from mouse and human tissues and concluded that most of the newly identified non-coding RNAs were actually due to technical artefacts and/or biological noise. Even though that view has already been challenged (Clark et al., 2011), it is understandable that not all researchers are quickly convinced by the new high-throughput techniques, especially when the data they produce will challenge decades of molecular biology research.

In this chapter the transcription profile of the *GAL* genes was investigated. It was found that there are far more transcripts being produced at the *GAL* locus than have been published so far (Figure 21). Several novel transcripts were identified in induced and repressed conditions. The identification of these transcripts was done by a combination of different techniques including Northern blotting, RACE and tiling arrays. Each technique has its advantages and disadvantages and their combination leads to a better understanding and validation of the results achieved.

In this study, it was found that, even though the *GAL* genes are referred to as "repressed" when cells are grown in glucose media, they are transcriptionally active with the production of several non-coding transcripts. When in inducible conditions, i.e. galactose media, there are even more ncRNAs, with the prevalence of sense-antisense pairs in coding regions (SAPs) and long read-through transcripts spanning two or more genes. Some characteristics of the ncRNAs in the *GAL* locus are that they are present at low steady-state levels compared to the extremely high steady-state levels of the coding

transcripts, they have defined start and end sites, revealed by the tight distinct bands on the Northern blots, and that most of them are polyadenylated, since the they were detected in the poly (A)+ tiling array. Only two transcripts (transcript **i** and **j**, Figure 20) were found to be very unstable, being degraded by the cytoplasmic 5'-3' exoribonuclease, Kem1p, while two more (transcript **f** and **g**, Figure 20) were found to be slightly unstable. These data are in agreement with van Dijk et al. Even though Xu et al. have annotated a *GAL1* antisense transcript (transcript **i**) as being a CUT, under the conditions used in this study, it was found that the exosome was not involved in degrading any of the transcripts, to a large extent, at the *GAL* locus.

The presence of different types of transcripts in the same region raises the question of what makes them different. Are the SUTs, CUTs and XUTs regulated in the same way? If so, how does the cell know which ones to keep and which ones to degrade and by which mechanism? Also, do different types of transcripts have different functions, i.e. are some repressive and others activating?

The possible role and function of the non-coding transcripts present at the *GAL* genes is the subject of the next chapters.

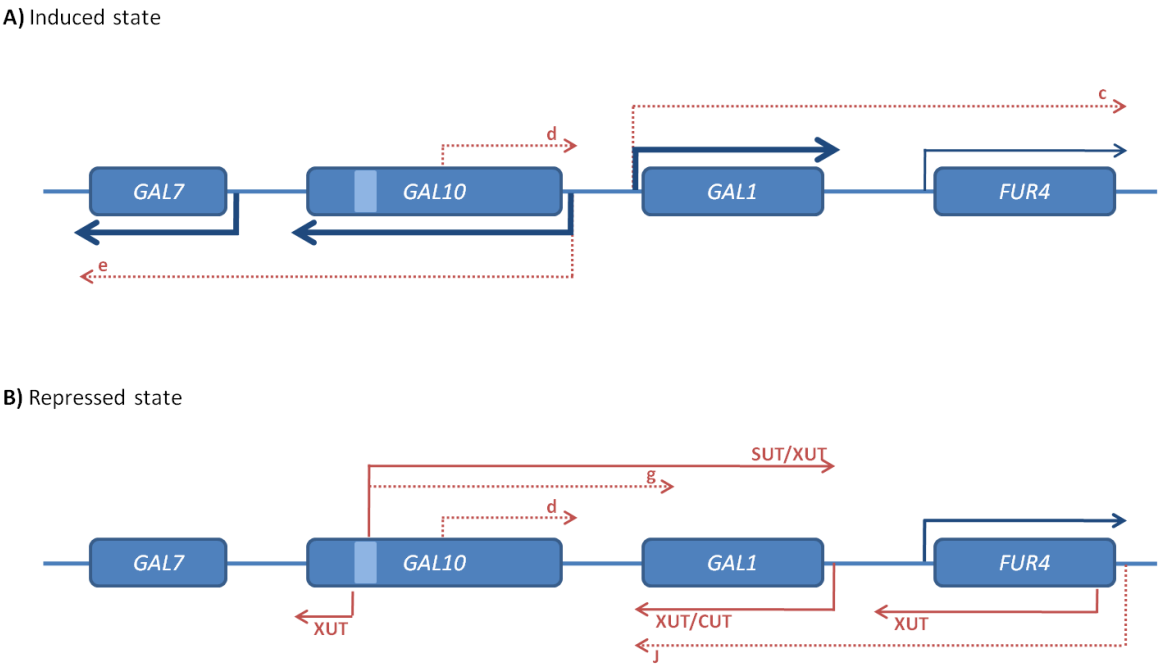


Figure 21. Schematic showing the published transcripts and the novel ones unveiled in this study. Transcripts present in the *GAL* locus in induced and repressed states. The red solid lines represent the transcripts published, SUTs, CUTs or XUTs. The red dotted lines represent the new transcripts unveiled in this study.

Chapter 4

Cellular localization of sense and antisense transcripts

4.1 Introduction

Organisms with the same genetic material and grown in the same conditions were thought to have identical expression of individual genes. However, it was found that gene expression, even in genetically identical organisms, can be rather diverse (Clark et al., 2011; Larson et al., 2009; Raj and van Oudenaarden, 2008). The fact that gene expression fluctuates between cells in a population has been known for a long time (Novick and Weiner, 1957; Raj and van Oudenaarden, 2008), however, the lack of methodology for proper single-cell analysis has meant that this field has only started to develop recently. Since then, several studies, from bacteria to higher eukaryotes, have established that gene expression is stochastic, i.e., noisy. It can be classified as extrinsic noise, when variations in the levels of the transcription machinery affects gene expression, or intrinsic, due to the variations in the random microscopic chemical reactions. Through experimental and mathematical studies it was concluded that in a given cell, both types of noise are required in order to have variability in gene expression (Elowitz et al., 2002). It has been suggested that this variability could drive processes such as the cell-cycle, metabolism, stress-response, development and circadian rhythms. However, good experimental data supporting this are lacking. After transcription, a pre-mRNA is processed, assembled into an mRNP, exported, translated and degraded. The regulation of each of these steps has been the subject of many years of research. Until recently the study of RNA biology has been restricted to whole populations, limiting the type of information that could be acquired. For instance, information on the spatial localization of the RNAs and the variation in expression between cells were lost. With the development of single-cell techniques, these questions can now be investigated. Among

these, RNA FISH (fluorescent *in situ* hybridization) and the MS2 or related systems are used for the study of gene expression. The MS2 system consists of a sequence motif that is inserted at the 3' end of the gene of interest that can be bound by a fluorescent protein. The advantages are that it can be used in live cells and is the only technique that can visualize the kinetics of gene expression in real time. In contrast, the main advantage of FISH is that is does not require any genetic manipulations and it is quantitative. It consists of the hybridization of probes to the mRNA of interest. However, it can only be performed in fixed cells and is therefore limited to the detection of the RNAs at single time points (Larson et al., 2009).

The observation that there is a considerate amount of stable ncRNAs being produced in a cell has raised the question of what regulation, if any, these RNAs are subjected to. It is known that a large number of ncRNAs are processed by the addition of a polyA tail and therefore it can be assumed that these are exported to the cytoplasm. However, no studies have yet addressed the question of the cellular localization of the ncRNAs. In this chapter, the detection of the *GAL1* antisense transcripts through the use of the RNA FISH methodology will be tested and analysed.

4.2 The *GAL1* sense transcript is easily detectable by FISH

To answer the question of where the RNAs are spatially localized, the technique of RNA fluorescent *in situ* hybridization (RNA FISH) was used. The protocol used was adapted from Zenklusen et al. (Zenklusen and Singer, 2010), using probes to detect transcripts from sense or antisense strand of *GAL1* designed for this study. Four probes were

designed for each strand. The DNA probes of ~50 nt and ~50 % GC content were carefully selected to avoid cross-hybridization to *GAL3* and five modified bases (amino-allyl dT) spaced by about ~10 nt, for the incorporation of the fluorophore, were added. The fluorophores used for this experiment were cyanine-5 (Cy5) for *GAL1* sense and cyanine-3 (Cy3) for *GAL1* antisense detection.

Firstly, the technique was tested using probes for the *GAL1* sense transcript in WT cells, which were grown to logarithmic phase in glucose or induced for 30 minutes and 3 hours in galactose. This allows a comparison to be made between signals in the repressed state and two times points during induction. The cells were grown and induced as described in section 2.3, cross-linked with paraformaldehyde and spheroplasts obtained with lyticase enzyme, until >80% digestion of the cell wall was achieved. The cells were then washed and attached to poly-L-lysine treated coverslips. The sense probes, at a concentration of 0.5 ng each, were then hybridized to the cells on the slides which were washed, DAPI treated and imaged. The imaging was done in collaboration with the laboratory of Ilan Davis using a quantitative microscope (DeltaVision CORE: Wide-field fluorescence deconvolution imaging).

Figure 22 (A), shows examples of the images obtained with DAPI staining and the *GAL1* sense Cy5 labeled probe (average projections of three planes). DAPI (blue) is used to demarcate the spatial localization of the nucleus by the strong binding of the fluorophore to AT rich regions of the DNA. The top left panel shows a DIC (differential interference contrast) image of spheroplasted cells with no RNA FISH probes, merged with DAPI, independently taken with a Zeiss Axioplan 2 Imaging microscope. This 3D-view of the cells obtained with polarized light allows observation of the contours of the cells as well as the

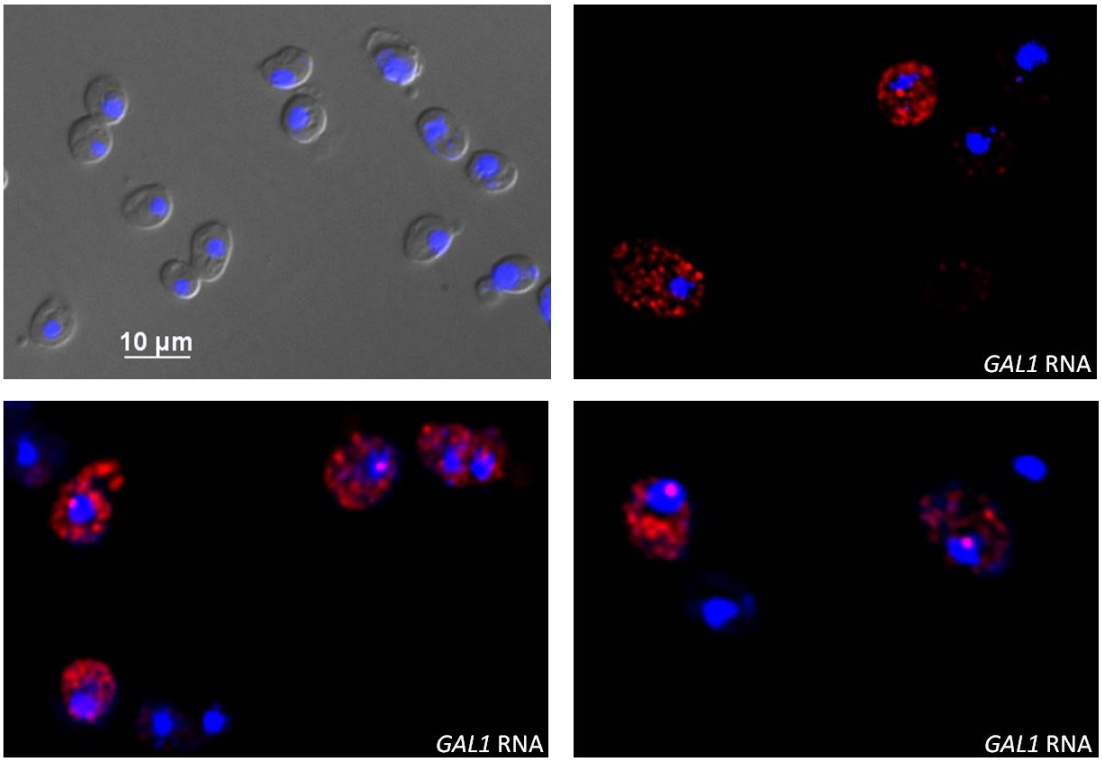
internal morphology, and when merged with DAPI, the demarcation of the nucleus. The remaining images are representative merged views of the DAPI staining and the Cy5 labelled *GAL1* sense probes (red) on cells induced for 3 hours with galactose. The quantitation of the RNA FISH raw data was done using the ImageJ software. For each intact cell, the plane with the maximum signal in the Cy3 and Cy5 (for double labelling experiments) channel was selected and the mean signals measured and recorded. Alternatively, a selection of slices (planes 10-25) which included the nucleus and most of the cytoplasm was maximally projected to retrieve the maximum signals. The error bars were calculated from standard deviations using the T-test and a confidence interval of 0.01 as indicated.

The detection of *GAL1* RNA with the labelled probe was successful by the observation of multiple red speckles. The first observation is the fact that the speckles are observed throughout the cell, an indication that the RNA is present in the cytoplasm, as expected. The speckles in the cytoplasm could either be the site of translation or P-bodies (processing bodies). Speckles of higher intensity, indicating the presence of more than one transcript, are also observed, with some co-localizing with the DAPI signal, indicating that they might be inside the nucleus. Previously published results for the *GAL* genes have identified these speckles as being localized to the nucleus and the nuclear periphery, where they are known as dot RNAs (Vodala et al., 2008), and for the purpose of this study, speckles that clearly co-localize with DAPI staining will be referred to as being in the nucleus (Abruzzi et al., 2006; Gandhi et al., 2011). The second important observation is the fact that not all cells are positive for the *GAL1* transcript. Although the lack of signal in some cells could be explained by inefficient spheroplasting, other studies

have shown that *GAL1* expression is not observable in all cells of a population (Larson et al., 2009; Raj and van Oudenaarden, 2008; Zenklusen et al., 2008).

Figure 22 (B) shows the result of the quantitation analysis for *GAL1* sense expression between cells induced for 30 minutes or 3 hours in galactose and cells grown in glucose for the same time period (images not shown). In glucose medium and after 30 minutes of induction, the *GAL1* mRNA transcript is not detectable by Northern blots (see for example Figure 29). However, the ncRNA transcribed antisense to *GAL10* but sense to *GAL1* (described as transcript **f** in the previous chapter) will also be detectable with the *GAL1* sense RNA FISH probes. Thus a signal is expected in these cells but should increase after 3 hours of induction in galactose. The mean intensity detected for the *GAL1* sense transcripts in cells induced for 3 hours is higher than at 30 minutes and in glucose, meaning that the RNA FISH methodology is capable of detecting the differences in transcript intensity also detectable by Northern blot.

A)



B)

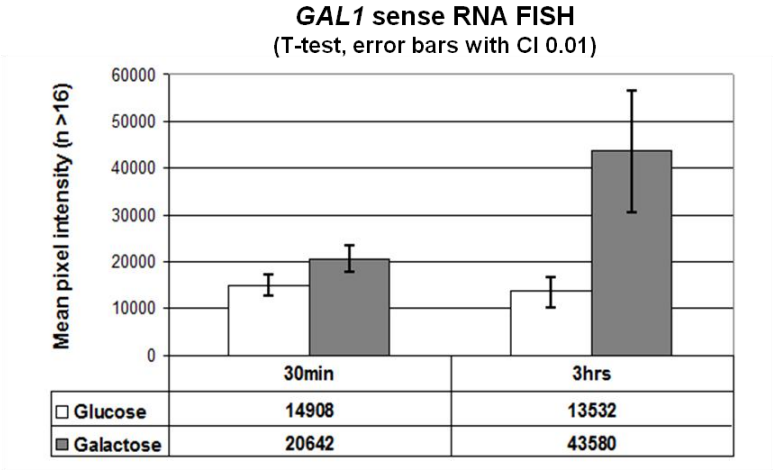


Figure 22. *GAL1* sense RNA FISH in induced and repressed cells. WT cells were grown in glucose as described in section 2.3 and harvested at zero minutes (glucose time point), 30 minutes and 3 hours of galactose induction. **(A)** GAL1 sense transcripts (red), induced for 3 hours, were detected with cyanine 5 labelled probes. Three different images of the same experiment shown, annotated with *GAL1* RNA. DAPI (blue) was used to identify the nucleus. Images were obtained with an average projection of 3 layers. Top left panel shows a differential interference contrast (DIC) image of spheroplasted cells with no RNA FISH probes, merged with DAPI, independently taken. **(B)** Comparison of the mean intensity signal of GAL1 sense transcripts in induced versus repressed conditions and 30 minutes versus 3 hours of induction. Error bars with confidence interval of 0.01 were calculated using the T-test. RNA FISH done by Simon Haenni.

4.3 Detection of ncRNAs at *GAL1* in a *kem1Δ* strain

Since the induced sense transcripts were successfully detected by RNA FISH the next step was to investigate whether it was possible to detect the ncRNAs transcribed on the sense or antisense strands of *GAL1*. Since most of the antisense transcripts have been described as being polyadenylated (this study and (Xu, 2009)) it is then possible that they are also exported. Since the previous data have shown that not all cells in a population induce the *GAL* genes in a synchronized manner, the same question could be posed for the antisense transcripts. Are they produced in only some cells of the population as observed for the sense transcripts? Are they all exported or does a proportion remain in the nucleus? Since most of the antisense transcripts are present at low steady-state levels, to increase the probability of detection the *kem1Δ* strain was used. As shown in chapter 3 (Figure 20), two *GAL1* antisense transcripts (transcripts **i** and **j**) as well as the *GAL10* antisense-GAL1 sense transcript (transcript **f**) were stabilized in the *kem1Δ* strain.

For the RNA FISH experiment, both sense and antisense probes were used in cells grown to log phase in glucose and induced for 90 minutes in galactose. To control for autofluorescence (Neg Control), cells grown in glucose were used, but no probe was added to the hybridization mixture. Figure 23 shows that in induced cells (*kem1Δ* Gal) the *GAL1* sense transcripts (red) are visible throughout the cell, as in the WT strain (Figure 22), indicating that the transcripts are exported as expected. From the analysis (Figure 23 B) it is possible to see that the *GAL1* sense transcript represents the highest signal detected. Not all the cells in the population have high levels of *GAL1* RNA, as discussed before. The *GAL1* antisense transcripts (green) are just detectable in the images (Figure 2A) and the analysis (Figure 23 B) shows that the signal is lower than for the sense

transcript, as anticipated, although higher than the control, indicating that it is above background. As known by the Northern blotting experiments, the *GAL1* antisense transcripts are XUTs and although they cannot be detected after three hours in galactose, they are present in glucose and during galactose induction.

In repressed conditions (*kem1Δ* Glu), both sense and antisense transcripts are detected with similar levels. Although the *GAL1* mRNA is not present in glucose, the *GAL10-1* antisense transcript (transcript **f**) that reads through *GAL1* on the sense strand is present and is stabilized in a *kem1Δ* strain. In fact, preliminary data recently obtained in the laboratory by Simon Haenni using *GAL10* antisense and *GAL1* sense RNA FISH probes indicated co-localization of the two signals on the same transcript (data not shown). For the *GAL1* antisense transcripts, the signal is similar to that of the *GAL1* sense transcripts in glucose and the antisense in induced conditions. This indicates that the detection of the antisense transcripts by RNA FISH was achieved.

In summary, the *GAL1* antisense transcripts in a *kem1Δ* strain can be detected with RNA FISH probes, although further optimization of the technique to reduce the background noise is required. Interestingly, the visual analysis of the data obtained and further data obtained in the laboratory by Simon Haenni, suggests that cells containing high levels of sense transcripts have very low signals for antisense transcripts, while cells with low sense transcript level have higher antisense signals. This might suggest reciprocal expression of sense and antisense transcripts. However, the fact that this technique measures the amount of transcripts present in a cell in a given time and not the rate of transcription *per se* need to be taken into account. Further experiments will be done to

establish if in fact there is an anti-correlation between cells expressing sense transcripts and cells expressing antisense transcripts.

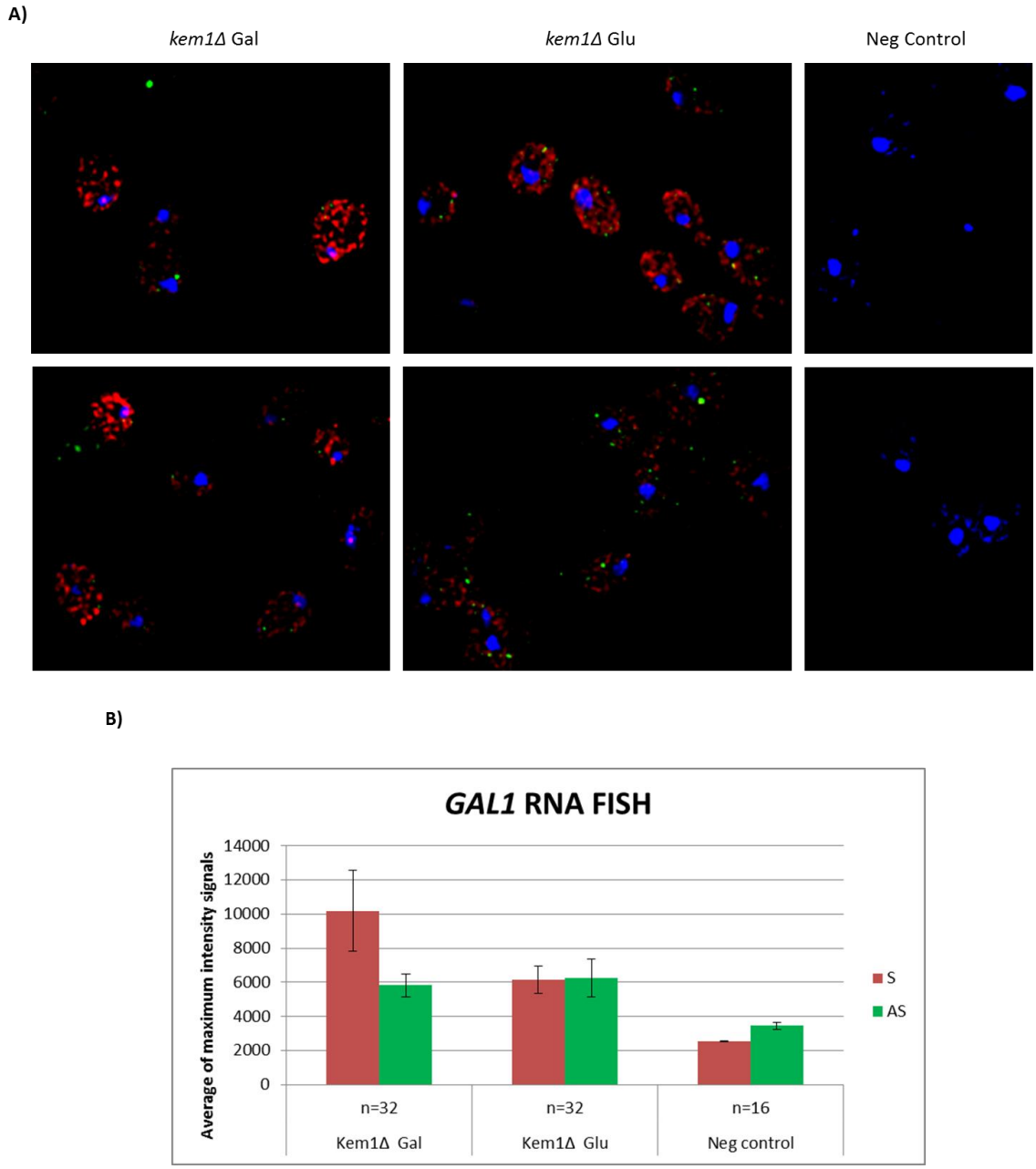


Figure 23. *GAL1* sense and antisense RNA FISH in a *kem1Δ* strain. Cells were grown as described in section 2.3. *kem1Δ* cells were harvested after growth to steady state in glucose (*kem1Δ* Glu) and after 90 minutes of galactose (*kem1Δ* Gal) induction. **(A)** *GAL1* sense (red) and *GAL1* antisense transcripts (green) were co-detected with cyanine 5 and cyanine 3 labelled probes, respectively. DAPI (blue) was used to identify the nucleus. Two separate images, for each condition, obtained with an average projection of 3 layers. **(B)** Comparison of the average of maximum intensity signal of planes 10-25 of *GAL1* sense and antisense transcripts in induced versus repressed conditions. Number of cells analysed identified by the n=x. Error bars obtained from standard deviation with confidence interval of 0.01.

4.4 Identification of sense and antisense nuclear dot RNA

In the RNA FISH experiments shown here, distinct fluorescent speckles were observed in co-localization with the DAPI staining, and as discussed before, these have been described as being inside the nucleus (Abruzzi et al., 2006; Gandhi et al., 2011). Their nature, however, is not fully understood, with the Singer laboratory claiming that they are the transcription site (TS) (Gandhi et al., 2011; Zenklusen et al., 2008; Zenklusen and Singer, 2010), while the Rosbash laboratory says they localize "near" the TS but that it contains not only nascent but also retained transcripts (Abruzzi et al., 2006; Vodala et al., 2008). The presence or absence of dot RNA could be indicative of a different fate for the transcripts. Transcripts characterised as XUTs would not be expected to be retained in the nucleus, as they are degraded in the cytoplasm in P bodies. However the *GAL10-1* long non coding RNA is both a SUT and a XUT and thus the SUT fraction could be retained in the nucleus. In contrast the *GAL1* antisense transcripts are mainly XUTs and would not be expected to be retained. Figure 24 shows the images obtained for *kem1Δ* cells in induced and repressed conditions with the brightest speckles. In induced conditions the co-localization of a *GAL1* sense dot RNA (indicated by the white arrows) with the DAPI staining is easily observed. In glucose, dot RNA is not observed, suggesting that the *GAL10-1* antisense ncRNA is not retained in the nucleus. As expected the *GAL1* antisense signal was not detected in the nucleus in either inducing or repressing conditions, consistent with it being annotated as a XUT. Bright spots in the cytoplasm (yellow arrows) are consistent with localization of the *GAL1* antisense transcript in P bodies. Although these data suggests that transcripts detected by *GAL1* sense or antisense probes are not found as nuclear speckles or dot RNAs, other data obtained in the laboratory by Simon

Haenni suggests that the *GAL10* antisense probes do detect nuclear dot RNA in glucose grown cells (data not shown).

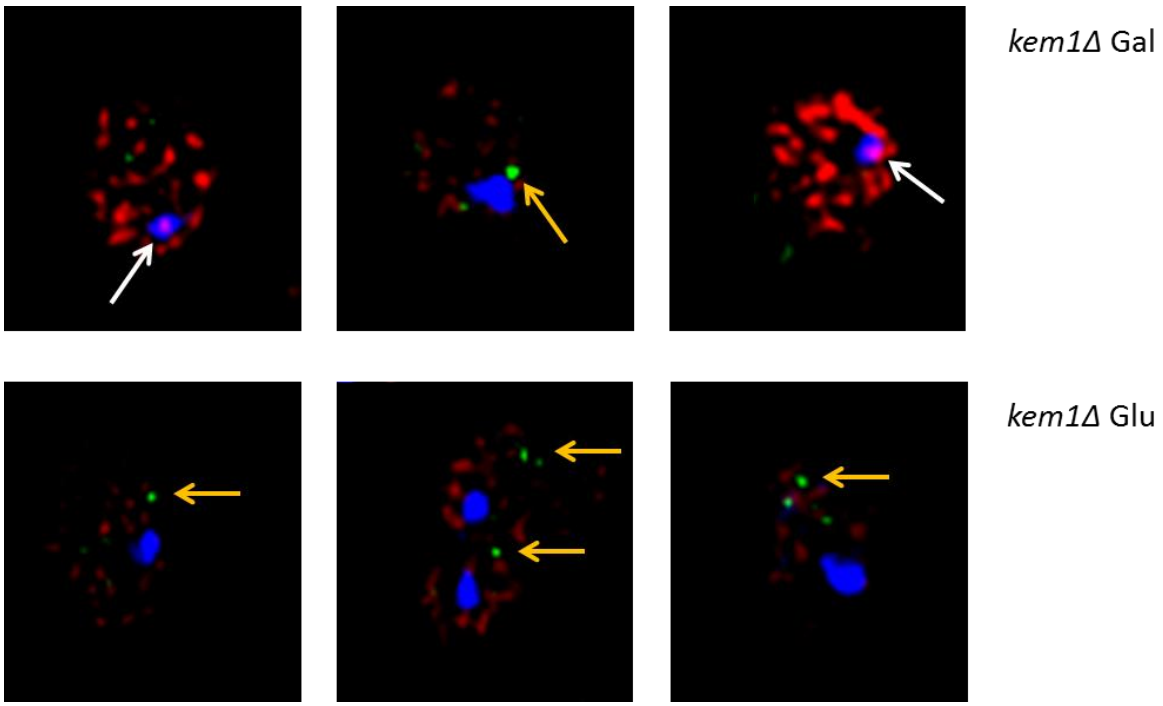


Figure 24. *GAL1* sense nuclear RNA dots in a *kem1Δ* strain. Cells were grown as described in section 2.3. *kem1Δ* cells were harvested after growth to steady state in glucose and after 90 minutes of galactose induction. *GAL1* sense (red) and *GAL1* antisense transcripts (green) were detected with cyanine 5 and cyanine 3 labelled probes, respectively. DAPI (blue) was used to identify the nucleus. White arrows identify *GAL1* nuclear RNA dots and yellow arrows identify antisense RNA dots. Images obtained with an average projection of 3 layers.

4.5 Discussion

Recently, methods for single-cell transcriptional studies have been developed and with the new evidence for stable ncRNAs present throughout the genome, including antisense transcripts, the question of where these transcripts localize in a cell, whether they are exported and if they are expressed in all cells in a population are novel questions that have not yet been addressed. This chapter introduced the early developments made in order to answer these questions. The RNA FISH methodology was chosen since it is the only technique that can detect both sense and antisense transcripts at the same time and without any genetic manipulations.

It was concluded that RNA FISH can indeed detect low abundance transcripts, as claimed by Zenklusen et al., such as the *GAL* antisense RNA (transcript **i**), and probably the *GAL10-1* antisense ncRNA (transcript **f**). These initial experiments were performed using the *kem1Δ* strain, since the antisense transcripts are stabilized, therefore, increasing the chance of detection. The results also support the presence of *GAL1* sense nuclear dot RNAs, identified in several studies (Abruzzi et al., 2006; Gandhi et al., 2011; Zenklusen et al., 2008), whose properties are still debatable particularly the disputed roles in transcription versus RNA export.

The results also hint at another interesting phenomenon; anti-correlative expression of the antisense and sense transcripts, meaning that one cell at a given time will be transcribing either the sense transcript or the antisense transcript but not both. This would eliminate the issue raised about the extremely complex regulation that the cells would have to have in order to prevent, for example, colliding RNA polymerases (Saeki and Svejstrup, 2009). This also raises the question of whether this anti-correlation might

be due to a regulatory mechanism exerted either by the sense or antisense transcription in repressing the other, making sure that they are not produced at the same time.

More experiments need to be done, with different sets of sense and antisense transcripts, for example, *GAL10*, which produces several antisense transcripts as described in the previous chapter.

Many more questions still remain unanswered, for example, if cells only express sense or antisense transcripts at one given time and, if so, if it might be due to alternating bursts of sense and antisense transcription, and, whether there are cells that once committed to sense transcription are no longer able to switch back to antisense as long as inducing medium is present. Other questions, as, why are the ncRNAs exported from the nucleus; are there different types of ncRNAs based on whether they are exported or not / polyadenylated or not and what dictates the production of a sense or antisense transcript in a given cell, have not yet been addressed.

Chapter 5

Transcriptional memory at the *GAL* locus

5.1 Introduction

Unicellular organisms, like *S. cerevisiae*, need to change their transcription profile rapidly in response to environment cues. The ability to maintain this acquired transcription profile through several generations might provide the progeny with a selective advantage. The capacity to remember a previous transcriptional state is known as transcriptional memory. How the cells inherit this state remains controversial and several different mechanisms have been proposed (Brickner et al., 2007; Halley et al., 2010; Kundu et al., 2007; Kundu and Peterson, 2010; Morschhäuser, 2010; Zacharioudakis et al., 2007). The transcriptional memory mechanisms proposed can be divided into cytoplasmic inheritance, when it is a protein or small molecule present in the cytoplasm that provides memory, or nuclear inheritance, when the changes occur at the chromatin level. Changes in chromatin states can be accomplished by post-translational modifications of histones tails such as methylation, acetylation, phosphorylation, ubiquitination and SUMOylation or by deposition of histone variants (*e.g.* H2A.Z), among others. The histone modifications are associated with transcription states (*e.g.* activation or repression) and can be passed on to the daughter cells upon replication. The exact mechanism through which the histones and their modifications are propagated is not yet known (Kundu and Peterson, 2009). One example of cellular memory is the "white-opaque" cell type switching in *C. albicans*, in which the cells can reversibly switch from their normal morphology (white) to an elongated form (opaque), an adaptive response that contributes to the infection and colonization of many body locations. The phenotypic change is accompanied by a genome-wide alteration in the transcription profile that seems to be regulated by the histone deacetylases Hda1p and Rpd3p (Morschhäuser, 2010). Another example of

cellular memory is the rapid re-activation of transcription of the *GAL* genes in *S. cerevisiae*. If the cells are grown in galactose-containing media, switched to glucose then back in galactose, they will activate the *GAL* genes much faster than if they had had no previous exposure to galactose. Until recently, there were three different mechanisms proposed for the transcriptional memory exhibited by the *GAL* genes including; the histone remodelling enzyme, SWI/SNF (Kundu et al., 2007); H2A.Z-mediated localization to the nuclear periphery (Brickner et al., 2007); and the cytoplasmic enzyme Gal1p (Zacharioudakis et al., 2007). Zacharioudakis et al. have suggested that the Gal1 protein is solely responsible for galactose memory (Zacharioudakis et al., 2007). They suggest that the Gal1 protein can be passed on through 6 to 7 generations, from the cytoplasm of the mother to the daughter cell and quickly reactivate transcription of the *GAL* genes through Gal4p. In light of this data the Peterson group, responsible for the SWI/SNF mechanism, have withdrawn their previous conclusions to support the inheritance of the cytoplasmic Gal1p (Kundu and Peterson, 2010). Furthermore, another publication by Rine's group has refuted the H2A.Z-mediated memory mechanism (Halley et al., 2010). This highlights the complexity and difficulties within this field. Moreover, the Gal1 protein inheritance mechanism, which is currently favoured, has not been tested rigorously. For example, the study of Zacharioudakis et al. (Zacharioudakis et al., 2007) failed to investigate the half-life of Gal1p and whether it was still present after 6-7 generations to exert the memory effect.

In this chapter, the concept that the Gal1 protein is required for memory is tested. There are a number of transcripts that could potentially give rise to the Gal1 protein that might be involved in transcriptional memory. These include the *GAL1* mRNA itself, as well as the

ncRNAs transcribed over the *GAL1* gene in repressed conditions and early during induction (the *GAL10-1* antisense transcript) or during induction (the *GAL1-FUR4* read-through transcript). To test the idea that the Gal1 protein, as opposed to the *GAL1* RNA or some other feature of the chromatin, is required for transcriptional memory, two different approaches were taken; the introduction of a frameshift mutation in the *GAL1* gene, prematurely truncating the Gal1 protein but leaving the *GAL1* mRNA nearly intact or the premature truncation of the Gal1 protein and RNA by insertion of a transcriptional terminator.

5.2 Establishing conditions to demonstrate transcriptional memory

The rapid induction pattern associated with transcriptional memory when cells are re-exposed to galactose was first examined in a WT strain. For this experiment, the WT strain was inoculated overnight in YP supplemented with 2% galactose (YPG), washed and then allowed to grow in YP supplemented with 2% glucose (YPD) for 7 hours, before being re-exposed to YPG and harvested at the time points indicated. The expression of the *GAL1* RNA in the non-memory and memory conditions was then assessed. Upon a first exposure to galactose (Figure 25, Non-memory), *GAL1* induced slowly with the transcript starting to appear within one hour of induction. Also very clear in Figure 25 (Non-memory) is transcript **f**, which initiates at the *GAL10* internal promoter and terminates at the end of *GAL1* (Figure 25 C).

The second exposure to galactose should give rise to a faster induction consistent with a galactose memory phenotype, and in fact, as Figure 25 (Memory) illustrates, *GAL1* induces within 5 minutes of exposure to galactose. Furthermore, the expression levels increase rapidly and by 1 hour the levels of induction are very high. Again, transcript **f** is visible and so is transcript **c**, which starts at the *GAL1* promoter and terminates at the end of *FUR4* (Figure 25 B). Transcript **c** is visible after 5 minutes of induction and its levels increase throughout the induction. This indicates that the expression of this transcript increases along with the expression of *GAL1*, leading to the idea that it is regulated by the same mechanisms as *GAL1*.

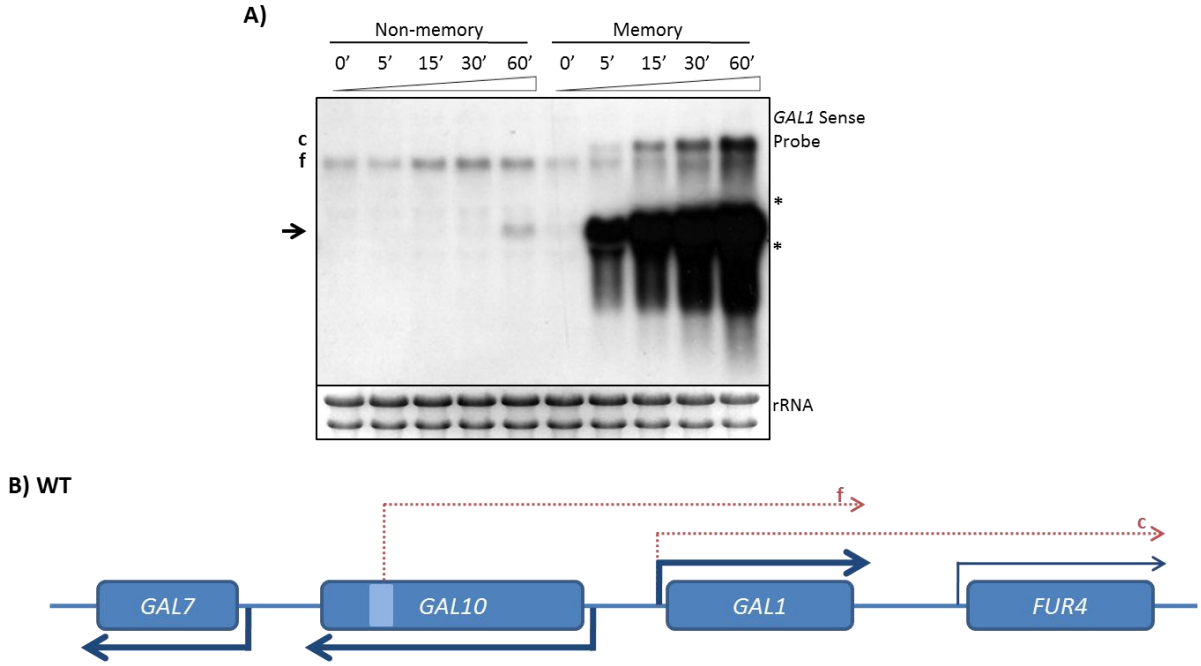


Figure 25. Galactose induction and re-induction of *GAL1* in the WT strain. (A) Northern blot analysis of total RNA extracted from a WT strain in a galactose induction (Non-Memory) and re-induction (Memory) at the time points indicated. For the galactose re-induction, a WT strain was inoculated overnight in 2% YPG, washed and then allowed to grow in 2% YPD for 7 hours before being re-exposed to 2% YPD and harvested at the time points indicated. The GAL1 coding transcript is identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gel (rRNA) is shown for loading control. **(B)** Schematic of the profile of transcripts identified in the Northern blot above. Transcript start and end sites are only approximate.

Since the *GAL10-1* ncRNA, which is transcribed through the entire *GAL1* gene, is expressed in glucose and remains present during the induction for an unknown amount of time, it was wondered whether this RNA could be the source of "memory" during glucose conditions through a mechanism in *trans*. The initial experimental design envisaged the expression of the *GAL10* antisense ncRNA (transcript **f**) from the *ADH1* promoter on a centromeric plasmid. However, this sequence proved difficult to clone and the *GAL1* coding region and 3'UTR was used instead to ask whether it is the Gal1 protein or the *GAL1* RNA that mediates transcriptional memory.

5.3 The Gal1 protein is required for fast induction of the *GAL* locus

To differentiate between the Gal1 protein and the RNA, two plasmids were created; one carrying the wild-type *GAL1* and another carrying a mutant form of *GAL1*, where a unique restriction site (AvaI) at +88 bp, from the ATG codon, was used to insert four extra nucleotides, resulting in a frameshift mutation (cloning performed by Peter Dudek). Consequently, this mutation resulted in altered translation of the mRNA, which caused premature translation termination and resulted in a severely truncated protein. Importantly, the RNA sequence remained unaltered apart from the four inserted nucleotides. The wild-type and mutated *GAL1* genes were placed under the control of the *ADH1* promoter, so that they would be expressed in both glucose- and galactose-containing media. Transcriptional memory was assessed by measuring the rate of induction of the *GAL7* RNA.

The observations from Zacharioudakis et al. (Zacharioudakis et al., 2007) were tested by expressing WT *GAL1* from a plasmid under the control of the constitutively active *ADH1* promoter (*pADH1 GAL1*) in the WT strain. As a control, the vector alone was used in the same strain. Figure 26 (A, vector and *pADH1 GAL1*) shows that constitutive expression of Gal1p from the plasmid causes a rapid induction of *GAL7* compared to the control. This result is in agreement with the data shown in Zacharioudakis et al. (Zacharioudakis et al., 2007) and suggests that this system is appropriate to monitor transcriptional memory.

To control for the expression of *GAL1* from the plasmid, a *GAL1* probe, which hybridizes to the *GAL1* on the plasmid, as well as to the genomic *GAL1*, was used. The result showed the expression of *GAL1* (*pADH1 GAL1*), even in glucose, since the expression of the plasmid copy of *GAL1* is independent of the carbon source. The increase in expression observed over time following addition of galactose probably corresponds to the induction of genomic *GAL1.*

Low levels of the *GAL7* transcript was detected in glucose in the *pADH1 GAL1* strain, consistent with the previous findings (Chapter 3, Figure 8), however, at higher levels than in the vector control, perhaps caused by the constitutive expression of *GAL1* and consequent activation of the *GAL7* promoter.

This data supports a trans-acting factor, Gal1 protein or *GAL1* RNA, mediating transcriptional memory. In order to differentiate the role of Gal1p from the role of its RNA in the fast induction of the *GAL* locus, *GAL1* with a frameshift mutation was expressed from the *ADH1* promoter (*pADH1 GAL1*_FS) in a WT and a *gal1Δ* strain. This allows the discrimination between the WT RNA produced from the genomic *GAL1* and the mutated form from the plasmid. If the fast induction is due only to the activity of the Gal1

protein and not to sequences in the *GAL1* RNA, then expressing a mutated form should have no effect on induction. The vector alone was transformed into both backgrounds (WT and *gal1Δ*) and used as a control. When expressing the mutated form of *GAL1* (*pADH1 GAL1*_FS) in the WT and *gal1Δ* strains (Figure 26 A and B), the first observation is that *GAL7* does not induce as fast as when expressing the WT *GAL1* gene on the plasmid (*pADH1 GAL1*). This suggests that the frameshift mutation does not produce a functional protein and supports the Gal1 protein and not the RNA as being responsible for the fast induction. Intringuingly, the *GAL7* kinetics vary slightly between strains which could be explained by the fact that this is an heterologous population in which not all cells will have the plasmids.

Analysis of the data for the *GAL1* transcript in the strains expressing *pADH1 GAL1*_FS (Figure 26 A and B) revealed an interesting difference. In the *gal1Δ pADH1 GAL1*_FS strain, consitutive expression of the transcript from the plasmid is detectable. This suggests that even though the translation is prematurely terminated, the transcript is not targeted for nonsense mediated decay in glucose (Mühlemann et al., 2008). However, levels of the transcript gradually decrease over the induction period in the *gal1Δ pADH1 GAL1*_FS strain but the reason for this is not known. In contrast, in the WT background there is no evidence for the expression of *GAL1*_FS RNA from the plasmid. It is not clear if this is a result of a technical issue or some more complex trans-silencing of expression (Berretta et al., 2008). In conclusion, in the absence of genomic *GAL1*, expression of *GAL1*_FS from a plasmid is not sufficient for fast induction of *GAL7* expression. This strongly supports a role for the Gal1 protein in transcriptional memory.

Analysis of the data in Figure 26 (A and B) raised another concern. The expression profile of *pADH1-GAL1* from the plasmid in the *gal1Δ* background shows steady state levels of *GAL1* in glucose and early during induction, consistent with the constitutive expression from the *ADH1* promoter. However, at 3 hours of induction there is a large increase in the steady-state levels of this *pADH1-GAL1* transcript. One explanation is that this transcript is stabilized and accumulates later in induction. However, although this signal is not due to genomic *GAL1* (since the control with the vector alone shows that no *GAL1* RNA is detected), it could be due to a cross-hybridization with *GAL3*, since the long *GAL1* probe was used. This concern about whether the *GAL1* probe could be hybridizing to *GAL3*, led to the use of further controls including a *gal3Δ* strain and a double delete *gal1Δgal3Δ* strain.

In Figure 26 (C and D) the first observation was that there was no expression of either *GAL1* or *GAL7* in the *gal3Δ* and *gal1Δgal3Δ* strains with the vector alone when induced. This is in accordance with the literature which shows that a *gal3Δ* strain displays a very slow *GAL* gene induction over several hours and that a *gal1Δgal3Δ* strain is non-inducible (Broach, 1979; Torchia and Hopper, 1986). Secondly, *GAL7* expression was recovered in these strains by expressing the WT *GAL1* but not the mutated form, further confirming that the mutated *GAL1* RNA is not responsible for transcriptional memory. Thirdly, Figure 26 (C and D) shows that *GAL1* expression from the *ADH1* promoter on the plasmid is sufficient to overcome the *gal3Δ* slow induction phenotype and the *gal1Δgal3Δ* non-induction phenotype (Bhat et al., 1990). Finally, the increase in plasmid-expressed WT *GAL1* RNA observed at 3 hours of induction in the *gal1Δgal3Δ* strain (Figure 26 D, *pADH1 GAL1*) suggests that this is independent of genomic *GAL3*. Interestingly, this increase in

steady-state levels was not observed for the mutant *GAL1* RNA in the *gal1Δgal3Δ* strain (Figure 26 D) or the *gal1Δ* strain (Figure 26 B), supporting a role for the Gal1 protein in this phenomenon. When analysing results obtained by Northern blotting it is necessary to remember that it shows the total amount of RNA in a population at a given time, meaning that it might reflect the number of cells responding in a population and/or changes to the stability of transcripts but not always the act of transcription *per se*. In this experiment, the increase in levels of the *GAL1* transcript at 3 hours could then be explained if, in a Gal1p-dependent way, the transcript accumulates in the cells.

In summary, these data a) confirm the involvement of Gal1p in the transcriptional memory; and b) suggest that *GAL1* RNA is not responsible for a fast induction of the *GAL* gene cluster, although it should be pointed out that the RNA expressed from the mutant gene (*pADH1 GAL1*_FS) contains a 4 nucleotides insertion and is not wild-type. Further experiments need to be done in order to elucidate the relationship between the genomic and plasmid-bound copies of *GAL1* and the involvement of *GAL3*. It would be important to characterize the expression profile of the plasmid-bound copy of *GAL1* in detail, to assess whether there are effects on RNA stability during induction and importantly to assess the levels and localization of the Gal1 protein in the various genetic backgrounds.

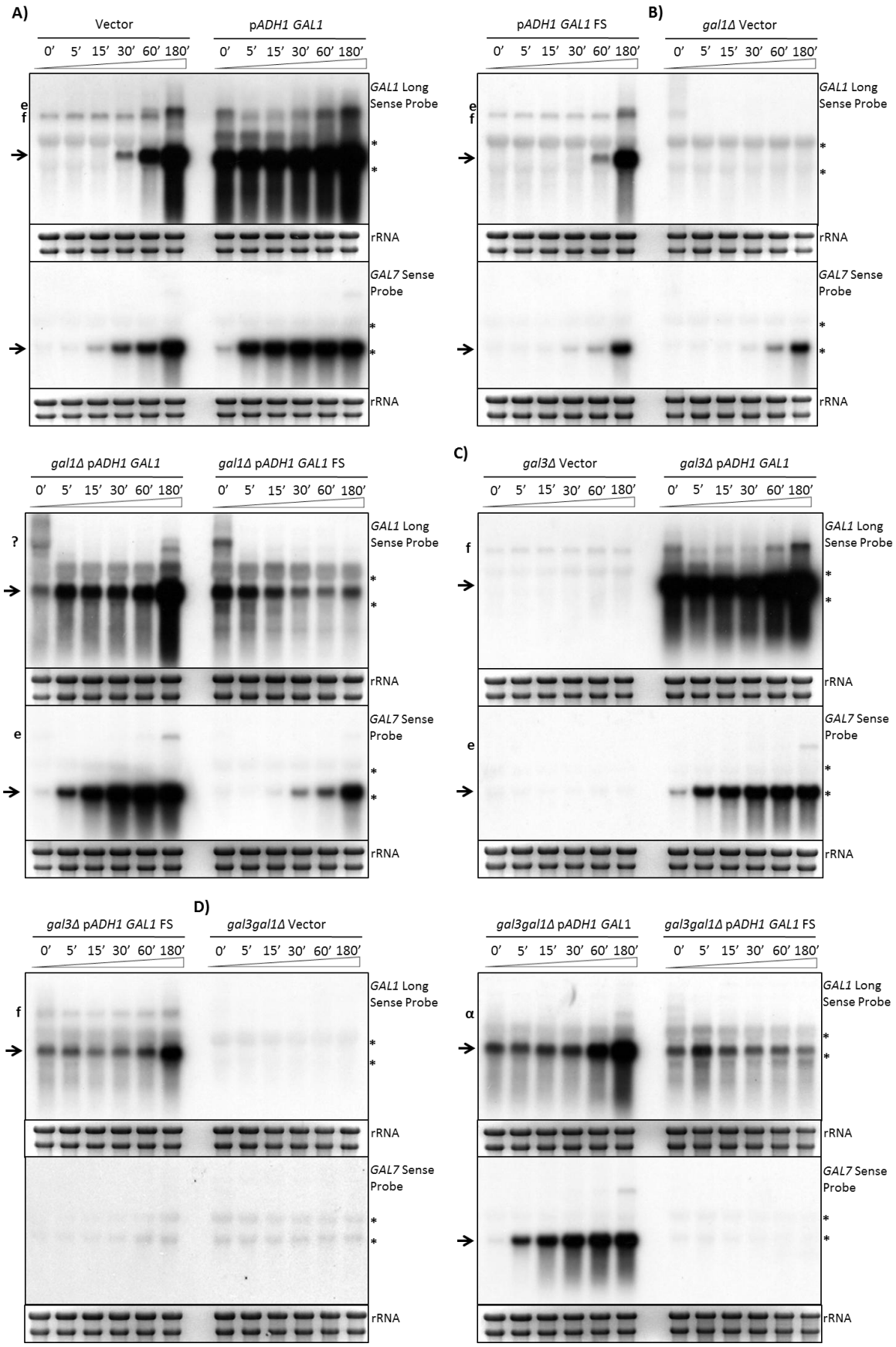

A)
Vector
pADH1 GAL1
0′ 5′ 15′ 30′ 60′ 180′
GAL1 Long Sense Probe
rRNA
GAL7 Sense Probe
B)
pADH1 GAL1 FS
gal1Δ Vector
C)
gal1Δ pADH1 GAL1
gal1Δ pADH1 GAL1 FS
gal3Δ Vector
gal3Δ pADH1 GAL1
D)
gal3Δ pADH1 GAL1 FS
gal3gal1Δ Vector
gal3gal1Δ pADH1 GAL1
gal3gal1Δ pADH1 GAL1 FS

Figure 26. The *GAL1* RNA is not involved in the transcriptional mechanism. Northern blot analysis of RNA extracted from cells induced for the time points indicated and probed for *GAL1* and *GAL7* sense transcripts. The cells were grown following the same method described in 2.3 but minimal media was used instead of YP. Different strains were transformed with a plasmid containing WT *GAL1* (p*ADH1 GAL1*) or *GAL1* with a frameshift mutation (p*ADH1 GAL1*_FS) under the control of the *ADH1* promoter. The vector alone was used as a control. The coding transcripts are identified by a black arrow. Transcripts **c**, **e** and **f** are described in chapter 3. **α** corresponds to a truncated version of transcript **f**. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. **(A)** Effect of expression of WT *GAL1* and mutated *GAL1* in a WT strain. **(B)** Effect of expression of WT *GAL1* and mutated *GAL1* in a *gal1Δ* strain. **(C)** Control for cross-reactivity of the *GAL1* probe with *GAL3* using a *gal3Δ* strain. **(D)** Control for interaction between *GAL1* and *GAL3* by using a *gal1Δgal3Δ* strain.

5.4 The COOH terminal region of Gal1 protein is not required for transcriptional memory.

The results presented in the previous section strongly support a role for the Gal1 protein, and not the *GAL1* RNA in transcriptional memory *in trans*. However, the possibility that the sequence of the RNA around which the frameshift was created is important for memory needed to be addressed. To do this, transcriptional memory was tested using a strain expressing a genomic truncation of the *GAL1* gene made by inserting the *ADH1* terminator (*ADH1*^{T}) at +757bp from the *GAL1* ATG codon. In this strain, the 5' region of the *GAL1* RNA will be expressed intact without full length Gal1 protein. The Gal1 protein is 528 amino acids long and the insertion of the *ADH1*^{T} results in a truncation at the residue Glu-252. Since the truncation is located in the middle of the protein it is predicted to disrupt the active site, which even though located predominantly in the N-terminal domain also contains residues from the C-terminal domain (Thoden et al., 2005). To confirm that the strain expressing genomic *GAL1 ADH1*^{T} is unable to catabolise galactose, equal concentrations of 1:10 dilutions of a WT strain and the *GAL1 ADH1*^{T}strain were plated onto an YPD and YPG-containing agar and allowed to grow for 2 days. Figure 27

shows that both the WT and *GAL1 ADH1*^{T} strains grow well and at the same rate in YPD (glucose). However, the *GAL1 ADH1*^{T} strain grows much more slowly on YPG (galactose) compared to the WT strain. This result indicates that the *GAL1 ADH1*^{T} strain is deficient for growth in galactose, indicating that either it has lost the galactokinase activity or it is severely impaired.

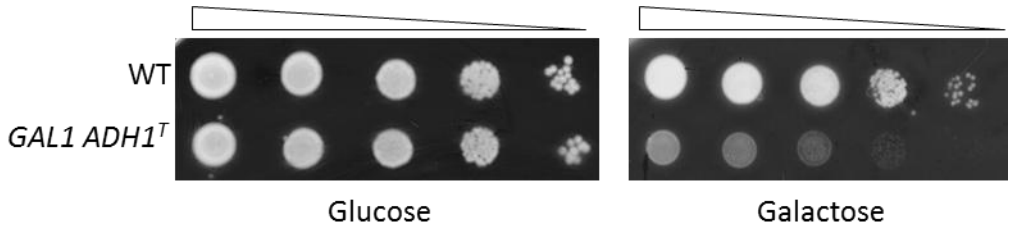


Figure 27. The *GAL1 ADH1*^{T} strain growth is impaired in galactose-containing media. The WT and *GAL1 ADH1*^{T} strains were spotted onto YPD (Glucose) or YPG (Galactose) plates and grown for 2 days. The cells were grown overnight in YPD, washed and then the same concentration of several serial 1:10 dilutions were performed and spotted onto the plates.

Next the *GAL1 ADH1*^{T} strain was tested for transcriptional memory by performing a galactose induction in "non-memory" conditions, where the cells have never been exposed to galactose or in "memory" conditions, where they have been pre-exposed to the sugar before being grown in glucose for 7 hours and then re-induced in galactose media. Figure 28 shows the galactose induction and re-induction of the *GAL1 ADH1*^{T} truncated strain. It is observable that when the cells have not been pre-exposed to galactose (non-memory conditions), the truncated *GAL1* induces at a normal rate, as compared to the WT strain (Figure 25), with the transcript appearing by 1 hour of induction. When the cells have been previously exposed to galactose, the truncated *GAL1* transcript is induced much faster and is observed at 5 minutes of induction. Even though

the levels, compared to the WT (Figure 25), are lower throughout the induction, the rapid kinetics observed is a characteristic of transcriptional memory. The *GAL10-1* antisense ncRNA is visibly truncated (transcript **f^**) and its levels remain constant in the *GAL1 ADH1*^{T} strain in the two growth conditions.

These results, together with the inability of the *GAL1 ADH1*^{T} strain to grow on galactose-containing medium, indicate that either the 5' region of the *GAL1* gene, the 5' region of the *GAL1* RNA or a truncated kinase-deficient Gal1 protein are sufficient for transcriptional memory when expressed at the *GAL* locus. It remains to be determined whether the galactose and ATP-binding domains of the truncated Gal1 protein are properly structured and capable of binding to their substrates. This result is quite unexpected and suggests that the requirements for transcriptional memory have yet to be fully explained.

Several experiments need to be done in order to answer some of the new questions raised by this experimental data. To completely rule out the involvement of Gal3p in transcriptional memory, *GAL3* should be deleted in the *GAL1 ADH1*^{T} strain and the strain tested for fast induction. A fast induction of the truncated *GAL1* would be expected, based on the fact that *gal3Δ* cells exhibit transcriptional memory when the *GAL1* ORF is expressed in *trans*. To further investigate the importance of the first 757bp of *GAL1*, more truncations closer to the *GAL1* ATG should be done to establish the minimum *GAL1* sequence required for transcriptional memory. To investigate whether it is the truncated protein that is responsible for the memory, the truncated and WT gene should be tagged and a Western blot performed to check protein levels. Finally, the possibility that there is

more than one mechanism by which transcriptional memory can be maintained needs to be considered.

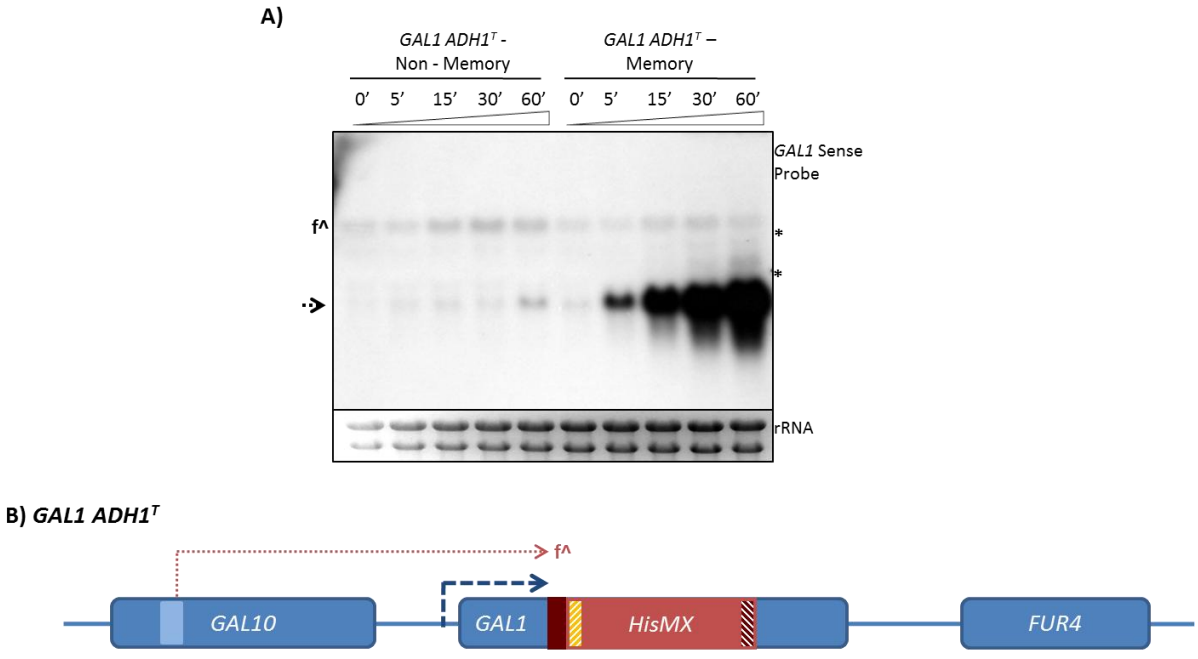


Figure 28. Galactose induction and re-induction of a truncated *GAL1* strain. (A) Northern blot analysis of total RNA extracted from a GAL1 *ADH1*^{T} strain in a galactose induction (Non-Memory) and re-induction (Memory) at the time points indicated. For the galactose re-induction, the GAL1 *ADH1*^{T} strain was inoculated overnight in 2% YPG, washed and then allowed to grow in 2% YPD for 7 hours before being re-exposed to 2% YPD and harvested at the time points indicated. The *GAL1* truncated transcript is identified by a dotted black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gel (rRNA) is shown as a loading control. **(B)** Schematic of the profile of transcripts identified in the Northern blotabove. Transcript start and end sites are only approximate.

5.5 Discussion

In this chapter, the mechanism underlying transcriptional memory was investigated. One of the experiments here shown supported a role for the Gal1 protein, acting *in trans*, in transcriptional memory. The second experiment showed clearly that the intact Gal1 protein is not required for memory. One possibility to reconcile these differences is that there is, in fact, more than one mechanism by which transcriptional memory can be maintained and this would unify the diverse observations in the published literature. Currently, four mechanisms, acting alone or in cooperation, could explain the results obtained here:

a) The intact Gal1 protein, with galactokinase activity, functions *in trans* to activate Gal4 and is sufficient for transcriptional memory and fast induction.
b) Only the first 252aa of Gal1p, lacking the galactokinase activity but with the ability to bind ATP and galactose, functions *in trans* to activate Gal4 and is sufficient for transcriptional memory and fast induction.
c) Transcription (mRNA or ncRNA) of the promoter, 5'UTR and/or first 757bp of *GAL1* and the subsequent effects on the chromatin structure, is sufficient for transcriptional memory and fast induction.
d) A mechanism involving only the chromatin structure spanning the promoter, 5'UTR and/or first 757bp of *GAL1* acting *in cis* at the *GAL* locus, is sufficient for transcriptional memory and fast induction (Zhou and Zhou, 2011).

The 252 aa truncated protein has an impaired galactokinase activity since it is not able to grow on galactose. However, if the truncated protein is folded properly, it is possible that the "Gal3p-like" domains, present in the N-terminal of Gal1p, are still able to bind ATP

and galactose and are capable of de-repressing the Gal4p-Gal80p complex and promoting rapid induction of the *GAL* locus. Furthermore, the result obtained with the *GAL1 ADH1*^{T} strain could be explained by the possibility that the *GAL1* RNA itself, the act of transcribing the RNA (either ncRNA or mRNA), or the chromatin at the *GAL1* 5' region is involved in the transcriptional memory mechanism by an *in cis* effect. Since the published work mentioned before (Zacharioudakis et al., 2007) substituted the *GAL1* ORF for the *GFP* ORF leaving the *GAL1* promoter intact in the genome but failed to show memory activity, this would suggest that the *GAL1* promoter alone is not involved in any *in cis* mechanism at the locus. Thus, if there does turn out to be a role for transcription-dependent or independent chromatin formation in memory, it is likely that the 5' region of *GAL1* is also required, supporting the model that Htz1 contributes to the transcriptional memory mechanism (Brickner et al., 2007).

To begin to test these different possibilities, the *ADH1*^{T} should be should be introduced into the *pADH1-GAL1* plasmid at position +757 in a *gal1Δ* strain to test whether the ability to maintain memory is still present. This would confirm whether memory is a function of the first part of the *GAL1* coding region and does not require the non-coding or promoter sequences which are not present on the plasmid. Subsequent deletion or mutational analyses either in the plasmid or the genome should begin to define the sequences required.

Chapter 6

The role of the non-coding transcripts in transcriptional induction at the GAL locus

6.1 Introduction

The *GAL* genes are activated in the presence of galactose and their mRNAs accumulate to very high levels. However, the activation kinetics, the point at which RNA starts to accumulate, varies between several minutes to a few hours depending on the growth conditions. The cells require the enzymes involved in the galactose metabolism in order to metabolize and grow in this carbon source. The most important factor for the activation of the *GAL* genes is Gal4p. Gal4p is a transcriptional activator with DNA-binding and activation domains (Ptashne, 1988; Struhl, 1987). It binds an upstream activating sequence (UAS_{GAL}) present throughout the genome with about 300 potential sites identified (Traven et al., 2006). A Gal4p dimer, when bound to the UAS_{GAL}, recruits co-activators including the SAGA (Spt-Ada-Gcn5 acetyltransferase) complex, which in turn recruit the pre-initiation complex (PIC), resulting in transcription initiation. Under non-inducible conditions Gal4p is repressed by the binding of Gal80p to its activation domain, thus preventing the interaction of Gal4p with the transcriptional machinery. In the presence of galactose, the repression of Gal4p is relieved by the interaction of Gal80p with Gal3p. There are two main theories to explain how this interaction results in the activation of Gal4p: the non-dissociation and the dissociation model. The non-dissociation model proposes that Gal3p binds to the Gal80p-Gal4p repressive complex and causes a change in Gal4p conformation that exposes the Gal4 activator domain (Leuther and Johnston, 1992; Platt and Reece, 1998), whereas the dissociation model suggests that Gal3p sequesters Gal80p in the cytoplasm, physically freeing Gal4p (Peng and Hopper, 2000, 2002). These two models not only disagree in the mechanism of activation of Gal4p but also disagree in the localization of Gal3p, with some studies

suggesting it is only cytoplasmic (Peng and Hopper, 2000) whilst other studies defend the presence of Gal3p in both the cytoplasm and the nucleus (Wightman et al., 2008). Recent studies using newly developed techniques are in favour of the presence of Gal3p in both cellular compartments (Egriboz et al., 2011; Wightman et al., 2008)

It has been known for decades that cells with *GAL3* mutations display a severe phenotype known as long term adaptation (LTA), in which induction of the *GAL* genes takes considerably longer than in WT cells (Bajwa et al., 1988; Bhat et al., 1990; Broach, 1979; Nogi, 1986; Torchia and Hopper, 1986). Gal3p is the galactose sensor that binds the sugar in the presence of ATP. In the absence of Gal3p it is thought that Gal1p replaces it in the sensory function (Bhat et al., 1990; Sellick et al., 2009). Since both *GAL1* and *GAL3* are repressed in glucose it is thought that residual levels of Gal1p and Gal3p together with a mitochondrial function, which signals the energy status of the cell, are sufficient for an initial activation that will lead to the production of more protein (Bhat et al., 1990). It is also thought that Gal1p is not as efficient as an inducer as Gal3p and this results in the strong phenotype displayed by the *gal3Δ* strains. However, the actual mechanism that causes this phenotype is not yet understood.

Many more questions still remain concerning the complex mechanism of transcriptional regulation of the *GAL* genes and the discovery, described here, of several ncRNAs present in that region possibly add an extra layer of complexity. The transcription of ncRNAs through the *GAL10-1* and *GAL7* promoters could potentially constitute an event that interferes with the process of induction of the *GAL* genes. In this chapter, the involvement of these non-coding transcripts in the induction kinetics of the *GAL* genes will be investigated.

6.2 Inducibility of the *GAL* genes and expression of the proteins involved in the galactose metabolism

The kinetics of induction of the *GAL* genes is highly dependent on the conditions used, as demonstrated by the differences between data in the published studies. It is therefore essential to determine the induction profile of the *GAL* genes under the standard growth conditions used in this study. These conditions comprise, unless otherwise stated, an overnight culture in 2% YPD, which is then diluted in 2% YPD to an OD_{600} of 0.2 and allowed to grow to 0.4 - 0.6 OD_{600}. At this point the cells are either harvested, for studies in repressive conditions, or washed and inoculated without dilution in YP supplemented with 2% galactose, and harvested at the required time points.

As shown in Figure 29. (A), the *GAL1* sense transcript is detectable within 30 minutes to 1 hour of induction, reaches maximum expression by 3 hours and is no longer detected after overnight growth in galactose-containing medium. In contrast, the *GAL7* sense transcript is detectable in glucose, albeit at extremely low levels (Figure 29 B, lane 1), consistent with the tiling array data previously shown (Chapter 3, Figure 8). *GAL7* starts to induce within 15 minutes of exposure to galactose medium and, similar to the *GAL1* sense transcript, it is also no longer detectable after an overnight growth in galactose medium (Figure 29. B). The repression of the *GAL1* and *GAL7* genes at the overnight time point is likely to be due to the exhaustion of galactose in the medium, however, in order to be certain the levels of galactose would have to be assessed, or the medium replenished frequently during the induction to check whether or not this repression would still be observed.

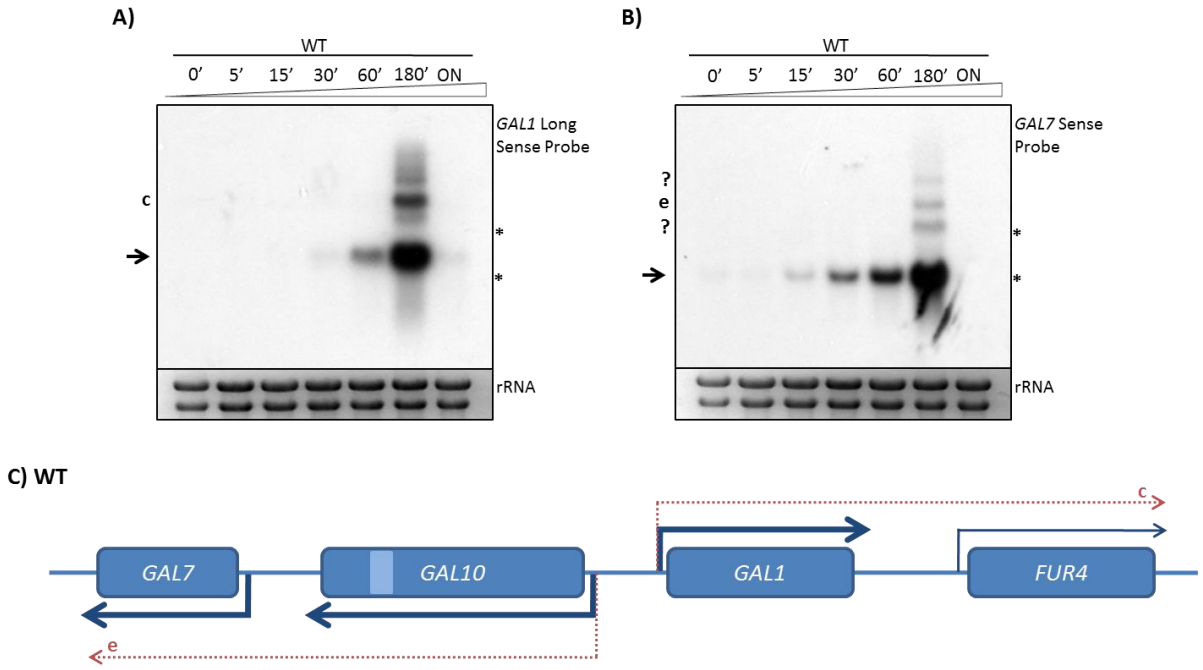


Figure 29. Induction kinetics of the *GAL1* and *GAL7* genes. Northern blot analysis of RNA extracted from WT cells induced for the time points indicated and probed for *GAL1* **(A)** and *GAL7* **(B)** sense. The cells were grown following the same method described in section 2.3. The coding transcripts are identified by a black arrow. Longer transcripts identified with the *GAL7* probe and marked with a question mark are unknown. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown below the blot as loading controls. **(C)** Schematic of the profile of transcripts identified in the Northern blots above. Transcript start and end sites are only approximate.

The growth conditions not only affect the *GAL7*, *10* and *1* cluster but also the remaining machinery necessary for the galactose metabolism. Apart from the already mentioned regulatory and structural proteins, like Gal4p, Gal3p, Gal80p and Gal2p, more proteins are involved in the transcriptional regulation of the *GAL* genes, including Pgm2p, Sin4p, Gal11p and Gal83p. Pgm2p (or Gal5p), is the phosphoglucomutase and converts glucose-1-phosphate into glucose-6-phosphate (Torchia and Hopper, 1986). *PGM2* has one putative Gal4p binding site (Ideker T, 2001). Sin4p is part of the mediator complex and is involved in both positive and negative regulation of transcription (Li et al., 1995; Mizuno and Harashima, 2003). Gal11p is a general activator of basal transcription, regulated by Sin4p, and involved in the phosphorylation of Gal4p (Long et al., 1991). Lap3p (or Gal6p)

is a cysteine-peptidase bleomycin hydrolase, regulated by Gal4p and down-regulates the expression of the *GAL* genes like *GAL80* (Ostergaard et al., 2001; Zheng et al., 1997). Gal83p is a putative subunit of the Snf1 complex and is involved in glucose repression (Erickson and Johnston, 1993). The fine balance of the levels of the different proteins involved, tightly controls the induction and expression of the *GAL* genes. It is therefore important to remember that some induction phenotypes can be caused by alterations in expression/activity of these proteins and not necessarily a direct effect on the *GAL* gene cluster. The expression of the genes encoding these proteins in the growth conditions used throughout this study was assessed using the tiling array data set (Figure 30). It is possible to observe a clear difference between the two conditions tested, YPD and YPG, with most of the regulatory genes being expressed at higher levels in galactose than in glucose medium. The genes *GAL2*, *GAL3*, *GAL80*, *LAP3* and *PGM2* contain Gal4p binding sites and therefore their higher expression in galactose is not surprising (Griggs and Johnston, 1991). Even though *GAL4* is highly expressed in galactose compared to glucose, it is not due to the positive feedback loop caused by the presence of UAS_G (which binds Gal4p), but due to the fact that *GAL4* is repressed in glucose by Mig1p (Griggs and Johnston, 1991). The remaining genes, *GAL83*, *SIN4* and *GAL11* do not contain UAS_G and their change in expression is therefore not due to a direct effect of Gal4p.

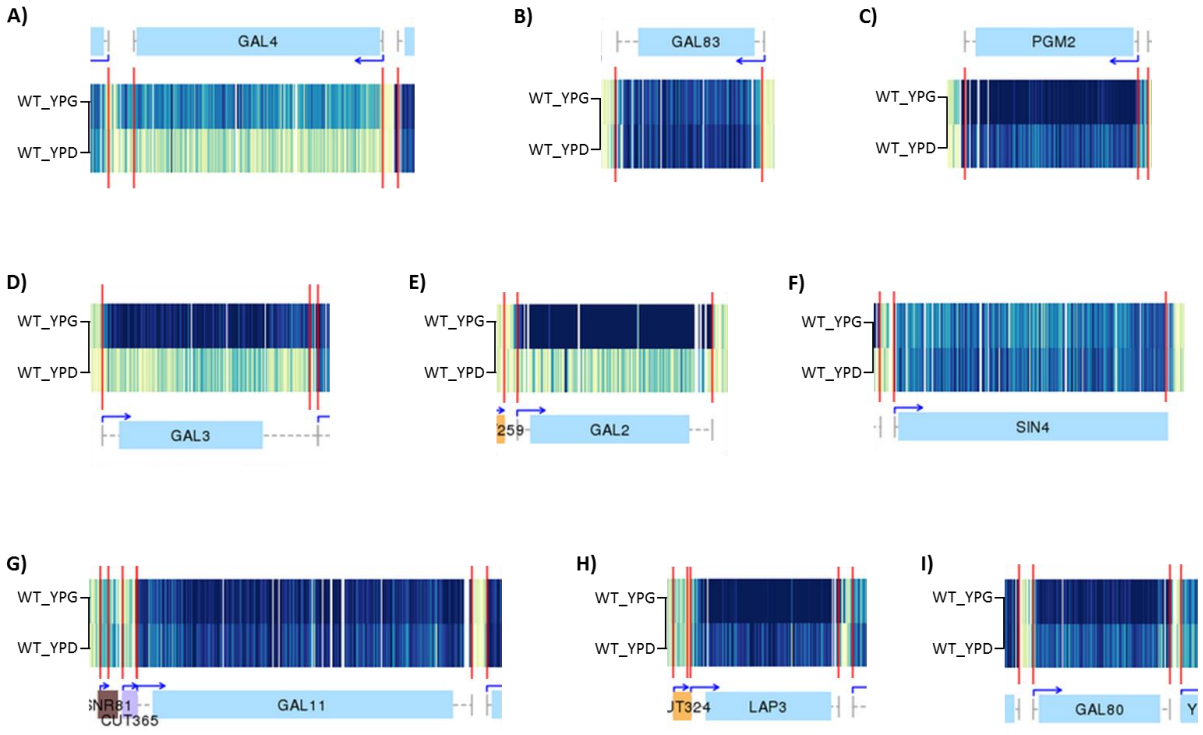


Figure 30. Tiling array of the proteins involved in the galactose metabolism. A tiling array experiment was performed in collaboration with Lars Steinmetz at EMBL. Two separate samples of a WT strain were grown as in 2.3 and then allowed to grow for a further 3 hours in fresh YPD (WT_YPD) or YPG (WT_YPG). The sense strands of *GAL4* **(A)**, *GAL83* **(B)**, *PGM2* **(C)**, *GAL3* **(D)**, *GAL2* **(E)**, *SIN4* **(F)**, *GAL11* **(G)**, *LAP3* **(H)** and *GAL80* **(I)** are shown. The colour code is light for low hybridization and dark for high.

6.3 The *GAL10* Reb1-dependent antisense transcripts are not responsible for the slow induction phenotype of a *GAL3* delete strain

The mechanism behind the slow induction LTA phenotype, displayed by strains lacking *GAL3* is not known (Bajwa et al., 1988; Bhat et al., 1990; Broach, 1979; Nogi, 1986; Torchia and Hopper, 1986). It was wondered whether the presence and/or transcription of the newly described ncRNAs has a repressive effect on the *GAL* genes that could explain the *gal3Δ* phenotype.

In order to address this question, the *GAL10-1* antisense transcript (transcript **f**) was considered a potential regulator given that it is transcribed over the *GAL10-1* promoter in glucose and early galactose induction. If the presence of *GAL3* somehow facilitated the switch between the down-regulation of the *GAL10-1* antisense transcript and the activation of the *GAL10* and *GAL1* sense transcription, then it would be expected that in the absence of the *GAL10-1* antisense transcript, these genes would be able to activate normally in the *gal3Δ* strain. Firstly, it was verified that under the growth conditions used, the *gal3Δ* strain had the expected phenotype. As shown in Figure 31, in a *gal3Δ* strain *GAL1* coding transcript is only present after an overnight induction with galactose compared to 3 hours in a WT strain. After a longer exposure of the film it was observed that transcript **f,** the *GAL10-1* antisense transcript, was being accumulated to higher than normal levels in the *gal3Δ* strain (Figure 31. B).

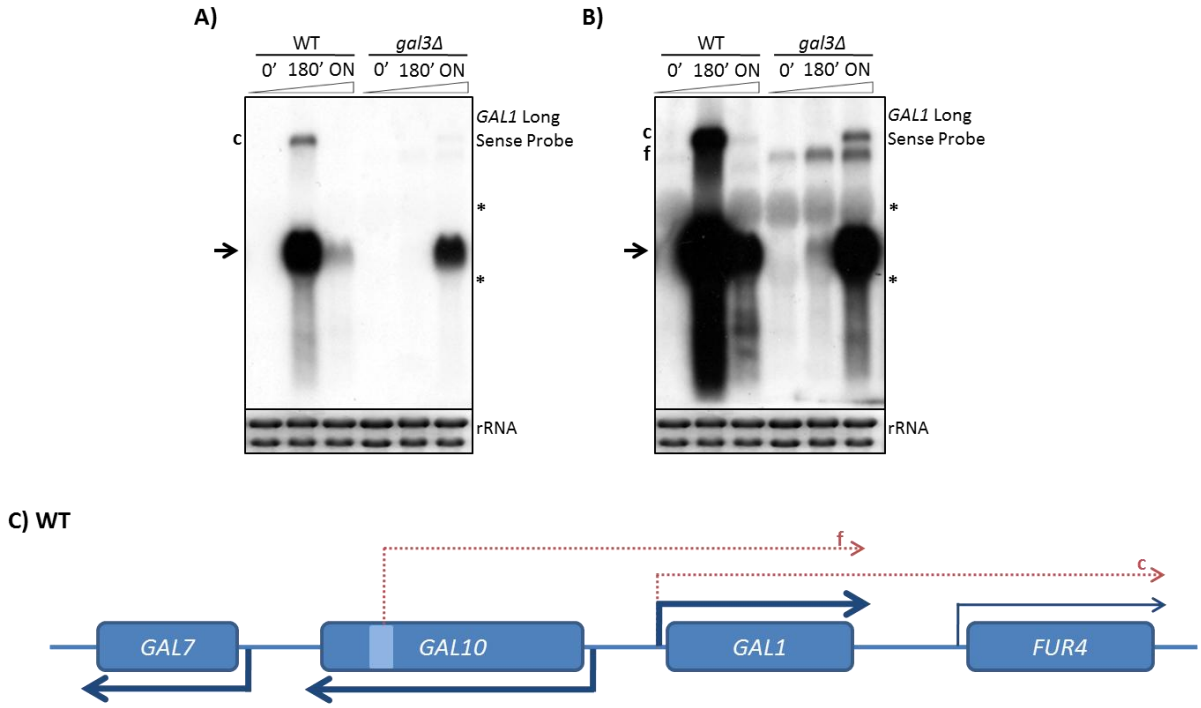


Figure 31. A *GAL3* deletion strain exhibits a very slow induction of *GAL1*. Northern blot analysis of total RNA extracted from WT and *gal3Δ* strain upon galactose induction at the time points indicated and probed for *GAL1* sense. A normal **(A)** and a longer **(B)** exposure time of the same membrane were taken. Indicated are the *GAL1* (black arrow), *GAL1-FUR4* (**c**) and *GAL10-GAL1* (**f**) transcripts. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. **(C)** Schematic of the profile of transcripts identified in the Northern blot above. Transcript start and end sites are only approximate.

It was further confirmed, using a probe homologous to the *GAL10-1* intergenic region, that both detectable *GAL10* antisense transcripts (**f** and **g**) were accumulating to higher levels in galactose in the *gal3Δ* strain (Figure 32). It is confirmed that both transcripts are originating from *GAL10*, since they are absent in a *gal10Δ* strain. Both of these transcripts are Reb1-dependent, as shown in Figure 32 (A), since the transcripts are no longer present in a strain that has all the Reb1 binding sites mutated (*Reb1_bsΔ*) (Houseley et al., 2008). This result suggested that the model mentioned before, whereby the antisense transcripts could be involved in the slow induction of a *gal3Δ* strain is more likely.

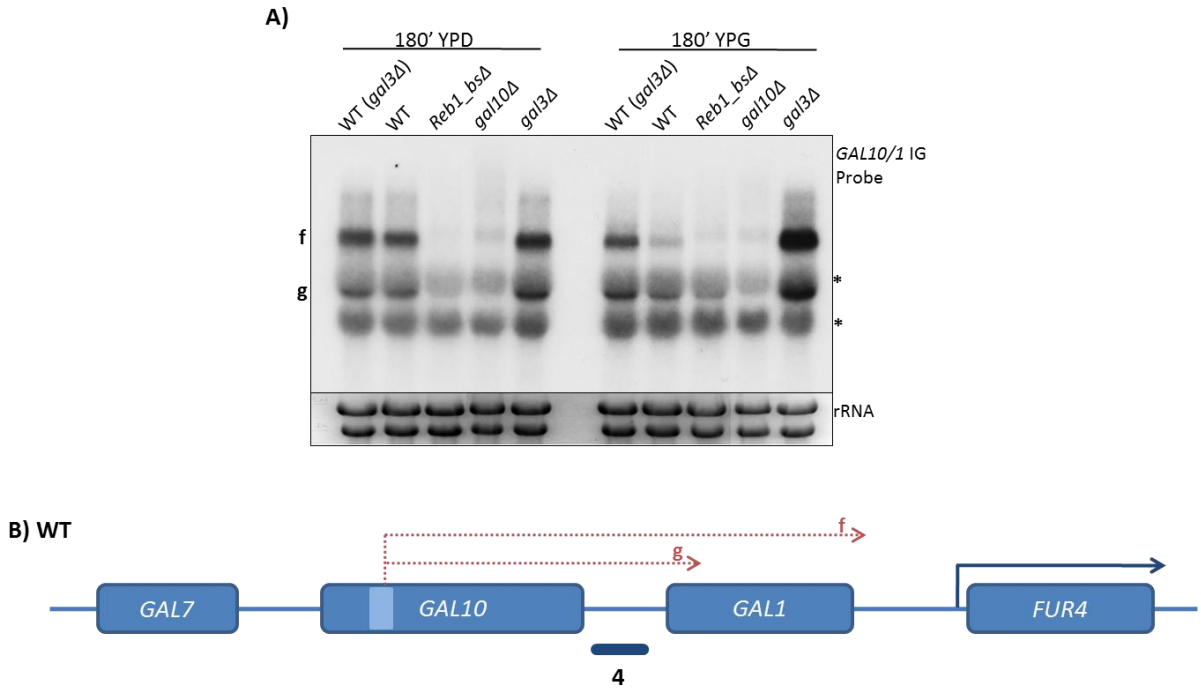


Figure 32. The *GAL10* antisense transcripts are increased in a *GAL3* deletion strain. (A) Northern blot analysis of total RNA extracted from the strains indicated and probed for the *GAL10/1* intergenic region, sense to *GAL1*. Cells were grown as in 2.3 and then re-incubated in fresh YPD or YPG for 3 hours. *GAL10* antisense transcripts (**f** and **g**) are indicated. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gel (rRNA) is shown as a loading control. **(B)** Schematic of the profile of transcripts identified in the Northern blot above. Transcript start and end sites are only approximate. Location of the probe *GAL10/1* IG (number 4) (Chapter 3, Figure 10) used is indicated.

To further test this hypothesis, the strain with the Reb1 binding sites mutated (*Reb1_bsΔ*) was used (Houseley et al., 2008). The transcripts that are produced from the *GAL10* internal bi-directional promoter (transcripts **f**, **g** and **h**) are Reb1-dependent, and are, therefore, absent. A *gal3Δ* mutation was then introduced into the WT and the *Reb1_bsΔ* strain and the induction kinetics were investigated. Figure 33 (A) shows that the induction profile of the *GAL1* transcript in the *gal3Δ* strain is identical to the one shown before (Figure 31) in a different strain background (FT4). Moreover, in the *Reb1_bsΔ* strain, the induction kinetics are the same as in the WT (Figure 33. B). The level of *GAL1* sense RNA is also similar, and the higher intensity observed is due to a technical problem. This result is

in accordance with the published data of Houseley et al, where he shows that this strain does not have a phenotype under normal growth conditions (2% YPG) (Houseley et al., 2008). Furthermore, when deleting *GAL3* in the *Reb1_bsΔ* strain, a very slow induction of *GAL1* is observed, identical to that in the *gal3Δ* strain. It was also observed that transcript **f** is not present in these strains and further confirmed in Figure 34, where the *GAL10-1* intergenic probe was used to control for possible transcripts present in this region. The absence of transcript **g** is also confirmed in Figure 34. Nevertheless, the slow induction of the *Reb1_bsΔgal3Δ* strain indicates that the non-coding transcripts produced from the *GAL10* internal bidirectional promoter are not responsible for the slow induction phenotype of *gal3Δ*.

In chapter 3, it was shown that transcript **d** (which starts in the middle of *GAL10* and possibly terminates at the *GAL10* promoter) is still present in minimal media, conditions that deplete the levels of Reb1p, therefore, confirming that it is Reb1-independent. Thus, it cannot be ruled out that this transcript could be transcribed through the *GAL10-1* promoter and be involved in the *gal3Δ* phenotype. In order to eliminate all the *GAL10* antisense transcripts a *gal10Δ* strain would be necessary, however, it is known that a *gal3Δgal10Δ* strain is non-inducible (Bhat et al., 1990; Nogi, 1986; Torchia and Hopper, 1986) and therefore it cannot be used for the purpose of this experiment.

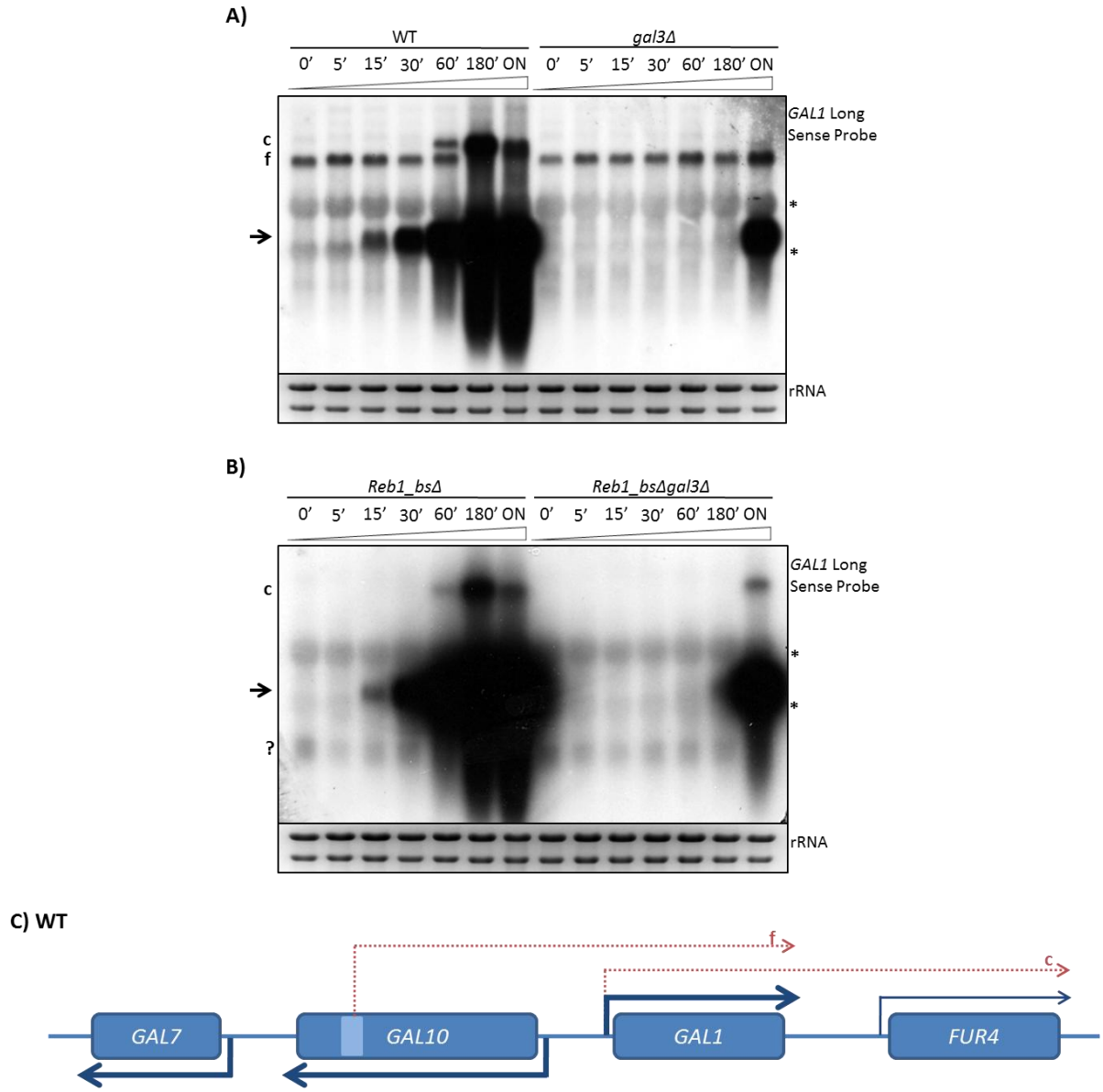


Figure 33. The long *GAL10* antisense transcript is not involved in the slow GAL gene induction exhibited by the *GAL3* deletion strain. Northern blot analysis of total RNA extracted from the strains indicated upon galactose induction at the time points specified and probed for *GAL1*. **(A)** Profile of induction of a *gal3Δ* strain compared with the WT (FT4 background). Transcripts **c** and **f** are indicated. **(B)** Comparison of the induction profiles of the *Reb1_bsΔ* and the double mutant *Reb1_bsΔgal3Δ*. Transcript c is indicated. An unknown transcript is marked with a question mark. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. The blur in (B) is due to a technical problem. **(C)** Schematic of the profile of transcripts identified in the Northern blots above. Transcript start and end sites are only approximate.

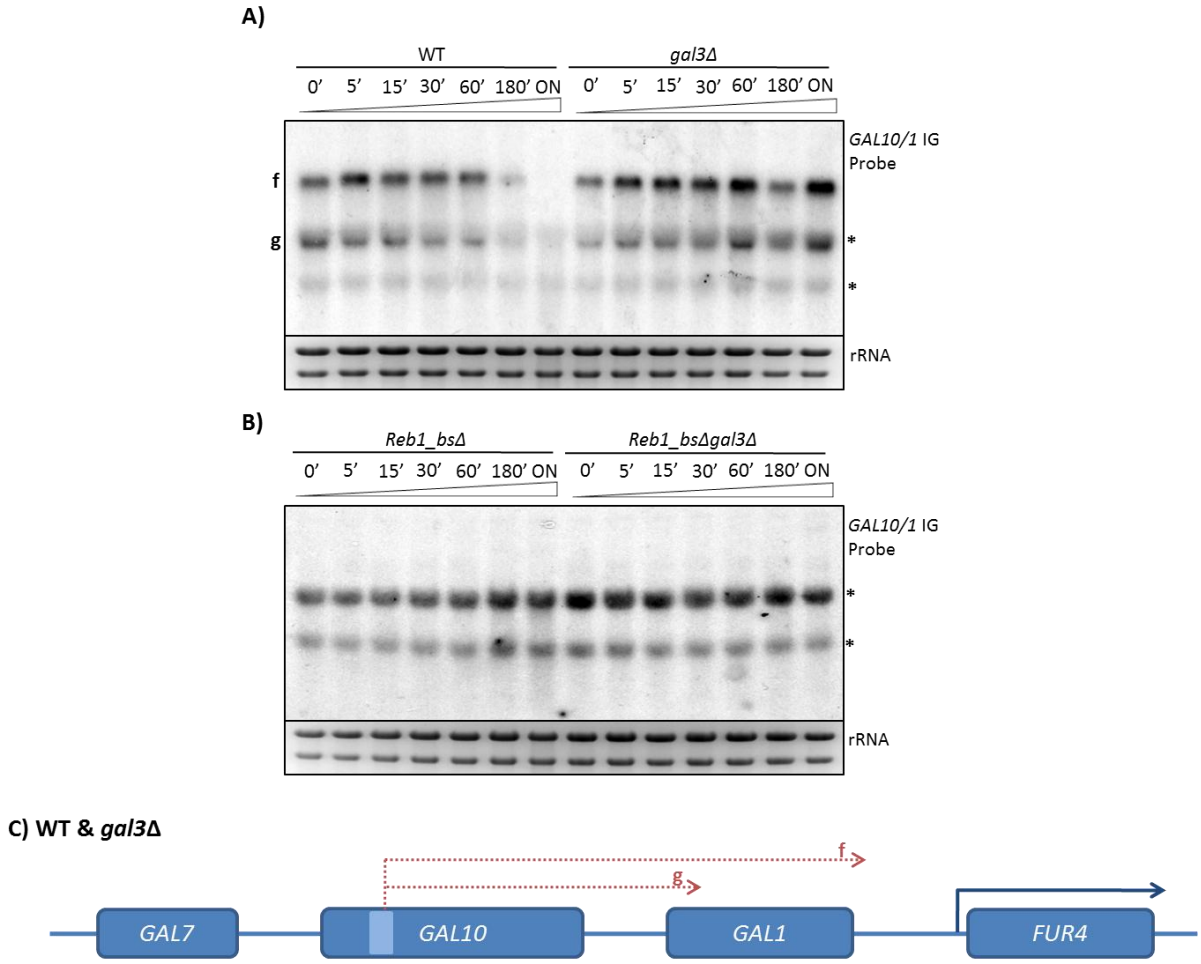


Figure 34. The mutation of the Reb1 binding sites abolishes the production of the *GAL10* transcripts produced from the internal promoter. Northern blot analysis of total RNA extracted from the strains indicated upon galactose induction at the time points specified and probed for *GAL10/1* intergenic region, sense to *GAL1*. **(A)** Profile of induction of a *gal3Δ* strain compared with the WT (FT4 background). Transcripts **f** and **g** are indicated. **(B)** Comparison of the induction profiles of the *Reb1_bsΔ* and the double mutant *Reb1_bsΔgal3Δ*. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. **(C)** Schematic of the profile of transcripts identified in the Northern blots above. Transcript start and end sites are only approximate.

6.4 Rrp6p and Kem1p involvement in transcriptional induction

In chapter 3, the effect of Rrp6p and Kem1p on the non-coding transcripts was investigated. It was shown that Rrp6p has a modest effect on the ncRNAs at the *GAL* locus in comparison to the much larger effect of Kem1p. However, in both cases there were increased levels of the mRNAs at 3 hours of induction (Chapter 3, Figure 18 and

Figure 20). The increased levels of the *GAL10* and *GAL1* sense RNAs by 1 hour of induction in the *kem1Δ* and *rrp6Δ* strains raised the question of whether these levels were due to an accumulation of stabilized transcripts, or if the increased levels, in fact, reflect faster induction kinetics in comparison to wild type. In order to answer this question the induction kinetics of *GAL10* and *GAL1* in the *kem1Δ* and *rrp6Δ* strains were analysed.

Figure 35 (A) shows the induction kinetics of the *GAL10* sense and antisense transcripts in the WT and *kem1Δ* strain. It was observed that the *GAL10* sense transcript is detected by 60 minutes of induction in a *kem1Δ* strain compared to 90 minutes in the wild type. Two read-through transcripts are also observed, with one corresponding to the *GAL10-7* transcript (transcript **e**) while the other is unknown and not always detected. All the *GAL10* antisense transcripts are stabilized by the lack of Kem1p, in particular transcript **f**, and their levels stays constant whereas in the WT they gradually get repressed and/or degraded as the *GAL10* sense induces.

Additionally, Figure 35 (B) shows the induction kinetics of *GAL1* sense and antisense transcripts in the same strains. The kinetics of the *GAL1* transcript is identical to that of *GAL10*, with it being detected by 60 minutes of induction in a *kem1Δ* strain compared to 90 minutes in the wild type. Transcript **f** (initiating from the *GAL10* internal promoter), which was also described as a substrate of Kem1p in the first chapter (Chapter3, Figure 20), is observed in the *kem1Δ* strain, and at much lower levels in the WT strain. The high level of the *GAL1-FUR4* read-through transcript (transcript **c**) observed at 90 minutes is probably due to a conditional effect of growth or media since both the WT and *kem1Δ* strains display it, however the cause for this high level is not known. Taking into account that new antisense transcripts (transcripts **i** and **j**) were revealed in the *kem1Δ* strain in

the first chapter (Chapter3, Figure 20), albeit only in glucose conditions, the antisense strand was also checked. Interestingly, it shows that the antisense transcripts detected in glucose, in particular transcript **j**, are present at early induction time points and become repressed later on. Unfortunately, it is not possible to determine whether these transcripts were present at this level before the cells were transferred to galactose or whether transcript **j** in particular was induced by the change in carbon source, without having the glucose time point to compare it to. Nevertheless, the antisense transcripts and/or transcription does not appear to influence the induction kinetics of the *GAL1* sense transcript.

In summary, Kem1p stabilizes several transcripts at *GAL10* and *GAL1* during galactose induction but that does not significantly influence the kinetics of induction.

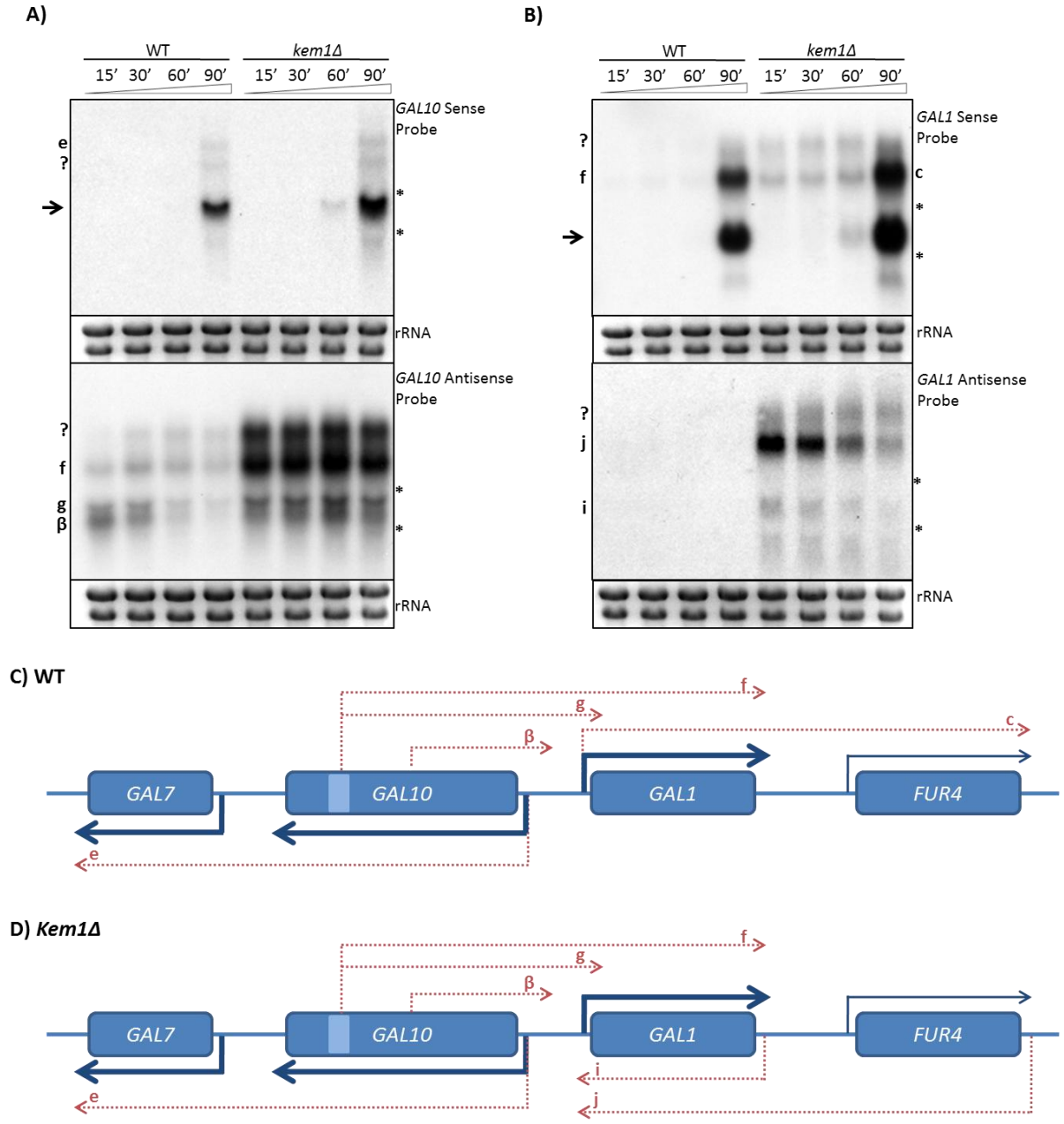


Figure 35. Kem1p is not involved in transcriptional induction. Northern blot analysis of RNA extracted from WT and *kem1Δ* cells, induced for the time points indicated and probed for *GAL10* **(A)** and *GAL1* **(B)** sense and antisense. The cells were grown following the same method described in section 2.3. The coding transcripts are identified by a black arrow. Transcripts identified with a question mark are unknown. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. Schematic of the strains and transcripts detected in the Northern blots above of the WT **(C)** and *kem1Δ* **(D)** strains. Transcript start and end sites are only approximate.

Next, the kinetics of galactose induction in an *rrp6Δ* strain were analysed. Figure 36 shows that the *GAL1* transcript is clearly detectable as early as 15 minutes after induction

in an *rrp6Δ* strain compared to 60 minutes in the WT. Also, the levels of *GAL1* expression in the *rrp6Δ* strain are considerably higher than the levels detected for the WT at 60 minutes. The differences in levels observed for transcripts **f** (all the time points) and **c** (60 minutes) are due to a technical issue with the Northern blot, and subsequent repeats show the induction of transcript **c** occurs at the same time as the *GAL1* coding transcript. The *GAL1* antisense strand was not analysed since, in the first chapter, it was concluded that the deletion of Rrp6 does not stabilize any antisense transcript to a great extent.

This result suggests that the increased levels of steady-state RNA at later time points in the *rrp6Δ* strain could be explained by faster induction of *GAL1* expression allowing more RNA to accumulate over time, suggesting, perhaps, a role for Rrp6p outside the context of the nuclear exosome. However, degradation of the *GAL1* transcripts by Rrp6p at the early and later time points during induction cannot be ruled out, although the biological reasoning for such an unproductive mechanism is difficult to comprehend.

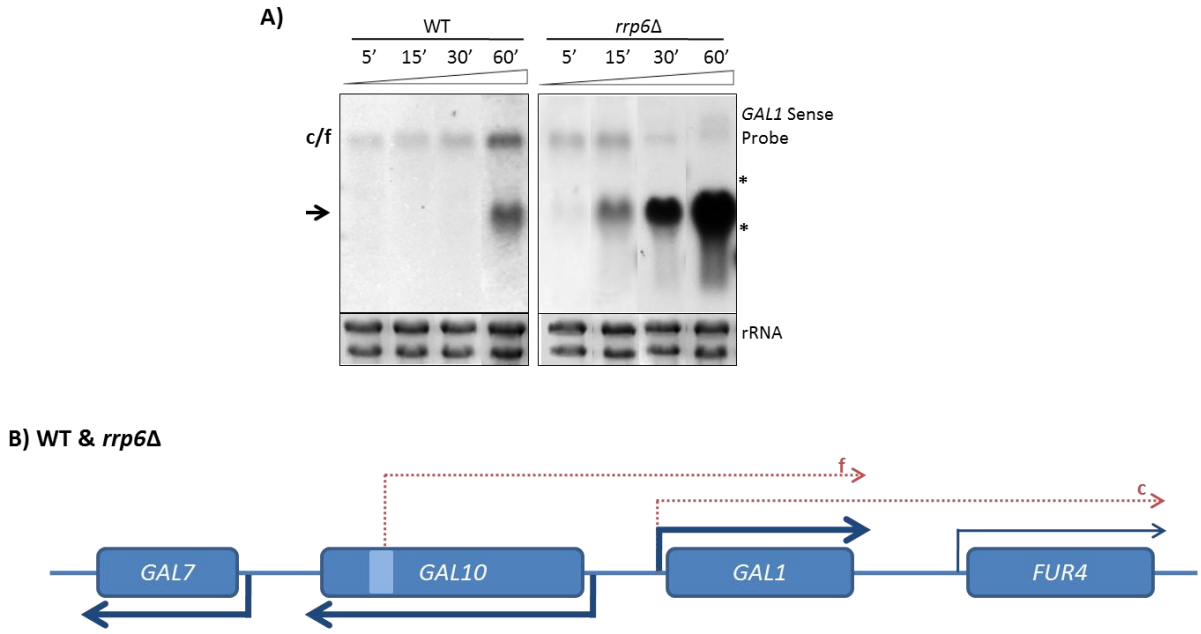


Figure 36. *GAL1* induces faster in the absence of Rrp6p. Northern blot analysis of RNA extracted from WT and *rrp6Δ* cells, induced for the time points indicated and probed for *GAL1* sense **(A)**. The cells were grown following the same method described in section 2.3. The coding transcript is identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA)are shown as loading controls. **(B)** Schematic of the profile of transcripts identified in the Northern blots above. Transcript start and end sites are only approximate.

As mentioned before, the induction of the *GAL* gene cluster is dependent on the fine balance between the factors that control galactose metabolism. Therefore, the expression of these factors should be assessed in a *kem1Δ* and, particularly, in a *rrp6Δ* strain in case the differences observed for the kinetics of induction are indirectly due to stabilization of the factors' mRNA and consequent rise in protein levels. Since analysing many genes by Northern blot is time consuming and an *rrp6Δ* strain was not used when the tiling array was performed in collaboration with the Steinmetz group, their tiling array data was used instead (Xu et al., 2009). The Steinmetz results need to be interpreted carefully since it is known that the growth conditions are different and the *rrp6Δ* cells were grown in minimal media, and this potentially alters the transcription profile.

Nevertheless it would be interesting to check whether any of the proteins involved in *GAL* gene regulation were regulated by Rrp6. The *kem1Δ* data set from the Morillon laboratory (van Dijk et al., 2011) was also checked, even though the effects on the transcriptional induction were not so substantial (Table 9).

Even though a statistical analysis would have to be done to conclude with certainty the difference in expression levels of the factors compared to the WT, it is compelling that some differences are visually evident on the tiling array map. The interesting observation is that *GAL4* is detected to higher levels than the WT in both the *rrp6Δ* and *kem1Δ* strains. An increase in *GAL4* expression/stabilization could potentially result in the faster induction observed in the *rrp6Δ* strain (Torchia and Hopper, 1986), although the increased levels of *GAL4* RNA do not influence induction in the *kem1Δ* strain. However, increased gene expression does not automatically result in increased protein levels, and so the levels of Gal4p in the *rrp6Δ* and *kem1Δ* strains would have to be assessed to infer the significance of the increased *GAL4* expression levels detected in the tiling array.

Genes involved in galactose metabolism	Expression profile in a *rrp6Δ* strain	Expression profile in a *kem1Δ* strain
GAL3	=	↑
GAL4	↑	↑
GAL80	=	=
GAL11	↑	=
GAL2	=	=
LAP3	↓	=
PGM2	=	↑
GAL83	=	=
SIN4	=	=

Table 9. Expression of the factors involved in the regulation of galactose metabolism in the *RRP6* and *KEM1* delete strains. Data visually compared in the data sets from http://vm-gb.curie.fr/XUT/images/YNL236W_1.htm (van Dijk et al., 2011) and http://steinmetzlab.embl.de/cgi-bin/viewNFRsharing.pl?showSamples=NFRsharing&gene=SIN4 (Xu et al., 2009). The ↑ means about 2-fold higher expression, ↓ about 2-fold lower expression and = means equal, compared to WT.

6.5 Discussion

In this chapter, the role of the non-coding transcripts in the *GAL* gene cluster in transcriptional induction was investigated. It was thought that since several ncRNAs are transcribed through the *GAL10-1* promoter in glucose and are gradually repressed in the presence of galactose, they could influence the induction kinetics. Houseley et al, have suggested that the *GAL10* antisense transcript (transcript **f**) is involved in repression of the *GAL10-1* promoter, however extremely small differences were observed and non-standard growth conditions had to be used (0.1 g l^{-1} galactose/0.2 g l^{-1} glucose). If indeed the ncRNAs have an effect on transcription induction, it would be expected to occur in every type of growth condition. Since Houseley et al only showed/detected the *GAL10* antisense transcript (transcript **f**) in their studies, and from chapter 3 it is known that several transcripts, stable and unstable, could potentially affect the *GAL10-1* promoter, the inducibility of the *GAL* genes in the *GAL3*, *KEM1* and *RRP6* deletion strains was investigated (Houseley et al., 2008).

The mechanism behind the very slow induction profile of a strain lacking Gal3p is still unknown. The evidence shown here, in which the ncRNAs transcribed through the *GAL10-1* promoter are up-regulated in a *gal3Δ* strain, raised the question of whether this slow induction could be due to the ncRNAs out-competing activators in the *GAL10-1* promoter. Overall, activation of the *GAL* genes in wild type cells could involve a fine balance between the repressive ncRNAs and the recruitment of activation factors, and the absence of Gal3p could disturb this equilibrium making it harder for the activating factors to overcome the repression from the ncRNAs. If so, then removal of the ncRNAs in a *gal3Δ* strain should result in a faster induction. The difficulty with this experiment is

that a method of abolishing all the ncRNAs has not been found. A strain with Reb1 binding sites mutated, and which therefore cannot produce the Reb1-dependent transcripts (transcripts **f**, **g** and **h**) has been used instead. No evidence was found that these transcripts are involved in the *gal3Δ* strain slow induction phenotype. However, it cannot be ruled out that other transcripts, such as the Reb1-independent transcripts), could be involved and until a way of abolishing all the ncRNAs in the *GAL* locus is found it will remain difficult to conclude whether they are involved in the regulation of transcription.

Given that the cytoplasmic Kem1p and nuclear Rrp6p enzymes are responsible for the degradation of some of the ncRNAs, it was thought that if these transcripts were involved in the transcriptional induction of the *GAL* genes, then a phenotype should be observed in the absence of these enzymes. In the absence of Kem1p no major differences in induction were observed. However, it was found that in the absence of Rrp6p the *GAL1* coding transcripts were observed at an earlier time point compared to the WT. Although, this could indicate that Rrp6p is degrading the *GAL1* transcripts produced early on in the induction, a biological reason for such mechanism is difficult to understand. A repression of induction of *GAL1* by Rrp6p is therefore favoured. At present there is no indication that this repression of induction is due to an ncRNA. The mechanism for how Rrp6p is controlling the induction of *GAL1* is not known, but some aspects of Rrp6 function should be considered. First of all, in the Steinmetz tiling array data set, *GAL4* expression seems up-regulated in the *rrp6Δ* strain, which could explain the fast induction observed in that strain. However the difference in the growth conditions used could influence the transcriptional expression of *GAL4* and therefore it is impossible to compare with the

data obtained here. Consequently, the expression of *GAL4* needs to be investigated under the growth conditions used in this study in the WT and *rrp6Δ* strain. Furthermore, an increase in expression levels might not reflect an increase in protein levels, and so this would also need to be investigated. Even though the main function of Rrp6p in the context of the nuclear exosome is the degradation of cryptic unstable transcripts (CUTs), its involvement in more subtle exonucleolytic processing of transcripts may be relevant here. Thus, its involvement in repressing transcriptional induction of *GAL1* could be a new functionality. For example, Vodala et al (Vodala et al., 2008) have proposed that the nuclear exosome, through Rrp6p, deadenylates a proportion of the *GAL1* sense transcripts which become concentrated in a "dot RNA" and it is this dot that tethers the *GAL1* gene to the nuclear periphery. Although in this study it was assumed that the substrate for the exosome was the *GAL1* mRNA, both the *GAL10-1* ncRNA and the *GAL-FUR4* RNA are transcribed over this region and could act as substrates for the exosome to make the dot RNA. As shown in chapter 4, there is preliminary evidence that ncRNAs also form dot RNAs in the nucleus, but further experiments to confirm are being performed. Association with the nuclear periphery has been implicated in rapid re-induction of the *GAL* genes (Brickner et al., 2007; Hampsey et al., 2011; Tan-Wong et al., 2009). The variable levels of some of the ncRNAs, sometimes observed in this work, might represent condition-specific processing, as opposed to complete degradation, by Rrp6p. Indeed, some transcripts, for example the *GAL1* antisense, are annotated as both a XUT and a CUT suggesting that they are subject to processing in both the nucleus and cytoplasm. Recent data obtained in this laboratory, has shown that triplex structures (DNA-DNA-RNA) can be formed, *in vitro*, in the *GAL10-1* promoter. Triplex structures are known to be repressive (Martianov et al., 2007). It is possible that the exosome, through Rrp6p,

trims the ncRNAs to shorter transcripts that could bind the promoter to form the triplex structures and therefore repress the promoter. However, more work needs to be done in order to establish the role of Rrp6p in the transcriptional induction of *GAL1*.

Chapter 7

Investigation of sense and antisense transcription units

7.1 Introduction

The transcript mapping data presented in Chapter 3 are summarised in Figure 37 and these data reveal three different transcription unit architectures at the *GAL* locus: conventional Promoter (P) to Terminator (T), Promoter to distal Terminator and Promoter to Promoter. This data also indicates the both conventional and non-conventional units can be transcribed on both strands. Although most transcription units (TU), including the many sense antisense pairs, fall into the conventional P-T class, studying a non-conventional TU, for example the P-P TU, could shed light on how the conventional units are controlled, particularly how the choice of whether to transcribe from the sense promoter or the antisense promoter is made and what factors influence this. This is the subject of this chapter.

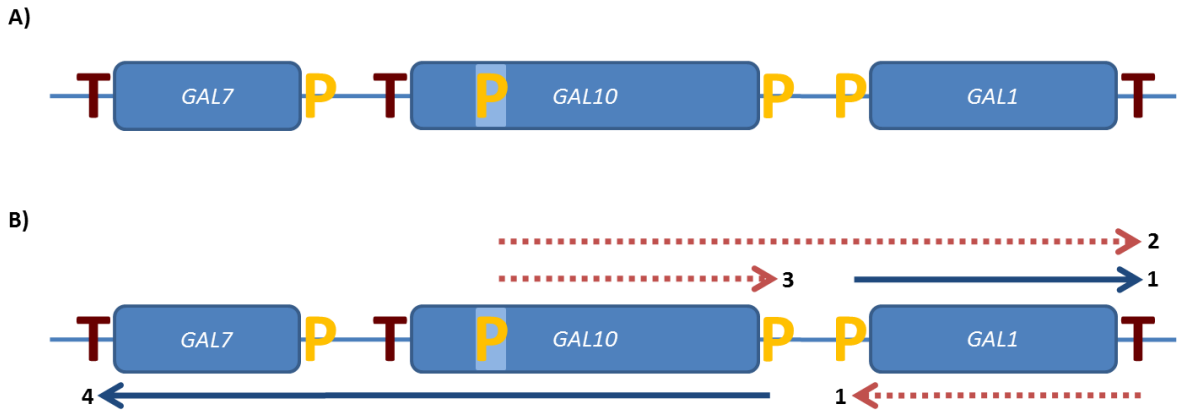


Figure 37. Schematic of the transcriptional architecture at the *GAL* genes. (A) Identification of the promoters (P) and terminators (T) present at the *GAL* locus. The *GAL10* internal bi-directional promoter is also shown. **(B)** Examples of types of transcription units that occur at the *GAL* locus: (1) Traditional organization, from a P to the proximal T. (2) and (4) P to distal T. (3) P - P.

The transcript mapping data raise questions about how pervasive transcription is initiated and terminated and how it influences transcription on the opposite strand. For example, are regions conventionally recognised as promoters also able to function as transcription terminators; do terminators also function as promoters of non-coding transcripts; are these non-coding transcripts polyadenylated and are there functionally different classes of non-coding transcripts? The recent data showing that promoters are functionally bidirectional in eukaryotes (Morris et al., 2008; Neil et al., 2009) provide one source of pervasive transcription (Xu et al., 2009) but published data, in addition to work performed as part of this thesis, suggest that regions conventionally recognised as terminators also possess properties normally associated with promoters (*e.g.* bound TBP, TFIIB and TFs) suggesting that under some circumstances, these regions may indeed act to initiate transcription (Murray et al., 2011). It is also not understood whether the non-coding transcripts that are defined as XUTs, CUTS or SUTs are functionally distinct, particularly in terms of the effect of their transcription on a conventional transcription unit. CUTs are likely to influence co-transcriptional events (chromatin marks for example) or RNA export from the nucleus (dot RNAs) as they are predominantly nuclear (Vodala et al., 2008; Wyers et al., 2005). XUTs are likely to influence transcription-related events on genes (such as chromatin architecture) and nuclear export (as they are efficiently exported and might aid export of other transcripts). SUTs could act *in cis* or *in trans* to influence transcription in the nucleus (Houseley et al., 2008; Pinskaya et al., 2009) or translation in the cytoplasm. XUTs (*e.g. GAL1* antisense), SUTs that are also XUTs (*e.g. GAL10-1* antisense), Rrp6p (nuclear exosome)-dependent 3' end processing and dot RNA formation (*e.g.*, *GAL1* sense transcript (Vodala et al., 2008)) have been identified at the *GAL* locus, but so far no conventional CUTs have been described.

In this chapter, a number of approaches are used to begin to address questions such as whether stable antisense transcription (to produce a SUT), arising from a bi-directional promoter, influences sense transcription from the conventional promoter differently from a stable antisense transcript arising from a terminator and to ask what features (both *cis*-acting DNA sequences and *trans*-acting factors such as Rrp6p) are responsible for this difference. To do this, different sequences were engineered into the middle of *GAL1* or *GAL10* so that promoter or terminator function could be assessed for its effect on the conventional *GAL1-10* promoter (either repressed or induced) without the complication of neighbouring gene regulatory sequences. Firstly, the results obtained with the *AgTEF* promoter inserted into the *GAL1* or *GAL10* coding regions will be presented followed by similar constructs with the *ScADH1* or *AgTEF* terminators.

7.1.1 Analysis of a promoter – promoter transcriptional organization and its effects on transcription

7.1.1.1 Promoters can initiate transcription bi-directionally.

Firstly, the question of whether a promoter placed within an ORF initiates transcription bi-directionally and, if so, the consequences of the antisense transcription/transcript on expression from the upstream promoter were addressed.

Insertions of the selection cassettes widely used for yeast genetic manipulations, which contain a strong promoter ($AgTEF^{P}$), were made at the genomic *GAL* locus. Two strains were engineered so that the cassette was inserted into *GAL10* and *GAL1*, with $AgTEF^{P}$ driving expression of a marker gene followed by $AgTEF^{T}$ in the same orientation as *GAL1*

or *GAL10*. At *GAL10* the *AgTEF*^{P}*:KanMX:AgTEF*^{T} cassette was inserted +1453 bp after the *GAL10* ATG codon, after deletion of the region between +1453-2007 bp to remove the internal bi-directional promoter (G10 *TEF*^{P}). At *GAL1* the *AgTEF*^{P}*:HisMX:AgTEF*^{T} cassette was inserted +757 bp after the *GAL1* ATG codon, without any deletions (G1 *TEF*^{P}). The schematic of the strains engineered is shown in Figure 38. The two selective markers used do not differentially influence the overall transcription profile, as concluded by the identical profiles after insertion of either marker at the same location in the same gene (data not shown).

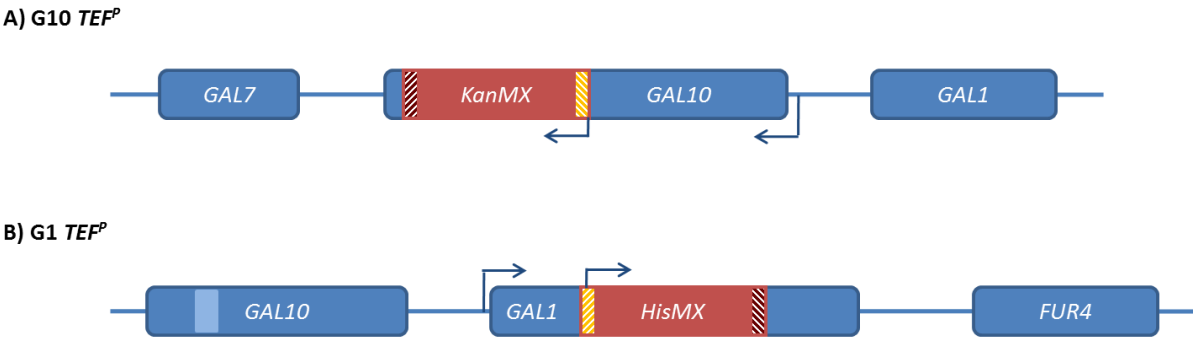


Figure 38. Schematic of the promoter constructs inserted at *GAL10* and at *GAL1*. Both strains were constructed as described in section 2.1. The *KanMX* and *HisMX* cassettes were used and are indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. The arrows indicate the direction of gene sense transcription. **(A)** The *AgTEF*^{P}*:KanMX:AgTEF*^{T} cassette was inserted in *GAL10* at +1453 bp, with the simultaneous deletion of the region between +1453-2007 bp, which corresponds to the internal bi-directional promoter. **(B)** The *AgTEF*^{P}*:HisMX:AgTEF*^{T} cassette was inserted in *GAL1* at +757 bp, without any deletions.

If all promoters are inherently bi-directional, then the insertion of the strong *AgTEF* promoter at *GAL10* and *GAL1* should give rise to an antisense transcript. The analysis of the sense and antisense transcript profiles is shown in Figure 39.

Firstly, the effect of the insertion on the sense transcripts is described after *GAL* gene induction in galactose for 3 hours. At both *GAL1* and *GAL10*, steady-state levels of sense transcripts are reduced by about ten-fold, and because of the absence of a terminator, truncated transcripts end in the vicinity of $AgTEF^{P}$ (transcripts **2** and **6**). Longer transcripts (read-through transcripts) extending over the marker cassette, presumably terminating at $AgTEF^{T}$, are also evident (transcripts **1** and **5**). This result raised the question of whether the low levels of sense transcript were due to inefficient transcription initiation or whether the transcripts were being degraded. To assess the stability of the sense transcripts, the catalytic subunit of the exosome, Rrp6p, was deleted. Additionally, the effect on transcription initiation was investigated by performing a chromatin immunoprecipitation (ChIP) of the RNA Polymerase II complex subunit Rbp1p. If the transcripts are being produced but actively degraded, then a WT occupancy of RNA Pol II should be expected. However, if the transcripts are stable, then the low level of expression could be explained by a lower amount of RNA Pol II in the gene due to inefficient transcription initiation. Figure 40 shows that the sense and antisense transcripts expressed in the G1 TEF^{P} strain are not further stabilized by the absence of the Rrp6p, since the levels of the transcripts are identical to those in the WT and *rrp6Δ* backgrounds. As changes in the stability of the steady-state RNA are not likely to explain the ten-fold reduction in levels, reduced occupancy of RNA Pol II at the gene was considered as an alternative explanation. Figure 41 shows the results of the RNA Pol II ChIP at 90 minutes of induction, for the $AgTEF^{P}$ insertions at *GAL10* (Figure 41 A) and *GAL1* (Figure 41 B). Indeed, at *GAL10* the levels of RNA Pol II at the 5' coding region are much lower in the G10 TEF^{P} compared to WT, indicating that transcription initiation has been disrupted. However, at *GAL1*, across the 5' coding region, the levels of RNA pol II in

the G1 *TEF*^{P} strain are much higher than in the WT strain. There are several reasons that could explain this result. Firstly, the sense and antisense transcripts could be degraded by another mechanism, for example by Kem1p, which was shown in Chapter 3 to degrade some of the transcripts in the *GAL* locus, especially at *GAL1*. Secondly, RNA Pol II could be paused in the gene, therefore increasing the probability of binding to the antibody, resulting in the higher levels detected by ChIP. This is supported by some researchers who claim that RNA Pol II can only be immunoprecipitated if paused or poised due to the, otherwise, fast kinetics of the enzyme. The reason why the levels of RNA Pol II at *GAL10* in the G10 *TEF*^{P} strain are reduced, compared to the unexpectedly high levels at *GAL1* in the G1 *TEF*^{P} strain, must indicate that, whatever the reason, the changes in RNA pol II distribution caused by the insertion of the *AgTEF*^{P}*:KanMX:AgTEF*^{T} cassette are gene-specific.

Next the antisense transcripts were assessed. At *GAL1*, where normally an antisense XUT is present, insertion of the cassette is associated with a stable antisense transcript (**7**) whose size is consistent with initiation close to *AgTEF*^{P} and termination at the *GAL10-1* promoter region (Figure 34 A and E). This strongly supports the hypothesis that *AgTEF*^{P} is bi-directional and capable of producing stable transcripts in both directions. The antisense transcript profile at WT *GAL10* is more complex with three transcripts evident; transcript **f** (~4 kb), which starts at the *GAL10* internal promoter and runs through to the end of *GAL1*; transcript **g** (~2.2 kb), which starts at the *GAL10* internal promoter and runs through the *GAL10-1* promoter region stopping at the beginning of *GAL1*; and transcript **β** (~1.8 kb), which is likely to be a longer variant of transcript **d** (~1.2 kb), starts in the middle of *GAL10* and runs into the *GAL10-1* intergenic region (see Chapter 3) (Figure 39 A

and C). The deletion of the *GAL10* internal bi-directional promoter and subsequent insertion of the *AgTEF*^{P}*:KanMX:AgTEF*^{T} cassette resulted in the shortening of transcripts **f** and **g** (transcripts **3** and **4** of lengths ~3.7 kb and ~1.9 kb respectively), suggesting that they are now being produced from or in the vicinity of the *AgTEF*^{P} and terminating at the same location as the WT transcripts. In addition, the G10 *TEF*^{P} strain also produces transcript **d**, and not **β**, which has not yet been mapped but an approximate position is shown on the map in Figure 39 (D).

Thus, as observed with *GAL1*, stable antisense transcripts, probably starting in the *GAL10* sequences, can be initiated by the *AgTEF*^{P} in the antisense direction. This result supports the idea that many promoters are bi-directional resulting in antisense transcription.

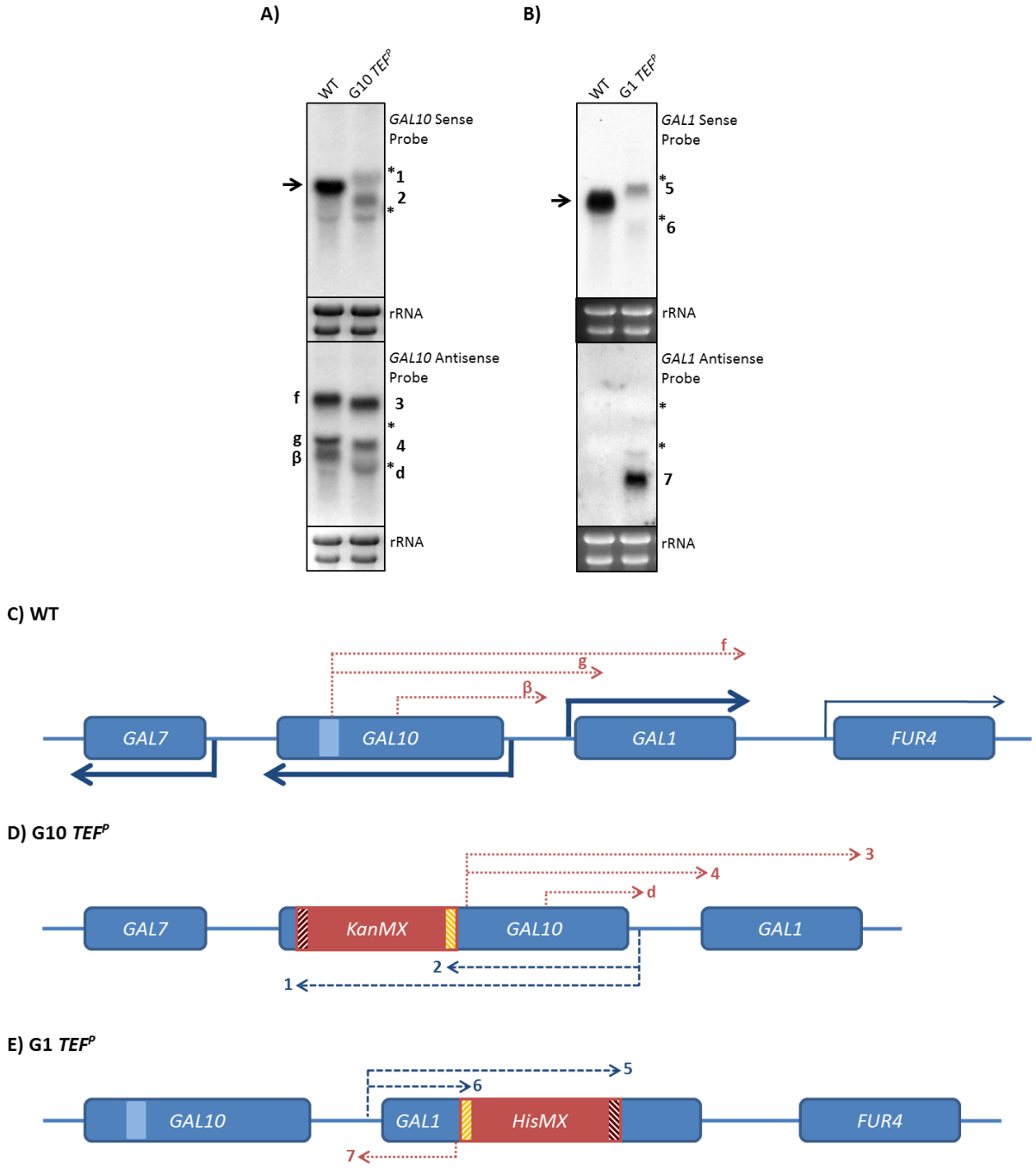


Figure 39. The promoter constructs inserted at *GAL10* and at *GAL1* drive antisense transcription. Northern blot analysis of total RNA extracted from the strains indicated and probed for *GAL10* **(A)** and *GAL1* **(B)** sense and antisense. The cells were induced in galactose for 1 hour (A) and 3 hours (B). The *KanMX* and *HisMX* cassettes used are indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematic below. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. Schematic of the transcripts detected in the Northern blots above of the WT **(C)**, the G10 *TEF*^{P} **(D)** and G1 *TEF*^{P} **(E)** strains.

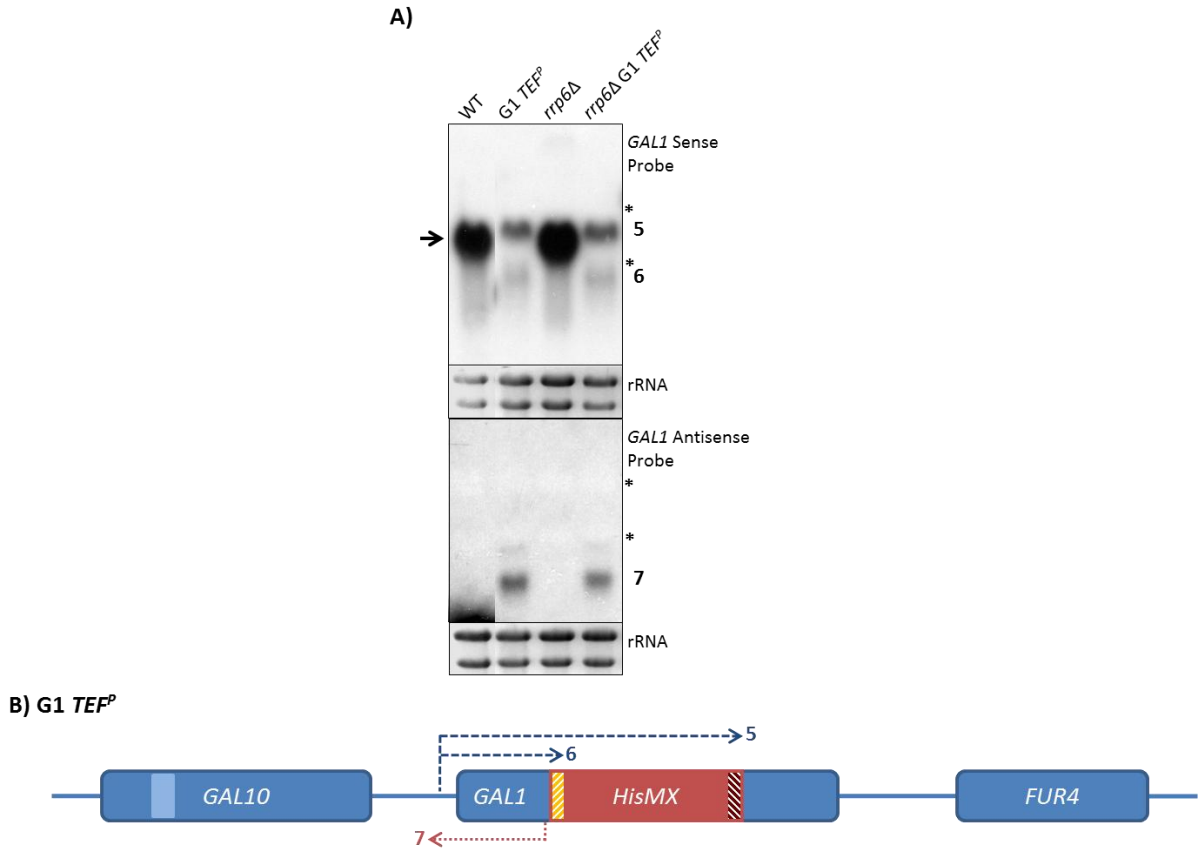


Figure 40. The transcripts expressed in the G1 *TEF*^{P} strain are not degraded by the exosome. (A) Northern blot analysis of total RNA extracted from the strains indicated and probed for *GAL1* sense and antisense. The cells were induced in galactose for 3 hours. The *HisMX* cassette is indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematic below. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. Schematic of the transcripts detected in the Northern blots above of the G1 *TEF*^{P} **(B)** strain.

So far it has been shown that the insertion of a promoter within either *GAL10* or *GAL1* resulted in the expression of new antisense transcripts and much reduced levels of the two sense transcripts initiating from the promoters and terminating either at the *AgTEF*^{P} or *AgTEF*^{T}. Furthermore, none of the transcripts at *GAL1* were found to be degraded by the nuclear exosome. At *GAL10* containing the *AgTEF*^{P}*:KanMX:AgTEF*^{T} cassette (G10 *TEF*^{P}), there is a good correlation between reduced levels of steady-state sense RNA and

reduced levels of RNA pol II, suggesting that RNA degradation is also unlikely here but this was not tested directly.

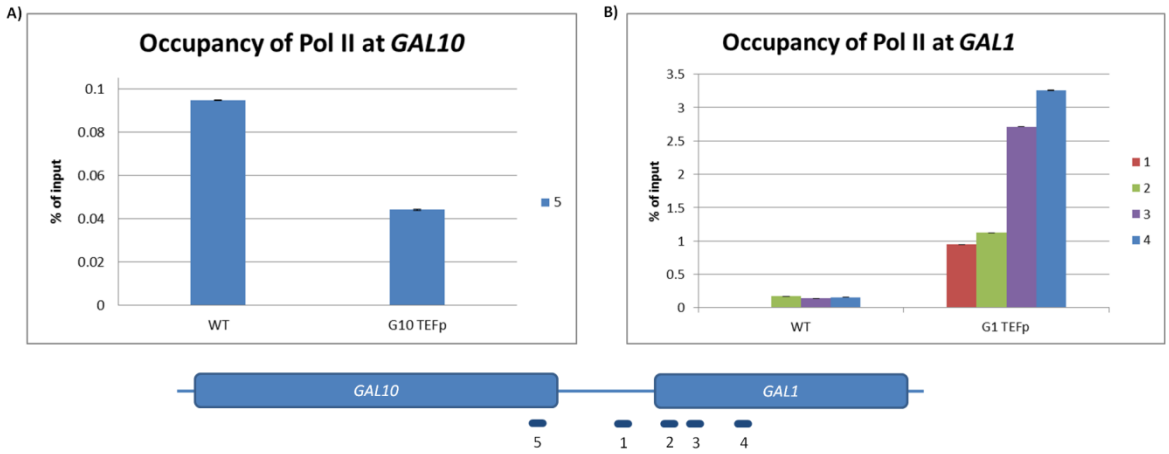


Figure 41. Frequency of RNA Polymerase II at *GAL10* and *GAL1* in the G10 *TEF*^{P} and G1 *TEF*^{P} strains. ChIP analysis of the RNA Polymerase II at *GAL10* **(A)** and *GAL1* **(B)** in the strains indicated was performed as described in section 2.6. The cells were induced in galactose for 90 minutes and then cross-linked. The primer pairs used for quantitative real-time PCR are indicated in the schematic below.

Next, it was important to check if the regulation of the *GAL10-1* promoter, particularly its inducibility, is altered due to the new transcriptional architecture resulting from the insertion of the expression cassette. The highly regulated process of *GAL* gene induction has been extensively covered in this study and this has provided important information regarding the regulation of these genes. For this reason the induction kinetics of the hybrid strains G1 *TEF*^{P} and G10 *TEF*^{P} were analysed (Figure 42). This analysis might shed light on the low level of expression from the promoter but also, perhaps, explain the unexpected RNA Pol II ChIP results at G1 *TEF*^{P}.

The strains G10 *TEF*^{P} and G1 *TEF*^{P} were induced in galactose, as normal, and harvested at several time points. In the WT strain, the *GAL10* coding sense transcript is induced 15 minutes after addition of galactose to the medium (Figure 42 A). However, for the G10

TEF^{P} strain, only transcript **2** is visible, at very low levels, at 30 minutes of induction. Steady-state levels of that transcript remain low at 3 hours of induction suggesting that the insertion does indeed affect promoter function, particularly its induction (Figure 42 A). Both the WT (transcripts **f**, **g** and **β**) and G10 *TEF*^{P} (transcripts **3**, **4** and **d**) antisense transcripts are observed in glucose and throughout the induction period.

For *GAL1* (Figure 42 B), the coding transcript is present after 60 minutes of induction, with transcripts **5** and **6** at G1 *TEF*^{P} inducing simultaneously. Consequently, there is no difference in the induction kinetics of the transcripts produced at *GAL1* in the G1 *TEF*^{P} strain compared to the WT. However, the overall levels of sense RNA remain low. This might be explained by reduced promoter function in all cells in the population or a reduction in the number of cells in a population producing transcript at any one time. Interestingly, the antisense transcript produced in that strain (transcript **7**) is present at low levels in repressed conditions and at higher levels when *GAL1* is induced by the addition of galactose. Levels of this antisense transcript (**7**) decrease as the sense transcripts begin to appear.

It is then concluded that there is no substantial effect of the *TEF*^{P} insertion on the induction kinetics at *GAL10* and *GAL1*, as the new transcripts are detected at similar time points. The antisense transcripts at G10 *TEF*^{P} do not change compared to the WT and do not vary in levels during induction. However, one interesting observation arising from the *GAL1* insertion experiments is the reciprocal nature of sense and antisense levels during induction, as sense transcript only appears once antisense levels begin to decrease. This raises the question as to whether the production of stable antisense RNA is associated

with repressing sense transcription and, if so, the nature of the mechanism by which sense transcription is reduced.

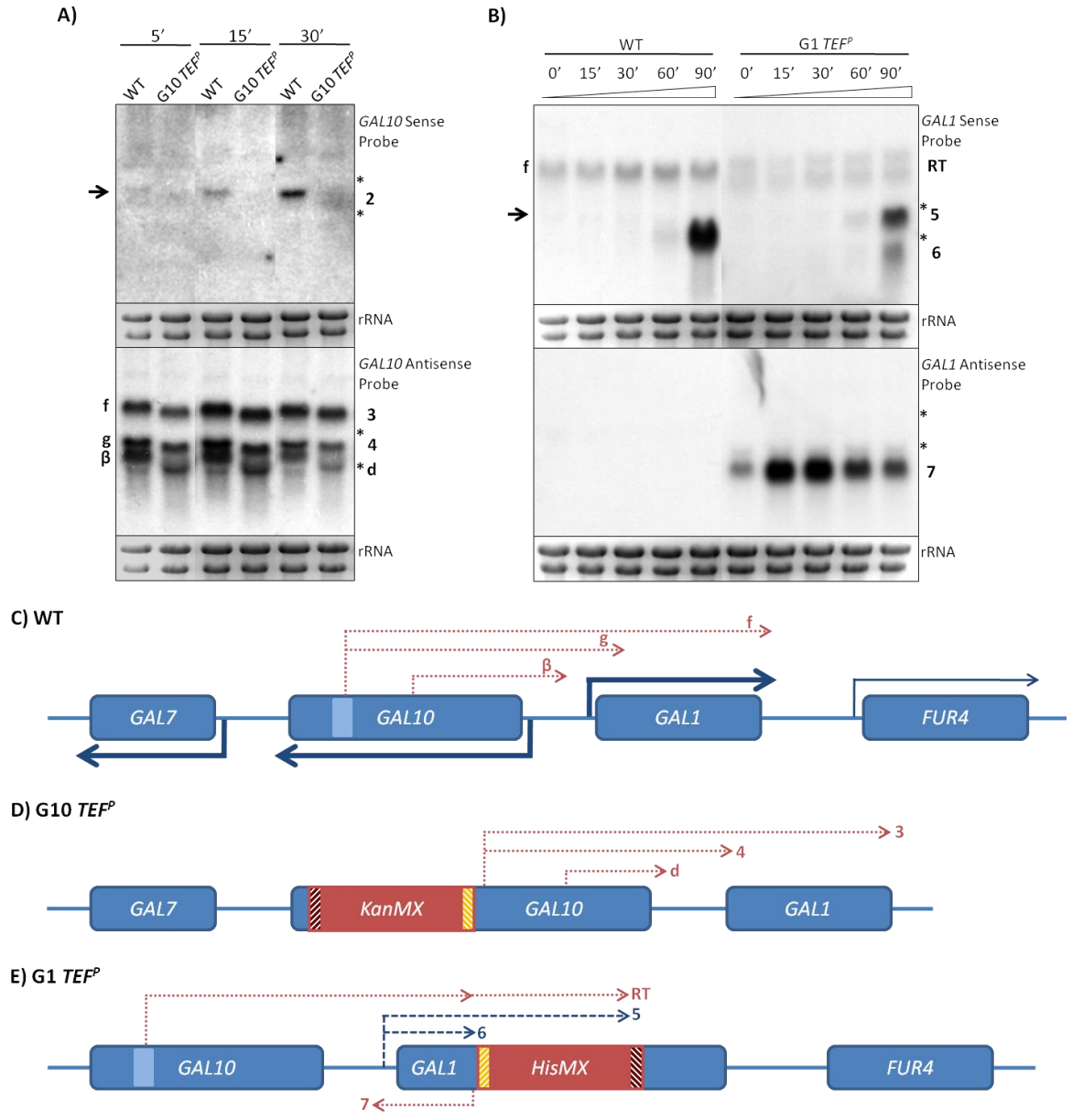


Figure 42. Induction kinetics of the *GAL10 TEF*^{P} and *GAL1 TEF*^{P} strains. Northern blot analysis of total RNA extracted from the strains designated, harvested at the time points indicated and probed for *GAL10* **(A)** and *GAL1* **(B)** sense and antisense. The *KanMX* and *HisMX* cassettes are indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematics below. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown for loading controls. Schematics of the transcripts detected in the Northern blots above of the WT **(C)**, the G10 *TEF*^{P} **(D)** and G1 *TEF*^{P} **(E)** strains.

7.1.1.2 Adenylation profile of the transcripts produced from the *AqTEF* promoter

It is known that most of the non-coding transcripts at the *GAL* locus are polyadenylated, from the signals in the tiling array hybridised with poly (A)+ enriched RNA (Chapter 3). Since the insertion of the TEF^P resulted in the production of new non-coding transcripts, it was investigated whether these are also polyadenylated. To this end, total RNA was mixed with oligo dT coated beads, washed and then eluted for the separation of poly (A)+ from poly (A)- transcripts. The non-bound fraction (A-) was kept and the RNA concentrated to use as a control. The G10 TEF^P strain was induced in galactose for 3 hours whereas the G1 TEF^P was induced for 15 minutes for higher expression of the antisense transcript (transcript **7**). The ribosomal RNA is not adenylated and therefore it is present at much reduced levels in the poly (A)+ enriched fraction. The presence of the rRNAs in the samples causes a slower migration of the mRNAs through an agarose gel, and therefore there is a difference in position of the transcripts between the poly (A)+ and the poly (A)- lanes (Figure 43). The poly (A)+ and poly (A)- transcripts of each strain came from the same total RNA sample and should therefore be directly comparable, taking into account the migration differences. The sense transcripts and the longer antisense RNA species in the *GAL10* strains were considered, whereas at *GAL1* the longer sense RNA species (the *GAL10* antisense RNA) and the antisense transcript were examined.

Figure 43 (A) shows the profile of the sense and antisense transcripts at *GAL10*. The *GAL10* sense transcripts, both in the WT and the G10 TEF^P strains, are equally distributed into the poly (A)+ versus poly (A)- fractions, although levels of these transcripts are much reduced for the G10 TEF^P strain. Interestingly, both the long read-through sense

transcript (**1**) and the short sense transcript (**2**) are equally polyadenylated in the G10 *TEF*^{P} strain, suggesting that a conventional terminator is not absolutely necessary for polyadenylation or that these transcripts are subject to polyadenylation by a different mechanism. The antisense transcripts (**f**) and (**3**) show quite a different profile, with the majority of transcript being in the poly (A)- fraction, albeit again that the overall levels are quite different.

The profile for the G1 *TEF*^{P} strain (Figure 43 B) shows a higher proportion of poly (A)- RNA for the sense transcripts **RT** and **5** early in induction. Transcripts **RT** and **f** both originate at the *GAL10* internal bi-directional promoter and share the same poly (A)+/(A)- distribution suggesting that they might be processed in the same way although they terminate in very different sequences (*GAL1*^{T} for **f** and *AgTEF*^{P}/*AgTEF*^{T} for **RT**). Finally the antisense transcript **7** is predominantly polyadenylated early in induction suggesting that this is the primary transcription unit on the G1 *TEF*^{P} gene early in induction. What is not yet clear is whether the sense and antisense transcripts are polyadenylated as default (regardless of their eventual fate as mRNA, CUT substrates for Rrp6p or dot RNAs) and subject to active de-adenylation to product the poly (A)- population or whether some transcripts are never adenylated. In steady-state it is clear that long transcripts, particularly if transcribed in the antisense direction and across more than one gene, tend to be enriched in the poly (A)- population.

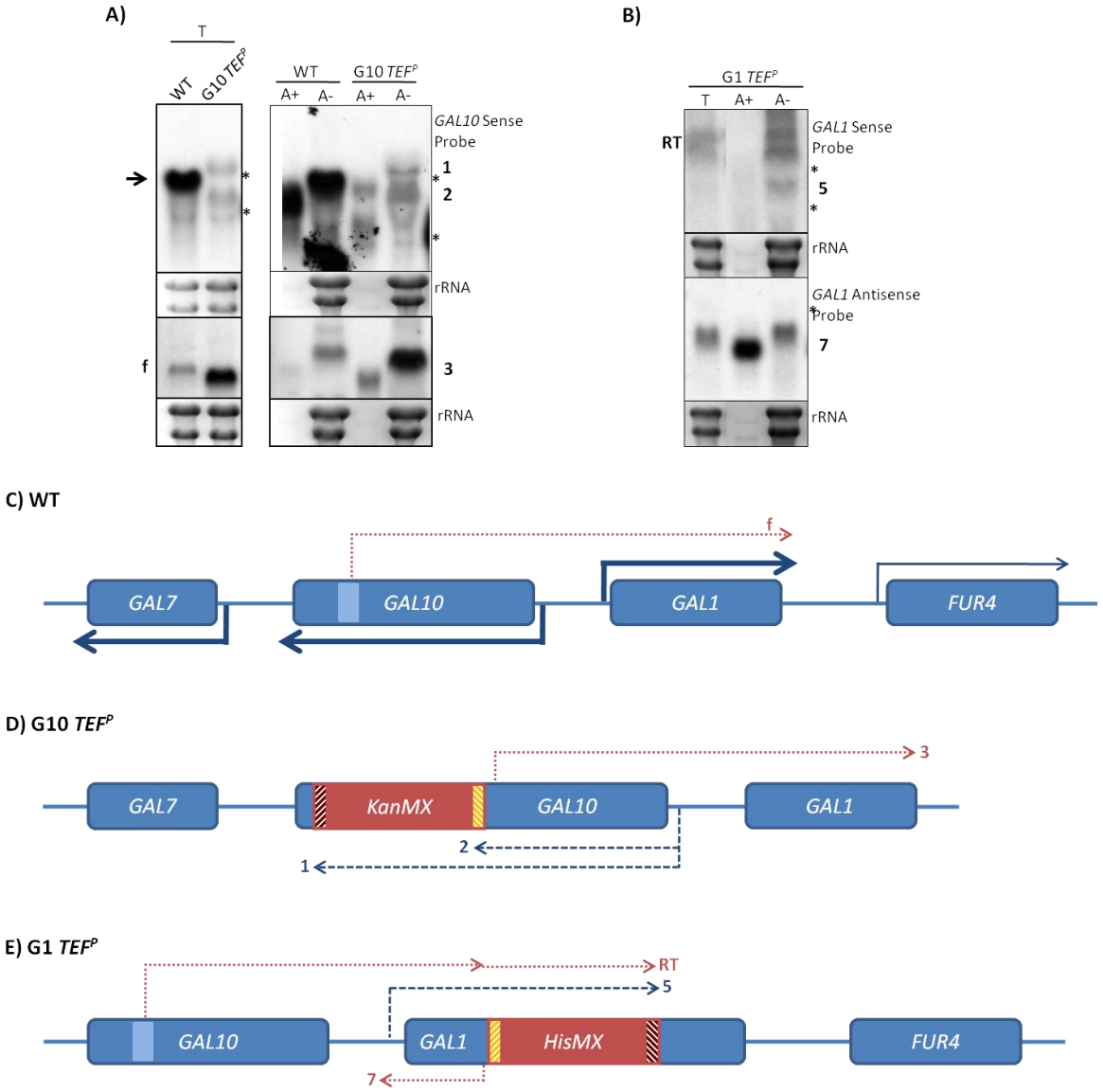


Figure 43. Different states of transcript polyadenylation. Northern blot analysis of total (T), poly (A)+ (A+) and poly (A)- (A-) enriched RNA extracted from the strains designated and probed for *GAL10* **(A)** and *GAL1* **(B)** sense and antisense. The cells were induced in galactose for 3 hours (A) and 15 minutes (B). The *KanMX* and *HisMX* cassettes are indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematics below. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown for loading controls. Schematics of the transcripts detected in the Northern blots above of the WT **(C)**, the G10 *TEF*^{P} **(D)** and G1 *TEF*^{P} **(E)** strains.

7.1.1.3 RNA FISH to determine the distribution of RNA in the nucleus and cytoplasm of WT and GAL10 TEF promoter strains

RNA FISH on fixed single cells can be used to ask if transcripts are retained in the nucleus or exported into the cytoplasm. Nuclear export is generally considered to be linked to the termination of transcription and polyadenylation of the transcript. Thus, the lack of a terminator/poly(A) site would lead to the lack of a poly (A)+ tail, which in turn would destabilize the improperly processed RNAs and/or impair export due to the well established coupling of 3' end formation and export (Brodsky and Silver, 2000; Straszer et al., 2002; Zenklusen et al., 2002). In the G10 TEF^{P} strain, the transcription unit lacks a terminator. The distribution of the sense transcripts that accumulate in the G10 TEF^{P} compared to WT, after 180 minutes induction, was assessed using RNA FISH. *GAL10* (green) RNA FISH sense probes were designed and hybridized to fixed G10 TEF^{P} and WT cells. Representative fields are shown in Figure 44. In the WT cells the signal is found throughout the cell with a bright dot coincident with the edge of the DAPI stained material and assumed to be nuclear. By contrast, most of the signal in the G10 TEF^{P} strain is concentrated into bright dots overlapping the DAPI stain with very little signal in the cytoplasm. The dots observed here are likely to be similar to the dot RNAs described by the Rosbash group (Vodala et al., 2008). Using RNA FISH, the Rosbash group described the accumulation of *GAL1* sense transcript near the site of transcription close to the nuclear periphery. Their work suggests that nuclear dots are a consequence of an accumulation of poly (A)- sense transcripts near the *GAL1* gene and this causes the tethering of the gene to the nuclear periphery (Vodala et al., 2008). Interestingly, the accumulation of poly (A)- RNA in the nucleus is dependent on the nuclear exosome,

which is involved in exonucleolytic processing of polyadenylated transcripts from their 3' region to create the de-adenylated RNA.

The images suggest that both the WT and G10 *TEF*^{P} strains have nuclear dot RNA, in addition, the WT strain has significant cytoplasmic staining, as expected. To quantify this, two different parameters were assessed (i) the overall amount of signal in a cell, including the cytoplasm and nucleus, that relates to steady-state levels of the given transcript and (ii) the intensity of the nuclear dot RNA (maximum signal) relative to the median signal in the cell, which indicates the relative number of transcripts located in a single dot, and is related not only to steady-state levels but also to the export efficiency and/or nuclear retention. As Figure 44 (B) shows, the relative intensity of the nuclear dot RNAs compared to the median signals (but not the absolute signals; not shown) was increased in the G10 *TEF*^{P} strains at the same time as the overall transcript levels were reduced by half, compared to the WT Figure 44 (C). This result is generally in agreement with the data obtained by Northern blotting (Figure 39) and suggests that in the G10 *TEF*^{P} strain, much of the RNA is retained as dot RNA in the nucleus.

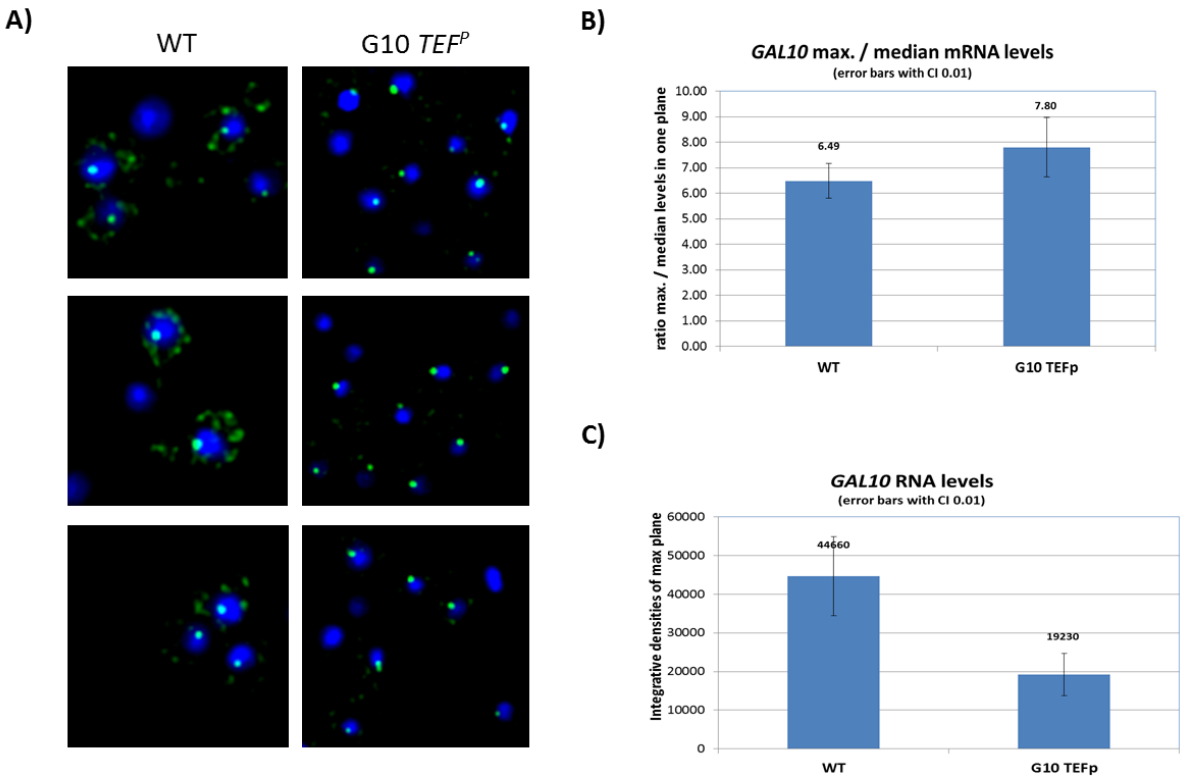


Figure 44. *GAL10* nuclear sense RNA dots in the WT and G10 *TEF*^{P} strains. Cells were grown as described in section 2.3. WT and G10 *TEF*^{P} cells were induced in galactose for 3 hours. **(A)** *GAL10* sense transcripts (green) were detected with cyanine 3 labelled probes. DAPI (blue) was used to identify the nucleus. Images were obtained with an average projection of 3 layers. **(B)** Quantitative analysis of the *GAL10* dot RNA intensity. Calculated as the maximum intensity signal relative to the media intensity signal. **(C)** Quantitative analysis of the overall signal of the *GAL10* sense transcripts. Intensity densities of the plane with the maximum signal. Error bars calculated with interval confidence of 0.01. Performed by Simon Haenni.

7.1.1.4 Insertions at *GAL1* do not influence expression of sense or antisense transcripts at *GAL10*.

The insertion of the *AgTEF*^{P}*:HisMX:AgTEF*^{T} cassette at *GAL1* resulted in a ten-fold decreased in *GAL1* steady-state sense transcript and the production of a stable antisense transcript. Such a pronounced decreased in transcription from the *GAL1* promoter raised the question concerning the possible consequences for transcription of the neighbouring *GAL10* gene.

Figure 45 (A, lanes 1 and 2), shows that there is no difference in the levels of the *GAL10* coding transcript between the WT and G1 TEF^{P} strains. This indicates that the decreased in transcription at *GAL1* does not affect the *GAL10* promoter.

Figure 45 shows that the *GAL10-1* antisense transcript (transcript **f**) is truncated at the $AgTEF^{P}$ and reads through the cassette until the $AgTEF^{T}$, as previously described for the *GAL1* sense transcripts. This occurs regardless of the position of the inserted cassette within the *GAL1* gene. Note that the cassette placed outside the *GAL1* transcription unit does not influence expression of *GAL1* or *GAL10*. This is discussed in more detail in the next section.

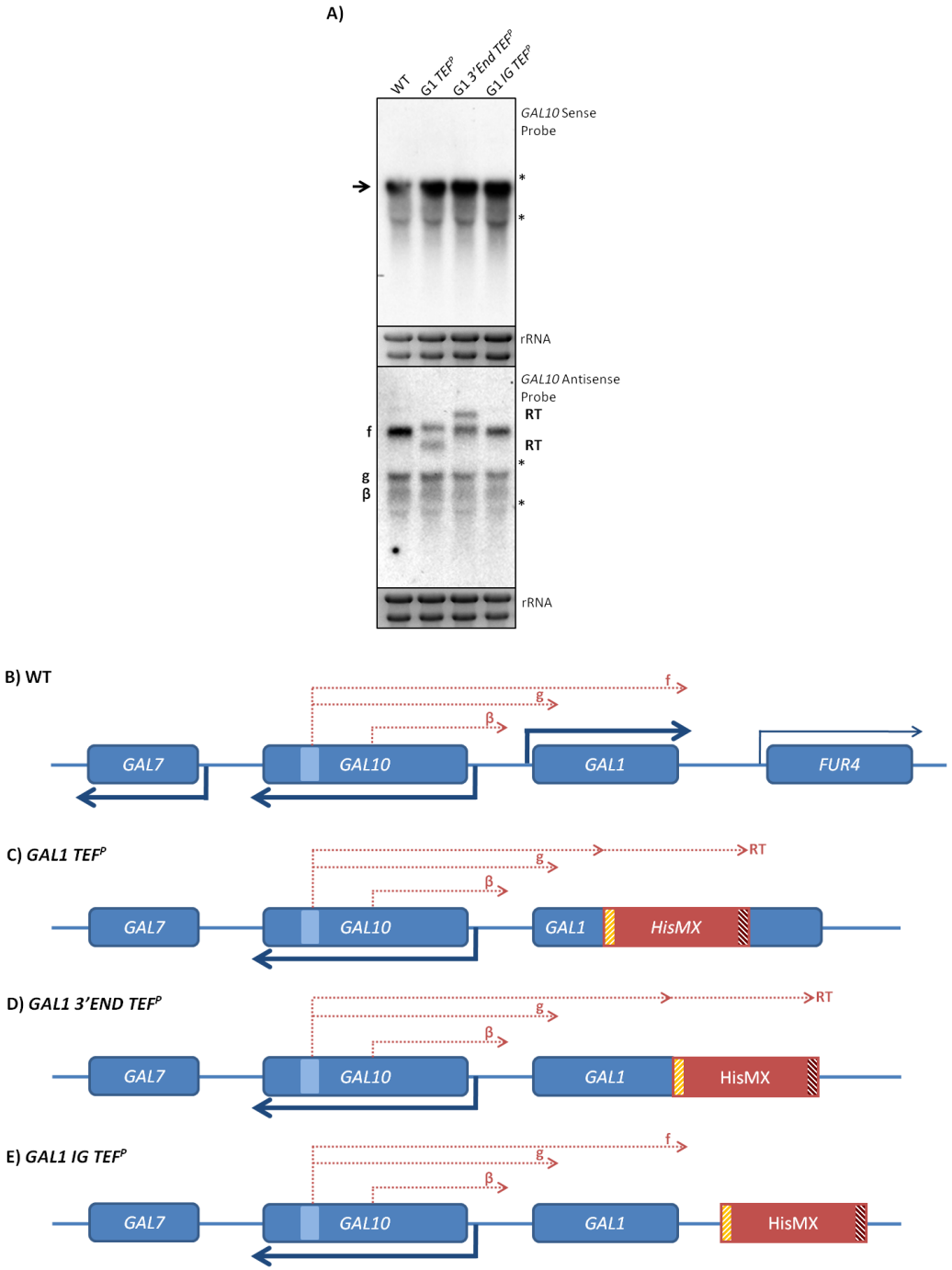


Figure 45. The promoter construct inserted at *GAL1* terminate the *GAL10* antisense transcripts. **(A)** Northern blot analysis of total RNA extracted from the strains indicated and probed for *GAL1* sense and antisense. The cells were induced in galactose for 3 hours. The *HisMX* cassette is indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematics below. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown for loading controls. Tiling array data for WT in YPD and YPG and Schematics of the strains and transcripts detected in the Northern blots above of the WT **(B)**, G1 *TEF*^{P} **(C)**, G1 *3'END TEF*^{P} **(D)** and G1 *IG TEF*^{P} **(E)** strains.

7.1.1.5 Insertion of the *AgTEF*^{P} at other locations at *GAL1* does not result in a stable AS transcript

To control for possible effects of the location of the insertions, two further strains were created, where the *AgTEF*^{P}*:HisMX:AgTEF*^{T} cassette was either inserted after the stop codon (G1 3'END *TEF*^{P}) or after the 3' UTR (G1 IG *TEF*^{P}) of *GAL1* and the transcription profiles were analysed by Northern blotting. As shown in Figure 46, the insertion of the cassette at the 3' end of *GAL1* (G1 3'END *TEF*^{P} strain) still resulted in decreased transcription from the *GAL1* promoter and the same pattern of transcripts as when the cassette was inserted +757 bp after the *GAL1* ATG. However, insertion in the intergenic region between *GAL1* and *FUR4* (G1 IG *TEF*^{P} strain) resulted in normal expression of *GAL1*, as expected.

Interestingly, analysis of the antisense strand revealed that only the G1 *TEF*^{P} strain produced a visible antisense transcript. It is not known whether there is no antisense transcription or whether the transcript is unstable. As described before (Chapter 3) there is an antisense XUT present at *GAL1*, making an unstable transcript a probable explanation. However, this remains to be tested experimentally.

Analysis of the *GAL10* coding transcript in these strains (Figure 45) revealed that expression was not compromised. Furthermore, the *GAL10* antisense transcript (transcript **f**) was visibly truncated in the G1 3'END *TEF*^{P} strain but not in the G1 IG *TEF*^{P} strain, where it was comparable to the WT transcript, confirming that this transcript terminates in the *GAL1* 3' UTR.

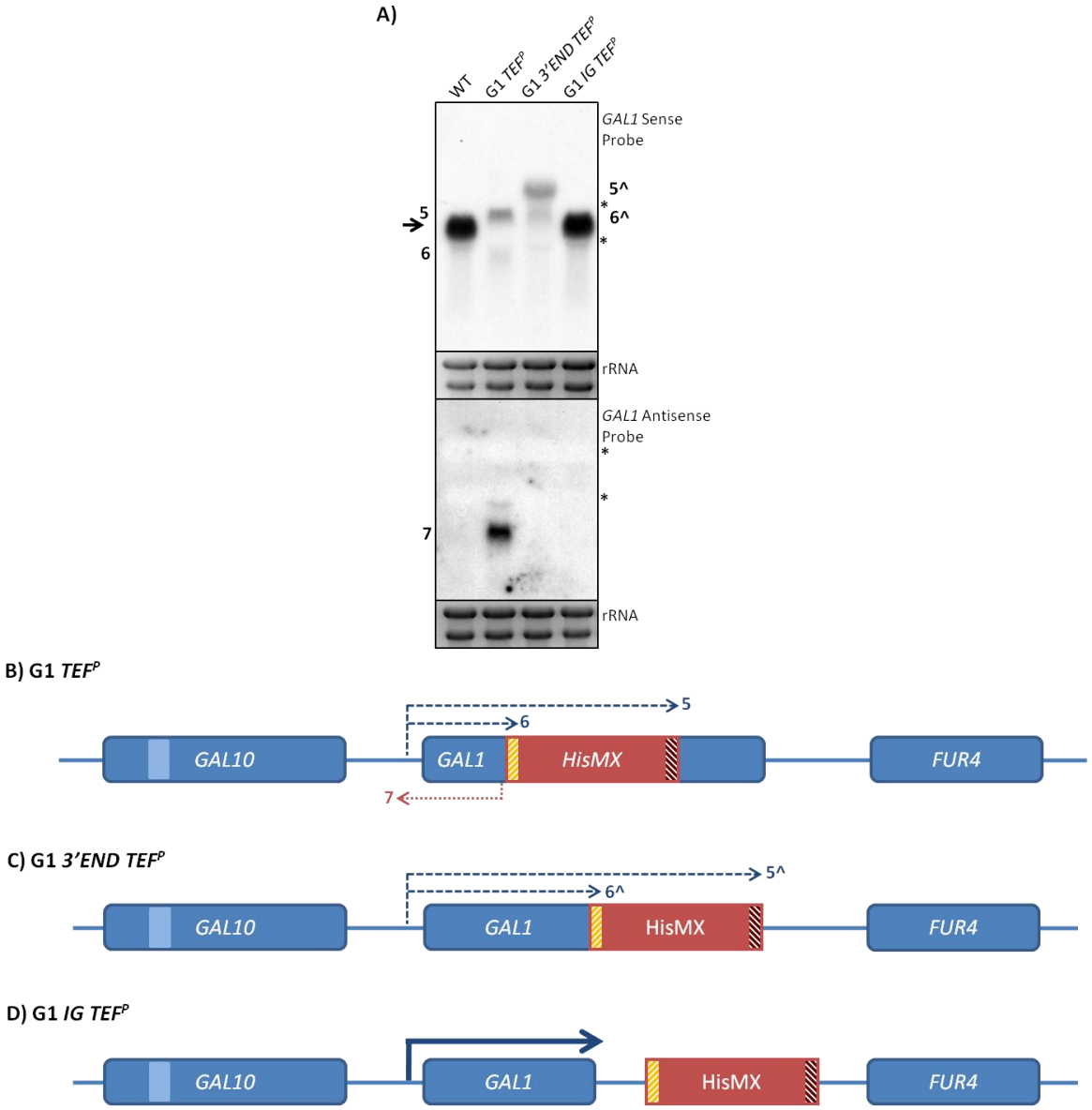


Figure 46. The promoter constructs inserted at different locations at *GAL1* do not drive stable antisense transcription. (A) Northern blot analysis of total RNA extracted from the strains indicated and probed for *GAL1* sense and antisense. The cells were induced in galactose for 3 hours. The *HisMX* cassette is indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematics below. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown for loading controls. Schematics of the strains and transcripts detected in the Northern blots above of the G1 *TEF*^{P} **(B)**, G1 *3'END TEF*^{P} **(C)** and G1 *IG TEF*^{P} **(D)** strains.

In summary, in this section the transcriptional architecture of transcription units was investigated. The insertion of a promoter within an ORF, creating a promoter - promoter

organization, resulted in the initiation of antisense transcription from the inserted promoter and the decreased expression from the natural promoter. The *AgTEF^{P}:HisMX:AgTEF^{T}* cassette was inserted at two genes, *GAL10* and *GAL1*, and the results obtained were identical. At both genes, the *AgTEF* promoter was capable of initiating antisense transcription, supporting the idea that many promoters are bi-directional. However, the effects of the antisense transcription on levels of the sense transcription are not easily discernible. At *GAL10*, the differences in expression from the *GAL10* promoter did not correlate with antisense transcription since no changes in expression were observed. At *GAL1* the expression of the stable antisense transcript produced from the *AgTEF^{P}* was induced in the presence of galactose and the levels decreased upon expression of the sense transcript. This raised the question of whether the stable antisense transcript is functionally connected to the reduced sense transcription.

None of the transcripts produced from the G10 *TEF^{P}* or G1 *TEF^{P}* hybrid genes appeared to be subject to degradation by the nuclear exosome, confirmed by the effect of a deletion of *RRP6* at the *GAL1* locus and the RNAPII levels at *GAL10* locus. Interestingly, levels of RNAPII at *GAL1* do not correlate with the transcription profile detected by Northern blot and further investigation needs to be done in order to address the cause.

Insertion of the *AgTEF^{P}:HisMX:AgTEF^{T}* cassette at other locations at *GAL1,* or in the intergenic region between *GAL1* and *FUR4,* resulted in no observable stable antisense transcripts. Although the presence of unstable antisense transcripts at these hybrid genes was not investigated, if this is the case then this raises the question of what makes a transcript stable or unstable and by what mechanism.

The nature of the long non-coding transcripts produced from these hybrid genes was investigated and it was found that many are non-adenylated. This raised the question of whether this is due to the mechanism described by Rosbash, whereby transcripts are processed from poly (A)+ to poly (A)- by Rrp6p and gathered into "dot RNA" at the nuclear periphery or by some other mechanism. Interestingly, the intensity of the dot RNA present in the strains expressing the hybrid *GAL10 TEF*^{P} gene is higher compared to WT. It is not clear if the dots contain the poly (A)- long non-coding RNA and/or the poly (A)- truncated sense transcripts. Nevertheless sequestration of RNA into dots could lead to inefficient RNA export from the nucleus and perhaps repression of expression from the major promoter. Indeed, it was found that, for both G10 *TEF*^{P} and G1 *TEF*^{P} strains, the expression of the truncated *GAL10* and *GAL1* transcripts were much reduced compared to the WT. Thus, steady-state levels of sense transcript are compromised in a Promoter-Promoter transcriptional architecture. This indicates that a Promoter-Promoter configuration is not as efficient as a Promoter-Terminator configuration, raising the question of whether there are sequences present in the terminator that are responsible for proper expression from the promoter. This is addressed in the next section.

7.1.2 Analysis of a Promoter – Terminator transcriptional organization and its effects on sense and antisense transcription

7.1.2.1 The 3' region of a gene can initiate antisense transcription

The next question to be addressed is whether a terminator placed within an ORF can initiate transcription of an antisense transcript (and terminate sense transcription) and, if so, what is the consequence of the antisense transcript on expression from the upstream promoter?

To answer this question, the same approach was used as in the previous section whereby a terminator construct was inserted within the ORF of *GAL10* and *GAL1*. The *ScADH1* terminator was used since it is a strong bi-directional terminator and it is usually used for genetic tagging of proteins (Longtine et al., 1998), and therefore is easy to manipulate. The $ScADH1^{T}$ and the selective marker cassette were inserted at *GAL10* and *GAL1* in the same positions as the previous experiments, with the terminator in the conventional orientation with respect to the *GAL1* or *GAL10* promoters. The $ScADH1^{T}$:$AgTEF^{P}$:*KanMX*:$AgTEF^{T}$ was inserted in *GAL10* +1453 bp after the *GAL10* ATG codon, after deletion of the region between +1453-2007 bp to remove the internal bi-directional promoter. At *GAL1* the $ScADH1^{T}$:$AgTEF^{P}$:*HisMX*:$AgTEF^{T}$ cassette was inserted +757 bp after the *GAL1* ATG codon, without any deletions. The schematic of the strains engineered is shown in Figure 47.

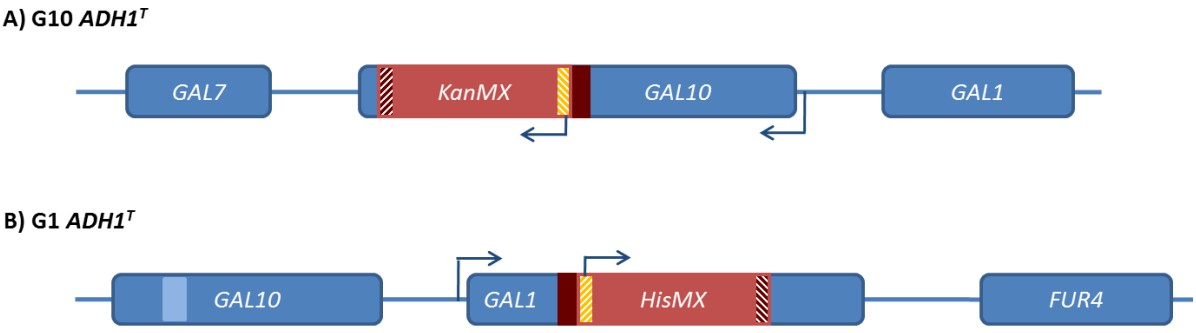


Figure 47. Schematic of the terminator constructs inserted at *GAL10* and at *GAL1*. Both strains were constructed as described in section 2.1. The *KanMX* and *HisMX* cassettes were used and are indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The *ScADH1* terminator is represented by the dark red box. The light blue box represents the *GAL10* internal bi-directional promoter. The arrows indicate the direction of transcription. **(A)** The *ScADH1*^{T}*:AgTEFP:KanMX:AgTEF*^{T} cassette was inserted in *GAL10* at +1453 bp with the simultaneous deletion of the region between +1453 and +2007 bp, which corresponds to the internal bi-directional promoter. **(B)** The *ScADH1*^{T}*:AgTEF*^{P}*:HisMX:AgTEF*^{T} cassette was inserted in *GAL1* at +757 bp, without any deletions.

Both strains were induced in galactose and the total RNA extracted for a Northern blotting experiment performed with *GAL10* and *GAL1* sense and antisense probes. Figure 48 shows that the insertion of the *ScADH1*^{T} in the *GAL10* and *GAL1* coding regions results in truncated sense transcripts, identifiable by their shorter size, which are expressed to similar steady-state levels as the WT strain. This indicates that the *ScADH1*^{T} sequences inserted are sufficient for transcription processing and termination, as expected.

Interestingly, the antisense transcripts produced in both strains (Figure 48) are similar to those produced as a result of insertion of the *AgTEF* promoter described in the previous section. At *GAL10* (Figure 48 A), the transcripts **3** (which starts from or in the vicinity of the *ScADH1*^{T} and terminates at the end of *GAL1*), **4** (which starts from or in the vicinity of the *ScADH1*^{T} and terminates at the *GAL1* promoter) and **d** (which starts in the middle of *GAL10* and terminates at the promoter) show the same general pattern but are expressed to lower levels than in the WT strain. At *GAL1* (Figure 48 B), transcript **7** (which

starts from or in the vicinity of the *ScADH1*^{T} and terminates at the *GAL1* promoter) is expressed at very low levels. Interestingly, the size of this transcript is very similar to the truncated *GAL1* sense transcript, making this a probable SAP (sense antisense pair).

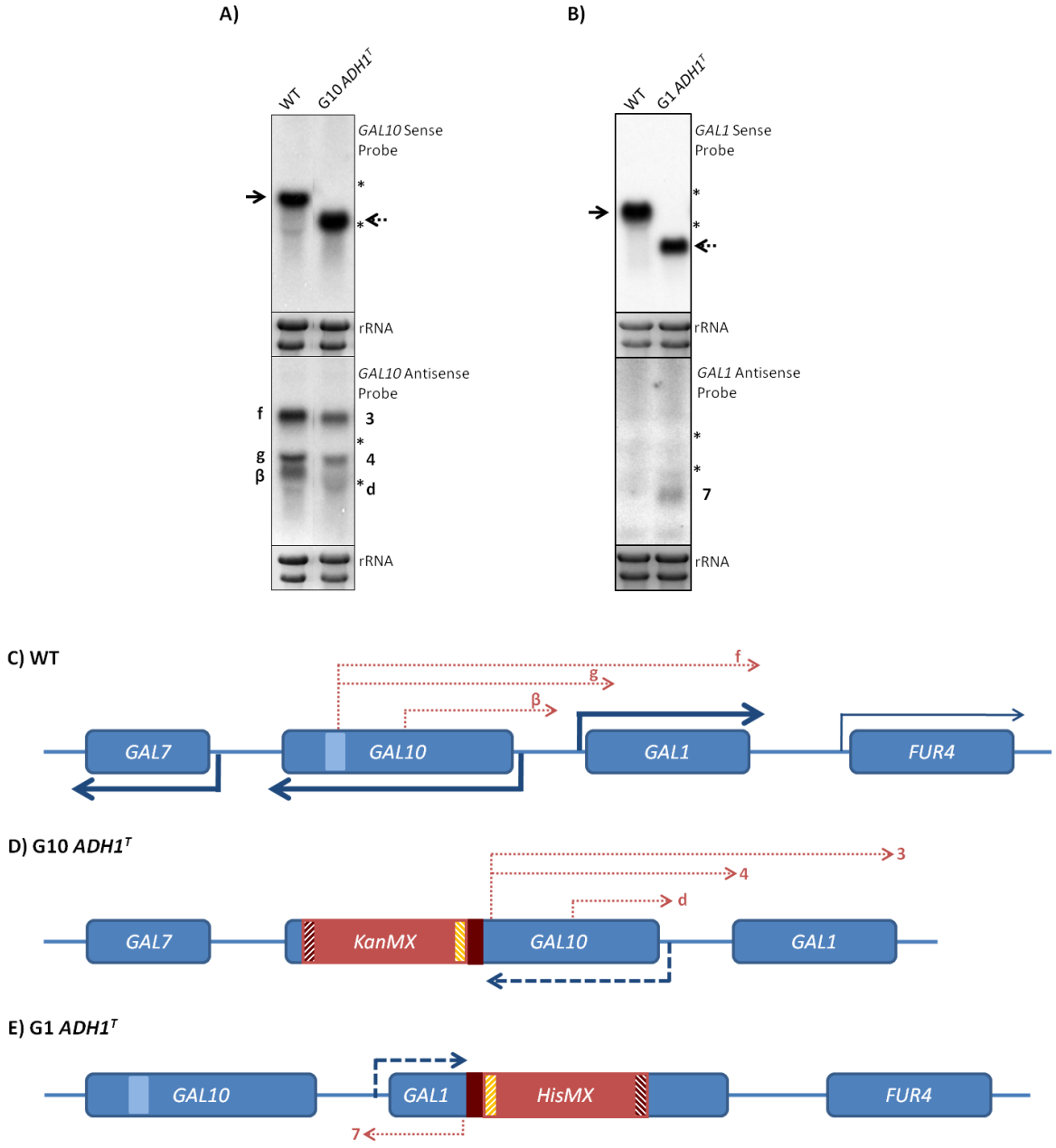


Figure 48. The terminator constructs inserted at *GAL10* and at *GAL1* terminate sense transcription and drive antisense transcription. Northern blot analysis of total RNA extracted from the strains indicated and probed for *GAL10* **(A)** and *GAL1* **(B)** sense and antisense. The cells were induced in galactose for 1 hour (A) and 3 hours (B). The *KanMX* and *HisMX* cassettes are indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The *ScADH1* terminator is represented by the dark red box. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematics below. The coding transcripts are identified by a black arrow and the truncated version by a dotted black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown for loading controls. Schematics of the transcripts detected in the Northern blots above of the WT **(C)**, the G10 *ADH1*^{T} **(D)** and G1 *ADH1*^{T} **(E)** strains.

To check if the truncated sense transcripts were still under the same mechanisms of regulation as the WT strain, the induction kinetics were investigated. The G10 *ADH1*^{T} and G1 *ADH1*^{T} strains were induced in galactose and harvested at several time points. Figure 49 (A), shows the results for the G10 *ADH1*^{T} strain, where the truncated sense transcript is expressed by 15 minutes of induction, equivalent to the WT strain. Both WT (transcripts **f**, **g** and **β**) and G10 *ADH1*^{T} (transcripts **3**, **4** and **d**) antisense transcripts show no difference during induction, apart from the slightly lower levels of transcripts **3** and **4** compared to **f** and **g**.

For the G1 *ADH1*^{T} strain (Figure 49 B), the truncated sense transcript also induces at the same time as the WT, being just visible at 1 hour of induction. The antisense transcript (transcript **7**) is present in repressed conditions (0' time point) and is rapidly induced in the presence of galactose and gradually repressed during the induction time. This profile is identical to the G1 *TEF*^{P} strain analysed in the previous section (Figure 42 B).

These results indicate that the presence of a terminator is sufficient for efficient expression from the *GAL10-1* promoter, manifested by the high levels of truncated sense transcripts. The *ScADH1*^{T} is also sufficient for the production of antisense transcription at both *GAL10* and *GAL1*.

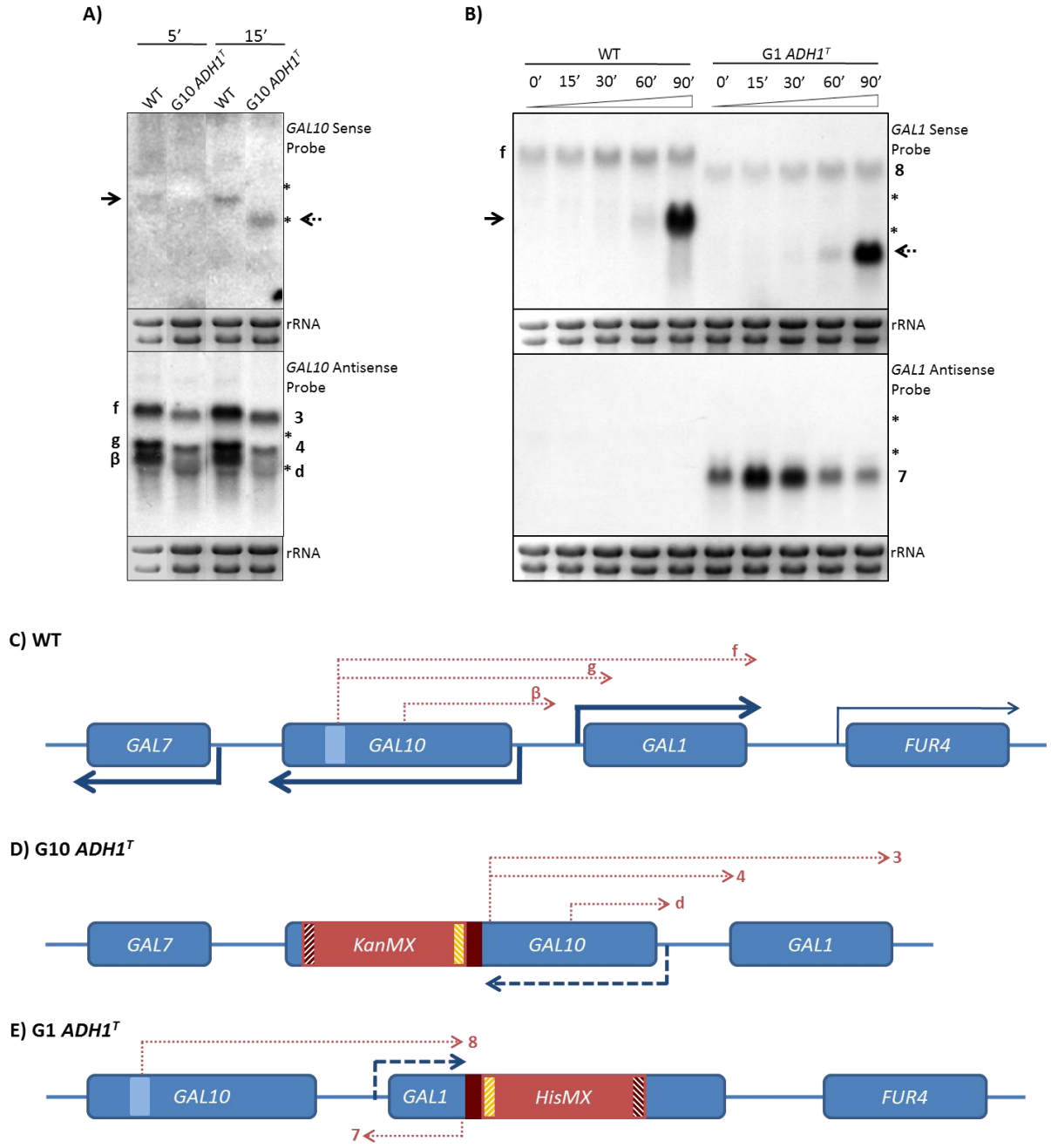


Figure 49. Induction kinetics of the *GAL10 ADH1*^{T} and *GAL1 ADH1*^{T} strains. Northern blot analysis of total RNA extracted from the strains designated, harvested in the time points indicated and probed for *GAL10* **(A)** and *GAL1* **(B)** sense and antisense. The *KanMX* and *HisMX* cassettes are indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The *ScADH1* terminator is represented by the dark red box. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematics below. The coding transcripts are identified by a black arrow and the truncated version by a dotted black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown for loading controls. Schematics of the transcripts detected in the Northern blots above of the WT **(C)**, the G10 *ADH1*^{T} **(D)** and G1 *ADH1*^{T} **(E)** strains.

7.1.2.2 Antisense transcription is not due to the neighbouring *AgTEF* promoter

Since the *ScADH1*^{T} insertion at *GAL10* and *GAL1* also contains the *AgTEF* expression cassette, it was important to show that the expression cassette does not influence transcription initiation directed by the terminator sequences. In order to control for this possibility, the *KanMX* cassette at *GAL10* was removed leaving the *ScADH1*^{T} (G10 *ADH1*^{T} *LoxP*) or the *AgTEF*^{T} (G10 *TEF*^{T} *LoxP*) (Figure 50). The cassettes were removed by Cre-lox recombination, which was achieved by the transformation of the G10 *ADH1*^{T} strain with the loxP-*LEU2*-loxP disruption cassette, ejecting the *kan*^{r} coding region and flanking *AgTEF*^{P} and *AgTEF*^{T} sequences, by targeted homologous recombination. The introduction of the Cre protein expressed on a his+ plasmid drives site-specific recombination between the introduced loxP sequences, excising the introduced *LEU2* marker gene leaving a 106 bp loxP signature on the 3' side of the *ScADH1* terminator. The *AgTEF*^{T} is the terminator present in the cassettes and the G10 *TEF*^{T} *LoxP* strain was engineered to answer the question of whether a different terminator is capable of driving antisense transcription. In this case the loxP sequences are upstream of *AgTEF*^{T}.

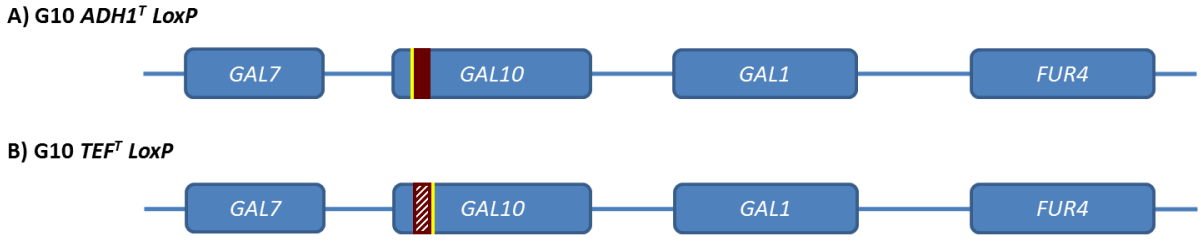


Figure 50. Schematic of the cassette-less terminator constructs inserted at *GAL10*. Both strains were constructed as described in section 2.1. The *KanMX* and *HisMX* cassettes were excised using the LoxP system. The *ScADH1* terminator is represented by the dark red box and the *AgTEF* terminator by the dark diagonal lines. The yellow line represents the residual loxP sequence. **(A)** The *ScADH1*^{T}*:AgTEFP:KanMX:AgTEF*^{T} cassette was inserted in *GAL10* at +1453 bp, with the simultaneous deletion of the region between +1453 and +2007 bp, which corresponds to the internal bi-directional promoter. **(B)** The *ScADH1*^{T}*:AgTEF*^{P}*:HisMX:AgTEF*^{T} cassette was inserted in *GAL1* at +757 bp, without any deletions. Strains constructed by David Brown.

The strains were induced in galactose, harvested at different time points and a Northern blotting experiment was performed with the total RNA extracted and probed for *GAL10* sense and antisense (Figure 51). The first observation is that in both LoxP strains the *GAL10* sense transcripts are truncated and are similar in size to those present in the G10 *ADH1*^{T} strain (Figure 48). Additionally, the levels and the induction kinetics for the sense transcripts are similar to those in the WT strain, with transcripts clearly visible at 1 hour in each of the three strains. This indicates that the *ScADH1*^{T} and *AgTEF*^{T} do not influence the regulation of the sense transcription differently from the WT, and that there is no influence of the *AgTEF*^{P}*:KanMX:AgTEF*^{T} expression cassette on *ScADH1*^{T} or *AgTEF*^{T} function.

The antisense transcript profile for the G10 *ADH1*^{T} *LoxP* strain looks very similar to that of the G10 *ADH1*^{T} strain (Figure 48), with transcripts **3**, **4** and **d** (described previously) visible. It is also noticeable that the levels of these antisense transcripts are lower than those of the WT. In the G10 *TEF*^{T} *LoxP* strain, transcripts **3** and **4** are longer, consistent with the presence of the residual loxP sequences and perhaps the use of a different transcription initiation site within *AgTEF*^{T}. However, the initiation site of these transcripts has not yet been mapped accurately and this remains to be proven. The levels are also lower than in the G10 *ADH1*^{T} strain and do not seem to be influenced by galactose.

This data suggests that the *AgTEF*^{P}*:KanMX:AgTEF*^{T} does not influence: (i) antisense transcription from *ScADH1*^{T} and *AgTEF*^{T} or (ii) sense transcription from the *GAL10* promoter when *ScADH1*^{T} and *AgTEF*^{T} are inserted at +1453 bp from the *GAL10* ATG codon.

These data also show that two different terminators are sufficient to generate antisense transcripts when located in the *GAL10* ORF. Interestingly, insertion of the $AgTEF^{P}$ at the same position also generates similar patterns of antisense transcripts supporting the ideas that (i) promoters are inherently bi-directional and (ii) that terminators also have promoter function. The major difference observed between the promoter and terminator function was between steady-state levels of RNA produced from the upstream *GAL10* or *GAL1* promoters; with a terminator ($ScADH1^{T}$ or $AgTEF^{T}$), high levels of transcript are observed (similar to WT) but with a promoter ($AgTEF^{P}$), levels of steady-state RNA are ten-fold lower than WT and this is evident even during induction. The conventional explanation for this observation is that in the absence of a terminator, the RNA produced is not adenylated and is degraded in the nucleus by the exosome. However, a proportion of both the major RNA species produced are polyadenylated and this is similar to the ratio of poly (A)+ to poly (A)- for the WT sense transcript. Moreover, deletion of the nuclear exosome function has little influence on steady-state RNA levels. It was observed, however, that sense transcripts expressed from the $AgTEF^{P}$ constructs accumulate in nuclear dots and are not efficiently exported to the cytoplasm, compared to the WT transcripts or those produced from $ScADH1^{T}$ (see next section). It has been proposed that inefficient export of transcripts from the nucleus leads to reduced expression from the promoter, although there is little experimental evidence to support this model (Mapendano et al., 2010). To address this, sequences in the *ScADH1* terminator required for efficient expression from the *GAL10-1* promoter will be defined by a deletion analysis in the next section.

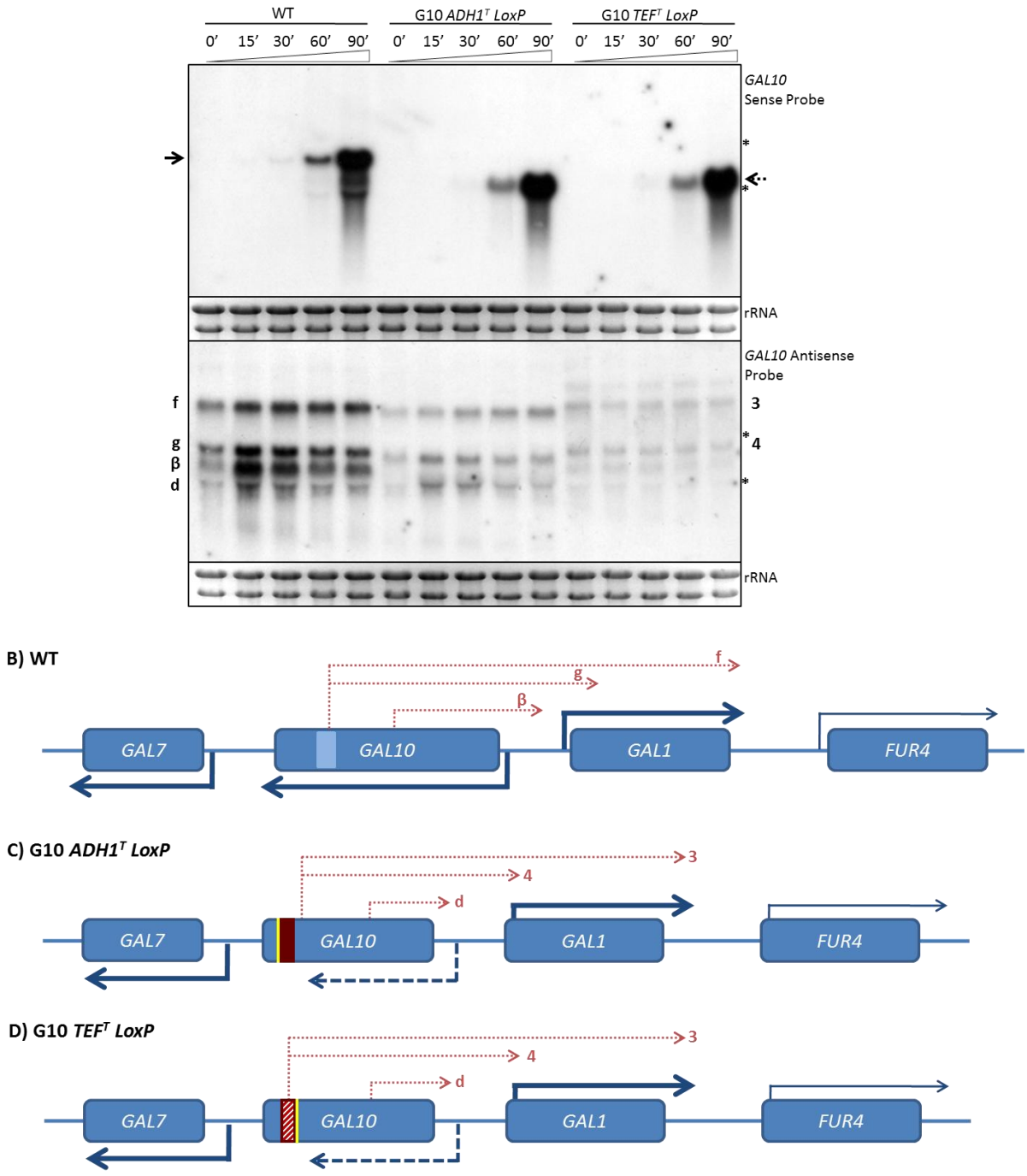


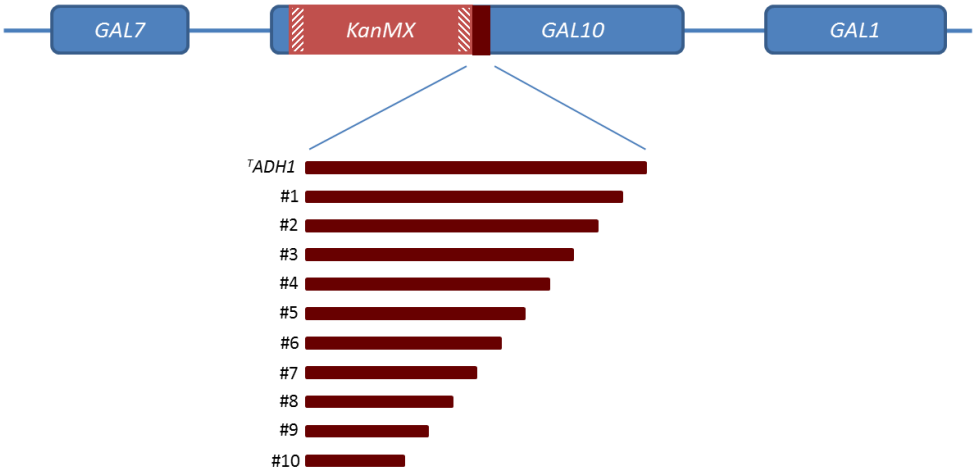
Figure 51. The cassette-less terminator constructs inserted at *GAL10* drive antisense transcription. (A) Northern blot analysis of total RNA extracted from the strains indicated and probed for *GAL10* sense and antisense. The cells were induced in galactose and harvested at the time points indicated. The dark diagonal lines represent the *AgTEF* terminator and the dark red box the *ScADH1* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematics below. The coding transcripts are identified by a black arrow and the truncated version by a dotted black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown for loading controls. Schematics of the transcripts detected in the Northern blots above of the WT **(B)**, the G10 *ADH1*^{T} *LoxP* **(C)** and G10 *TEF*^{T} *LoxP* **(D)** strains.

7.1.3 Analysis of the sequences in the *ScADH1*^{T} responsible for efficient expression from the promoter

7.1.3.1 Deletion analysis of the *ScADH1* terminator

To define sequences in the *ScADH1* terminator required for efficient expression from the *GAL1* or *GAL10* promoters, sequential deletions of 20 bp from the 5' to the 3' end of the *ScADH1*^{T} were made. Figure 52 shows a schematic of the deletions and the *ScADH1*^{T} sequence. The truncated *ScADH1*^{T} sequences were inserted at *GAL10* +1453 bp from the *GAL10* ATG codon, as in the previous strains. Examination of the *ScADH1*^{T} sequence reveals the presence of two TATA-like elements, a canonical 8 bp TATAAAA, a variant, TATACAAA, and four potential cleavage sites for polyadenylation and termination.

A) G10 *ADH1*^{T}



B) *ADH1*^{T}

5' - TGAGGCGCGCCACTTCTAAA **#1)** TAAGCGAATTTCTTATGATT **#2)** TATGATTTTTATTATTAAAT **#3)** AAGT**TATAAAAA**AAATAAGT

#4) G**TATACAAA**TTTTAAAGTGA **#5)** CTCTTAGGTTTTAAAACGAA **#6)** AATTCTTATTCTTGAGTAAC **#7)** TCTTTCCTGTAGGTCAGGTT

#8) GCTTTCTCAGGTATAGTATG **#9)** AGGTCGCTCTTATTGACCAC **#10)** ACCTCTACCGGCAGATCCGC **#11)** TAGGGATAACAGGGTAATAT - 3'

Figure 52. Schematic of the sequential deletions of the *ScADH1* terminator. (A) The *ScADH1*^{T}*:AgTEFP:KanMX:AgTEF*^{T} cassette, with different lengths of the *ScADH1*^{T}, was inserted in *GAL10* at +1453 bp with the simultaneous deletion of the region between +1453 and +2007 bp, which corresponds to the internal bi-directional promoter. **(B)** Sequence of the *ScADH1* terminator with the name (number) of the deletions annotated. For each strain, the sequence remaining is shown following the name. The TATA elements are indicated in bold and underlined, as well as the major poly (A)/cleavage site with a black line and the additional sites with dotted lines. Strains constructed by Struan Murray.

Analysis of the transcription profile of the sequential deletions of the *ScADH1*^{T} (Figure 53) was performed by Northern blotting. The G10 *TEF*^{P} strain was also used as a control since a similar transcript pattern is expected to that observed in this strain once all the sequences capable of poly(A)/termination in the *ScADH1*^{T} are deleted.

First, the sense transcription profile was examined at *GAL10*. The same transcript profile as seen previously (Figure 39 and Figure 48) is observed for the *AgTEF*^{P} and *ScADH1*^{T} insertions in *GAL10* (Figure 53). As for the deletion strains, in #1 to #4 the truncated

transcript is evident and at similar levels to the intact *ScADH1*^{T}, with the exception of the slightly lower levels detected for deletion #3. However, for deletions #5 and #6 very low levels of transcripts are detected, while deletions #7, #9 and #10 have a similar transcript profile to that of the G10 *TEF*^{P} strain, with the reappearance of transcripts **1** and **2**, consistent with loss of poly(A)/terminator function. From these observations several conclusions can be drawn for sense transcription: (i) the first 80 bp of the *ScADH1* terminator are not required for high steady-state levels of expression from the *GAL10* promoter (later in this section it is confirmed that transcripts from strain #4 are efficiently exported from the nucleus); (ii) the following 40 bp sequence is absolutely required for expression from the *GAL10* promoter, since only very low levels of transcripts are detected, but the residual sequence in these strains also seems to function as an insulator by blocking read-through transcription characteristic of the strain without the *ADH1*^{T} (G10 *TEF*^{P}); (iii) the remaining sequences in *ScADH1*^{T} do not contain the processing and termination abilities and transcripts **1** and **2**, distinctive of the G10 *TEF*^{P} strain, are again observed.

With such a contrasting sense transcription profile, the antisense profile was examined. As shown in Figure 53 with the *GAL10* antisense probe, the same transcripts as in the G10 *TEF*^{P} and G10 *ADH1*^{T} strains (transcripts **3**, **4** and **d**) are detected for every *ScADH1*^{T} deletion strain. The only noticeable difference is the levels of the transcripts **3** and **4**, which are much higher in deletions #4 to #7. However, this is not likely to be reciprocal with sense as levels of both sense and antisense are high in #4. A subtle but reproducible increase in the size of the transcript **f** is evident in #4 suggesting that *AgTEF*^{T} might be driving expression of the antisense in this and subsequent deletions.

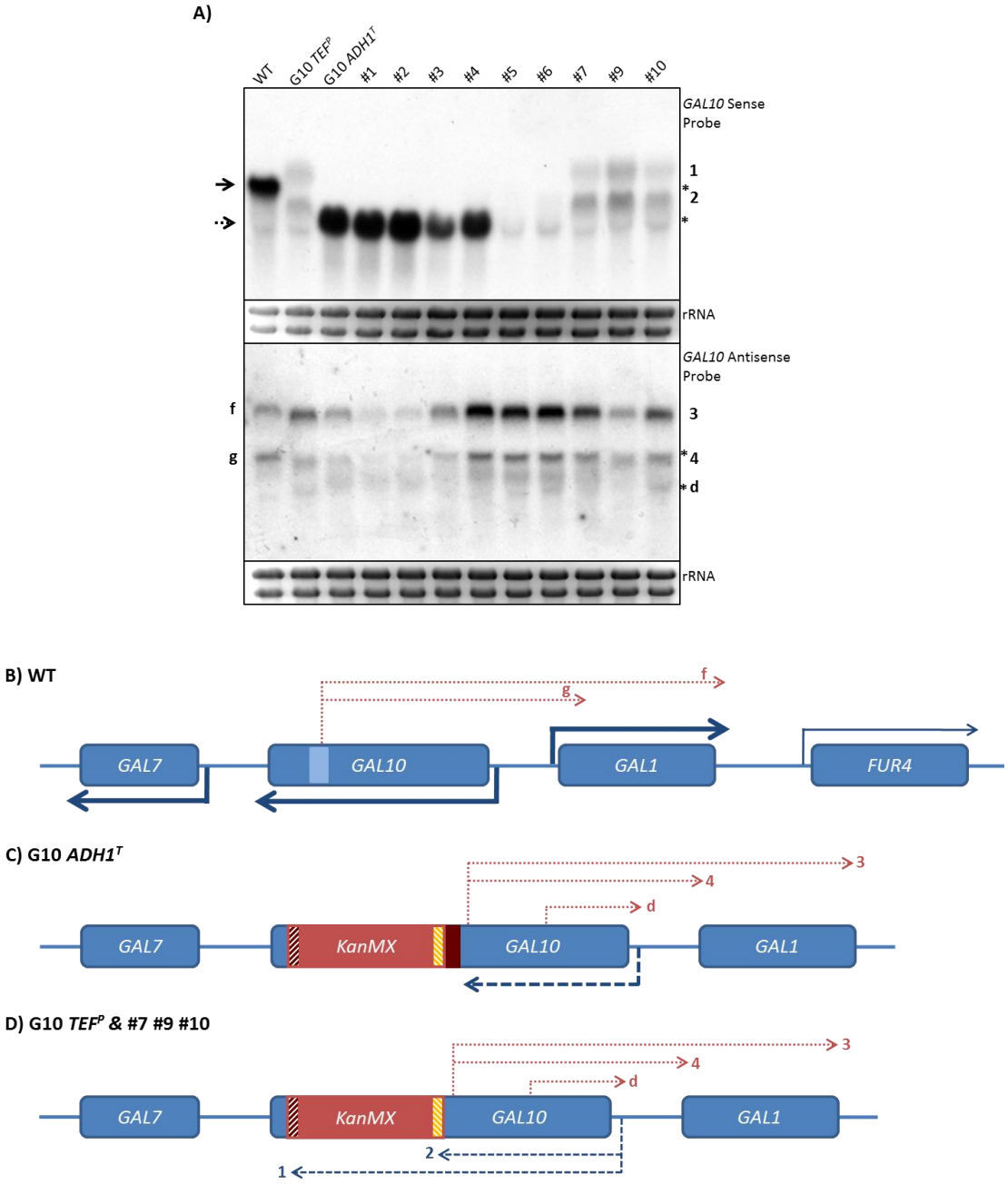


Figure 53. Analysis of the sequential deletions of the *ADH1* terminator. (A) Northern blot analysis of total RNA extracted from the strains indicated, harvested at 90 minutes of induction and probed for *GAL10* sense and antisense. The strains #1 to #10 are sequential deletions of the *ScADH1*^{T}, from the 5' end towards the 3' end, as shown in Figure 53. Deletion #8 is missing due to inefficient transformation. The *KanMX* cassette is indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The *ScADH1* terminator is represented by the dark red box. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematics below. The coding transcripts are identified by a black arrow and the truncated version by a dotted black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. Schematics of the transcripts detected in the Northern blots above of the WT **(B)**, the G10 *ADH1*^{T} **(C)** and G10 *TEF*^{P} & #7, #9 and #10 **(D)** strains.

The striking effect on the levels of sense transcription in deletions #5 and #6 prompted an investigation into the elements that could be having such a strong impact on transcription. As mentioned before, the *ScADH1*^{T} sequence contains two TATA-like elements, and curiously one of them was deleted in strain #4 and the other in strain #5. This led to the hypothesis that the TATA-like elements were responsible for proper expression from the *GAL10* promoter, since in the absence of both, no expression was detected. To see if the region containing these TATA-like elements is sufficient for efficient expression from the *GAL10* promoter, the 40 bp region (defined by deletions #4 and #5) that contains the two TATA-like boxes was inserted into the same position at *GAL10* to create strain 40. A second strain (40X) with the same 40 bp sequence but with the second TATA-like box scrambled was made. Figure 54 (B, D and E) shows the sequences used and the schematic of the strains.

Interestingly, the result shows that both 40 and 40X strains produce the sense read-through transcripts (**1** and **2**) (Figure 54 A) characteristic of strains lacking terminator function and further confirming that the insulator function, preventing read-through transcription, is likely to reside downstream of the 40 bp sequence, but before #7. Additionally, in the 40 strain, the polyadenylated truncated sense transcript is detected, albeit at low levels, but upon scrambling the consensus TATA element in the 40X strain, the truncated transcript is no longer detected (Figure 54 C). This indicates that the second TATA-like element is likely to be functional and capable of driving termination of a proportion of the sense transcription, however, not sufficient to drive high level expression from the *GAL10* promoter. There is no observable difference in the antisense transcript profiles between the 40 and 40X strains and therefore the TATA-like elements

are not required for antisense transcription. In summary, with the wild type *ScADH1*^{T} 40 bp sequence there is production of the truncated polyadenylated *GAL10* sense transcript, as well as the read-through transcripts.

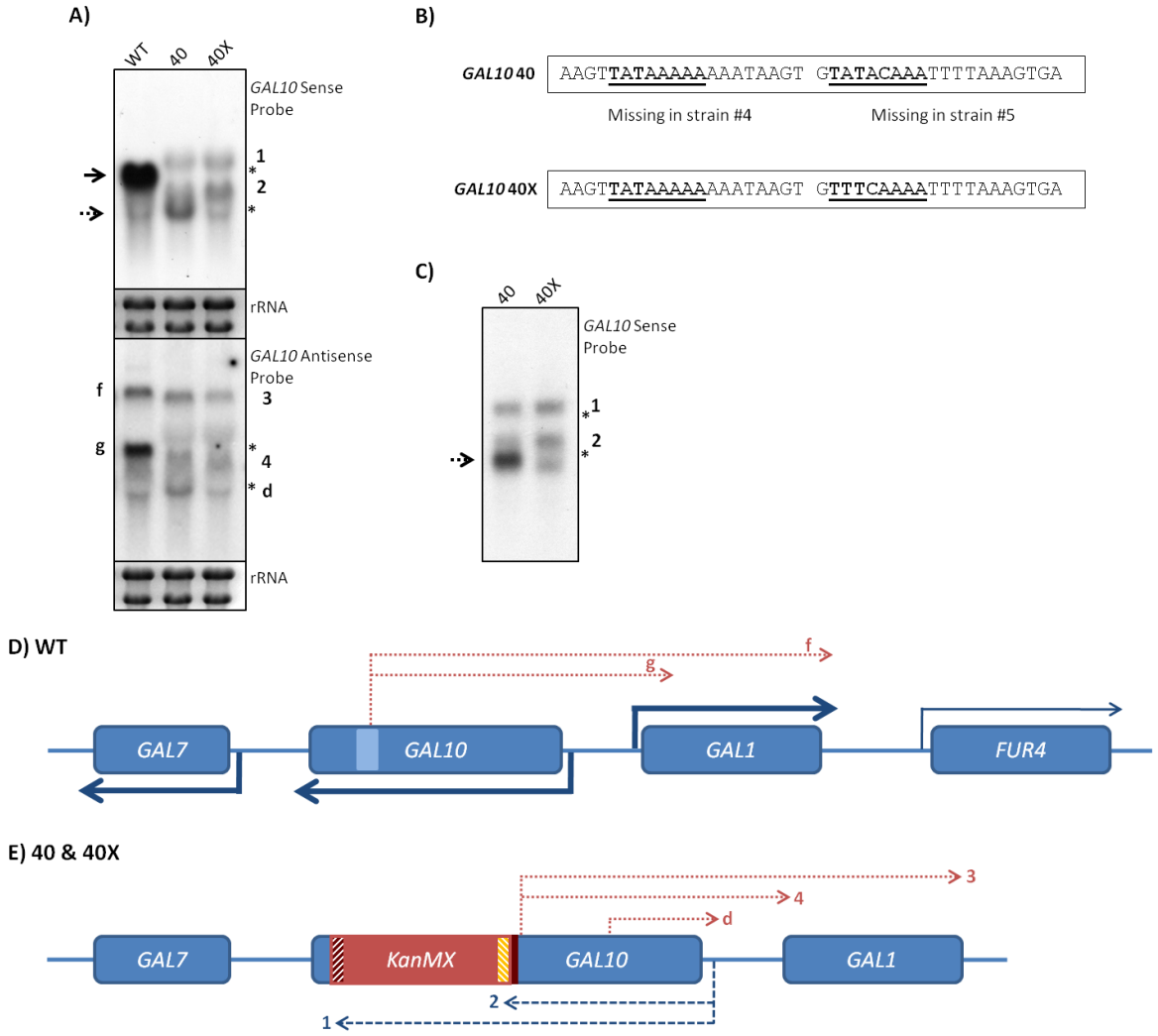


Figure 54. TATA elements important for transcription initiation from the *GAL10* promoter. Northern blot analysis of total RNA **(A)** and poly(A)+ RNA **(C)** extracted from the strains indicated, harvested at 90 minutes of induction and probed for *GAL10* sense and antisense. **(B)** Sequences of both strains with the TATA elements underlined and in bold. *GAL10* 40, refers to the 40 bp insertion that contains both TATA elements and *GAL10* 40X, to the 40 bp insertion that contains both TATA elements with the second one scrambled. Schematics of the transcripts detected in the Northern blots above of the WT **(D)**, the 40 and 40X **(E)** strains. Black arrow represents wild-type sense transcript and dashed arrows represent truncated sense transcript. Strains constructed by Struan Murray.

7.1.4 Deletion #5 in *ScADH^T* reduces sense transcript levels when inserted in *GAL1*

As described before, *GAL1* possesses a simpler transcription profile, and the insertion of the *ScADH1^T* resulted in the production of only one antisense transcript. Therefore, the *ScADH1^T* deletions were inserted in *GAL1* and the effect on sense and antisense transcription was monitored. The strains that showed the most difference in their transcript profiles, strains #4 and #5, were inserted +757 bp from the *GAL1* ATG codon, with the intact *ScADH1^T* at the same position as a control. The cells were induced in galactose for 3 hours and the RNA extracted and used for a Northern blotting experiment. As Figure 55 shows, the truncated *GAL1* sense transcript is observed for the intact *ScADH1^T* and deletion #4 strains, with similar levels to the WT coding transcript, whereas no transcript is observed in strain #5 at this exposure (see also Figure 56). This is similar to the sense transcript pattern observed at *GAL10*.

For the antisense transcript, very low levels of transcript **7** are visible in the strain with the intact *ADH1^T*. The signal for G1 #4 is also low and the transcript appears to be longer (see next section). For G1 #5, levels of antisense transcript are significantly higher, and the transcript is longer. The differences in levels and size of the antisense transcripts between the strains could be explained by a difference in the induction kinetics, since in the previously shown result for the G1 *ADH1^T* strain, the levels changed during the time course (Figure 49 B). Consequently, the induction kinetics of the strains were compared.

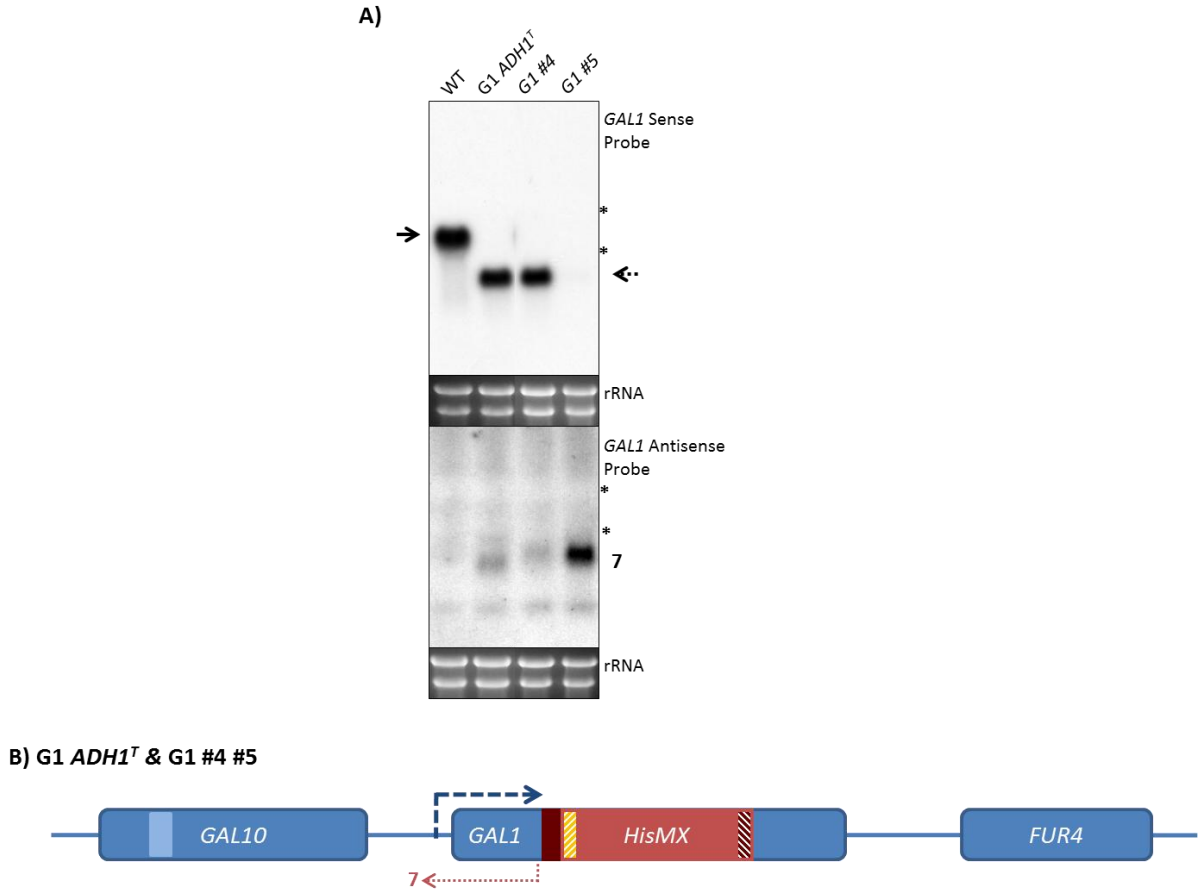


Figure 55. Insertion of the *ScADH1*^{T} deletion sequences at *GAL1* recapitulates the result at *GAL10*. (A) Northern blot analysis of total RNA extracted from the strains indicated, harvested at 3 hours of induction and probed for *GAL1* sense and antisense The *HisMX* cassette is coloured red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The *ScADH1* terminator is represented by the dark red box. The light blue box represents the *GAL10* internal bi-directional promoter. The coding transcripts are identified by a black arrow and the truncated version by a dotted black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. Schematic of the transcripts detected in the Northern blots above of the G1 *ADH1*^{T} and G1 #4 and #5 **(B)** strains.

In the sense direction, the G1 *ADH1*^{T} strain induces at the same time as the WT, at about 1 hour after addition of galactose to the medium, as previously described (Figure 49 B). However, the strain G1 #4 induces faster, with the truncated sense transcript appearing at about 15 minutes of induction (Figure 56). In the strain G1 #5 no transcript is observed. Several reasons could explain this result; (i) that this transcript is induced much slower; (ii) that it does not induce at all; (iii) that it is being degraded.

In the antisense direction all the *ADH1*^{T} strains produce transcript **7**, reaching highest levels at 15 minutes and gradually decreasing until 3 hours when it is barely evident (see Figure 55). Interestingly, the subtle differences in the size and levels of the antisense transcript (compare Figure 55, lane 3 and 4) anti-correlates with the sense levels, i.e. a strain showing higher levels of antisense transcript will, at that time point, show lower levels (if any) of sense transcript (see Figure 55 and Figure 56). Again, the difference in the size of transcript **7** is evident in Figure 56 moreover, the Northern blot for strain G1 #4 suggests that two transcripts could be present.

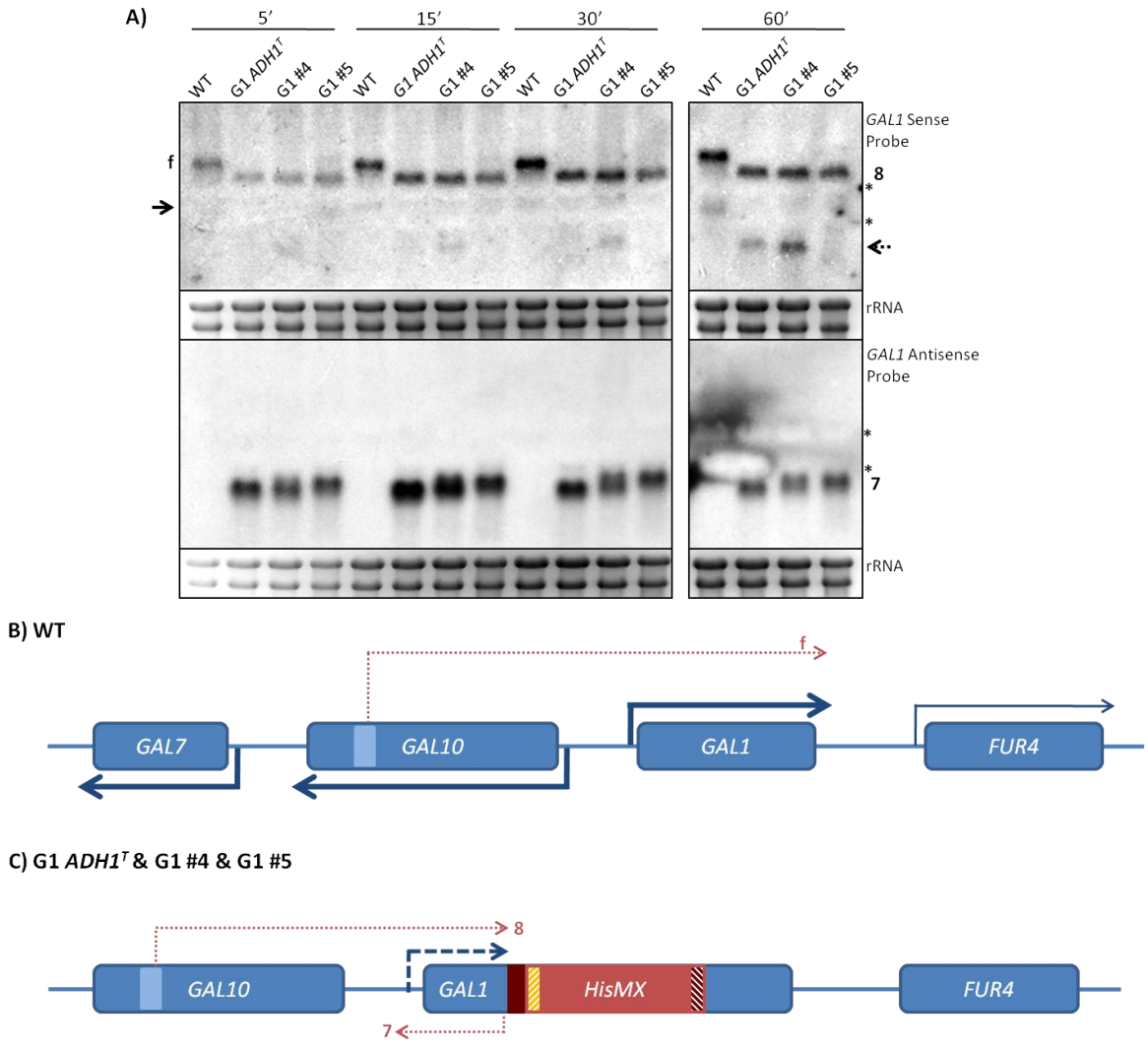


Figure 56. Induction kinetics of the G1 *ADH1*^{T}, G1 #4 and G1 #5 deletion strains. (A) Northern blot analysis of total RNA extracted from the strains designated, harvested in the time points indicated and probed for *GAL1* sense and antisense. The *HisMX* cassette is indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The *ScADH1* terminator is represented by the dark red box. The light blue box represents the *GAL10* internal bi-directional promoter. The coding transcripts are identified by a black arrow and the truncated version by a dotted black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown for loading controls. Schematics of the transcripts detected in the Northern blots above of the WT **(B)**, the G1 *ADH1*^{T} and G1 #4 & #5 **(C)** strains.

Also evident on the Northern blots in Figure 56 and Figure 57 is transcript **f**, the long non-coding transcript initiating at the internal promoter in *GAL10*. This transcript is present when the *GAL* locus is repressed and is detectable for the first few hours after induction. Importantly, in all the constructs, levels of this transcript are similar and terminate at the

end of *GAL1* (WT) or at the *ADH1* terminator (transcript **8**), resulting in a truncated transcript. This suggests that the residual sequences in deletion #5 are sufficient to drive termination of transcription initiated at a distal promoter (Figure 57) but not to maintain high-level expression from the proximal *GAL1* promoter.

As the *GAL1* and *GAL10* promoter is bi-directional, and insertion of *ADH1*^{T} #5 in *GAL1* results in apparent repression of the *GAL1* promoter, it was of interest to examine expression from the *GAL10* promoter (Figure 57). After three hours induction, levels of the *GAL10* sense transcript in G1 #5 are significantly reduced compared to those in G1 #4. As the stability, processing and export of *GAL10* sense transcript should not be affected by insertions at *GAL1*, this result supports a direct effect of sequences in the *ScADH1* terminator (specifically between #4 and #5) on the function of the bi-directional *GAL10* and *GAL1* promoter. It is interesting to note that the levels of *GAL10* sense RNA are also reduced in the G1 *ADH*^{T} strain but the reason is not understood. Perhaps it reflects the induction kinetics of this strain.

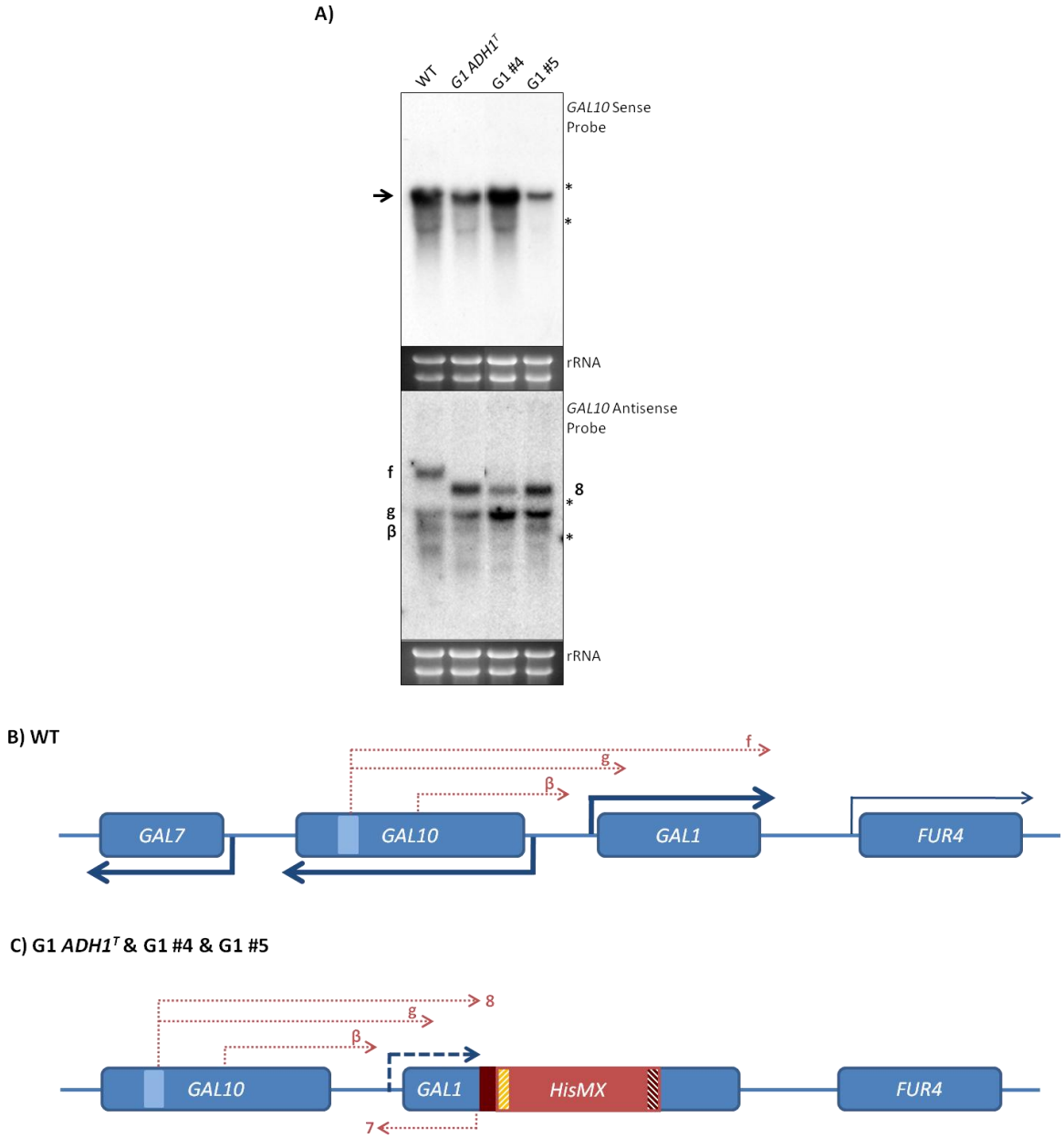


Figure 57. The terminator constructs inserted at *GAL1* down-regulates *GAL10* expression. (A) Northern blot analysis of total RNA extracted from the strains indicated and probed for *GAL10* sense and antisense. The cells were induced in galactose for 3 hours. The *HisMX* cassette is indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. All the transcripts detected by Northern blot are indicated in the schematics below. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown for loading controls. Schematics of the strains and transcripts detected in the Northern blots above of the WT **(B)**, G1 *ADH1*^{T}, G1 #4 and #5 **(C)** strains.

For further confirmation of this result, an RNA polymerase II ChIP was performed at *GAL10* in the strains containing truncations at *GAL1*, including G1 TEF^P, G1 $ADH1^T$, G1 #4 and G1 #5. This experiment should confirm if the reduced steady-state RNA levels at *GAL10* are due to repression of the *GAL10-1* promoter by the $ScADH1^T$ truncation #5 inserted in the *GAL1* coding region. The RNAPII ChIP was performed 90 minutes after galactose induction instead of the 3 hours induction as for the Northern blot. Figure 58 shows that, compared to the WT, the levels of RNA Pol II in the G1 $ADH1^T$ strain are lower and further decreased in the G1 #5 strain, compared to the WT, correlating with the results obtained by Northern blot. This strongly supports the hypothesis tested. Note that strains G1 TEF^P (see Figure 45) and G1 #4 have higher levels of RNA Pol II at *GAL10*, and although the Northern blot experiment shows no difference in transcript levels, the time of induction between the experiments needs to be taken into consideration. At present, the exact mechanisms regulating the expression of *GAL10* and *GAL1* in these hybrid strains is not known but a number of possibilities come to mind: i) it could be *in cis*, either due to an RNA interference mechanism or chromatin remodelling by the process of transcription through the *GAL10-1* promoter, therefore, involving the ncRNAs produced in these strains; or ii) *in trans*, by the physical interference/binding of a transcript homologous to the *GAL10-1* promoter, i.e. one of the ncRNAs.

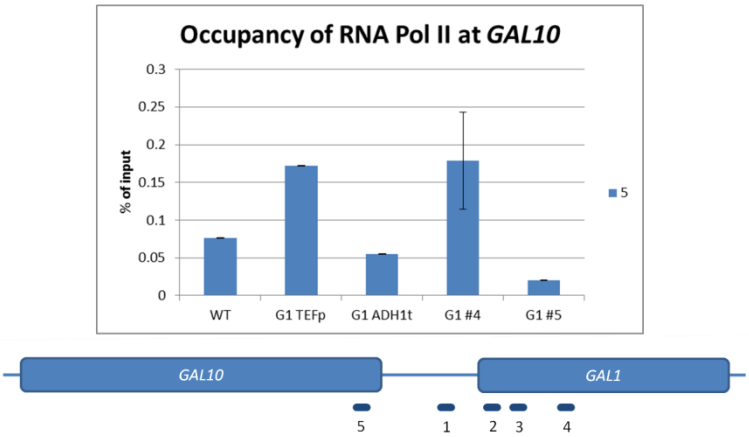


Figure 58. RNA Pol II levels at the *GAL10* gene in the *GAL1* engineered strains. ChIP analysis of the RNA Polymerase II in the strains indicated was performed as described in section 2.6. The cells were induced in galactose for 90 minutes and then cross-linked. The primer pairs used for quantitative real-time PCR are indicated in the schematic below.

7.1.4.1 Adenylation profile of the transcripts produced from the *ScADH1* terminator

Although it has been established that the residual sequences in the G1 *ADH1*^{T} #4 and G1 *ADH1*^{T} #5 strains are sufficient to terminate the *GAL10-1* antisense transcript (**8**), 3' end processing, in particular transcript adenylation, of the transcripts produced in these strains was investigated (Figure 59). At *GAL1*, polyadenylation of the longer sense RNA species (the *GAL10-1* antisense transcript, **8**) and the antisense transcript (**7**) were examined. At *GAL10,* the sense transcripts and the longer antisense RNA species (transcript **3**) in the G10 *ADH1*^{T} #4 and G10 *ADH1*^{T} #5 strains were also examined. As previously explained (section 7.1.1.2), the poly (A)+ enriched RNA runs faster through the agarose gel due to the absence of the rRNA, and so for the comparison with the total and non-adenylated RNA this needs to be taken into consideration. In addition, it is not known how efficiently poly (A)+ RNA binds to the oligo dT beads. Thus only differences in relative proportions of poly (A)+/(A)- in the two fractions are informative.

Figure 59 (A) shows the poly (A)+ and poly (A)- ratios in the *GAL10* terminator strains (G10 *ADH1*^{T}, G10 #4 and G10 #5), after three hours induction in galactose medium. The sense transcripts, in all the strains, are similarly distributed in the poly (A)+ and poly (A)- fractions, albeit with very different absolute levels of steady-state transcripts. This suggests that sequences required for polyadenylation of transcripts are present in deletions #4 and #5 and that a failure to polyadenylate is not the reason for low transcript levels. For the antisense transcripts **f** (WT) and **3** (G10 *ADH1*^{T}), very little signal is detectable in the poly (A)+ fraction suggesting that these transcripts are mostly non-adenylated. In the G10 #4 and G10 #5 strains, a small percentage of transcript **3** is poly (A)+, but this might simply reflect the higher overall levels of antisense transcripts in these stains.

For the *GAL1* strains, the cells were induced in galactose for 15 minutes only, in order to capture the time frame in which the antisense transcript is expressed to the highest levels. This means that only transcript (**8**) arising in *GAL10* and transcribed sense over *GAL1* is detected with the *GAL1* sense probe and this is distributed equally in the poly (A)+ and poly (A)- fractions. Similar levels of transcript **8** are present in all strains.

The antisense transcript **7** is mostly polyadenylated in all strains. Interestingly, the transcript appears to increase in size, being shortest in the G10 *ADH1*^{T} strain and longest in the G10 #5 strain, in both the adenylated and non-adenylated fractions. Whether the transcripts initiate and/or terminate at different sites is not known, neither is the length of the poly (A)+ tails, which could be different. It is possible that due to the deletions, the antisense transcript is now initiated under the influence of *AgTEF*^{P}, explaining the increase in length. Nevertheless, the longest antisense transcript, present in the G10 #5

strain, correlates with the reduced expression from the *GAL1-10* promoter, and a mechanism of transcriptional interference could, therefore, play a role.

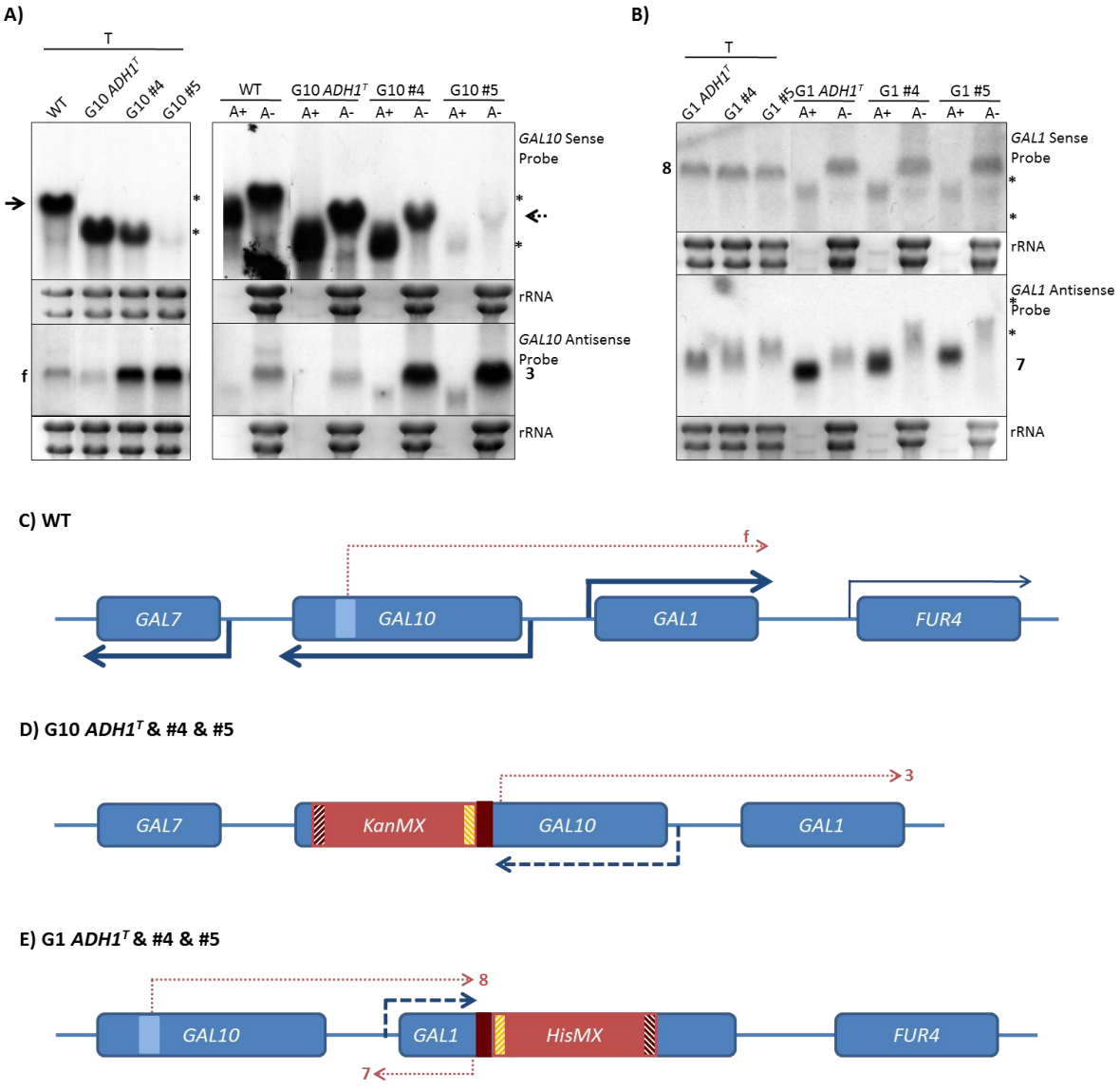


Figure 59. Different states of polyadenylation of the transcripts. Northern blot analysis of total (T), poly A+ (A+) and poly A- (A-) enriched RNA extracted from the strains designated and probed for *GAL10* **(A)** and *GAL1* **(B)** sense and antisense. The cells were induced in galactose for 3 hours (A) and 15 minutes (B). The *KanMX* and *HisMX* cassettes are indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The light blue box represents the *GAL10* internal bi-directional promoter. The coding transcripts are identified by a black arrow and the truncated transcripts by a dashed arrow. The ribosomal bands are marked with * for size guidance and the ethidium bromide-stained gels (rRNA) are shown as loading controls. Schematics of the transcripts detected in the Northern blots above of the WT **(C)**, the G10 *ADH1*^{T} **(D)** and G1 *ADH1*^{T} **(E)** strains.

7.1.4.2 RNA Polymerase II occupancy in the *ScADH1^{T}* strains

The difference in expression from the *GAL10-1* promoter in these engineered strains raises the question about the stability of the transcripts. As an indirect approach to assess whether degradation of the transcripts is responsible for the decreased expression observed in both G10 #5 and G1 #5 strains, a RNA Polymerase II ChIP was performed. The levels of RNAPII were assessed in the G10 *ADH1^{T}* and G1 *ADH1^{T}* strains at 90 minutes of induction by ChIP. Figure 60 (A) shows that for the *GAL10* terminator strains, the levels of RNA Pol II at the 5' coding region of *GAL10*, vary from similar to the WT, in the G10 *ADH1^{T}* strain, to very low, in the G10 #5 strain. In the G10 #4 strain it can only be said that the levels are between the ones observed for G10 *ADH1^{T}* and G10 #5 strains since the error associated with this sample makes it impossible to be accurate. Nevertheless, this ChIP result correlates with the results obtained by Northern blotting and indicates that transcription from the *GAL10* promoter is impaired in the G10 #5 strain.

In contrast, the results of the RNA Pol II ChIP for the *GAL1* terminator strains (Figure 60 B) do not correlate with steady-state sense RNA levels detected by Northern blotting, as previously observed for the G1 *TEF^{P}* strain (Figure 41). The levels of RNA Pol II across *GAL1* are much higher in the G1 #5 strain compared with the other strains. The G1 *ADH1^{T}* strain shows slightly higher levels of RNA Pol II compared to the WT, while the levels in the G1 #4 strain are between these two strains. This pattern of RNA Pol II occupancy does, however, correlate with the levels of antisense transcripts in these strains. ChIP does not differentiate between strands and it is known that by 90 minutes of galactose induction that the truncated sense transcripts are already being produced and this should contribute for the overall RNA Pol II signal. Since the levels of RNA Pol II detected by ChIP

do correlate at *GAL10* it suggests the result obtained for *GAL1* must be due to a specific but unknown feature of *GAL1*.

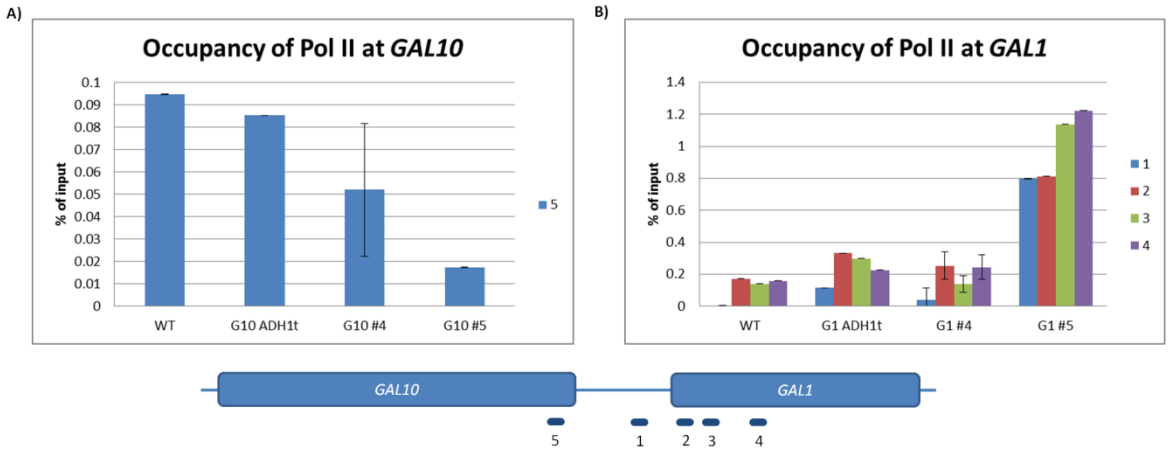


Figure 60. RNA Pol II levels at *GAL10* and *GAL1* in the terminator strains. ChIP analysis of the RNA Polymerase II at *GAL10* **(A)** and *GAL1* **(B)** in the strains indicated was performed as described in section 2.6. The cells were induced in galactose for 90 minutes and then cross-linked. The primer pairs used for quantitative real-time PCR are indicated in the schematic below.

7.1.4.3 Rrp6-dependence of the *GAL1 ADH1*^{T} strains.

In order to assess whether degradation of the transcript is responsible for the decreased steady-state transcript levels observed in the G1 #5 strain, the G1 *ADH1*^{T} constructs (G1 *ADH1*^{T}) were inserted into an *rrp6Δ* strain lacking nuclear exosome function. As Rrp6p is known to slow the induction kinetics at *GAL1* (Chapter 6, section 6.4), the induction profile was assessed for these strains. Figure 62 confirms the early induction for *rrp6Δ GAL1* and shows that *rrp6Δ* G1 *ADH1*^{T} also has a strong early induction profile.

Remarkably, the induction kinetics of *rrp6Δ* G1 #4 over the first hour in galactose are much slower than the kinetics observed in *rrp6Δ* G1 *ADH1*^{T} and *rrp6Δ* G1 #5 (Figure 62). However, by 3 hours of induction, *rrp6Δ* G1 #4 has similar transcript levels to *rrp6Δ* G1 *ADH1*^{T} (Figure 61 A and B) and the transcripts are equivalently polyadenylated (Figure 61 C). This suggests that induction of G1 #4 is insensitive to loss of Rrp6p. Interestingly, both

rrp6Δ G1 *ADH1*^{T} and *rrp6Δ* G1 #4 strains show an antisense transcript (**7**) throughout the induction period, although the transcript appears longer in *rrp6Δ* G1 #4, similar to that observed in the *RRP6* background. It is possible that this longer antisense transcript is maintaining the promoter in a repressed state for longer, even in the absence of Rrp6p. By three hours, the antisense is barely detectable and sense transcript levels have accumulated to similar levels as the *rrp6Δ* G1 *ADH1*^{T} and *rrp6Δ GAL1* strains. This implicates the first 80 bp of *ScADH1*^{T} as required for Rrp6-dependent processing of a transcript (possibly transcripts **7** or **8** that are present in glucose) leading to reduced accumulation of the sense transcript. It is assumed that a similar mechanism operates at *GAL1*^{T}. However, both ideas need to be tested experimentally. This was an unexpected observation as G1 #4 generally behaves in a similar way to G1 *ADH1*^{T}.

The results obtained with the *rrp6Δ* G1 #5 strain were also quite unexpected. Normally, G1 #5 maintains the long form of the antisense transcript (**7**) in glucose and throughout induction (Figure 56). Furthermore, levels of the truncated sense transcript are barely detectable even at 3 hours of induction (Figure 55). In the absence of Rrp6p there are major changes to the transcript profile for G1 #5 (Figure 61 and Figure 62). First, no antisense RNA (transcript **7**) is evident at any time during the induction (Figure 61 A, B and Figure 62). Second, transcript **8** (the *GAL10-1* antisense transcript) is no longer terminating at the remaining sequences in *ADH*^{T} #5 but reading though (Figure 61 A and B), as observed in the G1 *TEF*^{P} strain (Figure 46). As the long antisense RNA (transcript **7**) is implicated in repression of *GAL1*^{P}, induction of the sense transcript is now expected in the *rrp6Δ* G1 #5 strain and this is what is observed (Figure 61 and Figure 62). The sense transcripts appear with similar kinetics to the *rrp6Δ* G1 #4 strain. At sixty minutes of

induction, these appear to be predominantly read-through transcripts, but by three hours a shorter truncated transcript of the expected size is also evident (Figure 61 and Figure 62). Steady-state RNA levels at three hours are less than the *rrp6Δ* G1 #4 strain but significantly higher than in the *rrp6Δ* G1 #5 strain (Figure 61). At three hours, both sense transcripts in *rrp6Δ* G1 #5 are polyadenylated (Figure 61 C). It is important to stress that the transcript profile in the *rrp6Δ* G1 #5 strain is quite distinct from the *rrp6Δ* G1 TEF^{P} strain, which is not sensitive to Rrp6p (Figure 61 C). This implies that deletion #5 retains functional sequences. These data also implicate Rrp6p in recognising G1 #5 as a defective transcription unit and by means not fully understood, preventing the production of the sense transcript. A simple interpretation of the data would be that Rrp6p degrades the sense transcripts. However, studying the induction kinetics of this strain suggests a much more complex role for Rrp6p in subtle transcript processing, which helps to define the functionality of transcripts leading to repression of the sense transcript. It appears that Rrp6p, probably through a processing-dependent mechanism, indirectly involved in establishing the dominant transcription unit. In a WT situation, Rrp6p would degrade the read through transcripts produced in the G1 #5, and this would allow the expression of the antisense transcript. However in the absence of Rrp6p this de-repression of the antisense promoter does not occur and no antisense is produced. The slow kinetics of both G1 #4 and G1 #5 could be explained by an interference of Rrp6p until the dominant transcription unit is established.

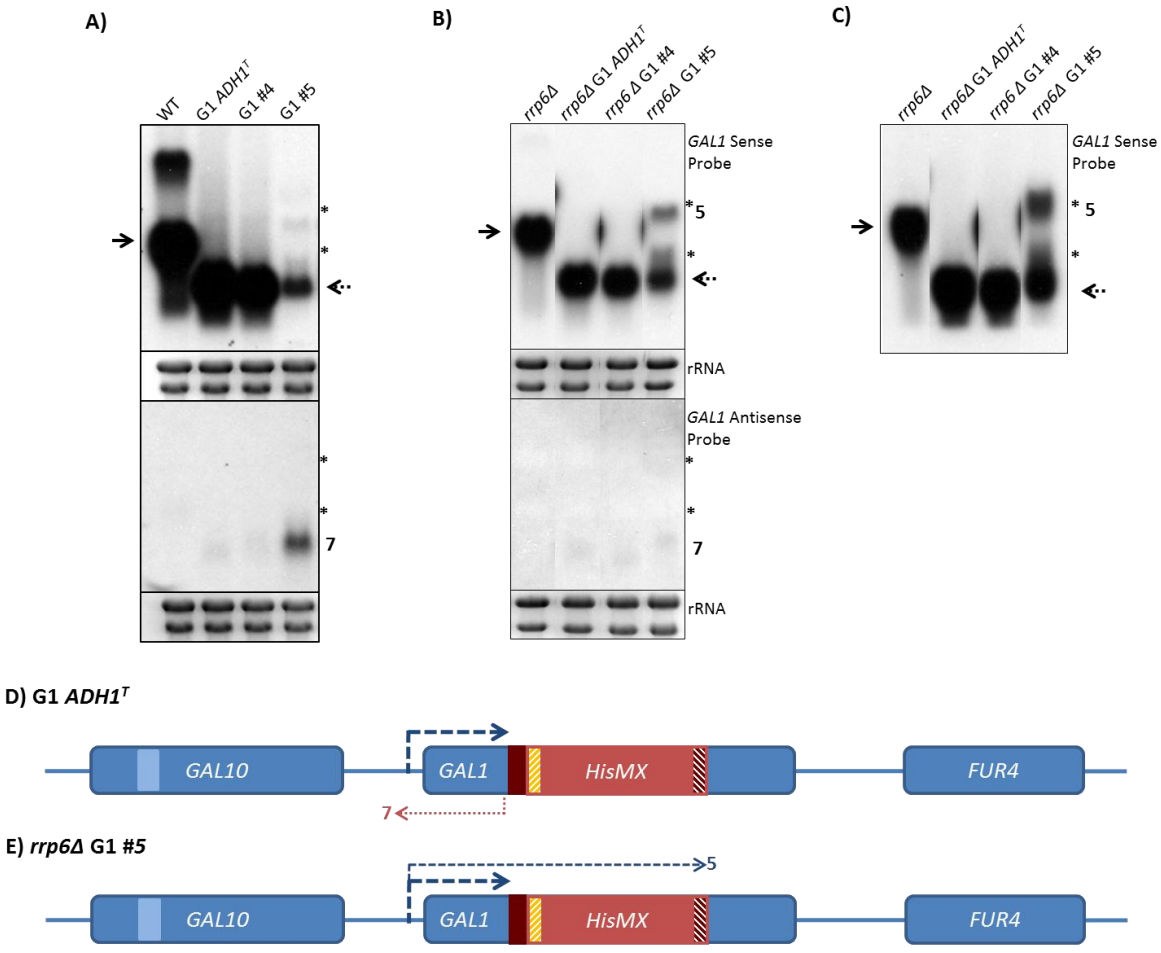


Figure 61. The transcription profile of G1 #5 in the absence of Rrp6p is altered. Northern blot analysis of total RNA extracted from the strains in the WT background **(A)** and in the *rrp6Δ* **(B)** and poly(A)+ RNA **(C)** from the *rrp6Δ* background and probed for *GAL1* sense and antisense. The cells were induced in galactose for 3 hours. The *HisMX* cassette used is indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The *ScADH1* terminator is represented by the dark red box. The light blue box represents the *GAL10* internal bidirectional promoter. All the transcripts detected by northern blot are indicated in the schematic below. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium-stained gels (rRNA) shown for loading control. Schematic of the transcripts detected in the northern blots above for the G1 *ADH1*^{T} **(D)** and *rrp6Δ* G1 #5 **(E)** strains.

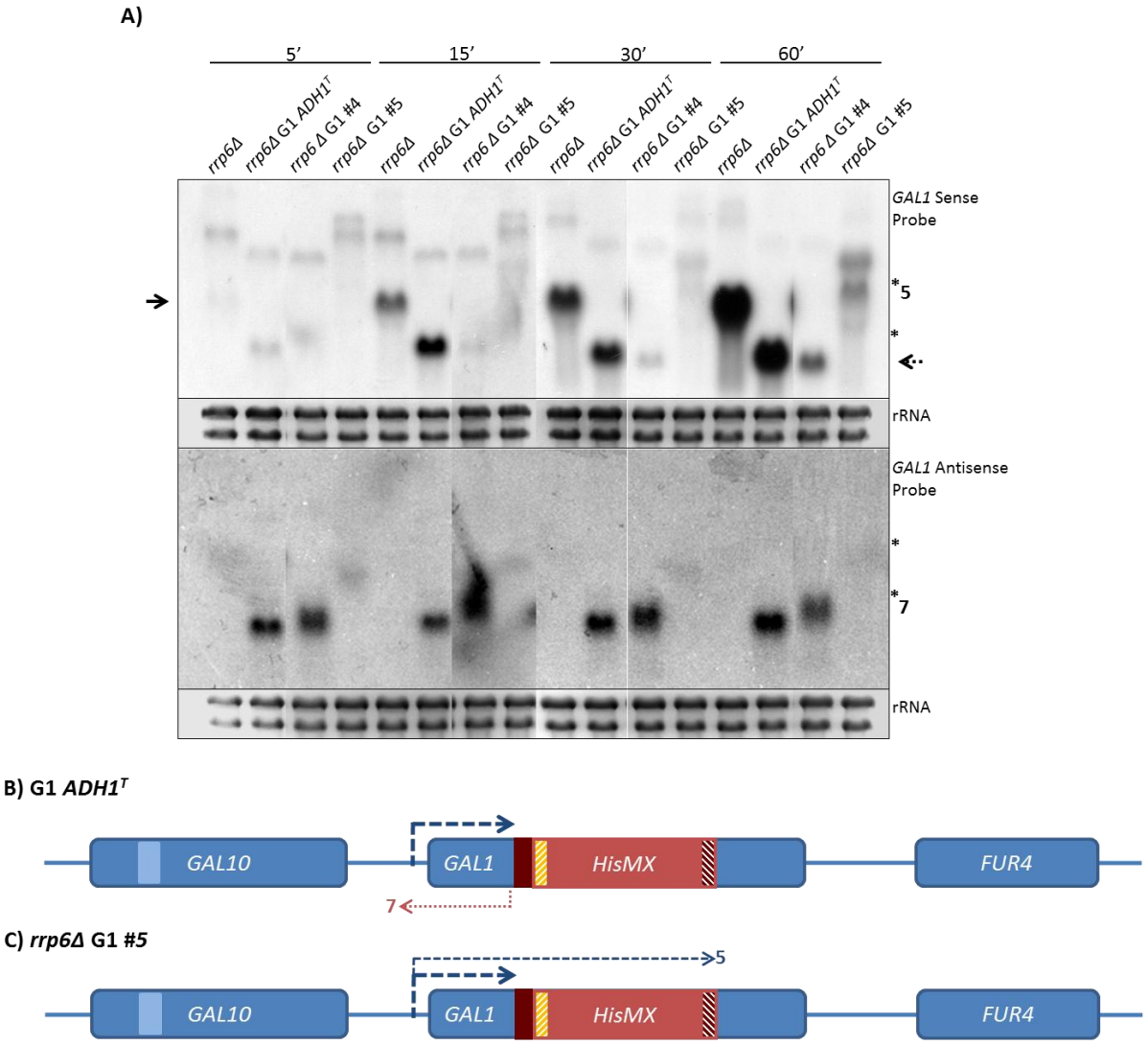


Figure 62. The transcription induction of the G1 *ADH1^T* and deletion strains #4 and #5. (A) Northern blot analysis of total RNA extracted from the strains and time points indicated and probed for *GAL1* sense and antisense. The *HisMX* cassette used is indicated in red, with the orange diagonal lines representing the *AgTEF* promoter and the dark diagonal lines representing the *AgTEF* terminator. The *ScADH1* terminator is represented by the dark red box. The light blue box represents the *GAL10* internal bidirectional promoter. All the transcripts detected by northern blot are indicated in the schematic below. The coding transcripts are identified by a black arrow. The ribosomal bands are marked with * for size guidance and the ethidium-stained gels (rRNA) shown for loading control. Schematic of the transcripts detected in the northern blots above of the *G1 TEF^P* **(B)** and *rrp6Δ* G1 #5 **(C)** strain.

7.1.4.4 Distribution of RNA in the *ScADH1* terminator strains

RNA FISH was used to assess the distribution of sense transcripts between the nucleus and cytoplasm in the *ScADH1* terminator strain. The RNA FISH was performed with *GAL1* sense probes (red) in the *GAL1* terminator strains (G1 $ADH1^{T}$, G1 #4 strains and G1 #5) and compared to the WT strain after 90 minutes of induction. Figure 63 A shows that high levels of *GAL1* sense transcripts were detected in the G1 $ADH1^{T}$ and G1 #4 strains while the intensity of the nuclear dot RNA was decreased. In the G1 #5 strain low levels of *GAL1* sense expression and RNA dots were detected consistent with the low levels of transcripts observed by Northern blot.

The overall high expression and the lower intensity of the dot RNA in the G1 $ADH1^{T}$ and G1 #4 strains, compared to the WT (Figure 63 B), indicates that the sense transcripts are being efficiently exported. Firstly, this suggests that the remaining sequences of the $ScADH1^{T}$ in the G1 #4 strain and are sufficient for efficient processing and export of the transcripts to the cytoplasm. Secondly, it might indicate that the transcripts in these strains are being exported faster than in the WT. This could be explained by a difference between the *GAL1* 3'end region and the $ScADH1^{T}$.

For the G1 #5 strain, the low level of overall expression and low level of nuclear dot RNA intensity can be explained by the low-level transcription observed by Northern blotting and therefore nothing can be said about the remaining $ScADH1^{T}$ sequences.

This result, in comparison to the one obtained for the G1 TEF^{P} (see section 7.1.1.3), indicates that the presence of a terminator is required for efficient export to the cytoplasm, as expected. It would therefore be expected that accumulation of transcripts in the nucleus in the G1 TEF^{P} strains would be a result of a higher amount of non-

adenylated transcripts. However, comparing the results of poly (A)+ versus poly (A)- transcripts obtained for the G1 *TEF*^{P} (Figure 43) and G1 *ADH1*^{T} (Figure 59) strains, no big differences are observed. The relation of the dot RNAs and the presence or absence of non-adenylated transcripts needs to be assessed in an *rrp6Δ* strain. It cannot be ruled out that the accumulation of the sense transcripts in nuclear dot RNAs in the G1 *TEF*^{P} strain results in the low expression from the promoter.

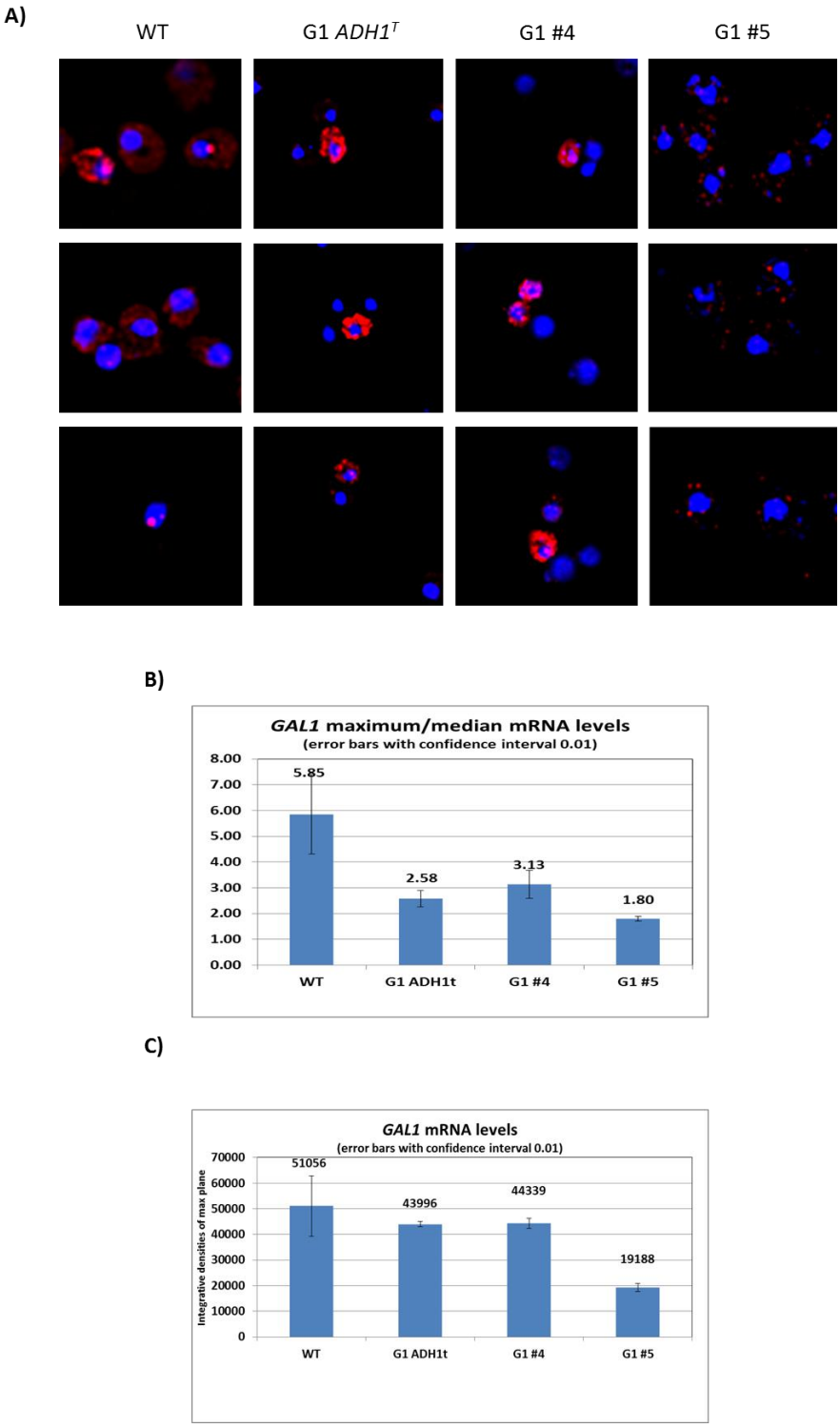


Figure 63. *GAL1* nuclear sense RNA dots in the G1 *ADH1*^{T}, G1 #4 and G1 #5 strains. (A) RNA FISH of the strains indicated. Cells were induced in galactose for 90 minutes. *GAL1* sense (red) transcripts were detected with cyanine 5 labelled probes. DAPI (blue) was used to identify the nucleus. Images obtained with an average projection of 3 layers. **(B)** Quantitative analysis of the *GAL1* dot RNA intensity. Calculated as the maximum intensity signal relative to the mediam intensity signal. **(C)** Quantitative analysis of the overall signal of the *GAL1* sense transcripts. Intensity densities of the plane with the maximum signal. Error bars calculated with interval confidence of 0.01.

7.2 Discussion

In this chapter the sense and antisense transcriptional architecture was investigated. The suggestion that antisense transcription can be not only produced from the inherent bi-directionality of the promoters but also from terminators raises a number of questions. Some of the questions addressed here include:

i) Is a promoter capable of initiating antisense transcription?

ii) Is a terminator without any neighbouring promoters capable of initiating antisense transcription?

iii) What is the effect of antisense transcription produced from a promoter or terminator on the sense promoter?

iv) What are the sequences responsible for antisense transcription in a terminator?

It was here unveiled, that both promoters and terminators are capable of producing antisense transcripts without the presence of neighbouring elements. The creation of a Promoter – Promoter transcriptional configuration resulted in low-level expression from the sense promoter. The deletion of the exonucleolytic component of the exosome (Rrp6p) and a RNAPII ChIP revealed that the observed low-level expression was not due to degradation of the truncated transcripts. Instead it was observed that the sense transcripts were being retained in a nuclear "dot RNA" and their export to the cytoplasm was reduced. This has led to the idea that either the production of non-adenylated transcripts due to the lack of a 3' end processing region or the exosome processing of poly (A)+ species to poly (A)-, results in accumulation of the transcripts in nuclear "dot RNAs" repressing expression from the promoter. However, the fact that deletion of the

RRP6 did not restore high expression from the sense promoter indicates that Rrp6p is not involved in the repression mechanism. So either the dot RNAs are not directly involved in the repression mechanism or Rrp6p is not involved in the formation of the dot RNA observed here. Also, the sense transcripts produced in the Promoter – Promoter configuration strains were not differentially polyadenylated compared to the WT. Therefore, it raises a question about the nature of the transcripts present in the dot RNA.

The insertion of a terminator, creating a conventional Promoter – Terminator transcriptional organization restored the expression from the sense promoter. It was also found that different types of terminators (*ScADH1*^{T} or *AgTEF*^{T}) result in the same outcome. Moreover, a deletion analysis of the *ScADH1*^{T} led to the finding that a 40 bp region is absolutely required for expression from the sense promoter in both genes tested but is not necessary for antisense transcription, unless *RRP6* is also deleted. Furthermore, the presence of the *ScADH1* terminator resulted in efficient export from the nucleus to the cytoplasm of the sense transcripts and decreased dot RNA, indicating that these are intrinsically related.

The antisense transcripts produced from both the *AgTEF*^{P} and *ScADH1*^{T} at *GAL10* did not seem to be implicated in sense transcription, whereas at *GAL1* some differences were observed not only in the levels of the antisense transcript but also in the length. It was also observed that the *ADH1*^{T} and *ADH1*^{T} #5 insertions in *GAL1* resulted in reduced expression from the *GAL10* promoter, possibly implicating the antisense transcript in a repressive mechanism.

Chapter 8

Discussion

In this thesis the presence, role and transcription of non-coding RNAs in the *GAL* locus was investigated. Two main topics were addressed: i) the description and biology of the ncRNAs and their implication in the regulation of the *GAL* genes, including induction and transcriptional memory; ii) the characterization of sense and antisense transcription units by the creation of galactose inducible hybrid genes at the *GAL* locus. The main findings for the first topic addressed were:

a) Identifying and mapping novel non-coding transcripts at the *GAL* locus in glucose or during induction with galactose.
b) Demonstrating the role of Rrp6p in regulating levels of the *GAL1* transcript during induction with galactose.
c) Demonstrating that expressing the first 757bp of *GAL1* in the context of the *GAL* locus, potentially encoding the first 252aa (of 528) of Gal1p, is sufficient for transcriptional memory.

The mechanism underlying transcriptional memory, which is a form of very fast induction, has been suggested to be solely due to the Gal1 protein (Zacharioudakis et al., 2007), however the work here suggests that the intact Gal1 protein is not required for memory. A strain with a transcriptional terminator inserted into the middle of *GAL1* in the context of the *GAL1* locus (and potentially producing a truncated version of Gal1p deficient in its kinase activity) retains transcriptional memory. Several mechanisms could be proposed to explain the results obtained; firstly, if produced the first 252aa of Gal1p could be sufficient for proper folding and may retain the ability to bind ATP and galactose, and thus relieve Gal80p repression of Gal4p; secondly, transcription either of the mRNA or a ncRNA through the *GAL10-1* promoter and/or first 757bp of *GAL1* may act *in trans* to

mediate memory; and thirdly, a mechanism involving only the chromatin structure at the promoter and/or first 757bp of *GAL1*. Further experiments are being designed to address these proposed mechanisms. It is likely that more than one of these mechanisms is responsible for transcriptional memory and this novel observation has uncovered the need to re-visit the previously suggested (Brickner et al., 2007; Kundu et al., 2007) and recently published (Zhou and Zhou, 2011) models.

The functionality of the ncRNAs is usually linked to regulatory events either by *cis* or *trans* acting mechanisms in mammalian cells (Martianov et al., 2007; Mattick, 2009; Mercer et al., 2009). However, in yeast there are only a few studied cases of regulation by ncRNAs and no general mechanism of action has been found (Berretta et al., 2008; Camblong et al., 2009; Camblong et al., 2007b; Gelfand et al., 2011; Houseley et al., 2008; Pinskaya et al., 2009; Uhler et al., 2007). In this thesis, possible mechanisms through which the ncRNAs influence transcription of the coding genes were investigated. Several novel transcripts were unveiled at the *GAL* locus by a combination of different techniques including RACE, RNA tiling arrays and Northern blotting. These ncRNAs were not only expressed in induced (galactose) but also in repressive conditions (glucose). Different types of transcripts were detected, including stable unannotated transcripts (SUTs) and Xrn1-sensitive unstable transcripts (XUTs), although no cryptic unstable transcripts (CUTs) were found at steady-state. The transcripts detected were initiated not only from promoters, such as the *GAL10-1* promoter, the *GAL10* internal bi-directional promoter and other cryptic internal promoters as in *GAL10*, but also from terminators.

The role of the ncRNAs was tested in some aspects of transcriptional induction of the *GAL* gene and it was found that in the absence of Rrp6p, *GAL1* transcripts accumulate faster, although presently there is no direct evidence that the mechanism at the native locus involves the ncRNAs. There is, however, some preliminary evidence to link Rrp6p to coding and ncRNAs involved in regulation of the *GAL* locus. First, the absence of Rrp6p might stabilize the *GAL4* RNA and this could lead to increased amounts of the Gal4p, which would likely to result in a faster induction of the *GAL* genes. Second, the exonucleolytic processing of *GAL1*-homologous transcripts by Rrp6p to poly(A-) and their accumulation into nuclear dot RNAs (Vodala et al., 2008) could lead to repression of the *GAL10-1* promoter. By its ability to process transcripts, Rrp6p could be involved in a different mode of repression, not yet described in yeast, but present in mammalian cells, involving the association of RNA to the homologous DNA sequence (Martianov et al., 2007). These structures have been recently shown in this laboratory to form *in vitro* at the *GAL10-1* promoter. The Rrp6p could trim any of the transcripts homologous to the promoter, to a shorter length capable of binding to the promoter resulting in repression due to inability of binding of either Gal4p or the PIC.

The presence of non-coding transcripts produced from various locations in the genome in both sense and antisense directions in relation to ORFs leads to an interleaved transcription profile. Such a complex transcription profile challenges the unidirectional view of transcription initiating at a promoter and terminating at a terminator. In fact, for example, antisense transcription usually initiates from a terminator and terminates at a promoter. The establishment and regulation of these unusual transcription units was investigated in the second topic of this thesis. The main findings for this topic were:

a) Many promoters are bi-directional and can initiate antisense transcription
b) Promoter – Promoter configuration results in decreased expression from the sense promoter and increased nuclear dot RNA
c) Terminators can initiate antisense transcription
d) Terminators contain sequences responsible for expression from the promoter

To start to address the many questions posed by unusual transcription units, genomic hybrid genes were created whereby a promoter or a terminator was inserted within an ORF at the *GAL* locus so that they are inducible with galactose.

The bi-directionality of many promoters (Morris et al., 2008; Neil et al., 2009) was further confirmed in this study, by the ability of a promoter inserted within an ORF to initiate antisense transcription. Furthermore, this Promoter – Promoter configuration resulted in reduced expression from the sense promoter, although there is no evidence to suggest that the stable antisense transcripts produced are directly responsible for this repression. It was also found that the sense transcripts produced were being accumulated in nuclear dot RNAs, suggesting perhaps, a mechanism through which the termination of transcripts at a region lacking the termination and/or processing elements leads to accumulation of these transcripts in dot RNAs resulting in repression of the sense promoter. Rrp6p was not found to be involved in any putative repressive mechanism by dot RNAs, since the low levels of sense transcript were not rescued in strains lacking this enzyme.

Furthermore, the insertion of a terminator within an ORF resulted in high expression from the promoter (assessed in terms of high levels of steady-state RNA), suggesting that low expression at the Promoter – Promoter configuration is indeed due to the lack of 3'end processing/termination region. Interesting, this high level of sense transcript from

the Promoter-Terminator configuration is sensitive to Rrp6p, while the low levels from the Promoter-Promoter configuration is not. It was also found that the terminator alone has the ability to initiate antisense transcription. A deletion analysis of the terminator unveiled sequences in the 3'UTR required for high levels of steady-state sense RNA, without any major effects on the antisense transcription. In this region two TATA-related elements were identified and implicated in maintaining high levels of sense transcript. Rrp6p was identified as acting through the TATA-related elements in the 3'UTR, determining whether the sense or the antisense transcript predominates at a bi-directional transcription unit. Additionally, the remaining sequences in the deleted terminator which resulted in very low levels of steady-state RNA were found to properly terminate an ncRNA indicating that the residual sequences retain the capacity to process/terminate the transcripts .

The investigation of sense and antisense transcription units is very complex and although significant contributions were made to better the understanding of this field, many intriguing questions remain. With the finding that terminators can promote initiation of transcription and sense promoters can terminate the antisense transcripts, the interesting question of what signals a terminator to function as a promoter and vice versa should be addressed. The preliminary findings here implicate Rrp6p and the 3'UTR in determining which transcript produced from bi-directional transcription units is dominant but no data exist about regulation of the promoters themselves. The early findings by RNA FISH that sense and antisense transcription might not occur in one cell at the same time, has led to the model that there is a preferred orientation of transcription, for a given period of time, perhaps depending on the cell cycle or the metabolic cycle, which in

some cases will be the sense transcription, while in others will be the antisense transcription. In a population, each cell will be in a different specific transcriptional state however most techniques are performed on a population of cells and therefore this effect will be evened out, except when the signal for one of the transcription unit becomes dominant. All these findings together highlight the need to re-assess what has been known about transcription.

References

Abruzzi, K.C., Belostotsky, D.A., Chekanova, J.A., Dower, K., and Rosbash, M. (2006). 3'-end formation signals modulate the association of genes with the nuclear periphery as well as mRNP dot formation. EMBO J *25*, 4253-4262.

Aranda, A., Pérez-Ortín, J.E., Moore, C., and del Olmo, M.l. (1998). The yeast FBP1 poly(A) signal functions in both orientations and overlaps with a gene promoter. Nucleic Acids Research *26*, 4588-4596.

Bajwa, W., Torchia, T.E., and Hopper, J.E. (1988). Yeast regulatory gene GAL3: carbon regulation; UASGal elements in common with GAL1, GAL2, GAL7, GAL10, GAL80, and MEL1; encoded protein strikingly similar to yeast and Escherichia coli galactokinases. Molecular and Cellular Biology *8*, 3439-3447.

Beck, C.F., and Warren, R.A. (1988). Divergent promoters, a common form of gene organization. Microbiol Rev.

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

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

Bhat, P., and Murthy, T. (2001). Transcriptional control of the GAL/MEL regulon of yeast Saccharomyces cerevisiae: mechanism of galactose-mediated signal transduction. Molecular Microbiology *40*, 1059-1066.

Bhat, P.J., Oh, D., and Hopper, J.E. (1990). Analysis of the GAL3 Signal Transduction Pathway Activating GAL4 Protein-Dependent Transcription in Saccharomyces cerevisiae. Genetics *125*, 281-291.

Bhaumik, S.R., Raha, T., Aiello, D.P., and Green, M.R. (2004). In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. Genes & Development *18*, 333-343.

Brickner, D.G., Cajigas, I., Fondufe-Mittendorf, Y., Ahmed, S., Lee, P.-C., Widom, J., and Brickner, J.H. (2007). H2A.Z-Mediated Localization of Genes at the Nuclear Periphery Confers Epigenetic Memory of Previous Transcriptional State. PLoS Biology *5*, e81.

Broach, J. (1979). Galactose regulation in Saccharomyces cerevisiae: The enzymes encoded by the GAL7, 10, 1 cluster are co-ordinately controlled and separately translated. Journal of Molecular Biology *131*, 41-53.

Brodsky, A.S., and Silver, P.A. (2000). Pre-mRNA processing factors are required for nuclear export. RNA *6*, 1737-1749.

Bryant, G.O., and Ptashne, M. (2003). Independent Recruitment In Vivo by Gal4 of Two Complexes Required for Transcription. Molecular Cell *11*, 1301-1309.

Buratowski, S. (2005). Connections between mRNA 3′ end processing and transcription termination. Current Opinion in Cell Biology *17*, 257-261.

Burkard, K.T.D., and Butler, J.S. (2000). A Nuclear 3′-5′ Exonuclease Involved in mRNA Degradation Interacts with Poly(A) Polymerase and the hnRNA Protein Npl3p. Molecular and Cellular Biology *20*, 604-616.

Butler, J.E.F., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & Development *16*, 2583-2592.

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

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007a). Antisense RNA Stabilization Induces Transcriptional Gene Silencing via Histone Deacetylation in S. cerevisiae. Cell *131*, 706-717.

Camblong, J., Iglesias, N., Fickentscher, C.l., Dieppois, G., and Stutz, F.o. (2007b). Antisense RNA Stabilization Induces Transcriptional Gene Silencing via Histone Deacetylation in S. cerevisiae. Cell *131*, 706-717.

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

Churchman, L.S., and Weissman, J.S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368-373.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368-373.

Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., Morillon, A.*, et al.* (2011). The Reality of Pervasive Transcription. PLoS Biol *9*, e1000625.

Culbertson, M. (1999). RNA surveillance: unforeseen consequences for gene expression, inherited genetic disorders and cancer. Trends in Genetics *15*, 74-80.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. PNAS *103*, 5320-5325.

Egriboz, O., Jiang, F., and Hopper, J.E. (2011). Rapid GAL Gene Switch of Saccharomyces cerevisiae Depends on Nuclear Gal3, Not Nucleocytoplasmic Trafficking of Gal3 and Gal80. Genetics *189*, 825-836.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic Gene Expression in a Single Cell. Science *297*, 1183-1186.

Erickson, J.R., and Johnston, M. (1993). Genetic and Molecular Characterization of GAL83: Its Interaction and Similarities With Other Genes Involved in Glucose Repression in Saccharomyces cerevisiae. Genetics *135*, 655-664.

Fan, X., Moqtaderi, Z., Jin, Y., Zhang, Y., Liu, X.S., and Struhl, K. (2010). Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. Proceedings of the National Academy of Sciences *107*, 17945-17950.

Fasken, M.B., and Corbett, A.H. (2005). Process or perish: quality control in mRNA biogenesis. Nat Struct Mol Biol *12*, 482-488.

Fasken, M.B., and Corbett, A.H. (2009). Mechanisms of nuclear mRNA quality control. RNA Biology *6*, 237-241.

Frischmeyer, P.A., and Dietz, H.C. (1999). Nonsense-Mediated mRNA Decay in Health and Disease. Human Molecular Genetics *8*, 1893-1900.

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

Gandhi, S.J., Zenklusen, D., Lionnet, T., and Singer, R.H. (2011). Transcription of functionally related constitutive genes is not coordinated. Nat Struct Mol Biol *18*, 27-34.

Geerlings, T.H., Vos, J.C., and Raue, H.A. (2000). The final step in the formation of 25S rRNA in saccharomyces cerevisiae is performed by 5' → 3' exonucleases. RNA *6*, 1698-1703.

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

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

Green, M. (2005). Eukaryotic Transcription Activation: Right on Target. Molecular Cell *18*, 399-402.

Greger, I.H., and Proudfoot, N.J. (1998). Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae. EMBO J *17*, 4771-4779.

Griggs, D.W., and Johnston, M. (1991). Regulated expression of the GAL4 activator gene in yeast provides a sensitive genetic switch for glucose repression. Proceedings of the National Academy of Sciences *88*, 8597-8601.

Gueldener, U., Heinisch, J., Koehler, G.J., Voss, D., and Hegemann, J.H. (2002). A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. Nucleic Acids Research *30*, e23.

Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M., and Martens, J.A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. Genes & Development *25*, 29-40.

Halley, J.E., Kaplan, T., Wang, A.Y., Kobor, M.S., and Rine, J. (2010). Roles for H2A.Z and Its Acetylation in *GAL1* Transcription and Gene Induction, but Not *GAL1*-Transcriptional Memory. PLoS Biol *8*, e1000401.

Hampsey, M., Singh, B.N., Ansari, A., Lainé, J.-P., and Krishnamurthy, S. (2011). Control of eukaryotic gene expression: gene loops and transcriptional memory. Advances in Enzyme Regulation *51*.

Hentze, M.W., and Kulozik, A.E. (1999). A Perfect Message: RNA Surveillance and Nonsense-Mediated Decay. Cell *96*, 307-310.

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

Hongay, C.F., Grisafi, P.L., Galitski, T., and Fink, G.R. (2006). Antisense Transcription Controls Cell Fate in Saccharomyces cerevisiae. Cell *127*, 735-745.

Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., and Vogelauer, M. (2008). A ncRNA Modulates Histone Modification and mRNA Induction in the Yeast GAL Gene Cluster. Molecular Cell *32*, 685-695.

Ideker T, T.V., Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science.

Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. Journal of bacteriology *153*, 163-168.

Ito, T., Miura, F., and Onda, M. (2008). Unexpected complexity of the budding yeast transcriptome. IUBMB Life *60*, 775-781.

Jansen, A., and Verstrepen, K.J. (2011). Nucleosome Positioning in Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews *75*, 301-320.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet *10*, 161-172.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental Biology *339*, 225-229.

Kaplan, C.D., Holland, M.J., and Winston, F. (2005). Interaction between Transcription Elongation Factors and mRNA 3′ -End Formation at the Saccharomyces cerevisiae GAL10-GAL7 Locus. Journal of Biological Chemistry *280*, 913-922.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J.*, et al.* (2005). Antisense Transcription in the Mammalian Transcriptome. Science *309*, 1564-1566.

Kornberg, R.D., and Lorch, Y. (1992). Chromatin Structure and Transcription. Annual Review of Cell Biology *8*, 563-587.

Kundu, S., Horn, P.J., and Peterson, C.L. (2007). SWI/SNF is required for transcriptional memory at the yeast GAL gene cluster. Genes & Development *21*, 997-1004.

Kundu, S., and Peterson, C.L. (2009). Role of chromatin states in transcriptional memory. Biochimica et Biophysica Acta (BBA) - General Subjects *1790*, 445-455.

Kundu, S., and Peterson, C.L. (2010). Dominant Role for Signal Transduction in the Transcriptional Memory of Yeast GAL Genes. Molecular and Cellular Biology *30*, 2330-2340.

Larschan, E., and Winston, F. (2005). The Saccharomyces cerevisiae Srb8-Srb11 Complex Functions with the SAGA Complex during Gal4-Activated Transcription. Mol Cell Biol *25*, 114-123.

Larson, D.R., Singer, R.H., and Zenklusen, D. (2009). A single molecule view of gene expression. Trends in Cell Biology *19*, 630-637.

Lavelle, C. (2007). Transcription elongation through a chromatin template. Biochimie *89*, 516-527.

Lennartsson, A., and Ekwall, K. (2009). Histone modification patterns and epigenetic codes. Biochimica et Biophysica Acta (BBA) - General Subjects *1790*, 863-868.

Leuther, K.K., and Johnston, S.A. (1992). Nondissociation of GAL4 and GAL80 in vivo after galactose induction. Science *256*, 1333-1335.

Li, Y., Bjorklund, S., Jiang, Y.W., Kim, Y.J., Lane, W.S., Stillman, D.J., and Kornberg, R.D. (1995). Yeast global transcriptional regulators Sin4 and Rgr1 are components of mediator complex/RNA polymerase II holoenzyme. Proceedings of the National Academy of Sciences *92*, 10864-10868.

Libri, D., Dower, K., Boulay, J., Thomsen, R., Rosbash, M., and Jensen, T.H. (2002). Interactions between mRNA Export Commitment, 3′-End Quality Control, and Nuclear Degradation. Molecular and Cellular Biology *22*, 8254-8266.

Lohr, D., Venkov, P., and Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J *9*, 777-787.

Long, R.M., and McNally, M.T. (2003). mRNA Decay: X (XRN1) Marks the Spot. Molecular Cell *11*, 1126-1128.

Long, R.M., Mylin, L.M., and Hopper, J.E. (1991). GAL11 (SPT13), a transcriptional regulator of diverse yeast genes, affects the phosphorylation state of GAL4, a highly specific transcriptional activator. Molecular and Cellular Biology *11*, 2311-2314.

Longman, D., Johnstone, I., and Cáceres, J. (2003). The Ref/Aly proteins are dispensable for mRNA export and development in Caenorhabditis elegans. RNA *9*, 881-891.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast (Chichester, England) *14*, 953-961.

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

Lykke-Andersen, S., Tomecki, R., Jensen, T.H., and Dziembowski, A. (2011). The eukaryotic RNA exosome: Same scaffold but variable catalytic subunits. RNA Biology *8*, 61-66.

Maniatis, T., and Fristish, E.F. (1982). Molecular Cloning: A Laboratory Manual.

Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E., and Jensen, T.H. (2010). Crosstalk between mRNA 3′ End Processing and Transcription Initiation. Molecular Cell *40*, 410-422.

Maquat, L.E. (2002). Skiing Toward Nonstop mRNA Decay. Science *295*, 2221-2222.

Maquat, L.E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol *5*, 89-99.

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

Martens, J.A., Wu, P.Y., and Winston, F. (2005). Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. Genes Dev *19*, 2695-2704.

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

Mattick, J.S. (2009). The Genetic Signatures of Noncoding RNAs. PLoS Genet *5*, e1000459.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Research *18*, 1073-1083.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nature Rev Genet *10*, 155-159.

Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. Nucleic Acids Research *38*, 2757-2774.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The Exosome: A Conserved Eukaryotic RNA Processing Complex Containing Multiple 3′→5′ Exoribonucleases. Cell *91*, 457-466.

Mizuno, T., and Harashima, S. (2003). Gal11 is a general activator of basal transcription, whose activity is regulated by the general repressor Sin4 in yeast. Molecular Genetics and Genomics *269*, 68-77.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. Cell *136*, 688-700.

Morillon, A., Karabetsou, N., O'Sullivan, J., Kent, N., Proudfoot, N., and Mellor, J. (2003). Isw1 Chromatin Remodeling ATPase Coordinates Transcription Elongation and Termination by RNA Polymerase II. Cell *115*, 425-435.

Morris, K.V., Santoso, S., Turner, A.M., Pastori, C., and Hawkins, P.G. (2008). Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet *4*, e1000258.

Morschhäuser, J. (2010). Regulation of white-opaque switching in <i>Candida albicans</i&gt. Medical Microbiology and Immunology *199*, 165-172.

Mühlemann, O., Eberle, A.B., Stalder, L., and Zamudio Orozco, R. (2008). Recognition and elimination of nonsense mRNA. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1779*, 538-549.

Murray, S.C., Serra Barros, A., Brown, D.A., Dudek, P., Ayling, J., and Mellor, J. (2011). A pre-initiation complex at the 3′ -end of genes drives antisense transcription independent of divergent sense transcription. Nucleic Acids Research *40*, 2432-2444.

Neil, H., Malabat, C., d/'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038-1042.

Nogi, Y. (1986). GAL3 gene product is required for maintenance of the induced state of the GAL cluster genes in Saccharomyces cerevisiae. Journal of bacteriology *165*, 101-106.

Novick, A., and Weiner, M. (1957). ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON. Proc Natl Acad Sci.

Ostergaard, S., Walløe, K.O., Gomes, C.S.G., Olsson, L., and Nielsen, J. (2001). The impact of GAL6, GAL80, and MIG1 on glucose control of the GAL system in Saccharomyces cerevisiae. FEMS Yeast Research *1*, 47-55.

Parthun, M.R., and Jaehning, J.A. (1992). A transcriptionally active form of GAL4 is phosphorylated and associated with GAL80. Mol Cell Biol *12*, 4981-4987.

Peng, G., and Hopper, J.E. (2000). Evidence for Gal3p's Cytoplasmic Location and Gal80p's Dual Cytoplasmic-Nuclear Location Implicates New Mechanisms for Controlling Gal4p Activity in Saccharomyces cerevisiae. Mol Cell Biol *20*, 5140-5148.

Peng, G., and Hopper, J.E. (2002). Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. PNAS *99*, 8548-8553.

Perocchi, F., Xu, Z., Clauder-MÃ¼nster, S., and Steinmetz, L.M. (2007). Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Nucleic Acids Research *35*, e128.

Pinskaya, M., Gourvennec, S., and Morillon, A. (2009). H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. EMBO J *28*, 1697-1707.

Platt, A., and Reece, R.J. (1998). The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex EMBO Journal *17*, 4086-4091.

Preker, P. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.

Proudfoot, N. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. Trends in Biochemical Sciences *14*, 105-110.

Pruneski, J.A., Hainer, S.J., Petrov, K.O., and Martens, J.A. (2011). The Paf1 Complex Represses SER3 Transcription in Saccharomyces cerevisiae by Facilitating Intergenic Transcription-Dependent Nucleosome Occupancy of the SER3 Promoter. Eukaryotic Cell *10*, 1283-1294.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature *335*, 683-689.

Radonjic, M., Andrau, J.-C., Lijnzaad, P., Kemmeren, P., Kockelkorn, T.T.J.P., van Leenen, D., van Berkum, N.L., and Holstege, F.C.P. (2005). Genome-Wide Analyses Reveal RNA Polymerase II

Located Upstream of Genes Poised for Rapid Response upon S. cerevisiae Stationary Phase Exit. Molecular Cell *18*, 171-183.

Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell *135*, 216-226.

Rodŕiguez-Navarro, S., Fischer, T., Luo, M.-J., Antúnez, O., Brettschneider, S., Lechner, J., Pérez-Ortín, J.E., Reed, R., and Hurt, E. (2004). Sus1, a Functional Component of the SAGA Histone Acetylase Complex and the Nuclear Pore-Associated mRNA Export Machinery. Cell *116*, 75-86.

Saeki, H., and Svejstrup, J.Q. (2009). Stability, Flexibility, and Dynamic Interactions of Colliding RNA Polymerase II Elongation Complexes. Molecular Cell *35*, 191-205.

Schmid, M., and Jensen, T. (2008). Quality control of mRNP in the nucleus. Chromosoma *117*, 419-429.

Seila, A.C. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: A new feature of active promoters. Cell Cycle *8*, 2557-2564.

Sellick, C.A., Jowitt, T.A., and Reece, R.J. (2009). The effect of ligand binding on the galactokinase activity of yeast Gal1p and its ability to activate transcription. Journal of Biological Chemistry *284*, 229-236.

Sellick, C.A., and Reece, R.J. (2005). Eukaryotic transcription factors as direct nutrient sensors. Trends in Biochemical Sciences *30*, 405-412.

Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. Current Opinion in Cell Biology *21*, 344-351.

Sil, A.K., Alam, S., Xin, P., Ma, L., Morgan, M., Lebo, C.M., Woods, M.P., and Hopper, J.E. (1999). The Gal3p-Gal80p-Gal4p Transcription Switch of Yeast: Gal3p Destabilizes the Gal80p-Gal4p Complex in Response to Galactose and ATP. Mol Cell Biol *19*, 7828-7840.

Singer, V.L., Wobbe, C.R., and Struhl, K. (1990). A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes & Development *4*, 636-645.

Straszer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R.*, et al.* (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304-308.

Struhl, K. (1987). Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. Cell *49*.

Struhl, K. (1995). Yeast Transcriptional Regulatory Mechanisms. Annual Review of Genetics *29*, 651-674.

Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R.R., Grimmond, S.M., Schroder, K.*, et al.* (2009). Tiny RNAs associated with transcription start sites in animals. Nat Genet *41*, 572-578.

Tan-Wong, S.M., Wijayatilake, H.D., and Proudfoot, N.J. (2009). Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. Genes & Development *23*, 2610-2624.

Thebault, P., Boutin, G., Bhat, W., Rufiange, A., Martens, J., and Nourani, A. (2011). Transcription Regulation by the Noncoding RNA SRG1 Requires Spt2-Dependent Chromatin Deposition in the Wake of RNA Polymerase II. Molecular and Cellular Biology *31*, 1288-1300.

Thoden, J.B., Sellick, C.A., Timson, D.J., Reece, R.J., and Holden, H.M. (2005). Molecular Structure of Saccharomyces cerevisiae Gal1p, a Bifunctional Galactokinase and Transcriptional Inducer. Journal of Biological Chemistry *280*, 36905-36911.

Thomas, M.C., and Chiang, C.-M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Research *18*, 1084-1091.

Tisseur, M., Kwapisz, M., and Morillon, A. (2011). Pervasive transcription - Lessons from yeast. Biochimie *93*, 1889-1896.

Torchia, T.E., and Hopper, J.E. (1986). Genetic and molecular analysis of the *GAL3* gene in the expression of the galactose/melibiose regulon of *Saccharomyces cerevisiae*. Genetics *113*, 229-246.

Traven, A., Jelicic, B., and Sopta, M. (2006). Yeast Gal4: a transcriptional paradigm revisited. EMBO Reports *7*, 496-499.

Uhler, J.P., Hertel, C., and Svejstrup, J.Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. Proceedings of the National Academy of Sciences *104*, 8011-8016.

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

Vasudevan, S., Peltz, S.W., and Wilusz, C.J. (2002). Non-stop decay—a new mRNA surveillance pathway. Bioessays *24*, 785-788.

Venters, B.J., and Pugh, B.F. (2009). A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. Genome Research *19*, 360-371.

Venters, B.J., and Pugh, B.F. (2010). How eukaryotic genes are transcribed. Crit Rev Biochem Mol Biol *44*, 117-141.

Vinciguerra, P., and Stutz, F. (2004). mRNA export: an assembly line from genes to nuclear pores. Current Opinion in Cell Biology *16*, 285-292.

Vodala, S., Abruzzi, K.C., and Rosbash, M. (2008). The Nuclear Exosome and Adenylation Regulate Posttranscriptional Tethering of Yeast GAL Genes to the Nuclear Periphery. Molecular Cell *31*, 104-113.

Wang, G.-Z., Lercher, M.J., and Hurst, L.D. (2011). Transcriptional Coupling of Neighboring Genes and Gene Expression Noise: Evidence that Gene Orientation and Noncoding Transcripts Are Modulators of Noise. Genome Biology and Evolution *3*, 320-331.

Wei, W., Pelechano, V., Järvelin, A.I., and Steinmetz, L.M. (2011). Functional consequences of bidirectional promoters. Trends in Genetics *27*, 267-276.

Wightman, R., Bell, R., and Reece, R.J. (2008). Localization and Interaction of the Proteins Constituting the GAL Genetic Switch in Saccharomyces cerevisiae. Eukaryotic Cell *7*, 2061-2068.

Workman, J.L., and Buchman, A.R. (1993). Multiple functions of nucleosomes and regulatory factors in transcription. Trends in Biochemical Sciences *18*, 90-95.

Woychik, N.A., Liao, S.M., Kolodziej, P.A., and Young, R.A. (1990). Subunits shared by eukaryotic nuclear RNA polymerases. Genes & Development *4*, 313-323.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B.*, et al.* (2005). Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. Cell *121*, 725-737.

Xu, Z. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033-1037.

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

Yassour, M., Pfiffner, J., Levin, J., Adiconis, X., Gnirke, A., Nusbaum, C., Thompson, D.-A., Friedman, N., and Regev, A. (2010). Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. Genome Biology *11*, R87.

Zacharioudakis, I., Gligoris, T., and Tzamarias, D. (2007). A Yeast Catabolic Enzyme Controls Transcriptional Memory. Current Biology *17*, 2041-2046.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol *15*, 1263-1271.

Zenklusen, D., and Singer, R.H. (2010). Analyzing mRNA Expression Using Single mRNA Resolution Fluorescent In Situ Hybridization. Methods Enzymol *470*.

Zenklusen, D., Vinciguerra, P., Wyss, J.-C., and Stutz, F. (2002). Stable mRNP Formation and Export Require Cotranscriptional Recruitment of the mRNA Export Factors Yra1p and Sub2p by Hpr1p. Molecular and Cellular Biology *22*, 8241-8253.

Zhang, Z., Fu, J., and Gilmour, D.S. (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′ -end processing factor, Pcf11. Genes & Development *19*, 1572-1580.

Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3′ Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. Microbiol Mol Biol Rev *63*, 405-445.

Zheng, W., Xu, H.E., and Johnston, S.A. (1997). The cysteine-peptidase bleomycin hydrolase is a member of the galactose regulon in yeast. J Biol Chem *272*, 30350-30355.

Zhou, B.O., and Zhou, J.-Q. (2011). Recent Transcription-induced Histone H3 Lysine 4 (H3K4) Methylation Inhibits Gene Reactivation. J Biol Chem *286*, 34770-34776.