Université de Montréal

Single molecule characterization of the roles of long non-coding RNAs in eukaryotic transcription regulation

par

Samir Rahman

Département de biochimie

Faculté de médecine

Thèse présentée à la Faculté de médecine en vue de
l'obtention du grade de doctorat en biochimie,
option genétique moleculaire

Mai 2017

Résumé

Récemment, des analyses dans divers organismes eucaryotes ont révélé que l'ensemble du génome est transcrit et produit en plus des ARNs messagers, une grande variété d'ARNs non codants de différentes longueurs. Les ARNs non codants de plus de 200 nucleotides, classés comme longs ARNs non codants (LARNnc), représentent la classe la plus abondante de transcripts non codants. Les études des fonctions des LARNnc suggèrent que beaucoup d'entre eux seraient impliqués dans la régulation de la transcription. L'objectif de ma thèse de doctorat était d'élucider les mécanismes de la régulation transcriptionnelle médiée par des LARNnc dans différents systèmes eucaryotes.

Dans mon premier projet, j'ai étudié le rôle d'un long ARN non codant antisens dans la régulation transcriptionnelle du gène *PHO84,* codant un transporteur de phosphate à haute affinité, chez *S. cerevisiae*. Des études antérieures ont montré que la suppression d'une proteine de l'exosome Rrp6 entraîne une augmentation de l'expression antisens et la répression de *PHO84*. Il a été suggéré que la perte de Rrp6 entraîne une stabilisation antisens au locus *PHO84*, entraînant le recrutement de l'histone de-acétylase Hda1 et la répression de *PHO84*. Cependant, le mécanisme par lequel Rrp6p régule la transcription de *PHO84* n'était pas connu. En combinant des méthodes à l'échelle de cellule unique, des approches biochimiques et génétiques, nous avons montré que les niveaux d'ARN antisens sont régulés principalement lors de l'élongation par le complexe Nrd1-Nab3-Sen1, qui nécessite Rrp6 pour un recrutement efficace à l`extrémité 3`de *PHO84*. De plus, nous révélons l'expression anticorrelé du sens et de l'antisens, En résumé, nos données suggèrent que la transcription antisens régule le seuil d'activation du promoteur *PHO84*.

Dans mon second projet, j'ai étudié les rôles des ARNs dérivés des amplificateurs (ARNa) dans la regulation de la transcription. En utilisant les cellules de cancer du sein MCF7 comme système modèle, nous avons cherché à déterminer comment les ARNa induits par l'oestrogène (E2) participent à la régulation de la transcription médiée par le recepteur d'oestrogène (ERα) au niveau de l'allèle unique. À l'aide de l'hybridation fluorescente à l'échelle de molécule unique (smFISH), nous avons révélé qu`après induction d'E2, les ARNa

sont induits avec une cinétique similaire à celle des ARNm cibles, sont localisés exclusivement dans le noyau, principalement associés à la chromatine, et sont moins abondants que les ARNm. De manière surprenante, nous avons constaté que les ARNa sont rarement co-transcrits avec leurs loci cibles, indiquant que la transcription active des gènes ne nécessite pas la synthèse continue ou l'accumulation d'ARNa sur l'amplificateur. En outre, en utilisant des mesures de la distance à sous-diffraction, nous avons démontré que la co-transcription des ARNa et des ARNm se produit rarement dans une boucle amplificateur-promoteur. De plus, nous avons révélé que la transcription basale d'ARNa n'exige pas ERα ou l'histone méthyltransférase MLL1 qui active l'amplificateur par la mono-méthylation H3K4. Dans l'ensemble, nos résultats ont montré que les ARNa peuvent jouer un rôle lors de l'activation du promoteur, mais ne sont pas nécessaires pour maintenir la transcription de l'ARNm ou pour stabiliser les interactions amplificateur-promoteur.

Mots clés: long ARN non-codant, transcription, smFISH, *S.cerevisiae*, *PHO84*, ARN antisens, MCF7, ERα, ARN amplificateur, MLL1

Abstract

Transcription is the initial step in gene expression and is subject to extensive regulation. Recently, analyses in diverse eukaryotes have revealed that in addition to protein coding genes, transcription occurs throughout the noncoding genome, producing non-coding RNAs of various lengths. Non-coding RNAs longer than 200 nucleotides, classified as long non-coding RNAs (lncRNAs), represent the most abundant class of non-coding transcripts, whose functions however are poorly understood. Recent studies suggest that many lncRNAs might have roles in transcription regulation. The goal of my PhD thesis was to elucidate the mechanisms of lncRNA mediated transcription regulation in different eukaryotic systems.

For my first project, I investigated the role of an antisense long noncoding RNA in transcription regulation of the high-affinity phosphate transporter gene *PHO84* in the unicellular eukaryote *S. cerevisiae*. Previous studies showed that deletion of the nuclear exosome component Rrp6 results in increased antisense expression and repression of *PHO84.* It was suggested that the loss of Rrp6 results in antisense stabilization at the *PHO84* locus, leading to recruitment of the histone de-acetylase Hda1 and repression of *PHO84*. However, most of the mechanistic details of how Rrp6p functions in regulating *PHO84* transcription were not understood. Combining single cell methods with biochemical and genetic approaches, we showed that antisense RNA levels are regulated primarily during transcriptional elongation by the Nrd1-Nab3-Sen1 complex, which requires Rrp6 for efficient recruitment to the 3'end of *PHO84*. Furthermore, we reveal anti-correlated expression of sense and antisense, which have distinct modes of transcription. In summary, our data suggest a model whereby antisense transcriptional read-through into the *PHO84* promoter regulates the activation threshold of the gene.

For my second project, I investigated the roles of enhancer derived RNAs (eRNAs). eRNAs are lncRNAs transcribed from enhancers that have been suggested to regulate transcription through different mechanisms, including enhancer-promoter looping, RNA polymerase elongation, and chromatin remodeling. However, no coherent model of eRNA function has yet emerged. Using MCF7 breast cancer cells as a model system, we sought to

determine how estrogen (E2) induced eRNAs participate in estrogen receptor alpha (ERα) mediated transcription regulation at the single allele level. Using single molecule fluorescent in situ hybridization (smFISH), we revealed that upon E2 induction eRNAs are induced with similar kinetics as target mRNAs, but are localized exclusively in the nucleus, mostly chromatin associated, and are less abundant than mRNAs. Surprisingly, we found that eRNAs are rarely co-transcribed with their target loci, indicating that active gene transcription does not require the continuous synthesis or accumulation of eRNAs at the enhancer. Furthermore, using sub-diffraction-limit distance measurements, we demonstrated that co-transcription of eRNAs and mRNAs rarely occurs within a closed enhancer-promoter loop. Moreover, we revealed that basal eRNA transcription does not require ERα or the histone methyltransferase MLL1, which activates the enhancer through H3K4 mono-methylation. Altogether, our findings showed that eRNAs may play a role during promoter activation, but are not required to sustain mRNA transcription or stabilize enhancer-promoter looping interactions.

Keywords: long non-coding RNA, transcription, smFISH, *S.cerevisiae*, *PHO84*, antisense RNA, MCF7, ERα, enhancer RNA, MLL1

Table of contents

List of tables

Chapter 2: Article 1

Chapter 3: Article 2

List of figures

List of abbreviations

3C	*chromosome conformation capture*
4C	*circularized chromosome conformation capture*
AP-1	*activator protein 1*
AR	*androgen receptor*
AS	*antisense*
ATP	*adenosine tri-phosphate*
BRD4	*bromodomain-containing protein 4*
CAGE	*cap analysis of gene expression*
Cas9	*CRISPR associated protein 9*
CBC	*cap-binding complex*
CBP	*CREB binding protein*
CBP20	*cap binding protein 20*
CBP80	*cap binding protein 80*
CDK	*cyclin-dependent kinase*
CF	*cleavage factor*
CGBP	*CpG binding protein*
CHA4	*catabolism of hydroxy amino acids 4*
ChIA-PET	*chromatin interaction analysis by paired-end tag sequencing*
ChIP	*chromatin immunoprecipitation*
ChIRP-MS	*chromatin isolation by RNA purification with mass spectrometry*
ChromRNA-seq	*chromatin-associated RNA-sequencing*
CPF	*cleavage and polyadenylation factor*
CRISPR	*clustered regularly interspaced short palindromic repeats*
CTCF	*CCCTC-binding factor*
CTD	*c-terminal domain*
CUT	*cryptic unstable transcript*
DAPI	*4',6-diamidino-2-phenylindole*
DHRS4	*dehydrogenase/reductase (SDR family) member 4*
DNA	*Deoxyribonucleic acid*
DRB	*5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole*
EGF	*epidermal growth factor*
ENCODE	*encyclopedia of DNA elements*
ERE	*estrogen response element*
eRNA	*enhancer RNA*
ERα	*estrogen receptor alpha*
FISH	*fluorescent in situ hybridization*

FOXA1	*forkhead box protein A1*
FOXC1	*forkhead box protein C1*
FRAP	*fluorescence recovery after photobleaching*
GFP	*green fluorescent protein*
GR	*glucocorticoid receptor*
GREB1	*growth regulation by estrogen in breast cancer 1*
GRO-seq	*global run-on sequencing*
GTF	*general transcription factor*
GTP	*guanosine tri-phosphate*
HDA1	*histone de-acetylase 1*
HDAC	*histone de-acetylase*
HIV-1	*human immunodeficiency virus-1*
HOTAIR	*HOX transcript antisense RNA*
HOXC	*homeobox C cluster*
iCLIP	*individual-nucleotide resolution UV crosslinking and immunoprecipitation*
ICR1	*interfering crick RNA 1*
IMD2	*inosine monophosphate dehydrogenase 2*
IME1	*inducer of meiosis 1*
IME4	*inducer of meiosis 4*
IRT1	*IME1 Regulatory Transcript*
lincRNA	*long intergenic non-coding RNA*
lncRNA	*long non-coding RNA*
LSD1	*lysine-specific histone demethylase 1*
MALAT1	*metastasis associated lung adenocarcinoma transcript 1*
MCF7	*michigan cancer foundation-7*
MCP	*MS2 coat protein*
MDN1	*Midasin AAA ATPase 1*
miRNA	*micro RNA*
MLL1	*mixed lineage leukemia 1*
mRNA	*messenger RNA*
mRNP	*messenger ribonucleoprotein*
NAB3	*nuclear polyadenylated RNA-binding 3*
ncRNA	*non-coding RNA*
ncRNA-a	*non-coding RNA-activating*
NDR	*nucleosome depleted region*
NEAT1	*nuclear paraspeckle assembly transcript 1*
NELF	*negative elongation factor*
NEXT	*nuclear exosome targeting complex*
NNS	*Nrd1-Nab3-Sen1*
NRD1	*nuclear pre-mRNA down-regulation 1*

ORF	*open reading frame*
P2RY2	*purinergic receptor P2Y2*
PAP1	*poly(A) polymerase 1*
PAR-CLIP	*photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation*
PIC	*pre-initiation complex*
PRC2	*polycomb repressive complex 2*
PROMPT	*promoter upstream transcript*
pTEFb	*positive transcription elongation factor b*
PWR1	*promoting watson RNA 1*
qPCR	*quantitative polymerase chain reaction*
RBM7	*RNA Binding Motif Protein 7*
REB1	*RNA polymerase I Enhancer Binding protein 1*
RIP-qPCR	*RNA immunoprecipitation qPCR*
RME2	*regulator of meiosis 2*
RNA	*Ribonucleic acid*
RNAPII	*RNA polymerase II*
RPD3	*Reduced Potassium Dependency 3*
RPL9B	*ribosomal protein of the large subunit 9B*
RRM	*RNA recognition motif*
rRNA	*ribosomal RNA*
RRP6	*ribosomal RNA processing 6*
RT-qPCR	*reverse transcriptase qPCR*
SAGA	*Spt-Ada-Gcn5-acetyltransferase*
SD	*synthetic defined*
SEN1	*splicing endonuclease 1*
SER3	*SERine requiring 3*
SET	*Su(var)3-9, Enhancer-of-zeste and Trithorax*
SFL1	*suppressor gene for flocculation 1*
sgRNA	*single-guide RNA*
SHAPE	*selective 2'-hydroxyl acylation analyzed by primer extension*
SHH	*sonic hedgehog*
siRNA	*small interfering RNA*
smFISH	*single molecule FISH*
snoRNA	*small nucleolar RNA*
snRNA	*small nuclear RNA*
snRNP	*small nuclear ribonucleoprotein*
SPT6	*suppressor of Ty insertion mutations 6*
SRG1	*SER3 regulatory gene 1*
SUT	*stable unannotated transcript*
SWI/SNF	*SWItch/sucrose non-fermentable*

TAD	*topologically associated domain*
TAR	*trans-activation response*
TARID	*TCF21 antisense RNA inducing promoter demethylation*
TBP	*TATA box binding protein*
TCF21	*transcription factor 21*
TF	*transcription factor*
TFF1	*trefoil factor 1*
TFIID	*transcription factor II D*
TIP60	*tat interacting protein 60 kD*
TRAMP	*Trf4/Air2/Mtr4p polyadenylation complex*
TRF4	*topoisomerase one-requiring function 4*
tRNA	*transfer RNA*
TS	*transcription site*
TSS	*transcription start site*
TTS	*transcription termination site*
UAS	*upstream activating sequence*
UTR	*untranslated region*
XIST	*X-inactive specific transcript*
XRN1	*exoribonuclease 1*
YEPD	*yeast extract peptone dextrose*
YY1	*Yin Yang 1*
ZRS	*zone of polarizing activity regulatory sequence*

Acknowledgements

First and foremost, I would like to thank my family and friends for their undying support and encouragement over the years. I thank my supervisor Dr.Daniel Zenklusen for inspiring me and challenging me throughout my academic career and offering invaluable advice and scientific insight, which has been instrumental in my progress as a graduate student. I would also like to thank my past and present colleagues and collaborators for their support, without whom none of the work that I have accomplished during my PhD would have been possible.

1 Introduction

1.1 Foreword

The central dogma of molecular biology proposes that a gene, the DNA sequence that encodes the blueprint for protein synthesis, is the most fundamental unit of heritable information. According to this conventional view, biological information is decoded through a linear process, whereby a gene is first transcribed into a messenger RNA, which is then translated into a protein [1]. Unexpectedly, studies within the last decade have found that at least three quarters of the human genome is transcribed, far exceeding the 1-2% of the human genome that is known to encode proteins [2–4]. This discovery challenged the long held notion depicting most non-coding regions as "junk" DNA because of their lack of protein coding potential. A myriad of transcriptome analyses following these initial studies revealed that all eukaryotes express a large ensemble of non-coding RNAs (ncRNAs) of various lengths. In addition to the well-defined classes of non-coding RNAs such as rRNAs, tRNAs, miRNAs, snRNAs, and snoRNAs, the subclass referred to as long non-coding RNAs (lncRNAs), consisting of ncRNAs longer than 200 nucleotides, represents the most abundant class of ncRNAs; whose function(s) however remain largely elusive. Various studies demonstrated that lncRNAs are often expressed in a cell type specific manner, implying that at least some of them might play a role in cellular differentiation or in regulating gene expression in response to cell type specific signalling pathways [2,3,5,6]. However, it is currently unknown what fraction of lncRNAs is functional, what their function is, and whether many might simply be by-products of 'transcription noise' [5–8]. Furthermore, biological complexity was shown to strongly correlate with the relative size of the noncoding genome rather than the number of protein-coding genes, suggesting that the noncoding fraction of the genome might play an important role in regulating eukaryotic development [9]. Therefore, understanding the role of lncRNAs is one of the most daunting yet exciting goals of biological research in the post-genomic era.

1.2 Chromatin remodeling and transcription initiation

The expansion of eukaryotic genomes during evolution has necessitated extensive DNA condensation. DNA compaction is facilitated by the binding of histone proteins to DNA,

resulting in the packaging of DNA into nucleosomes, often found in a "beads on a string" conformation where 146 nt of DNA is wrapped around a histone octamer consisting of 2 copies of the 4 core histone proteins histone H2A, H2B, H3 and H4 [10]. In eukaryotic chromatin, nucleosomes often occupy defined positions on DNA and occlude underlying regulatory sequences, such as transcription factor binding sites, from being accessed by the transcriptional machinery. To modulate DNA accessibility, eukaryotic organisms have evolved chromatin remodeling enzymes that rely on the energy of ATP to shift or evict nucleosomes from regulatory regions to allow access to the transcriptional machinery [11]. Transcription initiation is therefore a highly regulated multi-step process that involves the interaction of different protein complexes, including transcription factors, histone modifiers, and chromatin remodelers.

1.2.1 Nucleosome dynamics on regulatory regions in *S.cerevisiae*

The yeast *S.cerevisiae* has been a valuable model system to study the role of nucleosome remodeling in transcription regulation, demonstrating how signal-dependent loss or acquisition of histones at specific TF binding sites activates or represses transcription, respectively. Notably, the *PHO5* gene, which is induced in response to changing concentrations of extracellular phosphate, has been extensively studied as a model gene to investigate the role of signal-dependent nucleosome remodeling in transcription regulation [12–16]. *PHO5* encodes an acid phosphatase and is a member of the *PHO* family of genes, which are activated by the transcription factor Pho4p during inorganic phosphate starvation.

Pho4p localization and activity is regulated in response to the concentration of extracellular phosphate by the Pho80-Pho85 cyclin-CDK (cyclin dependent kinase) complex [17]. When cells are grown in a phosphate-rich medium, Pho4p is phosphorylated by Pho80-Pho85 and localized predominantly in the cytoplasm. Conversely, when cells are starved for phosphate, Pho4p is dephosphorylated and imported into the nucleus, where it is fully active. It was later shown that transcription activation by Pho4p is fine-tuned in response to varying concentrations of phosphate through intermediate levels of phosphorylation [18]. Furthermore, Springer and colleagues showed that a partially phosphorylated isoform of Pho4p is localized

in the nucleus and can activate a subset of phosphate-responsive genes when cells are grown in intermediate-phosphate conditions.

An early study on the *PHO5* promoter demonstrated that gene induction during phosphate starvation results in the disruption of well-positioned nucleosomes on the promoter [12]. It was later shown that transcription activation by Pho4p induces the displacement of histones from regions containing Pho4p binding sites [13]. Similarly, using a *Δpho80* mutant strain, in which dephosphorylated Pho4p localizes constitutively in the nucleus irrespective of phosphate levels, showed that transcription activation by Pho4p leads to a complete loss of nucleosomes on the *PHO5* promoter, recapitulating the disruption of chromatin that occurs in wild type cells in response to phosphate depletion [14]. Therefore, transcription factors can mediate chromatin accessibility by disrupting the nucleosome structure on target promoters.

In addition to transcription factors, chromatin accessibility is also modulated by histone acetyltransferases and chromatin remodeling complexes. Histone acetyltransferases neutralize positively charged amino acids on histones, such as lysine, by adding an acetyl group, which weakens interactions between histones and the negatively charged backbone of DNA, making regulatory sequences more accessible. Chromatin remodeling complexes, such as the SWI/SNF complex in yeast, utilize the energy of ATP to shift or evict nucleosomes from regulatory sequences.

A study by Reinke and colleagues revealed the importance of the SWI/SNF chromatin remodeling complex in *PHO5* activation by demonstrating that the mutant strain *Δsnf2* , which lacks the Snf2p component of SWI/SNF, shows delayed histone removal from the *PHO5* promoter during phosphate starvation. Moreover, they observed that histones on the *PHO5* promoter were hyper-acetylated prior to nucleosome loss, demonstrating a mechanistic link between histone acetylation and histone eviction. Interestingly, histone acetylation by the SAGA complex was previously shown to modulate the rate of chromatin remodeling and gene activation from the *PHO5* promoter [19]. SAGA also plays a distinct role in recruiting the TATA box binding protein (TBP), a component of the pre-initiation complex (PIC), onto the *PHO5* promoter TATA box motif , a core promoter sequence that has been evolutionarily conserved in eukaryotes [20]. Furthermore, SAGA activity was shown to be dependent on Pho4p. Altogether, studies on transcription using *PHO5* as a model demonstrate that transcription

factors work in concert with the PIC, as well as chromatin modifying and remodeling enzymes to evict nucleosomes from gene regulatory regions in a timely manner to activate transcription in response to extracellular signals. Conversely, it was shown that nucleosomes are re-assembled on the promoter by the SWI/SNF chromatin remodeling complex during re-repression of *PHO5* [16]. Therefore, histones can be evicted or re-assembled on regulatory regions in a dynamic manner to regulate transcription in response to specific signals.

Genome-wide analysis of histone dynamics in yeast has shown that histones are stably associated with genomic DNA within coding regions but rapidly dissociate and are replaced by new histones on promoters [21]. The dynamic turnover of histones occurred predominantly on promoters containing TATA box motifs recognized by SAGA/TBP. Intriguingly, another study on the dynamics of TBP binding across the yeast genome showed rapid turnover of TBP on SAGA/TBP bound promoters [22]. Surprisingly, only 10-20% of yeast genes contain TATA box motifs. These genes are rapidly activated in response to metabolic stress, as opposed to the majority of yeast genes that are TATA-less, which are mainly housekeeping genes that are constitutively expressed to maintain homeostasis. The rapid turnover of both TBP and histones on SAGA/TBP activated promoters may reflect a dynamic exchange between transcription factors and nucleosomes, which bind competitively on promoters, to maintain them in a transcriptionally active or inactive state, respectively.

1.2.2 Intercellular variability in gene expression

Gene expression is an inherently probabilistic process, as it is mediated by chemical reactions that rely on the availability of molecules which are present in small copy number and are diffusing in the nucleus (transcription factors, RNA polymerases, nucleotides, etc.). It has been shown that a genetically identical population of cells can show substantial intercellular variability in expression due to random fluctuations in the extrinsic factors mentioned above as well as intrinsic noise in promoter activity [23–27].

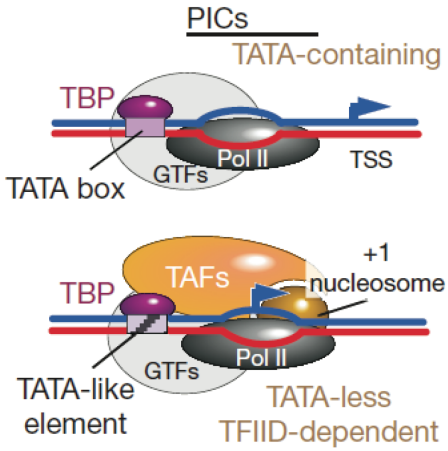
A study by Raser and O'Shea measured the intrinsic fluctuations in promoter activation using dual fluorescent protein reporters that monitored the expression of two alleles of the *PHO5* locus in a diploid yeast strain, and constructed a model of stochastic gene expression [23]. Their model predicts 3 different kinetic mechanisms of promoter activation, two

of which describe a 2 state model of expression, whereby the promoter remains inactive for most of the time and is activated infrequently in a random manner. Interestingly, although promoter activation is infrequent, every time the promoter is active, transcription can be re-initiated in rapid succession, possibly due to the stability of transcription factor-DNA interaction which may evict nucleosomes completely and facilitate the recruitment of multiple RNA polymerases. The two cases of this stochastic model both contribute to a large variability in mRNA levels within the population. In the first case, the activated promoter is very stable due to strong transcription factor-DNA interactions. Indeed, a subsequent study showed that a strong TBP-TATA box scaffold promotes strong "bursting" expression, whereby RNA pol II rapidly re-initiates transcription when the promoter is active, resulting in high variability in mRNA expression [28]. Mutations to the TATA box significantly reduced gene expression variability. The second case of stochastic activation described by Raser and O'Shea corresponds to a weaker "bursting" model caused by an active promoter state that is unstable, possibly due to nucleosome translocation on the promoter which might disassociate the TBP. Therefore, although this model is consistent with bursting expression, there are fewer transcription initiation events generated per burst, due to inefficient nucleosome remodeling by the PIC. The third kinetic mechanism of promoter activation described by Raser and O'Shea corresponds to a one state model of gene expression, whereby the promoter is always ON, allowing frequent activation. However, every time the promoter is activated, it does not result in efficient transcription, possibly due to inefficient binding of transcription factors. Consequently, gene expression does not occur in discontinuous bursts, but continuously, whereby only a fraction of transcription factor binding events on a continually accessible promoter will lead to transcription initiation. Therefore, in this model, single uncoordinated transcription initiation events are evenly distributed over time, producing little variability in expression within a population, allowing cells to maintain constant mRNA levels.

1.2.3 Nucleosome positioning on TATA-less vs TATA box-containing promoters

The inefficiency of transcription initiation in the one state model of expression may result from chromatin being in a partially accessible state that is intermediate between the nucleosome architectures of the ON and OFF states in the 2-state model. Interestingly, a

genome wide study on PIC assembly on yeast promoters characterized the differences in nucleosome architecture between TATA-box containing promoters and TATA-less promoters, which may explain why SAGA/TBP bound promoters would favor a bursting mode of expression, whereas TATA-less promoters would favor a continuous mode of expression [29] (**see Figure 1-1**). They show that TBP binds to the TATA-box motif significantly upstream of the transcription start site (TSS) within a large nucleosome free region that allows RNA pol II to scan a sufficient distance before initiating transcription. This increases the probability that every polymerase recruited by SAGA/TBP will successfully initiate transcription, resulting in bursting. Conversely, on TATA-less promoters, TBP is recruited to TATA-like motifs by transcription factor TFIID within close proximity to the TSS. In contrast to the dynamic exchange between histones and TBP on TATA-box promoters, TBP and histones do not compete for DNA access on TATA-less promoters, as TFIID/TBP was shown to cooperatively assemble with the +1 nucleosome flanking the TSS. The +1 nucleosome appears to restrict scanning by RNA pol II. Therefore, although the nucleosome architecture on TATA-less promoters never occludes binding of TBP, histones are not completely evicted by TBP, which impedes RNA pol II scanning, and reduces the probability that every RNA pol II recruited by TBP will successfully initiate transcription. Therefore TATA-less housekeeping genes may favor a continuous mode of expression, where the promoter is activated frequently, but transcription initiation is inefficient.



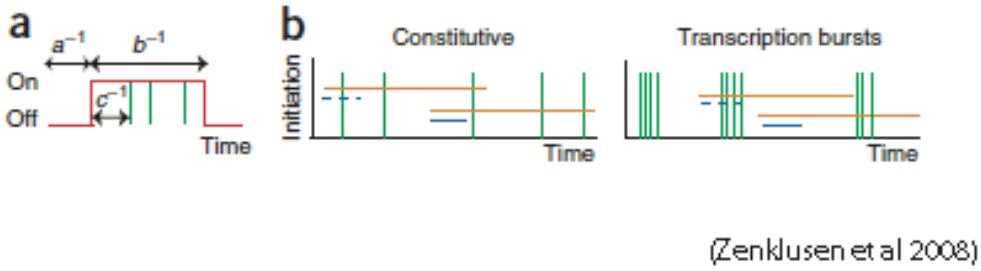
(Rhee and Pugh 2012)

Figure 1-1: TATA box vs TATA-less promoters

1.2.4 Single molecule studies of transcription kinetics

It was previously shown that TATA-less housekeeping genes and TATA box-containing genes are regulated by alternative modes of expression [27]. Using single molecule RNA fluorescent *in situ* hybridization (smFISH), Zenklusen and colleagues were able to quantify the absolute number of mRNAs per cell and monitor the transcriptional status of active alleles[27]. They showed that TATA-less housekeeping genes show little variability in mRNA numbers per cell and nascent transcripts per active allele. Conversely, the SAGA dependent *PDR5* gene showed high variability in number of mRNAs per cell and nascent transcripts per active allele. Simulations based on a stochastic mathematical model of gene activation and inactivation allowed them to distinguish between "bursting" and "constitutive" modes of expression (**see Figure 1-2**). The SAGA dependent TATA box-containing *PDR5* gene was expressed in a bursting mode, characterized by infrequent ON states that rapidly generate multiple transcripts, whereas the TATA-less genes were expressed in a "constitutive" mode from a promoter that is always ON, characterized by single uncoordinated transcription initiation events distributed in time. The bursting and constitutive modes of gene expression

might represent different biological strategies for survival. Whereas housekeeping genes would favor constitutive expression to maintain constant mRNA levels, genes that respond to fluctuating environmental conditions would favor transcriptional bursting in order to quickly reach the threshold of protein levels that is required to adapt to the novel stress. Interestingly, transcriptional bursting has also been described in mammalian cells using RNA FISH and stochastic modeling of gene activation and inactivation [26]. The rationale for bursting transcription may be very different in multicellular organisms vs unicellular organisms like yeast. While single celled organisms may utilize bursting to respond efficiently to an unpredictable environment, multicellular organisms may use bursting to achieve differential cellular function and behavior within relatively homogenous environments [26]. Bursting may also allow cells to rapidly respond to spatio-temporal cues during development to initiate cell fate specification. Altogether, single cell studies have shown that gene expression is not a deterministic process, but occurs stochastically due to the limited availability of molecules implicated in the transcription process, which results in variable expression levels in genetically identical populations of cells.



(Zenklusen et al 2008)

Figure 1-2: Constitutive vs bursting transcription

1.3 Transcription, termination, and degradation of non-coding RNAs

1.3.1 ncRNA transcription in *S. cerevisiae*

Similar to the tight regulation of transcription of protein-coding genes, non-coding RNA transcription appears to be extensively regulated, occurring from specific regulatory regions that were mapped in different eukaryotic genomes [30,31]. Interestingly, transcriptome studies in *S. cerevisiae,* which has a far more compact genome than mammalian cells, reveal a highly complex regulatory landscape, where transcription initiates bi-directionally from accessible chromatin in nucleosome depleted regions (NDRs) upstream and downstream of protein-coding genes [32,33] (**see Figure 1-3**). Distinct transcription pre-initiation complexes (PICs) bind within NDRs in opposite orientations, to initiate mRNA and divergent ncRNA transcription, respectively [29]. Since ncRNA transcription units often converge with protein-coding genes, mechanisms have evolved that induce premature transcription termination of ncRNAs to prevent transcriptional noise from interfering with gene expression.

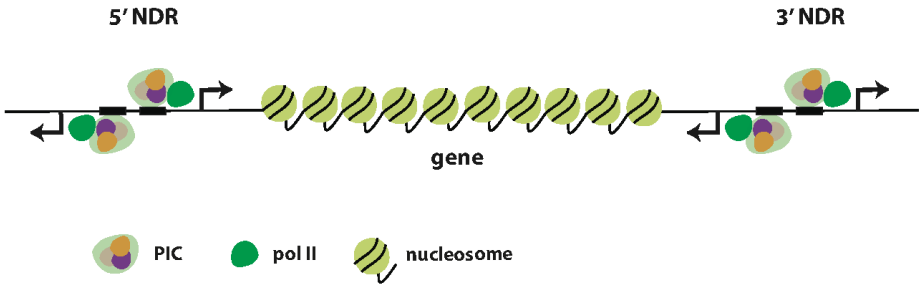


Figure 1-3: Bi-directional transcription from nucleosome-depleted regions (NDRs) in *S.cerevisiae*

1.3.1.1 Nrd1-Nab3-Sen1 mediated transcription termination of small ncRNAs

In *S. cerevisiae*, RNA pol II transcription termination occurs through at least two distinct mechanisms [30]. mRNA transcription is terminated by cleavage and polyadenylation mediated by the CPF/CF complex, whereas small noncoding RNAs such as sno/snRNAs are terminated by the Nrd1/Nab3/Sen1 complex **(see Figure 1-4)**. The Nrd1 complex recognizes sequence elements within sno/snRNAs and interacts with the RNA pol II c-terminal domain (CTD) to prevent transcriptional read-through of these noncoding RNAs into adjacent protein-coding genes by a polyadenylation independent mechanism [34]. Nrd1-Nab3-Sen1 and Pol II CTD mutants accumulate read-through snoRNA transcripts. Transcription termination is coupled to 3'end maturation of snoRNAs via cleavage by the endonuclease Rnt1, which is functionally analogous to 3'end cleavage of mRNAs by the CPF/CF complex. CPF/CF components were also shown to interact with the RNA pol II CTD, indicating that the termination of both mRNAs and sno/snRNAs requires the physical association of termination factors with the transcriptional machinery. However, whereas CPF/CF components interact with the elongating isoform of pol II phosphorylated at Ser2 in the CTD heptad repeat domain, the Nrd1 complex interacts with the transcription initiation isoform of RNA pol II phosphorylated at Ser5 [35]. Therefore the Nrd1 and CPF/CF complexes act at different stages during the transcription cycle of RNA pol II, whereby Nrd1 is recruited to nascent transcripts at an early step of transcription to inhibit elongation on Nrd1 regulated genes.

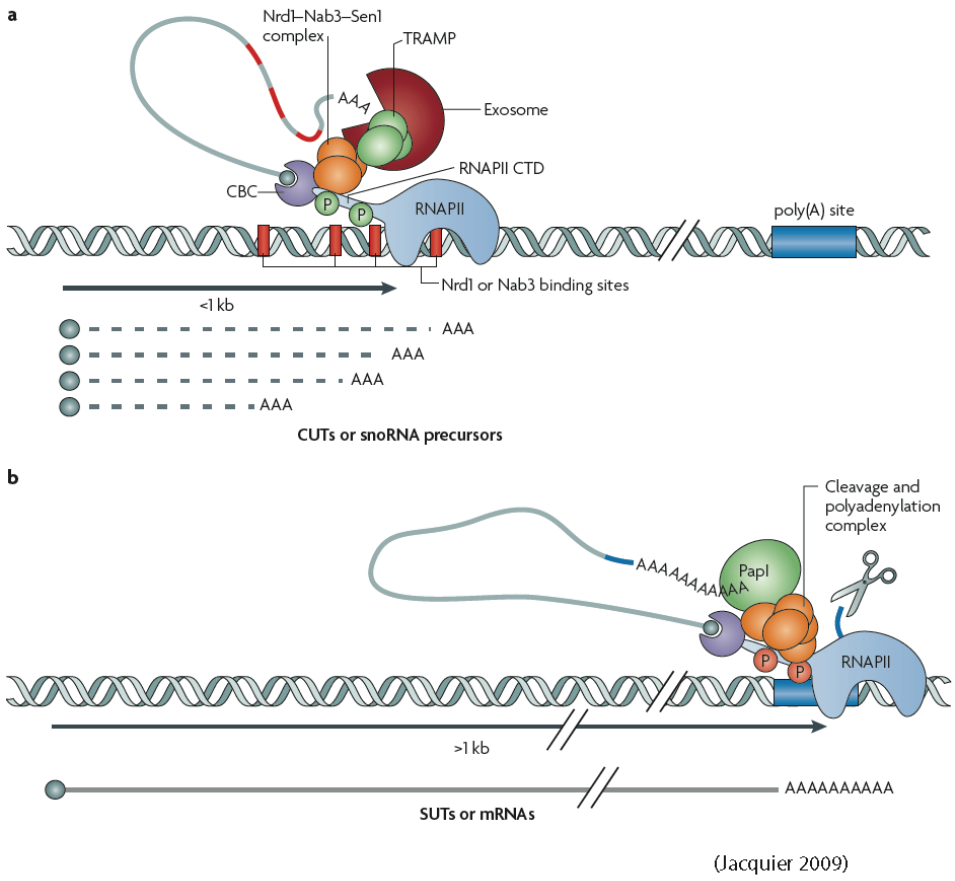


Figure 1-4: Nrd1-Nab3-Sen1 vs cleavage and polyadenylation mediated termination

Intriguingly, Nrd1 does not only act on ncRNAs, but has also been implicated in the regulation of a small set of protein coding genes [36]. The best studied example is the auto-regulation of its own mRNA through a cis element within its 5'UTR that is homologous to the Nrd1 responsive element in the snoRNA SNR13 [34]. In the temperature sensitive *nrd1*, *nab3,* and *sen3* mutants, *NRD1* mRNA levels are increased, suggesting that the Nrd1 complex can regulate the expression of its own transcript by modulating pol II elongation into the Nrd1 coding region. Nrd1 has also been implicated in the regulation of the *RPL9B* gene in response to a high copy number of the Rpl9 protein, encoded by *RPL9B* [37]. However, in contrast to Nrd1 mRNA regulation, *RPL9B* mRNA regulation occurs at the post-elongation stage. When Rpl9 is present in excess, it binds to an RNA hairpin motif in the 3'UTR near the polyA site,

which favours the use of downstream Nrd1 and Nab3 binding sites for alternative transcription termination by the Nrd1 complex to downregulate *RPL9B* mRNA levels.

Intriguingly, recent *in vivo* crosslinking approaches to identify Nrd1 and Nab3 binding sites revealed binding to many more mRNAs, identifying association to 20-30% of protein-coding transcripts in *S.cerevisiae* [36]. Most binding sites are localized at the 5'end of mRNAs, suggesting that they might mediate early termination at these genes. Furthermore, Nrd1 and Nab3 were also shown to bind throughout the coding region and 3'UTRs, suggesting that the Nrd1-Nab3-Sen1 complex might also be implicated in the alternative termination of mRNAs, possibly to downregulate their expression under certain conditions.

1.3.1.2 Coupling of ncRNA termination and degradation by Nrd1 and the nuclear exosome

In addition to promoting transcription termination and 3'end formation, the Nrd1 complex is also linked to RNA degradation by the nuclear exosome, an RNA surveillance complex that is implicated in the turnover of aberrant RNAs [38]. Proteomics analysis showed that Nrd1 is found in a complex with the nuclear exosome component Rrp6p [38]. Furthermore, *in vitro* RNA degradation assays using reporter constructs containing Nrd1 binding sites and purified RNA exosome from either a wild-type or *nrd1* mutant strains that lack an RNA-recognition motif (RRM) showed reduced exosome activity of the *nrd1* purified exosomes. Adding recombinant Nrd1 stimulated exosome-mediated degradation in both wild-type and mutant strains suggesting that Nrd1 recruits the exosome to its target substrates to induce RNA degradation. Consistent with such a model, mutagenesis of Nrd1 binding sites in the 5'UTR of *NRD1* mRNA made it more resistant to degradation suggesting that Nrd1 auto-regulates its own expression level by coupling transcription termination to exosome-mediated decay. Similarly, Gudipati et al., showed that the Nrd1 terminated *RPL9B* transcript was destabilized by the nuclear exosome [37]. However, the relationship between the RNA exosome and Nrd1 seems complex and not only linked to degradation. Studies from Vasiljeva and Buratowski demonstrated that the exosome also plays a role in transcription termination and 3'processing of snoRNAs, as both *nrd1* and Δ*rrp6* mutants show extended read-through transcription of

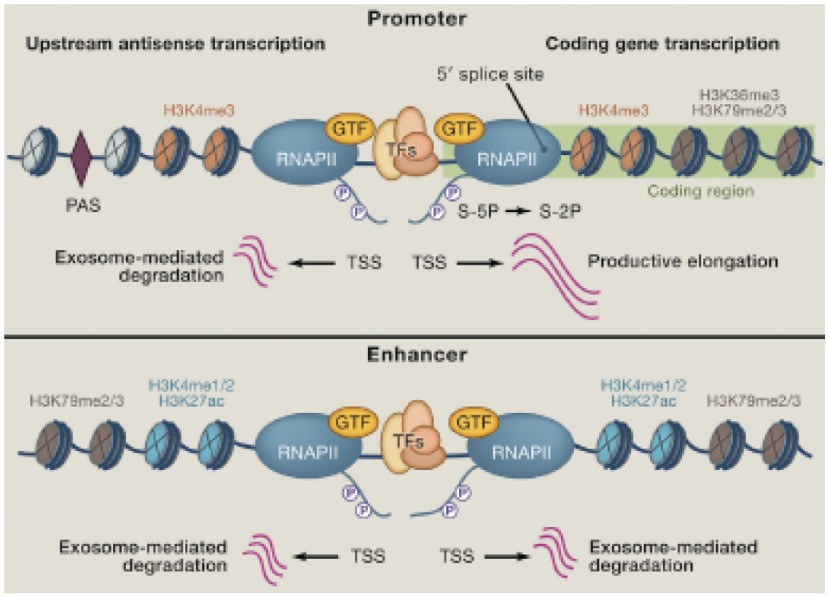
snoRNAs. In the context of 3'processing, Nrd1 may stimulate trimming of 3' UTR sequences by the nuclear exosome, without leading to complete 3'-5' degradation. Therefore, Nrd1 can recruit the exosome to nascent RNAs and target them towards a degradative or 3' processing pathway. Transcription termination by the Nrd1 complex has also been linked to the degradation of a class of non-coding transcripts in *S. cerevisiae* called cryptic unstable transcripts (CUTs) [39,40]. CUTs are transcribed from nucleosome depleted regions (NDRs) within intergenic and intragenic sequences both in sense and antisense to protein coding genes, and accumulate in strains depleted of the nuclear exosome component Rrp6p. [32,33]. As their transcriptional units often overlap with protein coding genes, mechanisms have evolved which prevent transcriptional read-through of these cryptic messages, which are subsequently targeted for rapid turnover. Furthermore, the degradation of CUTs was shown to be dependent on an alternative polyadenylation pathway involving the polyA polymerase Trf4 from the TRAMP complex [41]. Regulation of CUT transcription and turnover was extensively studied for the CUT *NEL025c*. *NEL025c* is present in two forms, a short form with heterogeneous 3'ends, and a long form. Both forms are only detectable in TRAMP/exosome mutants, with the short transcripts being vastly more abundant. While the long form remains polyadenylated in the Δ*trf4* mutant, the short form is non-polyadenylated and is produced through early transcription termination by the Nrd1 complex [39] . Mutagenesis of Nrd1 and Nab3 recognition sites and depletion of Nrd1p and Nab3p result in transcriptional read-through into the longer form, which is then polyadenylated by the canonical mRNA polyA polymerase Pap1p to produce a stable RNA that is not susceptible to exosome mediated degradation. Furthermore, the Nrd1 terminated short transcripts get targeted for degradation by the TRAMP/exosome complex. Similar results have been reported for other inter- and intragenic CUTs [40]. Together with previous proteomic analysis identifying components of the TRAMP complex to physically interact with Nrd1 and the exosome, these observations further emphasized that transcription termination of cryptic transcripts triggers their rapid turnover through a non-canonical polyadenylation pathway and the nuclear exosome [38].

Interestingly, there is a class of ncRNAs that is less susceptible to exosome mediated degradation than CUTs, referred to as stable unannotated transcripts (SUTs) [33]. SUTs are capped and polyadenlylated and shown to be destabilized by the cytoplasmic mRNA decay

machinery through decapping and degradation by the 5'-3' exonuclease Xrn1p [42,43]. Therefore, ncRNAs in *S. cerevisiae* can be processed by different regulatory pathways. This diversity in ncRNA processing mechanisms has been further expanded in higher eukaryotes where additional classes of ncRNAs are transcribed throughout the genome, in part from distal regulatory elements in intergenic regions.

1.3.2 ncRNA transcription from regulatory sequences in mammalian cells

In mammalian cells, RNA pol II and GTFs (general transcription factors) are not only recruited to promoters, but also to intergenic and intragenic regulatory sequences, often, but not always, corresponding to tissue-specific enhancers [44]. Enhancers and promoters share common transcription factor binding motifs and show similar patterns of bi-directional transcription, which initiates from the nucleosome boundaries of open chromatin regions [45,46] (**see Figure 1-5**). Bi-directionally transcribed regulatory regions express different types of RNAs with variable stabilities [47]. Canonical promoters typically produce bi-directionally or uni-directionally stable transcripts. These transcripts include protein coding mRNAs as well as long intergenic non-coding RNAs (lincRNAs). PROMPTs (promoter upstream transcripts) are unstable lncRNAs that are transcribed in antisense upstream of protein-coding genes, and are susceptible to early transcription termination [48–50]. Enhancers typically produce bi-directionally unstable enhancer RNAs whose short half-lives appears to be correlated with early termination similar to PROMPTs [45]. Interestingly, enhancers can occasionally produce unidirectional stable enhancer RNAs that are polyadenylated, which are often indistinguishable from lincRNAs. Unidirectional stable enhancer RNAs and lincRNAs are transcribed from loci that harbor the histone modifications H3K4me3/H3K36me3, which correspond respectively to the epigenetic signatures of transcription initiation and elongation that are also found on protein coding genes [51]. Therefore, despite the absence of open reading frames (ORFs), some lncRNA loci possess similar chromatin features as protein coding genes.



(Kim and Shiekhattar 2015)

Figure 1-5: Bi-directional transcription from enhancers and promoters

1.3.2.1 Transcription termination and degradation of PROMPTs and enhancer RNAs

How transcription of promoter and enhancer derived ncRNAs is terminated is still not fully understood. Some studies suggest that PROMPTs and enhancer RNAs may undergo premature transcription termination via similar mechanisms, mediated by the enrichment of early poly A signals which are absent in protein coding genes [45,47]. Protein coding genes are enriched in 5' splice sites close to the transcription start site that promote efficient splicing and productive Pol II elongation, which are followed by canonical polyA signals downstream that confer termination and are required for mRNA stability [52]. PROMPTs that are transcribed in antisense upstream of mRNA TSSs are enriched for polyA signals close to the TSS compared to their protein-coding counterparts [50,52]. Mutagenesis of these early polyA sites in PROMPTs induces transcriptional read-through and makes them less susceptible to nuclear exosome

mediated decay. Therefore, polyA signals proximal to PROMPT TSSs link PROMPT termination to degradation, which enforces promoter orientation towards protein coding genes, even though transcription initiates bi-directionally.

A specific complex called the nuclear exosome targeting complex (NEXT) was shown be implicated in the degradation of PROMPTs, targeting newly synthesized RNAs towards early degradation by the exosome [53]. Using iCLIP to map the genome-wide RNA targets of RBM7, the RNA binding component of NEXT, the authors discovered a correlation between RBM7 enrichment on eRNAs and PROMPTS and their sensitivity to exosome mediated degradation. Interestingly, PROMPTs were more efficiently stabilized than eRNAs in cells that were depleted of NEXT and exosome components. Depletion of CBC (cap binding complex) components decreased RBM7 binding on candidate PROMPTs, suggesting that NEXT may be recruited to newly synthesized PROMPTs by interacting with the cap structure, in particular since a physical link between the CBC and NEXT has been reported previously [54]. As eRNAs were also shown to be capped using CAGE-seq, their degradation may also be coupled to termination by the CBC-NEXT complex [55]. However, since RBM7 binds slightly less efficiently to eRNAs, as detected by iCLIP, and eRNAs are also less stabilized than PROMPTs upon exosome depletion, eRNA transcription termination and decay may require distinct factors that recognize specific motifs on nascent eRNAs, which are absent in PROMPTs.

Interestingly, the noncanonical polyadenylation pathway for CUT degradation in *S. cerevisiae* bears resemblance to the early polyA signals that link transcription termination of mammalian PROMPTs to their decay. Also, a study by Vasiljeva and Buratowski showed that the yeast cap binding proteins Cbp80p and Cbp20p co-purify with Nrd1p [38]. Therefore, the Nrd1-Nab3-Sen1 complex may play a role similar to the NEXT complex in mammalian cells, which was shown to physically link the cap structure with the exosome to target newly synthesized noncoding transcripts such as PROMPTs for rapid decay. Therefore, the mechanistic coupling of transcription termination and RNA degradation may have evolved early on to regulate both the size and abundance of ncRNAs, possibly to ensure that they do not intervene in the normal gene expression program. Nonetheless, even though many ncRNAs might be by-products of transcriptional noise, in part due to random collisions of

RNA pol II with accessible chromatin, others have been co-opted by natural selection to perform regulatory functions.

1.4 Evolution and functions of lncRNAs

In contrast to protein-coding mRNAs, the overwhelming majority of long non-coding RNAs are expressed at low levels; with many of them expressed at a single copy per cell [2]. Also, whereas many protein-coding mRNAs are constitutively expressed across cell types, lncRNAs show higher variability in expression and are often expressed in a cell type specific manner [2,3]. Cellular fractionation assays also showed that lncRNAs are predominantly enriched in the nuclear fraction, showing an inverse localization pattern to that of mRNAs, which are mainly cytoplasmic. This was confirmed by a recent imaging study that used single molecule FISH to investigate lncRNA localization and expression at the single molecule level, monitoring expression of different lncRNAs in different cell lines. This study revealed that lncRNAs are heterogeneous in their cellular localization, describing different localization patterns including lncRNAs that assemble into sub-nuclear foci, as well as predominantly cytoplasmic localization similar to mRNAs [56]. These observations suggest that lncRNAs function at different stages of gene regulation.

1.4.1 Evolutionary conservation and structural features of lncRNAs

Interestingly, long intergenic noncoding RNAs (lincRNAs) are often capped, spliced, and polyadenylated, showing similar transcriptional processing mechanisms as mRNAs. Largely based on these observations, one model proposes an evolutionary explanation of the origin of lincRNAs whereby many lincRNA genes may have started as protein-coding genes, but have gradually had their open reading frames (ORFs) eroded through mutations, and were later co-opted for non-coding functions [7]. Conversely, many lincRNA genes may serve as raw substrates for natural selection to evolve into *de novo* protein coding sequences. There is little evidence for evolutionary conservation of most lncRNAs, either at the level of primary sequence or secondary structure, which makes them a highly diverse class of RNAs with potentially heterogeneous functions, as opposed to highly conserved small non coding RNAs,

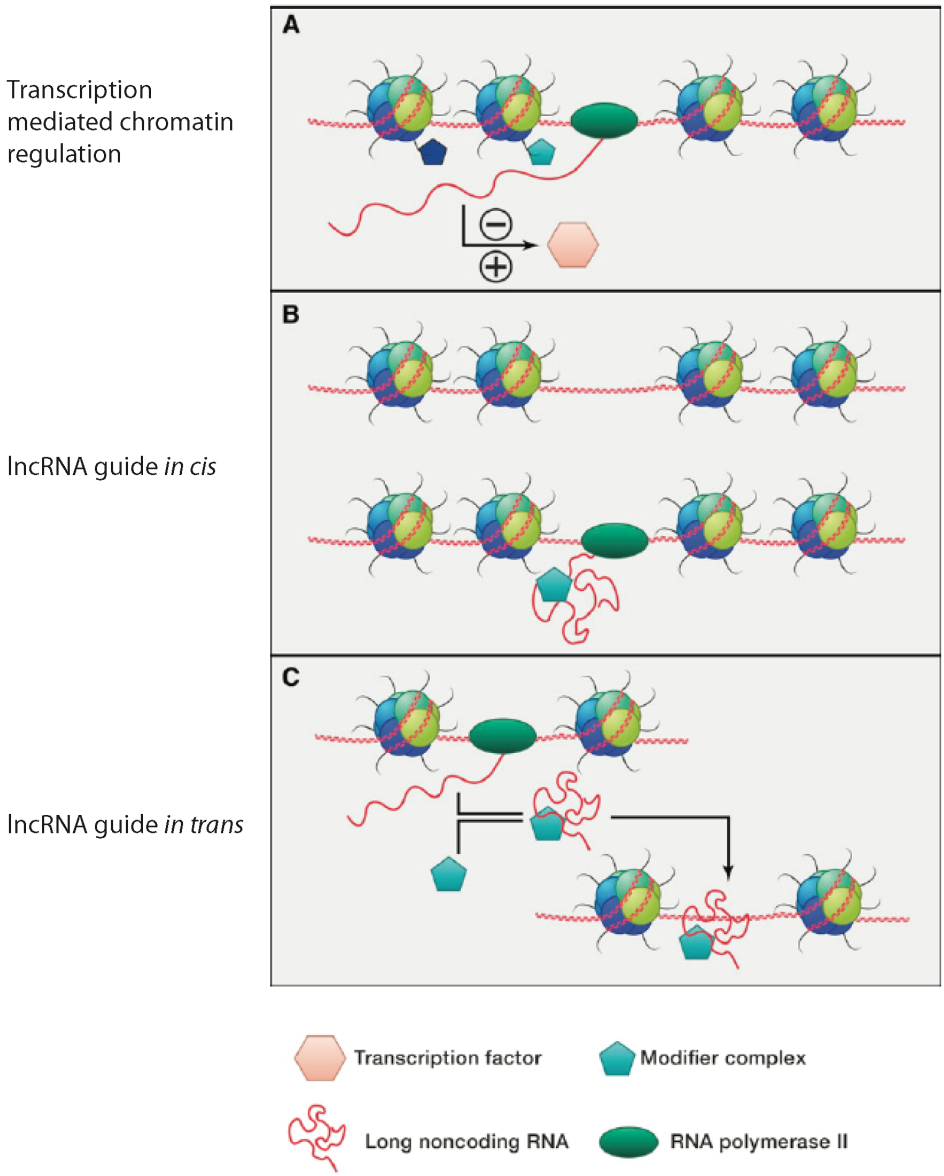
such as microRNAs, sn/snoRNAs, rRNAs and tRNAs, whose functions are narrowly defined [7].

One could postulate that lncRNAs should theoretically possess more conserved secondary structures than mRNAs, as they are predicted to have specific catalytic roles, whereas mRNAs merely encode information, which only requires the conservation of primary sequence. Surprisingly, RNA folding computational algorithms predict that lncRNAs would possess similar fractions of base-paired nucleotides as mRNAs, suggesting that lncRNAs are not more prone to form secondary structure than mRNAs [57]. In the absence of primary sequence conservation or secondary structure, it is also possible that the process of lncRNA transcription itself has a function in gene regulation, whereas the RNA molecule itself is inconsequential. While estimates from the ENCODE consortium reveal that ~75% of the human genome is transcribed, a smaller fraction of 62% shows evidence of processed transcripts. Therefore, there is a substantial fraction of the genome that shows transcriptional activity, but no clear evidence of post-transcriptional processing. In such cases, the process of transcription on noncoding regions may induce local changes to chromatin to modulate gene expression. This mechanism of regulation *in cis* might increase the cell's flexibility in responding to stimuli as it circumvents the time-consuming process of translating and importing transcription factors, which then need to scan the genome for their target regulatory sequences [58].

Despite the paucity of evolutionarily conserved consensus sequences, there is strong evidence that at least some lncRNAs are functional in a transcription independent manner, and that their functions are conserved across species. Most well-characterized lncRNAs are implicated in the regulation of transcription by recruiting chromatin modifying enzymes to target genomic loci, or in the organization of sub-nuclear domains that function in mRNA processing [7,58–60]. Furthermore, there are examples of long noncoding RNAs that are exported to the cytoplasm where they regulate the stability or translational efficiency of target mRNAs [61–64]. These examples show that in addition to the high diversity of ncRNA sequence, there might be many functional classes, whose common mechanistic principles still have to be determined.

1.4.2 Model LncRNAs and functional principles

In addition to the process of noncoding transcription regulating local changes to chromatin, lncRNAs can also act as guides to recruit chromatin modifying factors both *in cis* and *in trans* (**see Figure 1-6).** There are a few well-studied lncRNAs whose characterization has been particularly useful in establishing general models of lncRNA function. Canonical examples include *Xist*, which is implicated in mammalian X-chromosome inactivation, and *HOTAIR*, a lncRNA expressed from the *HOXC* locus, which is involved in anterior-posterior axis specification during development. *Xist* and HOTAIR act as guides that recruit the polycomb repressor complex PRC2, which deposits the repressive tri-methylation mark on histone H3K27 (H3K27me3) to inhibit transcription. Whereas *Xist* functions in *cis*, spreading along the entire X chromosome, HOTAIR can act in *trans*, as it was shown to recruit PRC2 to loci on multiple chromosomes [65]. Whether these lncRNAs act in *cis* or in *trans*, their essential role is to specify the genomic address code of chromatin modifying enzymes, which then regulate transcription of the target sequence.



(Batista et al 2013)

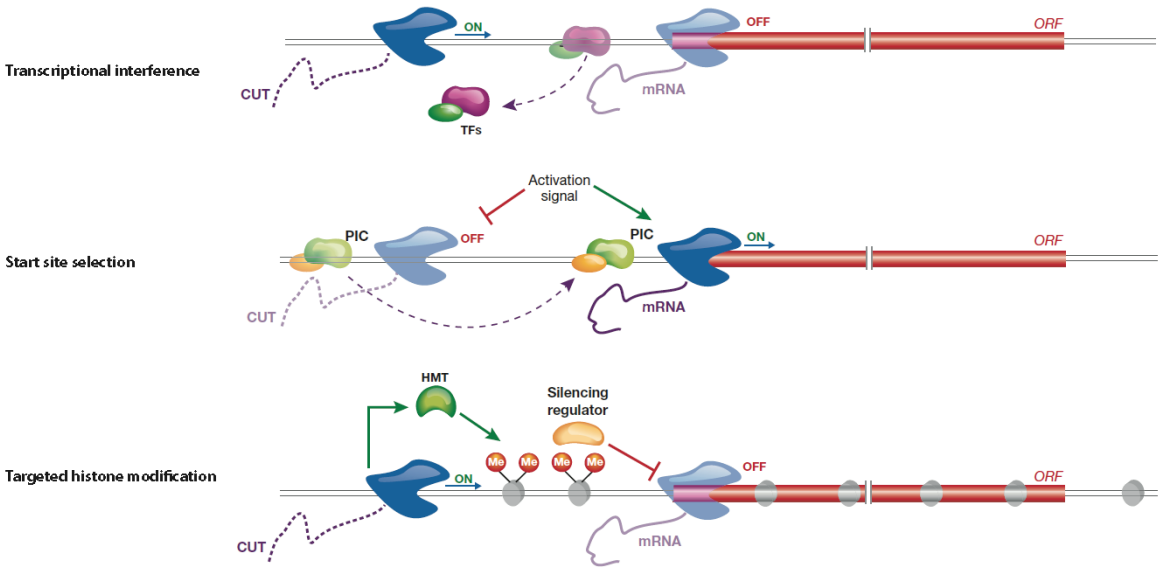
Figure 1-6: General principles of lncRNA function

Many other PRC2 recruiting lncRNAs have been identified in diverse cell types, showing that targeted epigenetic repression of transcription by lncRNAs is a common

regulatory mechanism [66,67]. LncRNAs may also act in a modular fashion, whereby distinct motifs within the RNA bind to different factors to regulate multiple enzymatic processes on chromatin [58]. For example, HOTAIR also recruits the LSD1 complex to demethylate H3K4, thereby removing an active histone mark while simultaneously depositing the repressive H3K27me3 mark via PRC2 [68]. Therefore, lncRNAs can potentially act as modular scaffolds to concentrate multiple factors on target genomic loci, which have distinct, but complementary roles. Other well characterized lncRNAs that might act in such a modular fashion are MALAT1 and NEAT1, which organize the structure of speckles and paraspeckles, respectively [69–71]. Speckles are sub-nuclear domains that contain mRNA splicing factors, whereas paraspeckles contain factors that are involved in mRNA editing. Interestingly, although MALAT1 is not well conserved at the primary sequence level, it was shown that other functional properties may be preserved in highly disparate eukaryotic organisms. For example, MALAT1 in zebrafish shares many features with its mammalian counterpart, such as high expression, lack of introns, a similar size of 7 kb, and a shared 3'end sequence, despite overall lack of sequence homology [7]. Interestingly, *Xist* shows very little sequence conservation among mammals, with a few ancient conserved motifs interspersed with species-specific sequences [72]. These findings reinforce the need to consider features other than primary sequence conservation to better elucidate lncRNA function. As imaging analyses and high-throughput sequencing studies have characterized a vast repertoire of lncRNAs that are enriched in the nucleus, there are many candidates awaiting further investigation, which may utilize the functional principles outlined above to regulate eukaryotic transcription or novel ones that remain to be discovered.

1.5 Mechanistic studies of ncRNA mediated transcription regulation in *S.cerevisiae*

The extensive convergence of ncRNA transcription with protein-coding genes in *S.cerevisiae* has stimulated investigation into the regulatory function of ncRNAs [32,33]. Various single locus studies have revealed distinct modes of gene regulation by ncRNAs, including start site selection dependent regulation, transcriptional interference, and targeted histone modification [30,31,73] (**see Figure 1-7**).



(Beretta and Morillon 2009)

Figure 1-7: Different mechanisms of transcription regulation by ncRNAs in *S.cerevisiae*

1.5.1 Start site selection dependent regulation

The mechanism of start site selection dependent regulation does not implicate the ncRNA directly in gene regulation, as the ncRNA is simply a by-product of a pre-initiation complex that is recruited to an alternate start site upstream of the mRNA start site under repressing conditions. Kuehner and Brow described such a mechanism in the regulation of *IMD2,* a gene that encodes inosine monophosphate dehydrogenase (IMPDH), which catalyzes the first step in GTP synthesis[74]. When GTP concentration is sufficient, RNA pol II initiates transcription from a start site upstream of the *IMD2* start site. Alternately, when GTP is deficient, RNA pol II initiates transcription from the *IMD2* start site. In this specific context, RNA pol II appears to act directly as a metabolic sensor, sensing GTP levels and responding appropriately to regulate *IMD2*. The ncRNA produced from the upstream site is a cryptic transcript, which itself has no function, and is terminated before the *IMD2* mRNA start site. This mechanism differs markedly from transcriptional interference, in which the process of ncRNA transcription directly obstructs activation from the mRNA TSS.

1.5.2 Transcriptional interference

Another mode of regulation is by 'transcriptional interference', whereby ncRNA transcription can impede transcription factor access to promoters, as described for *FLO11* [75]. *FLO11* is regulated by an intricate toggle switch between two ncRNAs, *ICR1* and *PWR1*, which have opposing roles in the activation of the gene. *ICR1* is transcribed upstream of *FLO11*, activated by the transcription factor Sfl1. *ICR1* transcription across the *FLO11* promoter blocks its access to transcriptional activators, thereby repressing *FLO11*. *PWR1* is transcribed upstream of *FLO11* as well, but in antisense to *ICR1*. *PWR1* is activated by the transcription factor Flo8, and interferes with the transcription of *ICR1*. It induces histone deacetylation by Rpd3L, preventing binding of Sfl1, and thereby inhibits *ICR1* transcription. This relieves promoter occlusion of *FLO11*, reverting it to its transcriptionally active state. Transcriptional interference was also initially suggested for the regulation of *SER3* by the upstream noncoding transcript *SRG1* [76,77]. *SER3* encodes a phosphoglycerate dehydrogenase

that catalyzes a step in serine biosynthesis. When serine is abundant, the gene is repressed by the upstream noncoding RNA *SRG1*. *SRG1* initiates from its own TATA box motif, and was shown to be activated by the transcription factor Cha4 which recruits the SAGA and Swi/Snf chromatin remodelers in response to high serine levels. It was proposed that *SRG1* transcription across the *SER3* promoter inhibits binding of transcriptional activators, therefore repressing *SER3*. The model of *SER3* regulation was further refined in a subsequent study which showed that transcriptional interference did not play a direct role in *SER3* repression [78]. It was demonstrated that *SRG1* transcription induces nucleosome assembly across the *SER3* promoter, which maintains a repressive chromatin structure.

1.5.3 ncRNA mediated gene activation through nucleosome remodeling

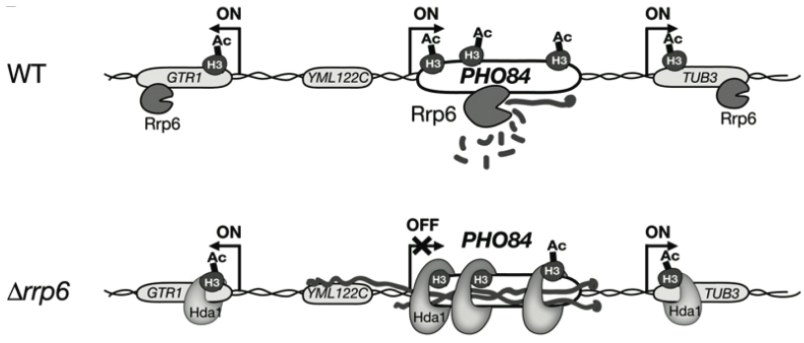
While non-coding RNA transcription may repress gene expression by inducing nucleosome assembly on the promoter, it may also activate gene expression by evicting nucleosomes from a promoter. Uhler et. al described such a mechanism of regulation for *PHO5*, showing that an antisense ncRNA transcribed from the 3' end of the *PHO5* gene evicts nucleosomes around the *PHO5* promoter, increasing accessibility to *PHO5* transcriptional activators, thereby enhancing *PHO5* activation [79]. However, anti-sense transcription mediated gene activation, as shown for *PHO5*, may be a rare phenomenon, as most genes that express AS RNAs are transcriptionally repressed [32,33].

1.5.4 Antisense RNA mediated gene repression through targeted histone modifications

Gene repression by anti-sense transcription has been described for the *GAL10* and *PHO84* loci. The *GAL10* gene is involved in galactose metabolism and is strongly induced by galactose, but repressed by glucose. Houseley and colleagues showed that glucose repression is mediated by an antisense lncRNA transcribed from the 3' end of the *GAL10* gene[80]. The *GAL10* antisense RNA recruits the histone methyltransferase Set2, which methylates lysine 36 on histone H3 across the *GAL10* locus, which then induces histone de-acetylation by Rpd3S, thereby repressing *GAL10* when cells are grown in glucose. The *GAL10* antisense ncRNA is activated by the transcription factor Reb1 which binds to the 3' end of *GAL10*.

The *PHO84* gene, which encodes a high affinity phosphate transporter, was shown to be repressed by an antisense lncRNA initiated from the *PHO84* 3' UTR [81]. This antisense transcript was expressed at very low steady state levels in a wild type strain. Deletion of the RNA exosome component *RRP6* results in increased antisense expression, which represses *PHO84* mRNA expression. Repression requires targeted histone de-acetylation at the *PHO84* promoter by Hda1, which was proposed to be recruited by *PHO84* antisense RNAs that accumulate at the *PHO84* locus upon stabilization after Rrp6 deletion (**see Figure 1-8**). However, many mechanistic details of this process were still unknown, including how Rrp6 participates in this process. The first part of my PhD thesis focused on using single cell and single molecule assays, in combination with biochemical and genetic tools to further dissect the mechanism of antisense RNA mediated regulation of *PHO84*.

The second part of my PhD focused on a different class of lncRNAs called enhancer RNAs (eRNAs), transcribed from mammalian regulatory elements called enhancers, as briefly described in a previous section. Specifically, I focused on eRNAs implicated in estrogen (E2) regulated transcription in breast cancer cells. The following sections will describe the regulation of the E2 signaling pathway, as well as general features of enhancers, chromatin topology in higher eukaryotes, and current models of eRNA function.



(Camblong et al 2007)

Figure 1-8: *PHO84* AS RNA represses *PHO84* by recruiting the histone de-acetylase Hda1

1.6 ERα mediated transcription regulation in breast cancer cells

1.6.1 Mechanism of action of ER alpha

Estrogen (E2) plays an important role in the normal growth of breast epithelial tissue as well as in breast cancer progression. Its mechanism of action is mediated primarily by the nuclear estrogen receptor ERα, which binds to cognate estrogen response element (ERE) sequences as a homodimer, and recruits co-factors to activate or repress transcription in response to E2 [82,83]. Genome-wide analyses on ERα and RNA pol II occupancy in response to E2 have shown that RNA pol II is recruited to promoter proximal regions of target genes, while ERα is primarily recruited to one or more distal cis regulatory sequences [84,85]. Interestingly, while ERα binds directly to EREs in the context of E2 activated genes, it is mostly recruited indirectly by other co-factors such as AP-1 in the context of E2 repressed genes. Many ERα binding sites coincide with Forkhead motifs recognized by the pioneer factor FOXA1, which remodels chromatin and recruits ERα upon E2 treatment. It was previously shown that knockdown of FOXA1 inhibits ERα recruitment and target gene expression [86]. The anti-estrogen compounds Tamoxifen and Fulvestrant (ICI 182 780) used in breast cancer therapy, which inhibit co-activator recruitment and ERα binding, respectively, repress gene expression by reducing the efficiency of RNA pol II recruitment on promoters [85]. Therefore, the E2 responsive gene expression program mediated by ERα is modulated primarily at the level of transcription initiation.

1.6.2 The kinetics of E2 induction

ERα was shown to promote a rapid, extensive, and transient transcriptional response upon E2 treatment, activating protein-coding as well as non-coding regions [87]. In this study, Hah and colleagues used GRO-seq to directly monitor the transcriptional activity of RNA polymerases at high resolution during early time points in order to measure the immediate cellular response to E2. In addition to demonstrating extensive upregulation of different classes of coding and non-coding transcripts synthesized by all three RNA polymerases, they show that the transcriptional response is rapid and transient, with many genes showing peak

transcription at 10 min or 40 min during an E2 time course, and returning to basal expression levels after 160 min. Interestingly, approximately half of the genes that show peak transcriptional activity at 10 or 40 min have TSSs for mRNA coding genes located within less than 10 kb from the ERα binding site, suggesting that transcribed regions proximal to ERα binding sites tend to show an immediate early response. Using GRO-seq also allowed them to measure RNA pol II dynamics during the E2 time course. They showed an increase in RNA pol II loading on TSSs, which is indicative of RNA polymerases being recruited at a faster rate than they are being released into the body of the gene. Therefore, the immediate regulatory effect of ERα binding is on the rate of transcription initiation by RNA pol II, rather than the rate at which RNA pol II transitions into an elongation competent isoform. These results corroborate with the effects of anti-estrogen compounds on RNA pol II recruitment to promoters.

1.6.3 Epigenetic modification of ER alpha responsive enhancers

ERα-bound distal *cis* regulatory sequences are mostly intergenic enhancer elements that are epigenetically marked by H3K4me1 and H3K27ac [88]. The deposition of H3K4me1, which is a canonical feature of active enhancers, is mediated primarily by the MLL family of histone methyltransferases. There are six well-characterized MLL histone methyl transferases: Set1A, Set1B, and MLL1-4 [89,90]. Notably, MLLs 1-4 have been implicated in depositing H3K4me1/2 on enhancers [91,92]. Depletion of MLL4 results in a significant decrease in H3K27ac levels on enhancers, as well as diminished recruitment of RNA Pol II and mediator, a protein complex that stabilizes interactions between transcription factors and RNA pol II[92]. These results establish MLL H3K4 methyltransferases as key players in enhancer activation. However, only MLL1/2 have been shown to interact with ERα by co-IP and to regulate estrogen-responsive gene expression programs [93,94]. These interactions may occur directly through a nuclear receptor binding domain within MLL1/2 or they could be mediated by other co-factors, such as Menin [93,94].

1.6.4 The role of MLL1 in enhancer activation

It was previously shown that the ERα co-factor FOXA1 requires the presence of H3K4me1 to bind to its target enhancers [95]. Recent studies have demonstrated that MLL1 is required for the deposition of H3K4me1 on the TFF1 enhancer and that depletion of MLL1 inhibits the recruitment of FOXA1 and ER alpha, thereby repressing target gene expression [96,97]. The mechanism by which MLL1 is recruited to target enhancers in response to estrogen treatment is currently unknown. It was shown that the CpG binding protein CGBP binds to MLL1 and that knockdown of CGBP leads to a decrease in the level of MLL1 recruited to the HOXA7 promoter as well as a decrease in the level of H3K4 methylation and gene expression [98]. Additionally, MLL1 was shown to bind directly to a CpG-rich region within the TFF1 enhancer [97] . It was demonstrated by an *in vitro* binding assay that MLL1 binds to the CpG rich region of the TFF1 enhancer when it is unmethylated, as methylation by the CpG specific DNA methyltransferase M.Sss1 inhibits MLL1 recruitment (Jeong et al 2014). Interestingly, it was shown that cell type specific CpG rich glucocorticoid receptor (GR) bound enhancers are made partially accessible by CpG demethylation, which might "pre-program" these enhancers for activity [99]. Therefore, CpG de-methylation of ERα target enhancers may represent the initial step in creating a partially accessible enhancer region to recruit MLL1, which then stimulates the recruitment of downstream factors that further remodel chromatin to activate transcription.

1.7 Enhancer function, chromatin topology, and transcriptional dynamics

1.7.1 Features of enhancers

Cellular differentiation during eukaryotic development requires precisely orchestrated spatiotemporal patterns of gene expression [100,101]. This regulation is mediated by enhancers which respond to tissue-specific signaling pathways. Typically located in non-coding intergenic DNA, enhancers contain binding sites for transcription factors, which they can bring into proximity towards gene promoters by a DNA looping process. How enhancers achieve this precise looping interaction with promoters has not been clearly elucidated, but

specific protein complexes such as cohesin and mediator are known to stabilize these long range chromatin interactions and facilitate RNA pol II recruitment to activate gene transcription in a cell type specific manner [101,102]. Tissue-specific enhancer-promoter interactions mediated by cohesin have been extensively mapped throughout the genome [103]. Enhancers also possess an open chromatin conformation which increases accessibility to transcription factors and RNA pol II. Genome wide analyses of epigenetic modifications have also shown that active enhancers are associated with distinct chromatin marks. Active enhancers are acetylated on H3K27 (H3K27ac) and mono-methylated on H3K4 (H3K4me1), which distinguishes them from active promoters, which are enriched in H3K4 trimethylation (H3K4me3). Enhancers are also bound by the transcriptional co-activators p300 and CREB binding protein (CBP), which are histone acetyltransferases, and are able to recruit RNA pol II. These features have been instrumental in the prediction and discovery of novel enhancer sequences throughout the genome.

1.7.2 Topologically associated domains

DNA is not randomly distributed in the nucleus. DNA FISH has revealed that individual chromosomes in G1 are not spread throughout the nucleus but preferentially occupy particular regions called chromosome territories [104]. Furthermore, more recent studies show that mammalian genomic DNA is further compacted into 'chromosome neighborhoods' called topologically associated domains (TADs), defined by regions of DNA within which physical interactions occur relatively frequently. TADs are largely cell type invariant, which confine enhancers and target promoters within higher order chromatin structures in the nucleus [105,106] (**see Figure 1-9**). These domains are separated by boundary elements which are bound by CTCF, a protein that plays a crucial role in establishing the architecture of the genome. A study by Lupiáñez and colleagues demonstrated that the boundary elements that segregate individual TADs are indispensable for normal gene expression regulation and development in humans and mice [107]. They investigated large scale genomic deletions and inversions overlapping two adjacent TADs that cause significant aberrations in human limb development. They showed that disrupting the boundary element between the two independent TADs caused pathological re-wiring of enhancer-promoter interactions. Enhancers from one TAD interacted inappropriately with promoters from the adjacent TAD inducing aberrant expression of those

genes. Altogether, this study suggests that TADs are stable structures conserved across species, and play a fundamental role in preventing aberrant enhancer-promoter interactions.

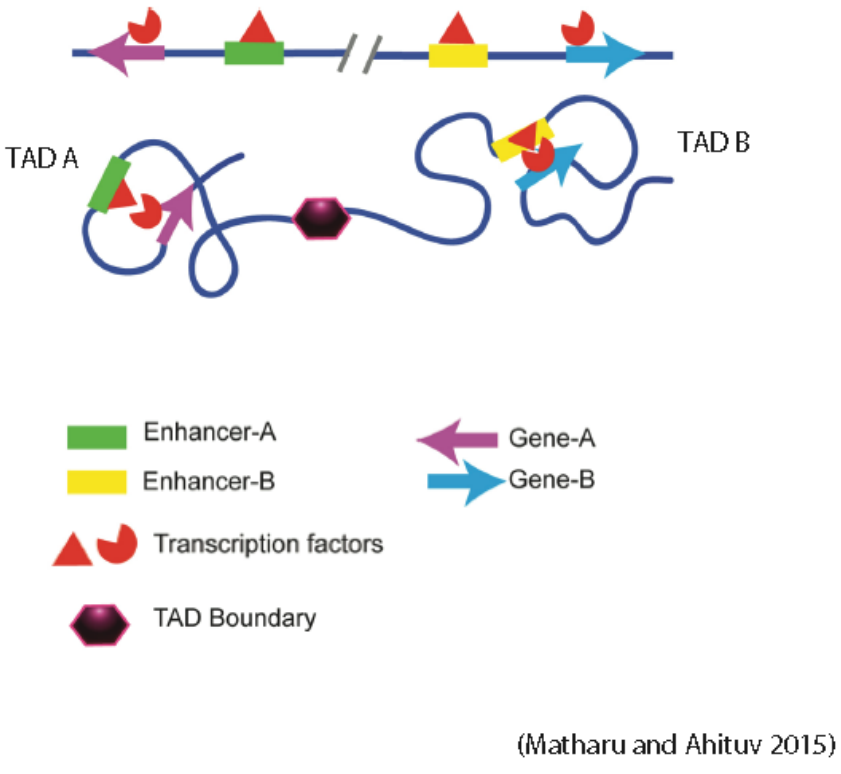


Figure 1-9: Topologically associated domains (TADs) regulate the specificity of enhancer-promoter interactions

1.7.3 Signal dependent enhancer-promoter looping interactions

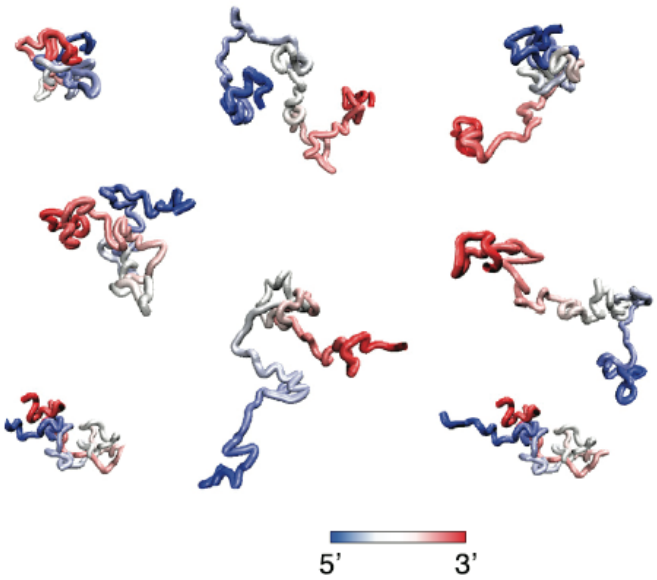
Enhancer-promoter interactions can be induced by transcription factors in response to specific signals. For example, it has been shown using different chromosome conformation capture assays (ChIA-PET, 3C, etc) that ER alpha maintains long range chromatin interactions between distal regulatory elements and promoters [108,109]. Fullwood et al demonstrated that most ERα-bound chromatin interactions are intra-chromosomal and occur within a genomic distance of 100 kb [109]. Interestingly, complex interaction hubs, where multiple enhancer-

promoter loops cluster together, show strong enrichment of RNA pol II on target promoters and high occupancy of the ERα co-factor FOXA1 on ERα-bound enhancers. Therefore, multiple genes that are involved in the same regulatory pathway may be activated within higher-order chromatin structures assembled by transcription factors that sequester regulatory elements and transcriptional machinery within sub-nuclear domains. Interestingly, a study by Ghavi-helm et al on developmental enhancers in *Drosophila* using 4C-seq showed a 3D chromatin topology analogous to that in humans, characterized by similar enhancer-to-promoter contact ratios [110]. Therefore, although the *Drosophila* genome is compact, with a noncoding fraction that is an order of magnitude smaller than that in humans, it shows similar levels of long range chromatin connectivity, suggesting that higher-order chromatin organization may have evolved early during eukaryotic evolution to ensure functional interactions between regulatory elements and genes and proper spatio-temporal control of gene expression. Interestingly, they also show that enhancer-promoter loops are pre-formed and remain relatively stable during development prior to gene activation and that RNA pol II remains associated with promoters in a paused state. Therefore, at least some stable long range chromatin interactions between enhancers and promoters may poise promoters for rapid activation in response to the appropriate developmental signaling pathways, which will then trigger the release of paused polymerases to activate transcription.

1.7.4 Dynamic changes in chromatin conformation

Long range genomic interaction profiles determined using chromatin conformation assays represent average measurements from cell populations, which do not necessarily imply stable loops, as chromatin fibers may adopt variable conformations in individual cells rather than forming static structures. Instead of forming a stable loop, an enhancer may interact transiently with its target promoter in response to a regulatory signal. Interestingly, simulation of intra-TAD chromatin fiber configurations using predictive polymer modeling suggests that chromatin fibers adopt various conformations, resulting in a broad range of distances between two loci on the same fiber, which was validated by 3D DNA FISH [111] (**see Figure 1-10**). Their modeling predicts that when two loci on an individual chromatin fiber are closer than 80 nm, they have the possibility to physically interact. Furthermore, using RNA FISH combined with

DNA FISH and super-resolution microscopy, they showed that an antisense RNA called Tsix, which represses Xist expression, is more strongly transcribed on the allele that physically interacts with the Tsix regulatory elements due to a more compact chromatin conformation. Therefore, transcription is regulated by the dynamic intra-TAD conformational changes that modulate interactions between regulatory sequences and their target promoters. Similarly, 3D DNA FISH showed that in the developing mouse embryo the Shh gene co-localizes with its enhancer ZRS within 200 nm specifically in tissues where the gene is expressed [112]. However, even in tissues where the gene is not expressed, Shh and ZRS reside within the same TAD and are proximal, with a median distance of 220-345 nm between Shh and ZRS in most tissues and developmental stages analyzed. Thus, TADs may harbor flexible chromatin fiber conformations that facilitate enhancer-promoter contacts to allow efficient activation of target gene transcription in response to developmental signals.



(Giorgetti et al 2014)

Figure 1-10: Variable chromatin fiber conformations within TADs

1.7.5 Transcriptional bursting

Although chromatin is packaged into stable TADs that maintain regulatory sequences in proximity to their target promoters, this does not necessarily ensure that gene expression will be a deterministic process. Previous work has shown that even though protein levels are stabilized in a genetically identical population of mammalian cells, transcription activation on individual alleles occurs largely at random, mostly due to intrinsic noise in the activity of the promoter, which stochastically switches between an active and an inactive state [26]. This results in discontinuous "bursting" transcription characterized by an ON state when RNA polymerases initiate multiple rounds of transcription in rapid succession, followed by a refractory OFF state, when transcription is completely abolished. Using RNA FISH, Raj and colleagues demonstrated that stochastic gene activation results in highly variable expression levels within a population of cells, whereby a fraction of cells accumulate a disproportionately high number of mRNAs due to the presence of "bursting" alleles with multiple nascent transcripts. Intriguingly, although modulating the concentrations of transcription factors or the number of transcription factor binding sites affected the burst size, i.e, number of nascent transcripts generated on an allele during an individual burst, the frequency at which the promoter was activated remained constant. Therefore, the frequency at which a promoter "bursts" is primarily due to intrinsic features of the promoter, such as the rate of chromatin remodeling, rather than extrinsic noise arising from the variable concentrations of transcription factors or the availability of RNA pol II. Interestingly, they also demonstrated that the variability in mRNA expression of *rpb1,* the largest subunit of RNA pol II, did not correlate with the variability in expression of the reporter genes under investigation. Therefore, although *rpb1* is transcribed in bursts, these bursts do not correlate with those of the target genes which depend on RNA pol II for transcription. Although it seems counter-intuitive that an essential gene such as *rpb1* would be activated stochastically, slow protein degradation would maintain similar RNA pol II levels within a cell population, as RNA polymerases produced from new "bursts" would contribute to the pool of polymerases available from previous "bursts".

1.7.6 Enhancer-promoter interactions and transcriptional bursting

Recent studies have shown that enhancers regulate gene expression levels by controlling the bursting frequency of their target promoters [113,114]. In a study by Fukaya et al, a reporter gene linked to different developmental enhancers was used to monitor transcriptional activity in live *Drosophila* embryos during nuclear cycle 14, the 1 hour interval prior to the onset of gastrulation when the fate map of the adult fly is established by the localized expression of patterning genes[113]. The reporter gene was tagged with a series of MS2 stem loops, which allowed the visualization of nascent RNAs in live cells using the MCP-GFP fusion proteins and the trajectory of the MS2 signal was recorded to measure different parameters of transcription. These measurements showed that the total transcriptional output of the reporter varies significantly when linked to different enhancers, and that this variability is correlated with burst frequency, but not the amplitude or duration of an individual burst, which remain largely invariant. Interestingly, the spatially restricted expression of patterning genes is also modulated by bursting frequency, as cells where the genes are active show a higher frequency of bursts. Furthermore, they demonstrated that introducing an insulator DNA sequence which selectively blocks interaction between an enhancer and a target promoter does not completely abolish reporter expression but reduces the frequency of bursts. These data suggest that chromatin topology is not static, and that an enhancer can dynamically re-position itself in proximity to its target promoter to activate transcriptional bursts. The insulator DNA may reduce bursting frequency by decreasing the probability of enhancer-promoter contacts.

Bartman and colleagues further showed that the role of enhancer-promoter contact is to modulate bursting frequency of the target locus [114]. They investigated the role of the β-globin enhancer in regulating the bursting parameters of the target β-globin gene during murine and human erythroid maturation. Using single molecule RNA FISH, they found that both the burst fraction, corresponding to the number of transcriptionally active alleles per cell, and burst size, corresponding to the number of nascent RNAs per allele, increased during erythroid maturation. However, the increase in burst size was more modest than the increase in burst fraction, implying that the β-globin enhancer may primarily regulate frequency of promoter activation rather than pol II density on an individual allele. In accordance with these results, targeted deletion of the β-globin enhancer reduced burst fraction more strongly than burst size.

In order to determine which bursting parameter is specifically regulated by enhancer-promoter looping, independently of other enhancer functions, they artificially forced enhancer-promoter contacts by tethering the β-globin enhancer to its target promoter. Forced chromatin looping increased the burst fraction, but not burst size, further validating that enhancer-promoter contacts modulate bursting frequency rather than the transcriptional output per individual burst. Interestingly, it was also shown that two target loci of the β-globin enhancer, β-globin and γ-globin, are not simultaneously transcribed at the single-allele level, even though their expression is correlated at the single cell level. This finding favors a promoter competition model, whereby the enhancer dynamically switches between two target promoters, resulting in uncoordinated onset of bursting on both genes. Intriguingly, the study by Fukaya et al showed synchronized bursting of two reporter loci tagged with MS2 and PP7 stem loops, respectively, that were positioned equidistantly to the enhancer. Perhaps, in this artificial reporter construct, the enhancer switched between the two promoters at much faster rate than that which would occur on endogenous loci, making the onset of bursting indistinguishable between the two reporters. Interestingly, an asymmetric dual reporter setup, whereby the PP7 reporter was positioned closer to the enhancer than the MS2 reporter, corroborates with a promoter competition model, as the PP7 reporter bursts more frequently than the MS2 reporter in this scenario. Altogether, these studies demonstrate that the bursting frequency of target genes is modulated by the frequency of enhancer-promoter looping interactions.

1.8 Enhancer RNAs

1.8.1 High-throughput studies of enhancer RNA transcription

Recent studies have discovered that in addition to being associated with specific epigenetic marks and transcription factors, enhancers are also actively transcribed in a tissue specific manner in response to signaling pathways, adding an additional layer of complexity to enhancer function [115–121]. A study by Kim et al in 2010 was among the first to characterize enhancer transcription on a genome wide scale. In this study they investigated transcriptional activation in mouse cortical neurons upon membrane depolarization by exposure to elevated potassium chloride [122]. Upon stimulation, the authors detected increased CBP binding to

H3K4me1 marked enhancers, which led to pol II recruitment and initiation of bidirectional transcription of non-polyadenylated enhancer RNAs (eRNAs). eRNA synthesis correlated with an increase in mRNA synthesis from neighboring genes, indicating that eRNA transcription marks active enhancers engaged in activating proximal target genes. Similarly, a study in 2011 by Wang et al in prostate cancer cells showed that depletion of the androgen receptor (AR) co-factor FOXA1 resulted in the alternative binding of AR to de novo enhancers, which activated a new transcription program that was associated with increased tumor progression [123]. The androgen receptor induced bidirectional eRNA transcription, which correlated with increased mRNA levels from target genes. Furthermore, De Santa et al 2010 investigated genome wide pol II recruitment in macrophages activated by endotoxin and discovered that upon stimulation, most of the extragenic pol II peaks coincided with enhancers [124]. These enhancers are transcribed into eRNAs that are induced early on during stimulation, followed by downstream mRNA expression. eRNA transcription also lead to histone acetylation upstream of the target gene, potentially increasing chromatin accessibility to pol II at the TSS of the gene.

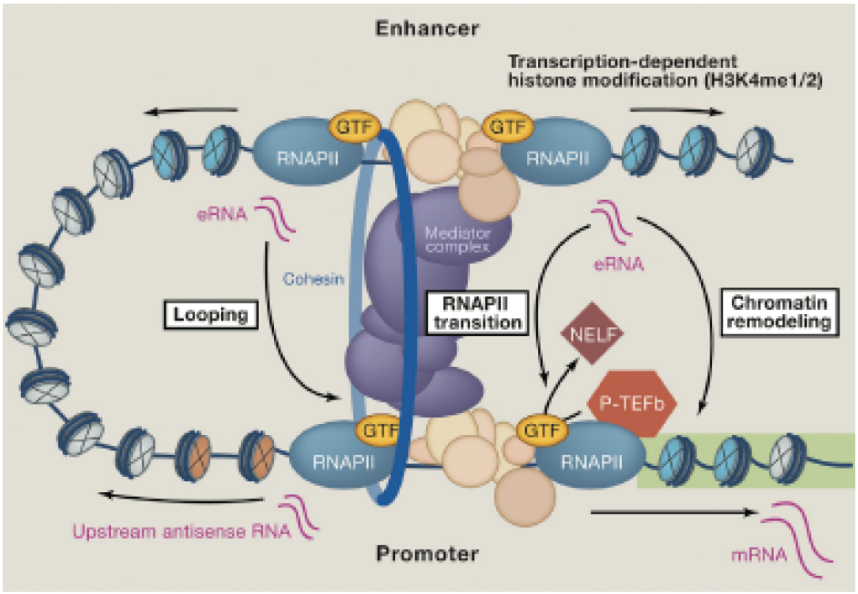
Following these initial studies describing transcription at enhancers, Zhu et al. showed that transcription at enhancers is a more robust predictor of enhancer activity than enhancer-specific histone marks, many of which are found on enhancers that are functionally inactive [125]. Intriguingly, the peak regions of eRNA transcription have been shown to converge with regions of high H3K4me1/2, suggesting a functional link between eRNA transcription and deposition of active histone marks on enhancers [91]. Such a model is further supported by the finding that the inhibition of eRNA synthesis by the transcription elongation inhibitor Flavopiridol reduces H3K4me2 levels on enhancers. Interestingly, the histone methyltransferase MLL was previously shown to interact and co-localize with RNA pol II on a subset of actively transcribed loci [126]. Therefore, RNA pol II elongation on enhancers may promote the recruitment of MLL, which is further substantiated by the correlation between eRNA transcript length and the spread of H3K4 methylation from the core enhancer.

A role for eRNAs in regulating transcription was further supported by a study in macrophages studying the nuclear receptors Rev Erb α and Rev Erb β that repress target genes by binding to distal enhancers and inhibit eRNA transcription [127]. The authors found that

targeted degradation of eRNAs transcribed from negatively regulated by Rev-Erbs enhancers reduced the expression of target genes, suggesting a functional role of eRNAs transcription regulation of target genes. They also showed that both the core transcription factor binding site on the enhancer and the sequences encoding eRNAs are required for robust promoter activation. Altogether, these studies support the idea that eRNA synthesis is an important component of enhancer function.

1.8.2 Models of eRNA function

Although several genome wide studies have observed correlations between eRNA and mRNA transcription, the mechanisms by which eRNAs activate target genes have yet to be fully elucidated. Three main lines of thought pertaining to the potential role of eRNAs have emerged in the last few years: i) The process of eRNA transcription through the enhancer region, but not eRNAs as functional RNAs, could play a role in chromatin remodeling, thereby stimulating transcription, and ii) eRNAs are functional molecules recruiting and/or binding various factors required for enhancer mediated transcription, acting in *cis* or *trans* and either stabilizing enhancer-promoter looping interactions, regulating the RNA pol II transcription cycle or regulating chromatin accessibility at promoters [51] (**see Figure 1-11**). iii) eRNAs could have no function in regulating transcription but represent transcriptional noise due to either random pol II collisions with the open and accessible chromatin regions of enhancers and/or as a consequence of transcription factor binding to enhancers. The following sections will discuss these different models.



(Kim and Shiekhattar 2015)

Figure 1-11: eRNA mechanisms of action

1.8.2.1 Regulation of chromatin accessibility at promoters.

A study investigating a gene regulatory network during skeletal muscle cell differentiation explored the role of eRNAs in activating master regulators of the myogenic program [128]. They demonstrated that the myogenic transcription factors MyoD and MyoG predominantly bind to enhancer sequences throughout the genome and that two different eRNAs transcribed within a large enhancer domain activate MyoD and MyoG transcription, respectively, resulting in a feed-forward loop. When these eRNAs were depleted, chromatin accessibility was reduced at the MyoD and MyoG promoters and this was paralleled with lower pol II occupancy at promoters, suggesting that eRNAs are implicated in remodeling chromatin at their target promoters, allowing pol II to access the promoter and initiate transcription of the target gene.

1.8.2.2 Regulation of enhancer-promoter looping

One prominent model for the role of eRNAs in transcription implicates eRNAs in mediating enhancer-promoter looping interactions [88,129,130]. In an initial study, Lai et al. characterized the function of enhancer-like RNAs referred to as ncRNA-activating (ncRNA-a)[129]. They demonstrated that these RNAs co-purify with the mediator complex and are required for enhancer-promoter looping. Furthermore, eRNAs were shown to stimulate the kinase activity of the mediator which phosphorylates histone H3 serine 10 on the target promoter required to activate transcription, suggesting a role of these RNAs in regulating chromatin architecture and transcription activation. Similarly, a study on androgen receptor (AR) induced transcription in prostate cancer cells shows that AR induced eRNAs facilitating the assembly of the AR-mediator complexes and stimulate enhancer-promoter looping [130].

Furthermore, Li et al investigated estrogen (E2) induced eRNAs in MCF7 breast cancer cells, showing that eRNA interaction with cohesin promotes enhance-promoter communication [88]. Exposure of MCF7 cells to E2 induces the binding of estrogen receptor alpha (ER alpha) to estrogen responsive enhancer sequences and results in the bidirectional transcription of eRNAs at these enhancers. eRNA transcription further correlates with an increase of mRNA transcription at target genes. Furthermore, RIP-qPCR experiments showed that eRNAs associate with the cohesin complex, which has been described previously to co-occupy ER alpha binding sites and to be implicated in enhancer promoter communication [102,131,132] . Knocking down components of cohesin resulted in decreased enhancer-promoter looping, and knocking down eRNAs inhibited recruitment of cohesin to enhancer sequences suggesting that E2 induced eRNAs may promote looping by recruiting cohesin. Surprisingly, a study investigating E2 induced transcription in MCF7 cells showed that inhibiting eRNA transcription elongation by DRB does not affect enhancer-promoter looping nor epigenetic modifications on enhancers, suggesting that eRNA synthesis is dispensable for the formation of an active enhancer-promoter complex [133]. The role of eRNAs in recruiting/stabilizing cohesin therefore still needs to be determined.

Li et al further investigated whether eRNA transcripts *per se* can activate gene transcription by tethering a sense eRNA transcribed from the enhancer targeting the FOXC1

gene to a luciferase reporter plasmid via the BoxB tethering system. Tethering of FOXC1 sense eRNA increased luciferase expression, showing that eRNA sequences themselves are functional independently of the process of enhancer transcription. This result is consistent with the findings of another study investigating eRNAs transcribed from p53 bound enhancers, in which tethering of an eRNA to a luciferase reporter plasmid using an MS2 based tethering system activated luciferase expression [134]. These eRNAs were shown to enhance transcription of target genes, and were required for p53 dependent cell cycle arrest. Furthermore, the authors suggested that since long range enhancer-promoter interactions were not p53 dependent, p53 induced eRNAs might act on pre-assembled enhancer-promoter complexes. Therefore, whether eRNAs initiate chromatin looping or simply stabilize pre-existing enhancer-promoter interactions is subject to debate [119].

1.8.2.3 Regulation of the RNA pol II transcription cycle

In addition to the studies implicating eRNAs in facilitating enhancer-promoter looping through scaffolding large protein complexes such as mediator or cohesin, other studies suggest that eRNAs activate the transcription cycle by promoting pol II elongation on target genes [135,136]. Using cultured cortical neurons as a model system, Schaukowitch and colleagues showed that eRNA knockdown does not affect enhancer-promoter looping but diminishes the release of the negative elongation factor (NELF) from target promoters. Therefore, in their system, in the absence of eRNAs, RNA pol II remains in a paused state on target promoters due to continued association with NELF. Furthermore, they showed that eRNAs interact with NELF, acting as a decoy to release it from paused pol II, and thereby stimulating productive elongation on the target gene. Interestingly, they also show that eRNAs are synthesized prior to gene activation, and are very unstable, having a half-life of less than 7.5 min. Therefore, their data suggests that eRNAs accumulate transiently on pre-established enhancer-promoter loops, are rapidly degraded following NELF release, preventing their diffusion and therefore likely restrict their activity to function in *cis*. Interestingly, eRNAs induced in macrophages are also transcribed prior to target gene induction, and were shown to have a similar short half-life less than 7.5 min [124].

High-throughput CAGE analysis in several human and mouse cell lines upon induction with different stimuli shows that eRNA transcription generally represents the immediate early response during cellular differentiation and activation followed by successive waves of transcriptional activation on target promoters [137]. Therefore, eRNA downregulation prior to pol II elongation on target genes may be a common regulatory pathway in different cell types. However, it was also shown that transcription elongation on both enhancers and promoters may be regulated by the same mechanism [138,139]. These studies showed that the Bromodomain-containing protein 4 (BRD4) is recruited onto acetylated histones on both enhancers and promoters during gene activation. BRD4 was shown to promote RNA pol II elongation on both enhancers and target genes, to synthesize eRNAs and mRNAs, respectively. The coordinated transcription regulation of eRNAs and mRNAs by BRD4 implies that eRNA expression may be temporally coupled with mRNA expression in certain contexts, instead of being downregulated prior to gene activation.

Interestingly, a recent study in prostate cancer cells showed that instead of acting as a decoy for NELF, eRNAs may also regulate pol II elongation on target genes by stimulating the activity of the positive transcription elongation factor pTEFb [136]. It was shown that a specific class of androgen receptor regulated eRNAs activate pTEFb, which phosphorylates Ser2 on pol II, resulting in transcription elongation on target genes, and increased cell growth. TALEN mediated gene editing revealed that a specific hairpin forming sequence within the eRNAs, which is similar to the HIV RNA TAR-L motif, and bound by CYCLIN T1, is required for pTEFb activation and target gene transcription. Whether eRNAs in other cell types possess similarly structured functional motifs is currently unknown.

1.8.3 Transcription termination and processing of eRNAs

Little is known about how transcription of eRNAs is terminated and whether it is coupled with 3'end processing, contributing to the lack of understanding of what constitutes a functional eRNA transcript. Transcriptome analyses in multiple human cell lines and primary tissues using CAGE have shown that eRNAs are capped, unspliced, and initiate bi-directionally from precisely mapped transcription start sites (TSSs) that delineate the boundaries of nucleosome-depleted core enhancers [55]. eRNAs have a median length of 300-

400 nucleotides, however, transcription termination sites (TTSs) have not been clearly defined. It is therefore not known whether eRNAs consist of a heterogeneous population of eRNAs of different lengths, or conversely a homogenous species with a uniform length. A recent study showed that eRNA function is contingent on transcription termination by the integrator complex [140]. Integrator was initially characterized as a metazoan-specific multi-protein complex that interacts with the CTD of RNA pol II to promote 3' processing of snRNAs [141]. Abrogation of integrator components resulted in the accumulation of extended primary transcripts due to defects in 3' end maturation of snRNAs and transcribed enhancers similarly show extended transcription at enhancers. This phenotype is reminiscent of the accumulation of read-through sn/snoRNA transcripts in *S.cerevisiae* cells depleted of the components of the Nrd1-Nab3-Sen1 complex, which also binds to the CTD of RNA pol II to facilitate 3'end maturation of small RNAs [34]. Integrator was also shown to play diverse roles in gene expression regulation, notably in RNA pol II pause release and elongation on protein coding genes [142–145].

In a first study linking integrator to eRNA transcription termination, Lai et al showed increased recruitment of the integrator complex to enhancers in HeLa cells upon epidermal growth factor (EGF) stimulation [140]. Surprisingly, depletion of integrator components decreased eRNA induction, as well as enhancer-promoter looping and target gene activation, but resulted in the accumulation of extended eRNA transcripts that are mostly polyadenylated and associated with RNA pol II. This corroborates with a previous study which showed that the loss of 3'end cleavage activity of the integrator resulted in increased levels of polyadenylated snRNA transcripts [146]. Therefore, this study proposed a model whereby 3'end cleavage by the integrator would terminate transcription and release eRNAs from RNA pol II. eRNAs released from RNA pol II by the integrator would then represent the functionally active transcripts with defined 5' and 3' ends. Supporting this, Northern blot analyses in several studies have shown that eRNAs migrate as single bands, suggesting that most eRNAs are uniform in length [130,134,147]. Intriguingly, expressing a catalytic mutant of the integrator subunit INTS11 that is defective in 3'processing led to decreased induction of both eRNAs and target mRNAs. Therefore, 3'processing of eRNAs by the integrator is somehow linked to their transcription, although the mechanistic basis of these findings remain unclear. Therefore,

the regulation of transcription termination of eRNAs may play an important role in eRNA function. However, whether other factors apart from the integrator are implicated in this process needs further investigation.

1.9 Objectives of Thesis

Transcription is the first step in gene expression and subject to extensive regulation. Many factors are implicated in transcription regulation, including transcription factors and chromatin remodelers, required to regulate promoter accessibility and recruitment of RNA polymerase, many of them well characterized. In recent years, lncRNAs emerged as new regulators of gene activation, but mechanistic understanding of the roles of lncRNAs is largely lacking. LncRNAs have the potential to regulate transcription in many different ways, and the sheer number and diversity of lncRNAs indicates that no uniform mechanism likely exists. Therefore, one of the most exciting questions of the post-genomic era is to decipher the roles of lncRNAs in regulating gene expression.

The general aim of my PhD thesis was to investigate distinct classes of lncRNAs in different eukaryotes, to elucidate their roles in transcription regulation and to characterize the factors that control their biogenesis. In particular, I combined high resolution microscopy tools with genetic and biochemical approaches to study lncRNA mediated gene regulation from a single cell and single molecule perspective. Historically, studies on transcription have relied on isolating RNA from large numbers of cells, and therefore could only describe the ensemble behavior of a population. However, individual cells within a population respond differentially to signaling pathways due to the probabilistic nature of the chemical processes that mediate transcription, resulting in intercellular variability in gene expression. Therefore, to glean novel insights into the mechanisms by which lncRNAs regulate gene transcription, my studies have a strong emphasis on using single molecule approaches to quantitatively measure gene activity at specific target loci in single cells. In this thesis, I present two studies where I investigated the roles of two different classes of lncRNAs in gene activation, using the yeast *S. cerevisiae* as well as human breast cancer cells as model systems.

1.10 Figure Legends

Figure 1-1: **TATA box vs. TATA-less promoters**[29]

This figure illustrates differences in nucleosome architecture and transcription start site position between TATA box and TATA-less promoters

Figure 1-2: Constitutive vs bursting transcription[27]

- **a.** Stochastic model of gene activation and inactivation
- **b.** Plots showing transcription initiation events over time for constitutive and bursting transcription

Figure 1-3: Bi-directional transcription from nucleosome-depleted regions (NDRs) in *S.cerevisiae*

This figure illustrates the recruitment of pre-initiation complexes to nucleosome-free regions upstream and downstream of protein-coding genes in *S.cerevisiae.*

Figure 1-4: Nrd1-Nab3-Sen1 vs cleavage and polyadenylation mediated termination[30]

- **a.** Nrd1-Nab3-Sen1 mediated termination of small ncRNAs
- **b.** Termination by the canonical cleavage and polyadenylation complex

Figure 1-5: Bi-directional transcription from enhancers and promoters[148]

This figure depicts the chromatin architecture and regulation of bi-directional transcription from enhancers and promoters.

Figure 1-6: General principles of lncRNA function[59]

A summary of the different modes of lncRNA mediated regulation, in which the RNAs can act as guides to recruit transcription co-factors, or the process of lncRNA transcription can change the local chromatin environment.

Figure 1-7: Different mechanisms of transcription regulation by ncRNAs in *S.cerevisiae*[31]

This figure depicts transcription regulatory mechanisms that have been identified for genes in yeast. The top image shows regulation by transcriptional interference, in which the transcription of an ncRNA through a gene promoter prevents binding of transcription factors. The image in the middle shows regulation by start site selection, in which an ncRNA is produced as a by-product of alternate transcription initiation upstream of the mRNA transcription start site. The bottom image shows regulation by targeted histone modification, in which an ncRNA serves as a guide to recruit chromatin modifying complexes.

Figure 1-8: *PHO84* AS RNA represses *PHO84* by recruiting the histone de-acetylase Hda1[81]

An early model of *PHO84* regulation by antisense RNA which proposes that antisense RNAs stabilize upon Rrp6 deletion and accumulate at the *PHO84* locus, recruiting the histone de-acetylase Hda1 to the promoter to repress transcription.

Figure 1-9: Topologically associated domains (TADs) regulate the specificity of enhancer-promoter interactions[149]

This figure illustrates the organization of genomes in higher eukaryotes into discrete neighbourhoods called topologically associated domains (TADs), which are mostly cell type invariant. TADs compartmentalize enhancers with their target promoters and are segregated from each other by defined boundary elements.

Figure 1-10: Variable chromatin fiber conformations within TADs[111]

This diagram shows different intra-TAD chromatin fiber conformations predicted using simulations based on polymer physics. These different conformations lead to dynamic interactions between different genomic regions.

Figure 1-11: eRNA mechanisms of action[148]

A summary of the suggested mechanisms of eRNA mediated transcription regulation, which focus on 3 main roles: enhancer-promoter looping, RNA pol II transition, and chromatin remodeling.

2 Bimodal expression of *PHO84* is modulated by early termination of antisense transcription

2.1 Aims of article 1

For my first project, I investigated a lncRNA transcribed in antisense to the yeast gene *PHO84*, which encodes a high affinity phosphate transporter. Previous studies had shown that deletion of Rrp6, a component of the nuclear RNA surveillance machinery, results in increased antisense expression and a decrease in *PHO84* expression. This led to a model that suggested that *PHO84AS* RNAs are stabilized in the absence of Rrp6 and accumulate at the *PHO84* locus, recruiting the histone de-acetylase Hda1 at the *PHO84* promoter to suppress *PHO84* transcription. However, many mechanistic details on how anti-sense mediated transcriptional repression is achieved were missing. Following up on these studies, we further elucidated the mechanisms of AS RNA biogenesis and function in repressing *PHO84* expression. The results of this study were published in *Nature Structural and Molecular Biology* in July 2013 where I was shared first author (Castelnuovo et al, "Bimodal expression of PHO84 is modulated by early termination of antisense transcription". Nat Struct Mol Biol. 20(7):851-8)).

2.2 Article 1

Bimodal expression of *PHO84* is modulated by early termination of antisense transcription

Castelnuovo, Manuele[1,3], Rahman, Samir[2,3], Guffanti, Elisa[1,3], Infantino, Valentina[1], Stutz, Françoise[1,4] and Zenklusen, Daniel [2,4]

[1] Department of Cell Biology and National Center of Competence in Research "Frontiers in Genetics", University of Geneva, Switzerland.

[2] Département de Biochimie, Université de Montréal, Canada

[3] Equal contribution

[4] Corresponding authors

E-mail: francoise.stutz@unige.ch daniel.r.zenklusen@umontreal.ca

Keywords: yeast, single molecule FISH, antisense RNA, Rrp6, Nrd1-Nab3-Sen1, transcription termination, antisense 3'end processing, *PHO84* regulation.

2.2.1 Abstract

Many *S. cerevisiae* genes encode antisense transcripts, some of which are unstable and degraded by the exosome component Rrp6. Loss of Rrp6 results in the accumulation of long *PHO84* antisense RNAs and repression of sense transcription through *PHO84* promoter deacetylation. We used single molecule resolution fluorescent in situ hybridization (smFISH) to investigate antisense-mediated transcription regulation. We show that *PHO84* antisense RNA acts as a bimodal switch, in which continuous low frequency antisense transcription represses sense expression within individual cells. Surprisingly, antisense RNAs do not accumulate at the *PHO84* gene but are exported to the cytoplasm. Furthermore, rather than stabilizing *PHO84* antisense RNA, the loss of Rrp6 favours its elongation by reducing early transcription termination by Nrd1-Nab3-Sen1. These observations suggest that *PHO84* silencing results from antisense transcription through the promoter rather than the static accumulation of antisense RNAs at the repressed gene.

2.2.2 Introduction

Genome-wide pervasive transcription has been reported in many eukaryotic organisms, producing hundreds of non protein-coding RNAs (ncRNAs). Even the small yeast genome encodes many intergenic, promoter-associated and antisense transcripts, some stable and others rapidly degraded and hence called cryptic unstable transcripts (CUTs) [1-3]. The degradation of these 200-600 bases long CUTs is in great part mediated by Rrp6, a 3'-5' exonuclease belonging to the nuclear exosome [4,5]. Exosome-mediated degradation is assisted by TRAMP, a surveillance complex containing the non-canonical polyA polymerase Trf4, while mRNAs are polyadenylated by Pap1, resulting in stable and export competent mRNPs [4,6-9].

The Nrd1-Nab3-Sen1 (NNS) complex mediates transcription termination of CUTs, snRNA, snoRNAs, and some mRNAs [7,10-13]. It is recruited to the 5'end of most RNA polymerase II (RNAPII) transcription units through interaction of Nrd1 with the Ser5/Ser7 phosphorylated RNAPII C-terminal domain (CTD) [14-16]. Transcription termination by NNS depends on the abundance of specific Nrd1 and Nab3 binding motifs on the nascent RNA and occurs primarily on short transcripts as the recruitment of NNS decreases towards the 3'end of long transcription units. Consistent with the physical interactions between the NNS, TRAMP and exosome complexes, CUT degradation has been directly linked to NNS-mediated early termination [4,7,10,11].

Genome-wide studies indicate that numerous genes produce upstream tandem or antisense transcripts [17,18], a fraction of which may function in gene regulation [19]. Transcription of an upstream tandem ncRNA was proposed to interfere with the expression of the *SER3* [20,21], *URA2* [22], *FLO11* [23] and *IME1* [24] genes through various mechanisms, including co-transcriptional chromatin modifications, that establish histone repositioning and a repressive chromatin state blocking access to transcription factors. While the *RME2* antisense RNA was proposed to repress the meiotic regulator *IME4* gene *via* transcription interference [25,26], antisense RNA transcription may also affect sense expression by influencing the epigenetic state of chromatin. Indeed, antisense RNA transcription originating within *GAL10* and running into the divergent *GAL1* gene in glucose deposits H3K4-me2/3 and H3K36-me3 by the Set1

and Set2 histone methyl transferases respectively. These marks signal the recruitment of the Rpd3S histone deacetylase (HDAC) attenuating *GAL1* gene expression [27,28]. H3K4me2 deposited by Set1 during noncoding transcription was also implicated in repression by signaling the recruitment of the Rpd3L and Set3 histone deacetylases to specific gene promoters [24,29,30].

Our earlier studies focused on the *PHO84* gene encoding a high-affinity phosphate transporter. *PHO84* transcription is induced by the activator Pho4 imported into the nucleus upon phosphate starvation [31]. The activation threshold of the *PHO84* promoter depends on the nuclear concentration of Pho4 and the accessibility of the Pho4 binding sites [32,33]. *PHO84* mRNA is weakly expressed in standard yeast media containing intermediate phosphate levels. In these conditions, *PHO84* also produces two antisense transcripts (*PHO84* AS) starting at its 3' end and extending into the *PHO84* promoter. Loss of Rrp6 increases *PHO84* AS levels and this accumulation is paralleled by the recruitment of the Hda1/2/3 histone deacetylase (HDAC) complex over the locus, histone deacetylation at the promoter and transcriptional repression. We proposed that stabilization and accumulation of antisense RNAs at the *PHO84* gene might facilitate Hda1 recruitment maintaining repression of sense transcription [34].

To further elucidate the mechanism of antisense-mediated transcription regulation, we used single molecule fluorescent *in situ* hybridization (smFISH) to detect individual sense and antisense RNAs [35-37]. We show that the presence of *PHO84* sense and antisense transcripts in single cells is strongly anti-correlated, suggesting a switch-like regulation mechanism. Our data provide evidence that Rrp6 does not degrade full-length antisense transcripts, but prevents antisense transcription elongation by favoring early termination by Nrd1-Nab3-Sen1, while the H3K4 methyl transferase Set1 may antagonize this event. These observations suggest that antisense-mediated silencing is regulated, at least in part, through transcription attenuation and that *PHO84* repression results from antisense transcription through the promoter, followed by rapid export of antisense RNA into the cytoplasm.

2.2.3 Results

Bimodal expression of *PHO84* sense and antisense transcripts

We have suggested that *PHO84* AS RNAs might stably associate, possibly in multiple copies, with the *PHO84* gene to help efficient recruitment of chromatin modifiers. Such a process would require only a single initial burst of antisense transcription to establish silencing of sense. Conflicting with such a model, Northern blot analysis of *PHO84* sense and antisense expression shows that low levels of antisense RNA can be detected in wild-type cells under conditions where *PHO84* sense is transcribed, indicating that very low expression of antisense might not be sufficient to repress sense transcription (Fig. 2.1a). Alternatively, low level of *PHO84* AS RNA expression in wild-type cells may continuously fine-tune *PHO84* sense expression, a process that may be regulated by Rrp6. However, different levels of antisense expression in wild-type versus *Δrrp6* cells may also reflect different subclasses of cells in a population that express either *PHO84* sense or antisense. Different models can therefore be suggested for how antisense-mediated silencing of the *PHO84* gene is established. Either the regulation occurs by a graded response, where increasing antisense levels lead to decreasing levels of sense transcription, or by a switch-like mechanism, where low level of antisense expression in a single cell is sufficient to down-regulate sense transcription (Fig. 2.1b).

To detect single RNA molecules, we designed smFISH probes targeted to the 5' region of sense and antisense *PHO84* transcripts. Probes were labeled with fluorescent dyes allowing to distinguish sense and antisense transcripts and hybridized to fixed yeast cells, followed by image acquisition. We first localized *PHO84* transcripts in wild-type cells under conditions where both sense and antisense RNAs are detected by Northern blotting (Fig. 2.1a). While both *PHO84* sense and antisense RNAs can be detected in wild-type cells (Fig. 2.1c), they are never co-expressed (Fig. 2.1d), suggesting that antisense-mediated repression of *PHO84* operates through a switch-like rather than a graded process. Consistent with the role of Rrp6 in modulating sense repression through antisense RNA, the fraction of cells expressing antisense

increases (from 28 to 55%) in a *Δrrp6* strain, whereas the percentage of sense expressing cells decreases (Fig. 2.1c, 2.1d and 2.1e).

At the single cell level, sense expression is much higher than antisense: large numbers of *PHO84* mRNAs are detected within individual cells suggesting that *PHO84* transcription occurs in strong bursts when repression is overcome. In contrast, *PHO84* AS expression levels are very low in individual wild-type cells, with most cells expressing no or only a single antisense RNA molecule. In the *Δrrp6* strain, antisense levels are higher and more cells express *PHO84* AS, however most cells still only contain 1-3 antisense RNA molecules and a substantial fraction of cells (40%) shows no signal (Fig. 2.1c, 2.1d and 2.1e). Double negative cells are not due to inability to detect RNAs in these cells, as double staining for the constitutively expressed *MDN1* RNAs shows expression of *MDN1* in all cells (Supplementary Fig. 1a). Thus, very low antisense expression appears sufficient to exert a repressive effect on *PHO84* transcription in individual cells. Unexpectedly, we did not observe a significant accumulation of antisense RNA in the nucleus (Fig. 2.2) as most antisense RNAs detected in wild-type and *Δrrp6* cells are found in the cytoplasm, suggesting that *PHO84* AS RNAs, like mRNAs, do not remain associated with the *PHO84* gene but are rapidly exported.

PHO84 antisense RNAs do not accumulate at the _PHO84_ locus

The fraction of antisense RNA molecules detected in the nucleus can represent nascent RNAs associated with the transcription machinery, RNAs diffusing in the nucleoplasm on their way to the cytoplasm, or antisense RNAs associated with the *PHO84* gene in a transcription independent manner. To distinguish between these possibilities, we further characterized the nuclear *PHO84* AS RNA signal. The quantitative nature of smFISH allows defining how many RNAs are present in a single RNA spot and we have shown that cytoplasmic mRNA spots have a uniform signal intensity representing single mRNAs [36,38]. Nuclear signals often show higher intensities as they represent sites of active transcription where multiple nascent mRNAs are associated with a gene. The frequency and number of nascent mRNAs detected for a specific gene depend on its transcription rate and length. If antisense RNAs accumulate in multiple copies at the *PHO84* gene, higher intensity nuclear signals compared to cytoplasmic signals should be detected. Furthermore, if antisense RNAs

stay associated at the gene for long periods of time, most cells with no sense expression should show a nuclear antisense signal. As shown in Figures 2.2a and 2.2b, nuclear signals corresponding to multiple nascent mRNAs are detected on the long, constitutively transcribed *MDN1* gene, however most nuclear *PHO84* AS RNA signals show the same intensity as single cytoplasmic antisense molecules, indicating that antisense transcripts do not accumulate at the *PHO84* gene. Furthermore, only 13% of WT and 20% of *Δrrp6* cells show nuclear signal, inconsistent with a model where antisense RNAs stay associated with the gene locus for a long time (Fig. 2.2c).

It is likely that most nuclear AS signals with an intensity of a single RNA represent nascent rather than freely diffusing nucleoplasmic antisense RNAs. Indeed, nuclear *PHO84* AS signals, like nascent *PHO84* mRNA signals are always located at the nuclear periphery, consistent with the subtelomeric position of *PHO84* on chromosome XIII locating the gene close to the nuclear periphery (Supplementary Fig. 1b). Furthermore, our earlier studies showed that mRNAs are rarely detected in the nucleoplasm except at the site of transcription, suggesting that mRNA export is fast, probably occurring within seconds after release from the site of transcription [36,39]. If *PHO84* AS RNAs transcribed at a low frequency behave like mRNAs, detecting antisense RNAs within the nucleus is likely a rare event, except when they are nascent. Thus, nuclear *PHO84* AS RNAs are likely to be nascent and to behave like mRNAs that rapidly dissociate from the locus after synthesis. These observations suggest that antisense transcription rather than antisense RNA accumulation at the gene may mediate *PHO84* gene silencing.

PHO84 antisense RNAs behave like mRNAs

To confirm that antisense transcripts behave like mRNAs, we first monitored antisense RNA distribution in a mutant for the poly(A) polymerase Pap1. mRNA cleavage and polyadenylation occurs co-transcriptionally and is required for nuclear export. The *pap1-1* and *pap1-1Δrrp6* temperature sensitive strains were grown at 25^{o}C and shifted to 37^{o}C before fixation. After a 1h heat-shock, *pap1-1* cells accumulate antisense RNAs in the nucleus and fewer transcripts are observed in the cytoplasm, a phenotype that was more pronounced in *pap1-1Δrrp6* (Fig. 2.3a and 2.3b). Antisense RNAs do not accumulate in one spot but

distribute throughout the nucleus, with a tendency to localize within the nucleolus (Supplementary Fig. 2a). The higher accumulation in *pap1-1 Δrrp6* compared to *pap1-1* suggests that antisense RNAs are degraded by Rrp6 when not polyadenylated by Pap1 and/or that a higher number of antisense RNAs is expressed in a *pap1-1 Δrrp6* background (Fig. 2.3b, and see below). Loss of the non-canonical polyA polymerases Trf4 and Trf5 did not reduce the amounts of polyadenylated *PHO84* AS RNAs, confirming their polyadenylation by Pap1 (Fig. 2.3c). Notably, shifting *pap1-1 Δrrp6* double, but not *pap1-1* single mutant cells to 37°C results in the accumulation of an elongated polyadenylated antisense RNA (Supplementary Fig. 2b). Together, these analyses suggest that when Pap1 is inactive, a single long antisense transcript is produced that remains in the nucleus and is degraded by Rrp6, presumably following polyadenylation by the non-canonical Trf4/5 polyA polymerase as a result of nuclear surveillance [40]. Thus the classical cleavage and polyadenylation machinery is required for 3'end processing and export of *PHO84* AS RNA confirming that these long ncRNAs behave like mRNAs. Accordingly, their nuclear export is mediated by the general mRNA export receptor Mex67, since *PHO84* AS transcripts accumulate in the nuclei of the *mex67-5* and even more in the *mex67-5Δrrp6* conditional mutants when shifted to 37°C (Supplementary Fig. 3a and 3b). Moreover, the number of cytoplasmic *PHO84* AS RNAs greatly increases in *Δxrn1* cells indicating that, like mRNAs, they undergo 5' to 3' exonucleolytic degradation in this compartment (Supplementary Fig. 3c).

Antisense RNA at *PHO84* gene requires active transcription

A feature of *bona fide* mRNAs is their rapid dissociation from the gene after transcription termination; nascent mRNA detection therefore requires ongoing transcription. To define whether detection of nuclear antisense RNAs requires transcription, we determined *PHO84* AS localization and abundance in the *rpb1-1* strain, containing a temperature sensitive mutation in the major RNAPII subunit [41]. To test the efficiency of transcription shutoff we simultaneously monitored *MDN1* mRNA distribution. Figure 2.4 shows that after 5 min at 37^{o}C, most cells have lost nuclear *MDN1* signal and mRNA abundance further declines over time, consistent with transcription shutoff. Similarly, nuclear *PHO84* AS signal is quickly lost and cytoplasmic RNA numbers subsequently decrease. Thus, ongoing transcription is required

to detect nuclear antisense RNA further indicating that *PHO84* AS RNA does not stay associated with the *PHO84* gene. The observation that the number of cells with antisense RNA increases in *Δrrp6* (Fig. 2.1d) and that antisense transcription rather than accumulation is required to mediate sense silencing (Fig. 2.1 and 2.2) suggest that loss of Rrp6 does not primarily affect antisense RNA stability, but may also influence its transcription.

To compare *PHO84* AS RNA turnover in wild-type and *Δrrp6* cells, we measured antisense levels at various times following inhibition of RNAPII transcription with phenanthroline (Fig. 2.5a) [42]. Surprisingly, *PHO84* AS RNA decays at a similar rate in both strains with a half-life of 11.4 min in wild-type and 12 min in *Δrrp6* cells (See Methods). In contrast, the half-life increased to 27.3 min in the *Δxrn1* strain, confirming 5' to 3' antisense RNA degradation in the cytoplasm as revealed by smFISH (Supplementary Fig. 3c). Since loss of Rrp6 does not substantially increase *PHO84* AS RNA half-life, these results indicate that the elevated levels of antisense RNA in *Δrrp6* (Fig. 2.5b) are due to increased antisense RNA production rather than stability.

Loss of Rrp6 increases antisense transcription

Increased *PHO84* AS transcription in *Δrrp6* predicts a higher number of nascent antisense RNAs in this strain versus wild type. Indeed, besides an increased number of both antisense producing cells and antisense RNA molecules per cell (Fig. 2.1e), more *Δrrp6* cells (20%) show nascent antisense RNAs compared to wild type (13%) consistent with higher transcription frequency in *Δrrp6* (Fig. 2.2c).

One hallmark of active transcription is K4 methylation on histone H3 by Set1, the only yeast H3K4 histone methyl transferase recruited to the 5' end of transcription units [43,44]. Most active genes show peaks of H3K4 trimethylation at the 5'end, di-methylation in the middle and monomethylation at the 3'end. We postulated that if loss of Rrp6 increases antisense transcription, Set1 dependent H3K4me3 should increase over the *PHO84* 3' end in *Δrrp6* versus wild type. We performed chromatin immunoprecipitation (ChIP) of tri- and dimethylated H3K4 in wild-type and *Δrrp6* cells also devoid of the transcription factor Pho4, completely abrogating sense transcription (Fig. 2.5c). In this setup H3K4 methylation derives

only from antisense transcription. Interestingly, we observe that the H3K4me3 and H3K4me2 peaks respectively at the 3' end and middle regions of *PHO84* are substantially increased upon loss of Rrp6. As a control, the *ACT1* gene showed the expected high level of H3K4me3 at its 5' end with no enrichment at the 3' end, consistent with the absence of antisense transcription on this gene. Due to the low antisense transcription frequency, RNAPII is barely detectable at the 3'end of *PHO84* in a *Δpho4* strain, yet the levels slightly increase in *Δrrp6Δpho4* (data not shown). The more efficient detection of H3K4 methylation suggests persistence of this histone mark between transcription events. These observations support the view that loss of Rrp6 increases antisense transcription.

***PHO84* antisense elongation is regulated by the NNS complex**

To investigate how loss of Rrp6 may increase transcription, we explored the physical and functional links of Rrp6 with the Nrd1-Nab3-Sen1 and TRAMP complexes [4,7,11]. Transcription termination by NNS is stimulated by Nrd1 and Nab3 binding motifs on the nascent RNA. Interestingly, several potential Nrd1-Nab3 binding sites are present within the 5' end of *PHO84* AS RNA (Fig. 2.6a and Supplementary Fig. 4). Furthermore, transcriptome-wide analyses of Nrd1-Nab3 bound RNA sequences revealed association with the 5' end of many antisense transcripts, including *PHO84* AS RNA, suggesting that these ncRNAs undergo early transcription termination [45,46]. Accordingly, depletion of the essential Nrd1 protein using the glucose repressible *GAL1* promoter leads to increased *PHO84* AS levels in wild-type cells and this effect is even more pronounced in *Δrrp6* (Fig. 2.6b). Moreover, a modified *PHO84* gene in which a number of putative Nrd1-Nab3 binding sites at the 5' end of the antisense RNA have been mutagenized, produces more antisense transcripts both in wild-type and *Δrrp6* cells. The relatively modest effect of the *cis*-mutations may be due to only partial removal of potential NNS binding sites to maintain the *PHO84* open reading frame intact (Supplementary Fig. 4). These observations confirm the role of Nrd1/Nab3/Sen1 in *PHO84* AS transcription attenuation.

Rrp6 and Set1 have opposite effects on early termination

To address whether absence of Rrp6 might increase antisense transcription elongation by affecting optimal NNS function, we monitored Nrd1 association with the 3' end of the *PHO84* gene by ChIP in wild-type or *Δrrp6* cells (Fig. 2.6c). While loss of Rrp6 does not affect Nrd1 protein levels, we observed a large decrease in Nrd1 binding at the *PHO84* 3' end in *Δrrp6*, suggesting that loss of Rrp6 may affect early termination by lowering the association of Nrd1-Nab3-Sen1. The additive effect on antisense RNA production of *Δrrp6* and Nrd1 depletion or Nrd1-Nab3 binding site mutagenesis (Fig. 2.6b and Supplementary Fig. 4b), situations that weaken but do not eliminate Nrd1-Nab3-Sen1 function, supports the notion that NNS and Rrp6 act in the same pathway.

Interestingly, Nrd1 association with the *PHO84* 3' end was slightly enhanced in *Δset1*, suggesting that in contrast to *Δrrp6*, loss of Set1 may increase early termination (Fig. 2.6c). A recent study similarly reported elevated Nrd1 binding in *Δset1* and correlated this phenotype with increased Ser5 phosphorylated RNAPII CTD, the mark implicated in NNS recruitment [16,47]. This is also in agreement with our earlier data showing reduced *PHO84* AS RNA production in *Δset1* [48]. Accordingly, smFISH analyses indicate reduced antisense expression in *Δset1* and restoration of antisense RNA levels in *Δset1Δrrp6* (Supplementary Fig. 5). Taken together, the data suggest that Rrp6 and Set1 have antagonistic effects in the regulation of antisense RNA production by respectively facilitating and interfering with early transcription termination by Nrd1-Nab3-Sen1.

2.2.4 Discussion

Expanding on an extensive list of *cis*- and *trans*-acting factors, recent studies have established ncRNAs as additional players in controlling the regulated expression of protein coding genes. Transcription regulation by ncRNAs is achieved by multiple ways, however in-depth mechanistic understanding is still missing. Our detailed analyses of *PHO84 cis*-acting antisense RNAs at a single cell and single molecule level indicate that low frequency antisense transcription, but not the antisense RNA itself, contributes to *PHO84* gene repression.

Our earlier studies showed that an extra *PHO84* gene copy induces repression of both the transgene and the endogenous copy, and suggested that *PHO84* AS RNAs may participate in a still poorly defined mechanism of silencing in *trans* independent of Hda1/2/3 and therefore distinct from silencing in *cis* [48]. Based on the rapid export of antisense RNAs revealed by smFISH, it seems unlikely that antisense RNAs act in *trans* by diffusing from one gene copy to the other, unless the two genes undergo pairing. The primarily cytoplasmic localization of *PHO84* AS RNAs suggests they are more likely to act in *trans* through an indirect mechanism. These possibilities should be investigated in the future.

smFISH reveals distinct sense and antisense expression modes

The single molecule microscopy approach revealed critical parameters on *PHO84* regulation that could not be obtained using classical ensemble measurements (Fig. 2.1). First, we showed that antisense-mediated regulation does not generate a gradual decrease of sense transcription but modulates the threshold of the *PHO84* activation switch. Second, smFISH revealed that sense and antisense expression are achieved through different modes, *PHO84* mRNA being transcribed in bursts that lead to a strong accumulation in a fraction of cells, whereas antisense RNA is transcribed constantly at a very low rate in most cells not expressing *PHO84* mRNA. Third, the ability to localize individual RNAs within different cellular compartments showed that *PHO84* AS RNA behaves like an mRNA that dissociates from the gene locus after polyadenylation by Pap1, leaves the nucleus using the canonical Mex67-dependent mRNA export pathway, and is eliminated by the cytoplasmic Xrn1-dependent RNA degradation machinery.

Loss of Rrp6 favours antisense transcription elongation

Consistent with the increased levels of antisense RNA observed in *Δrrp6* through classical RNA analyses (Fig 1a), smFISH revealed more antisense RNA molecules per cell as well as an increased number of cells with antisense RNA compared to wild type (Fig. 2.1d and 2.1e). Our observations indicate that loss of Rrp6 does not result in nuclear stabilization of full-length antisense RNAs but rather promotes antisense transcription followed by rapid export. First, although the number of cells showing nascent transcripts is increased in *Δrrp6*,

more than one molecule is rarely observed at the transcription site; moreover this nuclear signal is strictly dependent on ongoing transcription both in wild type and *Δrrp6* indicating that once made, antisense transcripts don't remain at the gene (Fig. 2.2, 2.3 and 2.4). Second, the antisense RNA turnover rate is comparable in wild-type and *Δrrp6* strains, supporting the view that the increased steady state levels in *Δrrp6* are due to enhanced antisense RNA production (Fig. 2.5a and 2.5b). Finally, H3K4 tri- and di-methylation at the 3' end and middle region of *PHO84* are higher in the absence of Rrp6 consistent with increased antisense transcription (Fig. 2.5c). Combining mean transcript values and half-life data (Fig. 2.1 and 2.5) indicates a *PHO84* AS RNA transcription frequency of only 1 and 3 RNAs per hour in wild-type and *Δrrp6* cells respectively (Supplementary Table 1). These numbers are consistent with the incidence of nascent transcripts, another measure for transcription frequency. In *Δrrp6*, 20% of cells show a nuclear *PHO84* AS signal (Fig. 2.2c), suggesting that a cell contains a nascent mRNA 20% of the time, i.e. for 12 min every hour. Assuming transcription of antisense RNA occurs at a rate similar to other low frequency transcribed genes (0.8kb/min) and termination/transcript release is a rate-limiting step as suggested for mRNAs, transcription of the 2.3kb antisense RNA takes almost 4 minutes to complete [36]. This fits well with a transcription frequency of 3 *PHO84* AS RNAs per hour, as a nascent antisense signal would be detected 3 times per hour for 4 min. Consistent with this finding, *pap1-1* Δ*rrp6* cells accumulate on average 3.7 AS RNAs after 1 hour heat shock (Fig. 2.3b). These data indicate that continuous but low frequency antisense RNA transcription occurs in cells not expressing sense.

Rrp6 and Set1 influence antisense early termination by Nrd1

Antisense RNA transcription frequency is increased in Δ*rrp6* compared to wild-type and accompanied by a higher fraction of cells with a repressed *PHO84* gene. Regulating antisense transcription frequency could therefore be a way to modulate the strength of repression. Transcription frequency of *PHO84* AS RNA appears to be controlled both at the level of initiation and through the regulation of elongation and termination efficiency of a short transcript by the NNS complex. It is unclear what controls initiation; the presence of a NFR in the 3'UTR of the *PHO84* gene may be sufficient to allow low frequency transcription

of antisense RNA [18]. This 'default' antisense transcription may be further controlled by the NNS termination pathway. Indeed, mutagenesis of Nrd1-Nab3 binding motifs or Nrd1 depletion results in increased antisense levels. Moreover the association of Nrd1 with the *PHO84* 3' end is strongly reduced in *Δrrp6* suggesting that Rrp6 may contribute to antisense early termination by favouring stable NNS complex association (Fig. 2.6). Notably, as recently observed [47], Set1 has opposite effects since its loss increases Nrd1 binding (Fig. 2.6c), suggesting that Set1 and/or H3K4 methylation may interfere with early termination efficiency. These observations are consistent with the positive effect of Set1 and H3K4 trimethylation on antisense RNA production at *PHO84* and other antisense-producing genes [48,49]. Interestingly, both gene-specific and genome-wide studies suggest that TRAMP and exosome components are required for snRNA/snoRNA transcription termination by Nrd1 and loss of Trf4 was shown to reduce Nrd1 binding to snRNA genes [50,51]. Together with our results, these observations support the view that both TRAMP and Rrp6 may more generally contribute to efficient NNS-dependent transcription termination. Since the activity of both Nrd1 and Rrp6 is regulated in different physiological conditions [52,53], genes like *PHO84* may be controlled in part through modulation of antisense transcription elongation.

A novel view on antisense-mediated gene repression

Our data show that *PHO84* transcription is regulated by a sensitive on-off switch where sense transcription is either completely turned off or strongly induced once the repression is overcome. The activation threshold of Pho4 regulated genes is defined both by the nuclear concentration of the Pho4 transcription factor and accessibility of Pho4 binding sites [32]. Antisense transcription may ensure that *PHO84* transcription is activated only in presence of a strong enough stimulus either by reducing Pho4 accessibility through promoter nucleosome rearrangement, and/or, as shown previously, by placing repressive histone marks [34]. Antisense transcription is not able to establish stable repressive marks, as cells rapidly induce *PHO84* sense expression when shifted from high phosphate, a condition where antisense RNA is abundant, to low phosphate medium (Supplementary Fig. 6). Antisense transcription might therefore act as a buffer, protecting cells from responding to weak signals.

H3K4 di-methylation deposited by Set1 during noncoding RNA transcription has been implicated in gene repression [49,54] by recruiting the histone deacetylases Set3 and Rpd3L at promoter regions [24,29,55]. Notably, we observed that in addition to Hda1, *PHO84* antisense-dependent repression similarly depends on Set1 and Rpd3L (J. Zaugg, M. C., N. Luscombe and F. Stutz, unpublished). Thus, besides promoting antisense production, Set1-dependent H3K4 methylation deposited during antisense transcription may also contribute to *PHO84* gene repression by enhancing HDAC recruitment to the sense promoter.

Recent global studies show that many chromatin regulators, including Set1, barely affect steady state gene expression, but are required for rapid transcriptional responses to environmental stresses. Many of these highly regulated genes are associated with distal or antisense ncRNA transcription [29,30,49]. Consistently, our large-scale search for *PHO84*-like genes, i.e. repressed by antisense transcription in *Δrrp6* in a process dependent on Set1 and the HDACs Rpd3 and Hda1, identified highly regulated TATA-box containing genes (J. Zaugg, M. C., N. Luscombe and F. Stutz unpublished). These genes are frequently expressed in transcription bursts and their promoters undergo important chromatin rearrangements upon activation or repression, as described for *PHO84* [32,33]. Thus, a larger picture emerges suggesting that the role of noncoding transcription may be to reinforce the rapid on-off switch of highly regulated genes by promoting the formation of repressive chromatin. This process occurs in wild-type cells and is enhanced in *Δrrp6*. Further studies will address how, following a sense transcription burst, low rate antisense transcription contributes to efficient nucleosome reassembly at the promoter preventing inappropriate transcription factor binding and firing of sense transcription.

2.2.5 Materials and Methods

Strains, media and culture conditions.

The yeast strains used in this study are listed in Supplementary Table 2. Yeast strains were streaked on YEPD plates at 25 °C. Liquid cultures were inoculated with cells taken from

plates and grown at 25 °C for 16–24 h under exponential growth conditions (OD_{600} < 0.8) in YEPD or synthetic complete (SC) minimal medium.

Fluorescent *in situ* hybridization.

Fluorescent in situ hybridization procedure. Twenty-nucleotide-long DNA oligonucleotides containing a single 3′ amine were labeled post-synthesis with amine-reactive fluorescent dyes and hybridized to paraformaldehyde-fixed yeast cells as described in refs. 36 and 37. Images were acquired using epifluorescence microscopy, and three-dimensional data sets were reduced to two-dimensional data sets for image analysis. Cell segmentation, single-RNA counting and quantification of nascent transcripts were done as described in ref. 36.

Probe design and labeling. Twenty-nucleotide-long DNA oligonucleotide probes were designed using the online software Stellaris Probe Designer version 2.0 at the Biosearch Technologies website. Probes typically have a 50% GC content; however, GC content can range from 40% to 55% (for probe sequences see Supplementary Table 2). Probes were synthetized containing a single 3′ amine that can be coupled to an amine-reactive fluorescent dye. For a typical labeling reaction, 20 μg of pooled probes (31 for *PHO84* antisense, 30 for *PHO84* sense and 48 probes for *MDN1*) were lyophilized, resuspended in labeling buffer (0.1 M sodium bicarbonate, pH 9.0) and mixed with a single reactive dye pack of amine-reactive dye (DyLight amine-reactive dyes: DyLight 550 (#62263), DyLight 594 (#46413) and DyLight 650 (#62266) (Thermo Scientific)). The reaction was carried out overnight in the dark at room temperature. Labeled probes were purified using the Quiagen QIAquick Nucleotide Removal columns (Qiagen #28304) according to the manufacturer's instructions. Probe concentration and labeling efficiency were measured using NanoPhotometer Pearl (Implen) and calculated as described in ref. 38. Probes were stored at −20 °C in the dark.

Cell fixation, preparation, storage and hybridization. Cells were grown in SD complete and 2% glucose at 25 °C overnight to mid-log phase (OD_{600} = 0.6–0.8) and fixed by adding paraformaldehyde (Electron Microscopy Science #15714) to a final concentration of 4% for 45 min at room temperature. Cells were subsequently washed 3× with 10 ml of Buffer B (1.2 M sorbitol, 100 mM $KHPO_4$, pH 7.5) and stored overnight at 4 °C in Buffer B. Cell walls were

then digested with lyticase (Sigma #L2524, dissolved in 1× PBS to 25,000 U/ml. Stored at −20 °C). Digested cells were plated on poly-L-lysine–treated coverslips and stored in 70% ethanol at −20 °C in 12-well cell culture plates. Cells can be stored in 70% ethanol for several months before hybridization. For hybridization, cells were removed from 70% ethanol, washed twice with 2× saline sodium citrate (SSC) and hydrated in 10% formamide/2× SSC. Labeled probes were resuspended in 10% (v/v) formamide, 2× SSC, 1 mg ml^{-1} BSA, 10 mM ribonucleoside vanadyl complex (NEB #S1402S), 5 mM $NaHPO_4$, pH7.5, 0.5 mg ml^{-1} *Escherichia coli* tRNA and 0.5 mg ml^{-1} single-stranded DNA and hybridized overnight at 37 °C. Cells were then washed in 10% formamide/2× SSC at 37 °C for 1 h, followed by a quick wash in 1× PBS at room temperature. The coverslips were quickly dried in 100% ethanol and mounted on glass slides using Prolong Gold with DAPI mounting medium (Invitrogen #P36935). For a more detailed protocol, see refs. 36 and 38.

Image acquisition and analysis. Images were acquired using an epifluorescence microscope, either a Nikon E800 upright microscope equipped with a Photometrics CoolSNAP HQ (CCD) camera or a Zeiss Axio Observer Z1 inverse microscope with a Zeiss AxioCam MRm camera, using 100× oil-objective and specific filter cubes (Chroma Filters 31000 (DAPI), 41001 (fluorescein isothiocyanate (FITC)), SP-102v1 (Cy3/DyLight550), SP-103v1 (Cy3.5/DyLight594) and CP-104 (Cy5/DyLight650) (Chroma Technology)) corresponding to the excitation and emission spectra of the smFISH probes used. Three-dimensional image data sets were acquired, with 200 mn z stacks covering the entire depth of cells. The z stacks were projected onto a two-dimensional plane by applying a maximum projection using ImageJ. RNA signals were detected and quantified using a spot-detection algorithm fitting a two-dimensional Gaussian mask implemented with custom-made software for the IDL platform (ITT Visual Information Solutions); cell and nuclear segmentation as well as quantification of nascent RNAs were performed all as described in ref. 36. For all quantifications, data from at least three different experiments were analyzed, each containing >100 cells.

Plasmid constructions.

The *PHO84* plasmid with the mutated Nrd1 and Nab3 motifs was obtained by first cloning the *PHO84* wild-type gene (from −1,000 bp to +350 bp) as a SalI fragment into pUC18 to

create pFS3594. A BglII-Nde1 DNA fragment spanning the *PHO84* 3′ end and downstream vector sequences was synthesized by mutagenizing the putative Nrd1 and Nab3 binding motifs encoded within the *PHO84* 3′ end on the antisense strand. The wild-type BglII-Nde1 fragment of pFS3521 was replaced by the synthetic mutant fragment to create pFS3644. Both the wild-type and mutant *PHO84* were subcloned into YCpLac111 as SalI fragments to generate pFS3521 and pFS3625, respectively.

Northern blot analysis and RT-qPCR.

Total RNA was prepared and analyzed by northern blotting using standard methods as described in ref. 34. For RT-qPCR quantifications, total RNA was treated with DNase (Ambion) to remove genomic DNA contamination. cDNAs of sense or antisense RNAs were generated by SuperScript II reverse transcriptase (Invitrogen) with 1 µg of DNase-treated total RNA using gene- and strand-specific primers. cDNAs were quantified by RT-qPCR (Bio-Rad). The same amplicon was used to quantify sense and antisense cDNA. The sequences of all the primers are listed in Supplementary Table 2.

Chromatin immunoprecipitation.

ChIPs were performed essentially as described previously[34]. Yeast strains were grown to OD_{600}= 0.8 in YEPD medium at 25 °C and cross-linked for 10 min by the addition of formaldehyde to a final concentration of 1.2%. Cross-linked and sonicated chromatin extracts from 1.5 mg of Bradford-quantified proteins were immunoprecipitated overnight in the presence of protein G Sepharose (Amersham, Pharmacia) with 5 µl of antibody against H3K4me3 (Abcam 8580), H3K4me2 (Abcam 32356), H3 (Abcam 1791, clone Y47) or HA epitope (Covance monoclonal antibody HA.11, clone 16B12) for the Nrd1–HA tagged strains. All immunoprecipitations were repeated at least three times with different chromatin extracts from independent cultures. Immunoprecipitated DNA was purified and quantified by qPCR with the primers listed in Supplementary Table 2 and expressed as the percent of input DNA or percent of input DNA normalized to H3.

Determination of decay rates.

Cells were grown to OD_{600} = 0.8 in YEPD medium. At T = 0, 100 µg/ml of 1,10-phenanthroline (Sigma) was added to the culture (as described in ref. 42 and references therein). Samples were taken at different time points and analyzed for their RNA expression by RT-qPCR as described above.

Half-lives were calculated by the equation $t_{1/2} = 0.693/k$, in which k is the rate constant for mRNA decay. Values of each time point are normalized for internal variations with *SCR1* RNA, a control that is still stable at the 30-min time point.

2.2.6 Acknowledgements

We thank T. H. Jensen (Aarhus University, Denmark), D. Libri (Centre National de la Recherche Scientifique, Gif-sur-Yvette, France), A. Morillon (Curie Institute, Paris, France) and D. Tollervey (University of Edinburgh, UK) for strains; K. Weis (University of California, Berkeley) for communicating data before publication, D. Larson (National Cancer Institute, Bethesda, USA) for updates of image analysis software and M. Oeffinger, F. Robert, O. Gahura, A. Maffioletti, C. Dargemont and V. Géli for discussions and critical reading of the manuscript. This work was supported by SystemsX and Novartis fellowships (MC), an EMBO fellowship (EG), the Swiss National Science Foundation (grant no. 31003A_130292), the National Center of Competence in Research "Frontiers in Genetics", IGE3 and the Canton of Geneva (FS), and the Canadian Institutes of Health Research (MOP-BMB-232642), the Canadian Foundation for Innovation and the 'Fonds de recherche du Québec – Santé' (Chercheur Boursier Junior I) (DZ).

2.2.7 Author contributions

MC performed ChiP and RNA analyses; SR performed the smFISH experiments; EG prepared strains and performed RNA analyses; VI made mutants and RNA analyses. MC, SR, DZ and FS analyzed the data; DZ and FS supervised the project and wrote the manuscript.

2.2.8 References

1. Neil, H. et al. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**, 1038-42 (2009).

2. David, L. et al. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A* **103**, 5320-5 (2006).

3. Xu, Z. et al. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033-7 (2009).

4. Jacquier, A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet* **10**, 833-44 (2009).

5. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763-76 (2009).

6. LaCava, J. et al. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713-24 (2005).

7. Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J. & Libri, D. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol Cell* **23**, 853-64 (2006).

8. Vanacova, S. et al. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol* **3**, e189 (2005).

9. Wyers, F. et al. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725-37 (2005).

10. Arigo, J.T., Eyler, D.E., Carroll, K.L. & Corden, J.L. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell* **23**, 841-51 (2006).

11. Vasiljeva, L. & Buratowski, S. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell* **21**, 239-48 (2006).

12. Carroll, K.L., Ghirlando, R., Ames, J.M. & Corden, J.L. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *Rna* **13**, 361-73 (2007).

13. Steinmetz, E.J., Conrad, N.K., Brow, D.A. & Corden, J.L. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413**, 327-31 (2001).

14. Gudipati, R.K., Villa, T., Boulay, J. & Libri, D. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat Struct Mol Biol* **15**, 786-94 (2008).

15. Kim, H. et al. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* **17**, 1279-86 (2010).

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

17. Xu, Z. et al. Antisense expression increases gene expression variability and locus interdependency. *Mol Syst Biol* **7**, 468 (2011).

18. Murray, S.C. et al. A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res* **40**, 2432-44 (2011).

19. Tisseur, M., Kwapisz, M. & Morillon, A. Pervasive transcription - Lessons from yeast. *Biochimie* **93**, 1889-96 (2011).

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

21. Martens, J.A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature* **429**, 571-4 (2004).

22. Thiebaut, M. et al. Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in S. cerevisiae. *Molecular Cell* **31**, 671-82 (2008).

23. Bumgarner, S.L. et al. Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment. *Mol Cell* **45**, 470-82 (2012).

24. van Werven, F.J. et al. Transcription of Two Long Noncoding RNAs Mediates Mating-Type Control of Gametogenesis in Budding Yeast. *Cell* (2012).

25. Gelfand, B. et al. Regulated antisense transcription controls expression of cell-type-specific genes in yeast. *Mol Cell Biol* **31**, 1701-9 (2011).

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

27. Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D. & Vogelauer, M. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Mol Cell* **32**, 685-95 (2008).

28. Pinskaya, M., Gourvennec, S. & Morillon, A. H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *Embo J* **28**, 1697-707 (2009).

29. Kim, T., Xu, Z., Clauder-Munster, S., Steinmetz, L.M. & Buratowski, S. Set3 HDAC Mediates Effects of Overlapping Noncoding Transcription on Gene Induction Kinetics. *Cell* (2012).

30. Weiner, A. et al. Systematic dissection of roles for chromatin regulators in a yeast stress response. *PLoS Biol* **10**, e1001369 (2012).

31. Komeili, A. & O'Shea, E.K. Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. *Science* **284**, 977-80 (1999).

32. Lam, F.H., Steger, D.J. & O'Shea, E.K. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246-50 (2008).

33. Wippo, C.J. et al. Differential cofactor requirements for histone eviction from two nucleosomes at the yeast PHO84 promoter are determined by intrinsic nucleosome stability. *Mol Cell Biol* **29**, 2960-81 (2009).

34. Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G. & Stutz, F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell* **131**, 706-17 (2007).

35. Femino, A.M., Fay, F.S., Fogarty, K. & Singer, R.H. Visualization of single RNA transcripts in situ. *Science* **280**, 585-90 (1998).

36. Zenklusen, D., Larson, D.R. & Singer, R.H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* **15**, 1263-71 (2008).

37. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5**, 877-9 (2008).

38. Zenklusen, D. & Singer, R.H. Analyzing mRNA expression using single mRNA resolution fluorescent in situ hybridization. *Methods Enzymol* **470**, 641-59 (2010).

39. Oeffinger, M. & Zenklusen, D. To the pore and through the pore: A story of mRNA export kinetics. *Biochim Biophys Acta* **1819**, 494-506 (2012).

40. Rougemaille, M. et al. Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. *EMBO J* **26**, 2317-26 (2007).

41. Nonet, M., Scafe, C., Sexton, J. & Young, R. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol Cell Biol* **7**, 1602-11 (1987).

42. Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M.D. & Hughes, T.R. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol* **24**, 5534-47 (2004).

43. Krogan, N.J. et al. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* **11**, 721-9 (2003).

44. Ng, H.H., Robert, F., Young, R.A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11**, 709-19 (2003).

45. Creamer, T.J. et al. Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet* **7**, e1002329 (2011).

46. Wlotzka, W., Kudla, G., Granneman, S. & Tollervey, D. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *Embo J* **30**, 1790-803 (2011).

47. Soares, L.M. & Buratowski, S. Yeast Swd2 Is Essential Because of Antagonism between Set1 Histone Methyltransferase Complex and APT (Associated with Pta1) Termination Factor. *J Biol Chem* **287**, 15219-31 (2012).

48. Camblong, J. et al. Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. *Genes Dev* **23**, 1534-45 (2009).

49. Margaritis, T. et al. Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. *PLoS Genet* **8**, e1002952 (2012).

50. Grzechnik, P. & Kufel, J. Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol Cell* **32**, 247-58 (2008).

51. Gudipati, R.K. et al. Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells. *Mol Cell* **48**, 409-421 (2012).

52. Darby, M.M., Serebreni, L., Pan, X., Boeke, J.D. & Corden, J.L. The S. cerevisiae Nrd1-Nab3 Transcription Termination Pathway Acts in Opposition to Ras Signaling and Mediates Response to Nutrient Depletion. *Mol Cell Biol* **32**, 1762-1775 (2012).

53. Lardenois, A. et al. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proc Natl Acad Sci U S A* **108**, 1058-63 (2011).

54. van Dijk, E.L. et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475**, 114-7 (2011).

55. Buratowski, S. & Kim, T. The role of cotranscriptional histone methylations. *Cold Spring Harb Symp Quant Biol* **75**, 95-102 (2010).

56. Holstege, F.C. et al. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717-28 (1998).

57. Wang, Y. et al. Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A* **99**, 5860-5 (2002)**.**

58. Jimeno, S., Rondon, A.G., Luna, R. & Aguilera, A. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *Embo J* **21**, 3526-35 (2002)**.**

59. Houseley, J. & Tollervey, D. Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Rep* **7**, 205-11 (2006).

2.2.9 Figure legends

Figure 2-1: *PHO84* sense and antisense expression are anti-correlated.

(a) Deletion of *RRP6* increases expression of *PHO84* antisense RNA. Northern blot (left) and RT-qPCR (right) analyses in WT and Δ*rrp6*. Error bars reflect standard deviations of an average from 3 independent experiments. **(b)** Two possible models for antisense-mediated gene silencing. In a graded response, gradual accumulation of *PHO84* AS RNAs leads to a gradual reduction in sense levels (Top left), whereas in a bimodal response, sense and antisense expression are anti-correlated (Top right). Changes in *PHO84* mRNA and antisense levels over time are represented as green and red lines respectively. Lower panels show *PHO84* sense expression resulting from graded or bimodal regulation at the single cell level. **(c)** Bimodal expression of *PHO84* sense and antisense RNAs. smFISH detects *PHO84* sense (green) and antisense RNAs (red) in individual WT and Δ*rrp6* cells. Nuclear DNA was stained using DAPI (blue), cellular outlines were visualized using DIC optics and the scale bar is 5μm. FISH probes positions are drawn at the top. **(d)** Less cells express *PHO84* sense in Δ*rrp6*. Frequency distribution of *PHO84* sense and AS expression in individual cells from (**c**). **(e)** Deletion of *RRP6* results in higher level of *PHO84* AS RNA in single cells. Frequency distribution of the number of *PHO84* AS RNAs per cell in WT and Δ*rrp6*. Error bars in (**d**) and (**e**) reflect standard deviations of an average of three independent experiments.

Figure 2-2: *PHO84* antisense RNAs do not accumulate at the *PHO84* locus.

(a) *PHO84* AS RNAs are exported to the cytoplasm. smFISH for *MDN1* mRNA (green) and *PHO84* AS RNA (red) in WT and *Δrrp6* cells. Scale bar is 5μm. The cartoon on the left illustrates the detection of nascent and cytoplasmic RNAs. **(b)** *PHO84* AS RNAs do not accumulate at the *PHO84* gene locus. Frequency distribution of the number of nascent RNAs for *MDN1* (WT only) and *PHO84* AS RNAs in WT and *Δrrp6*. **(c)** Deletion of *RRP6* leads to a higher frequency of cells showing nascent *PHO84* AS RNAs. Frequency distribution of cells containing nascent (nuclear) *PHO84* AS RNAs in WT and *Δrrp6*. Error bars in (**b**) and (**c**) reflect standard deviations of an average of three independent experiments.

Figure 2-3: *PHO84* antisense RNAs are polyadenylated by Pap1.

(a) Inactivation of Pap1 leads to nuclear accumulation of *PHO84* AS RNAs. smFISH using probes against *PHO84* AS RNAs (red) in *pap1-1* and *pap1-1Δrrp6* cells grown at 25°C and either directly fixed or shifted to 37°C for 1 hour prior to fixation. Nuclear DNA was stained using DAPI (blue), cellular outlines were visualized using DIC optics and the scale bar is 5μm. **(b)** *pap1-1Δrrp6* cells accumulate high numbers of *PHO84* AS RNA in the nucleus. Frequency distribution of the number of *PHO84* AS RNAs detected by smFISH in *pap1-1* and *pap1-1Δrrp6* cells after 1 hour shift to 37°C. **(c)** *PHO84* AS RNAs polyadenylation requires Pap1. Northern blot membranes with oligo dT purified total RNA were hybridized with *PHO84* AS specific probes. Strains were exponentially grown in SC medium 2% glucose (Glu; lanes 1-4) or 2% galactose (Gal; lane 5) followed by 20h in 2% glucose (Glu; lane 6) to deplete Trf5 as indicated. *ACT1* and *TRF4* mRNA specific probes were used to control for loading and *TRF4* deletion.

Figure 2-4: *PHO84* antisense nuclear detection needs ongoing transcription.

smFISH detecting *MDN1* mRNA (green) and *PHO84* AS RNA (red) in *rpb1-1* and *rpb1-1Δrrp6* cells grown at 25°C and shifted to 37°C for 5, 10 and 20 min prior to fixation. Nuclear DNA was stained using DAPI (blue), cellular outlines were visualized using DIC optics and the scale bar is 5μm.

Figure 2-5: Effect of *Δrrp6* on antisense RNA half-life and transcription.

(a) Deletion of *RRP6* does not alter *PHO84* AS RNA half-life. RT-PCR analysis measuring *PHO84* AS RNA decay rates in WT, *Δrrp6* and *Δxrn1* after transcription shut off by adding 100 μg/ml 1,10-Phenantroline to the medium. *PHO84* AS RNA levels were normalized to *SCR1* RNA, stable at 30 min. Data are expressed as a percentage of the amounts present before addition of the inhibitor. Error bars represent standard deviations for three independent experiments. **(b)** *PHO84* AS RNA levels are elevated in a *Δrrp6 and Δxrn1*. Antisense RNA levels were measured by RT-qPCR and expressed relative to the levels in WT that were set to 1. Error bars reflect standard deviations of an average obtained from three independent experiments. **(c)** Higher levels of H3K4 tri- and di-methylation at the 3' end of *PHO84* in

Δrrp6. Chromatin immunoprecipitation (ChIP) analysis of H3K4 tri- (top) and di-methylation (bottom) at the *PHO84* locus. ChIP with anti-H3K4me3, anti-H3K4me2 or anti-H3 antibodies from *Δpho4, Δpho4Δrrp6, Δpho4Δset1* and *Δpho4Δrrp6Δset1* strains. DNA quantified by real-time PCR with primers specific for the 5', middle and 3' regions of *PHO84* and *ACT1* (as indicated on top). H3K4me2/3 values were normalized to H3 values and the highest value was arbitrarily set to 1. Error bars reflect standard deviations of an average obtained from three independent experiments. Comparison of the mean differences was analysed by the Student-t test. P values <0.05 are indicated by (*).

Figure 2-6: *PHO84* antisense transcription is attenuated by NNS.

(a) NNS terminates short *PHO84* AS transcripts. Cartoon illustrating the role of NNS in *PHO84* AS transcription. Short antisense RNAs (red line) previously shown to be polyadenylated by Trf4 and degraded by Rrp6 [1] are proposed here to be terminated at Nrd1-Nab3 motifs (blue bars) by the NNS complex, while the long read-through antisense transcripts (green lines) are subjected to 3'end cleavage and polyadenylation by Pap1 before export into the cytoplasm. **(b)** Depletion of Nrd1 increases *PHO84* AS RNA levels. *PHO84* AS RNA levels were measured using RT-qPCR after in GAL-Nrd1 and GAL-Nrd1 Δ*rrp6* strains grown in medium containing 2% galactose (Gal) or shifted for 7h in 2% glucose (Glu) to deplete Nrd1. Error bars reflect standard deviations of an average from 3 independent experiments. Comparison of the mean differences was analysed by the Student-t test. Stars indicate the level of significance: p value <0.01 (**). The value of GAL-Nrd1 (Glu 7h) was arbitrarily set to 1. **(c)** Deletion of *RRP6* reduces Nrd1 recruitment. ChIP analysis of Nrd1-HA binding at *PHO84* 3' end quantified by qPCR and expressed as % of input. Three biological and two technical repeats were analysed, error bars reflect standard errors. Comparison of the mean differences was analysed by the Student-t test. Stars indicate the level of significance: p value <0.01 (**). The small panel shows Western blot analysis of Nrd1 protein levels in the strains used for the ChIP. Ssn6 was used as normalization control.

2.2.10 Figures

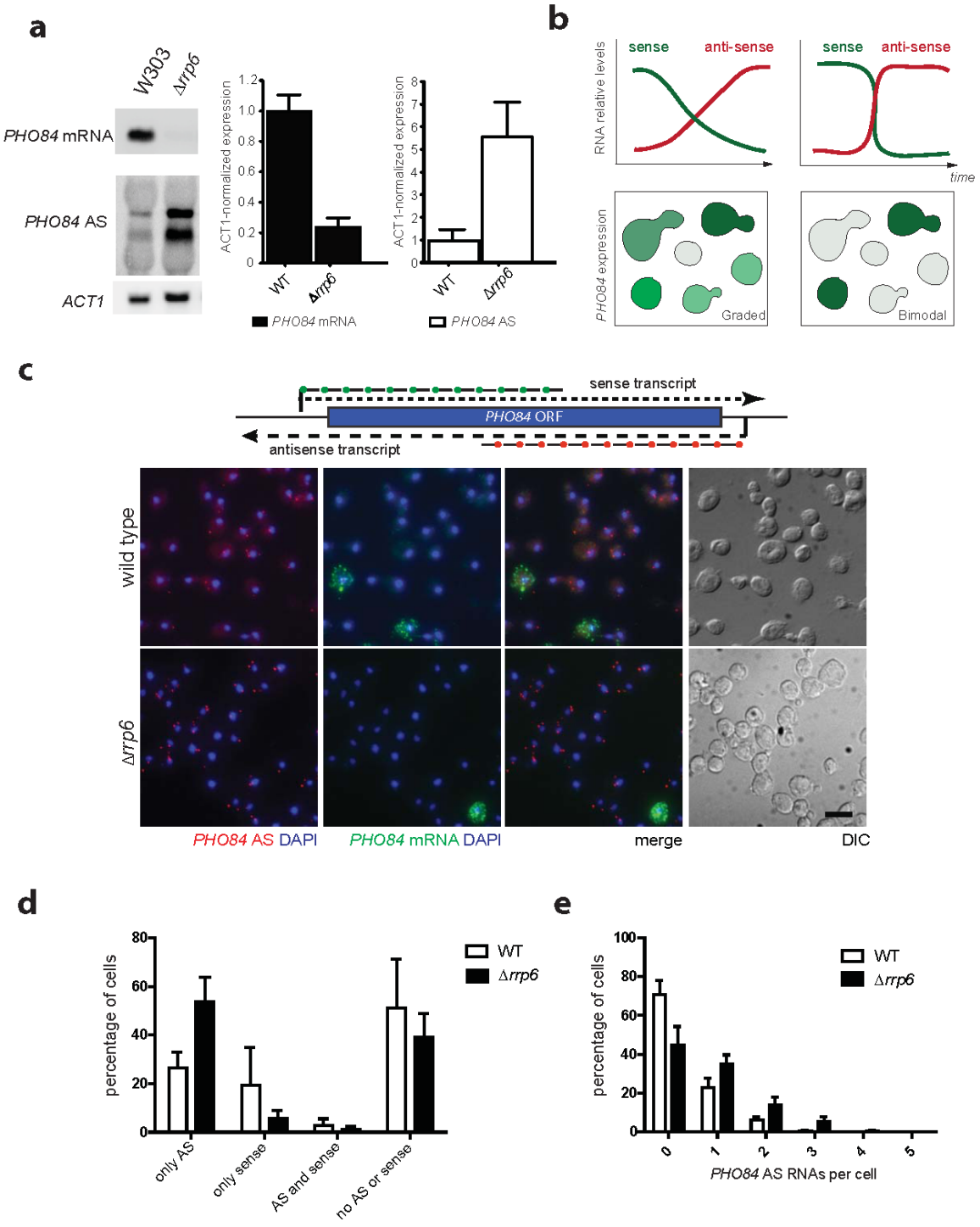


Figure 2-1: *PHO84* sense and antisense expression are anti-correlated

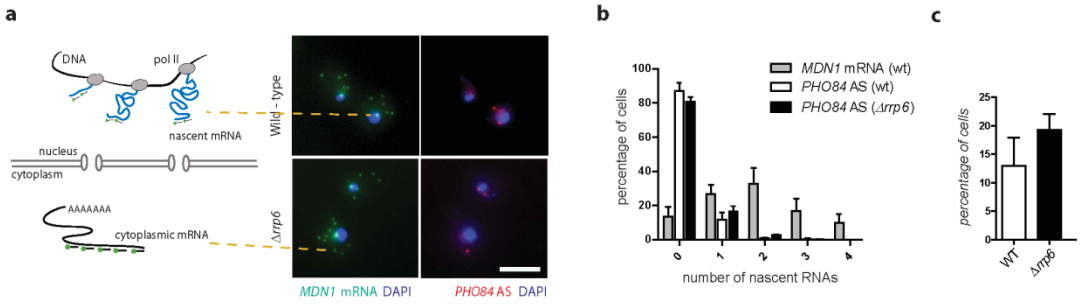


Figure 2-2: *PHO84* antisense RNAs do not accumulate at the PHO84 locus

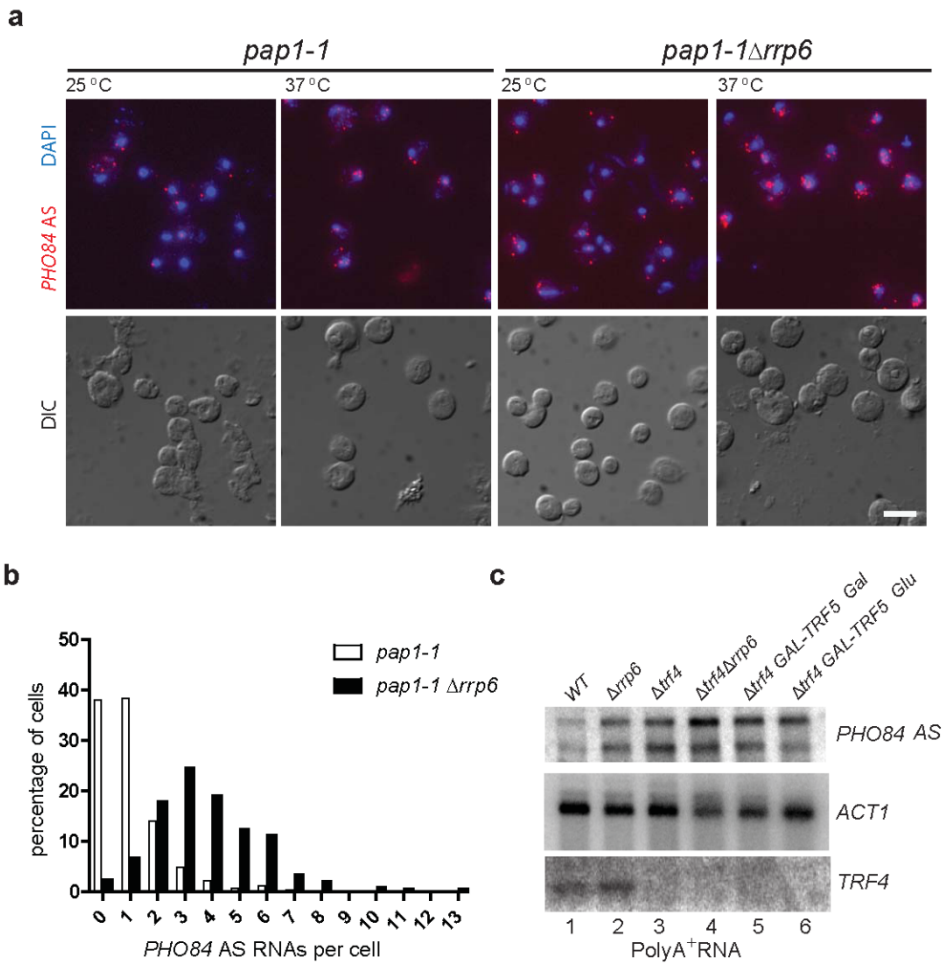


Figure 2-3: *PHO84* antisense RNAs are polyadenylated by Pap1

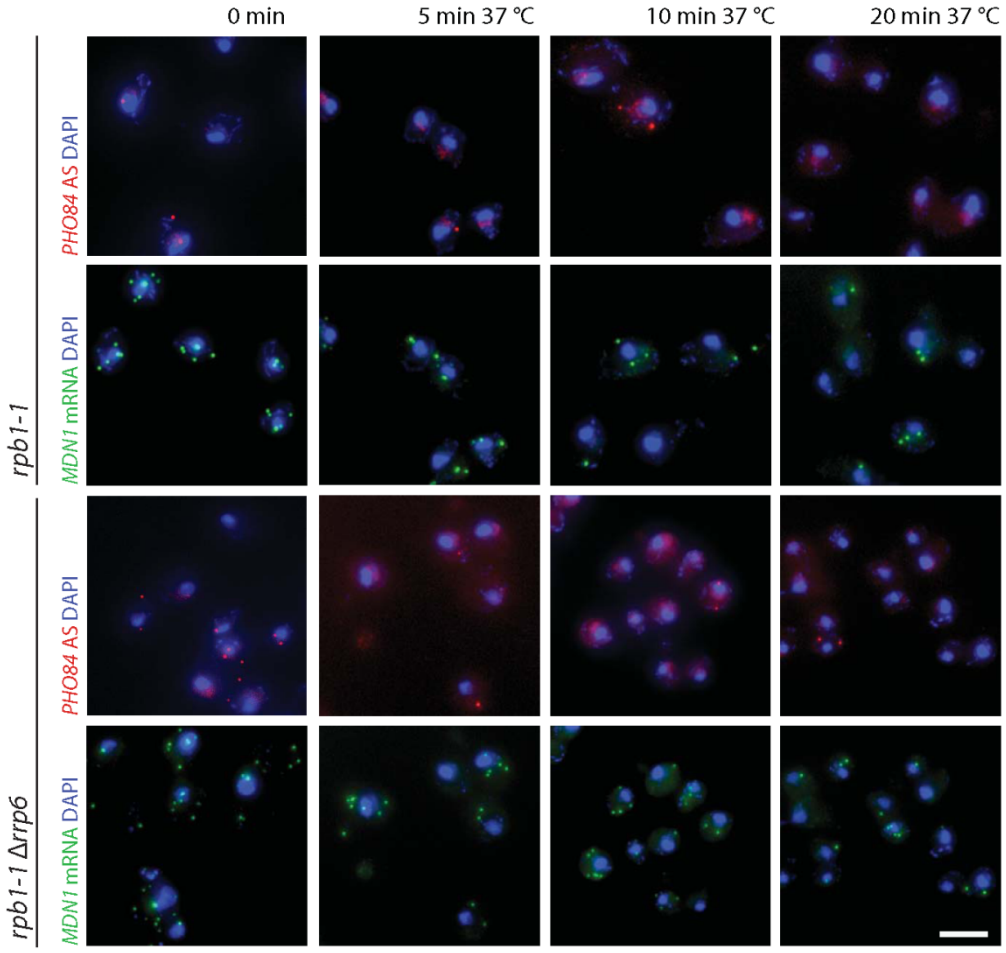


Figure 2-4: *PHO84* antisense nuclear detection needs ongoing transcription

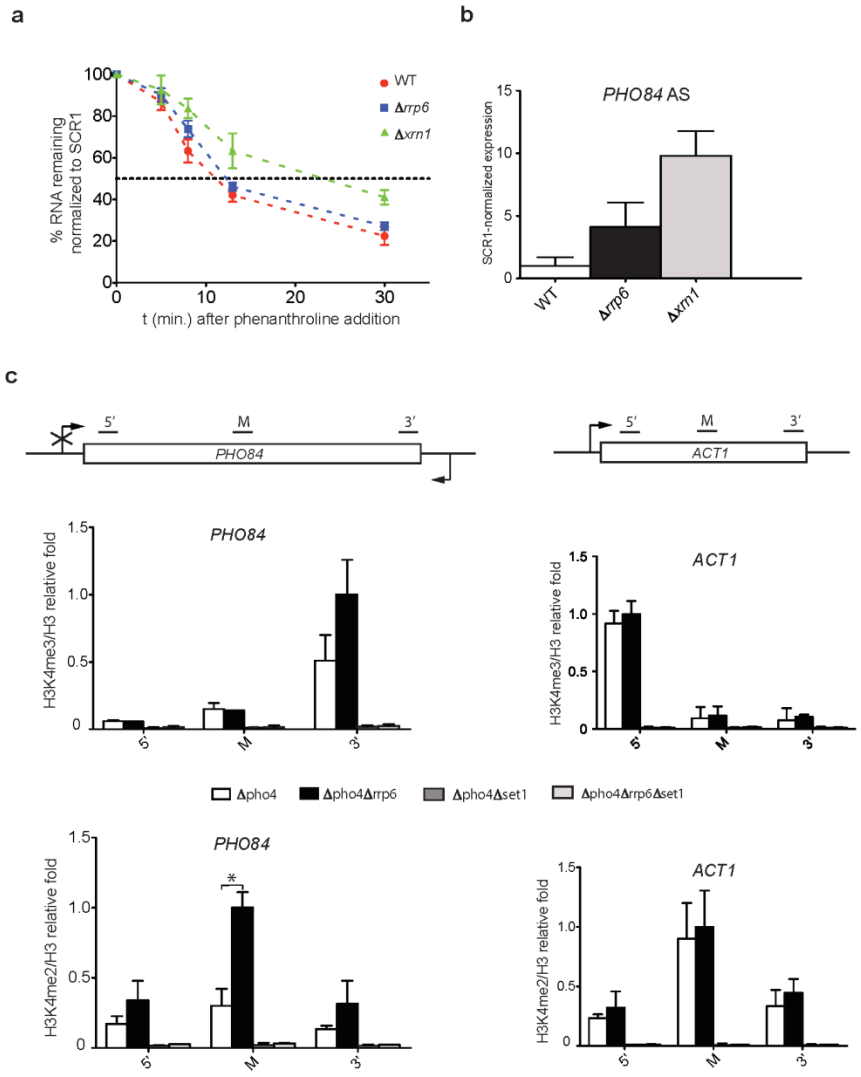


Figure 2-5: Effect of Δ*rrp6* on antisense RNA half-life and transcription

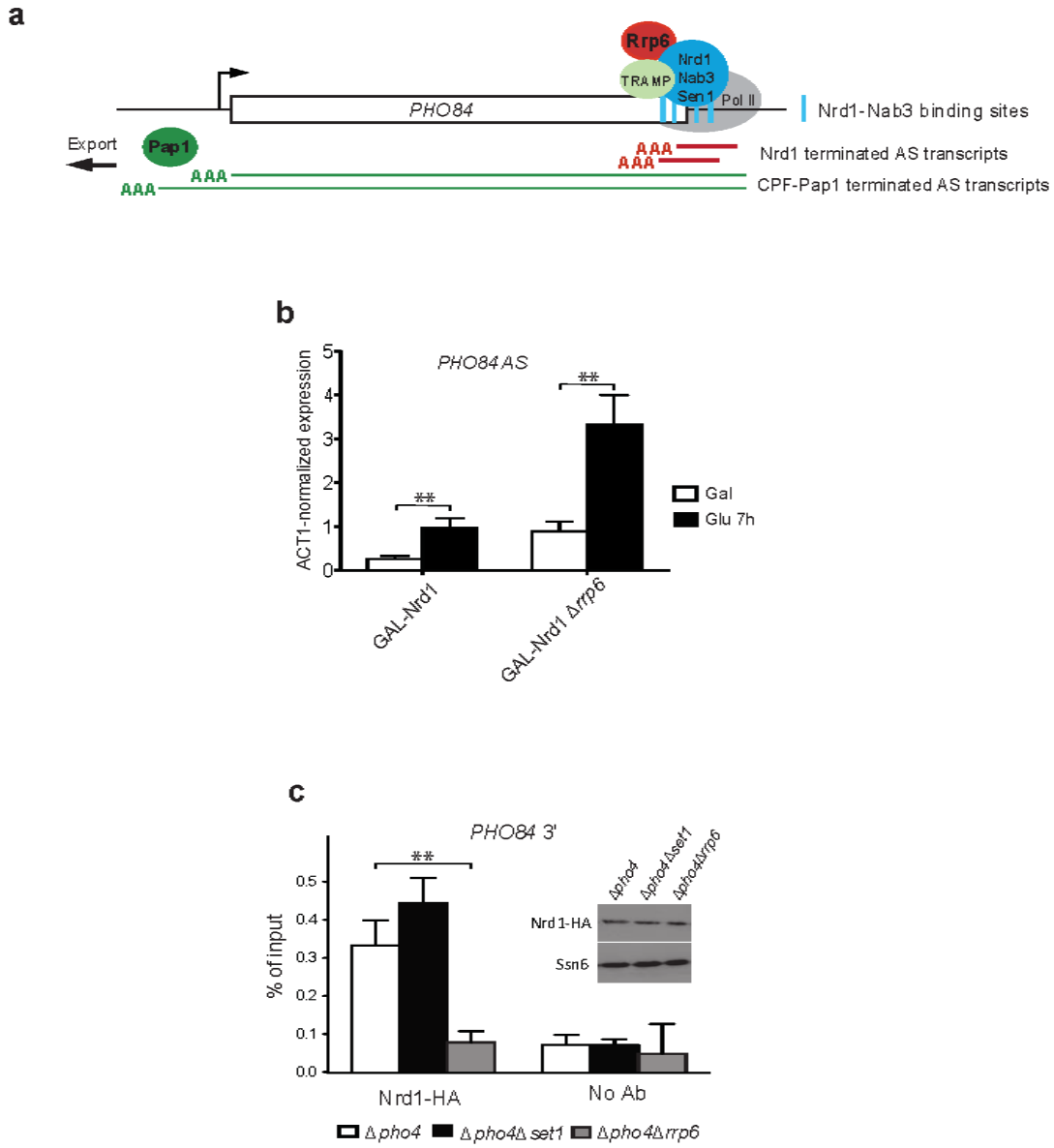
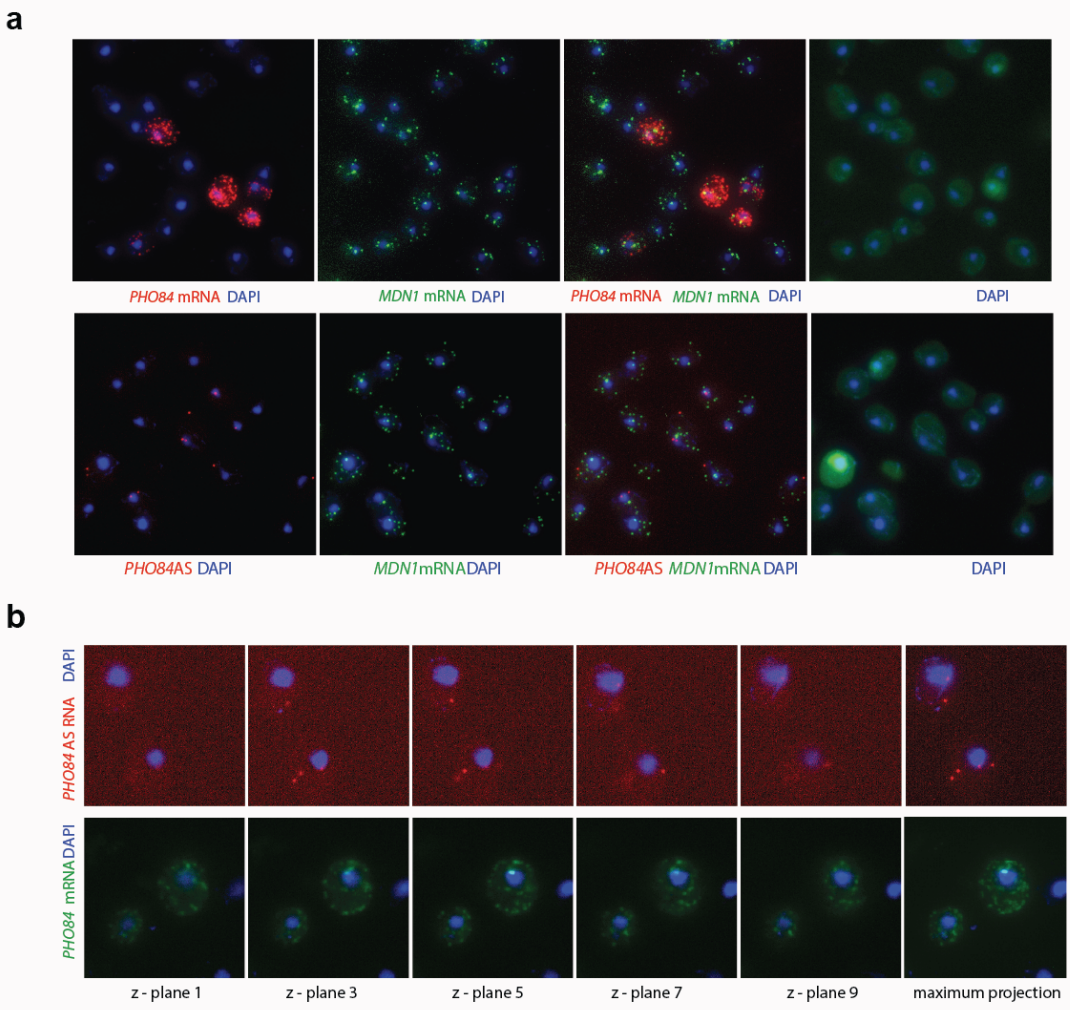


Figure 2-6: *PHO84* antisense transcription is attenuated by NNS

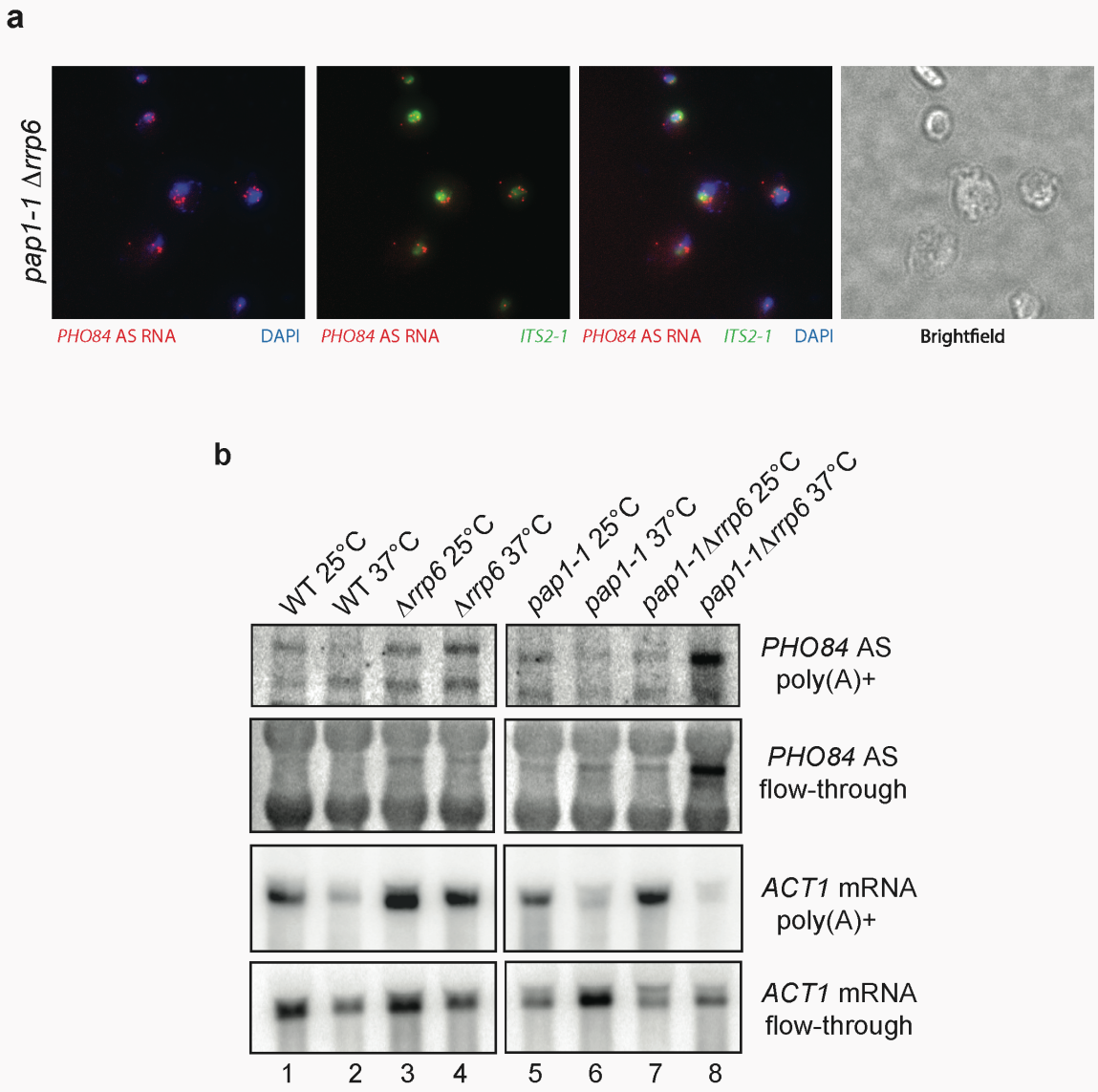
2.2.11 Supplementary data

Supplementary Figure 1



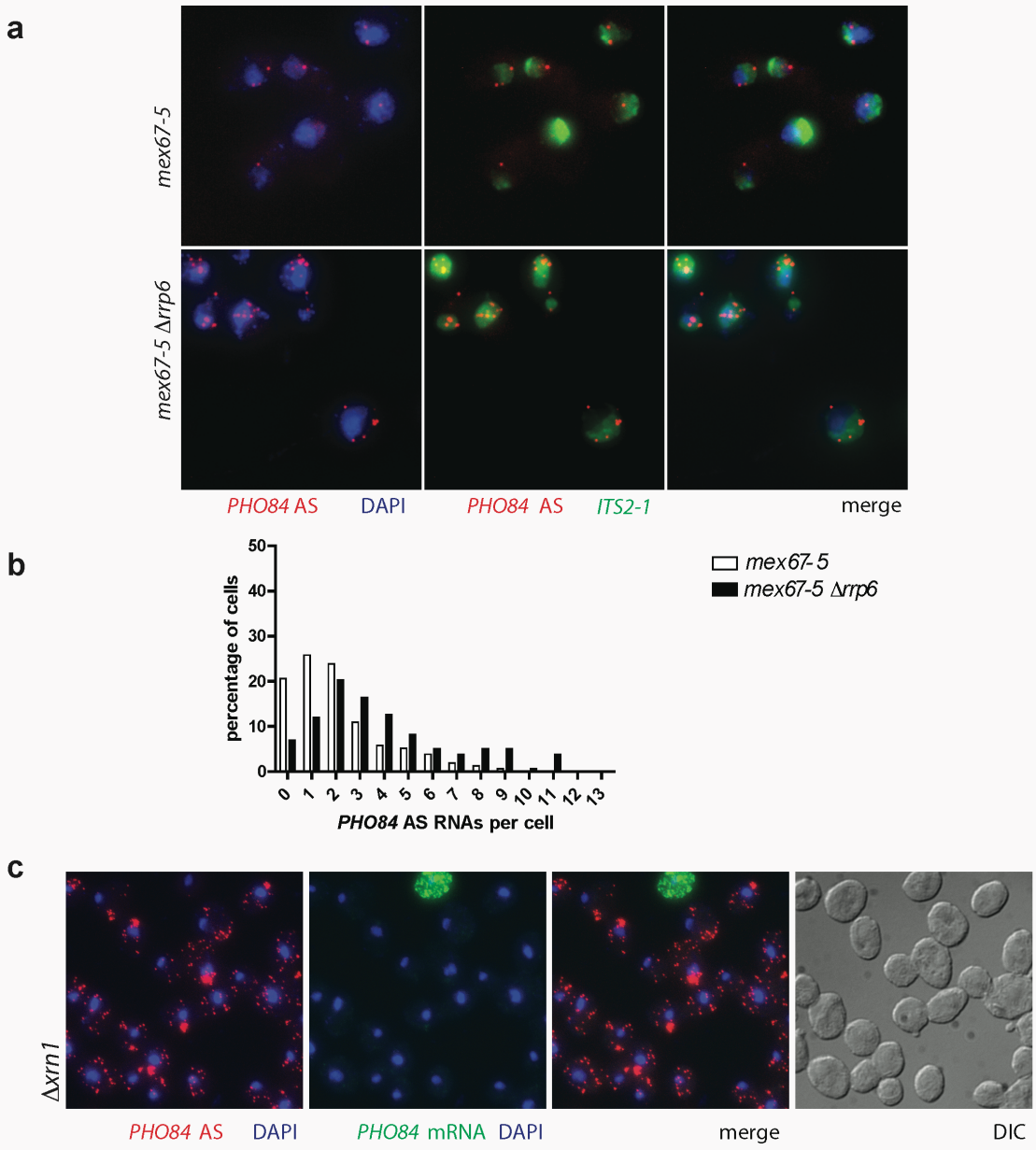
Supplementary Figure 1: smFISH of *MDN1* mRNAs and *PHO84* nascent transcripts.
a) *MDN1* mRNAs are detected in all cells, also in *PHO84* sense and/or *PHO84* AS RNA negative cells. smFISH using probes against *PHO84* sense (upper panel) or *PHO84* AS RNAs (lowe panel) and *MDN1* mRNAs were performed as shown in Figure 1. Nuclear DNA is stained using DAPI, cellular autofluorescence in the GFP chanel is used to show the cellular boundaries. **(b)** *PHO84* sense and AS RNA loci are always located at nuclear periphery, consistent with the location of the of the *PHO84* gene close to the subtelomeric region of Chromosome XIII. Different z-sections as well as the maximum projected image are shown. Images were acquired in 200nm steps, only every second z-plane is shown. Importantly, RNA signals are rarely observed in the nucleus, except at the siteof transcription, suggesting that RNA nuclear export is fast.

Supplementary Figure 2

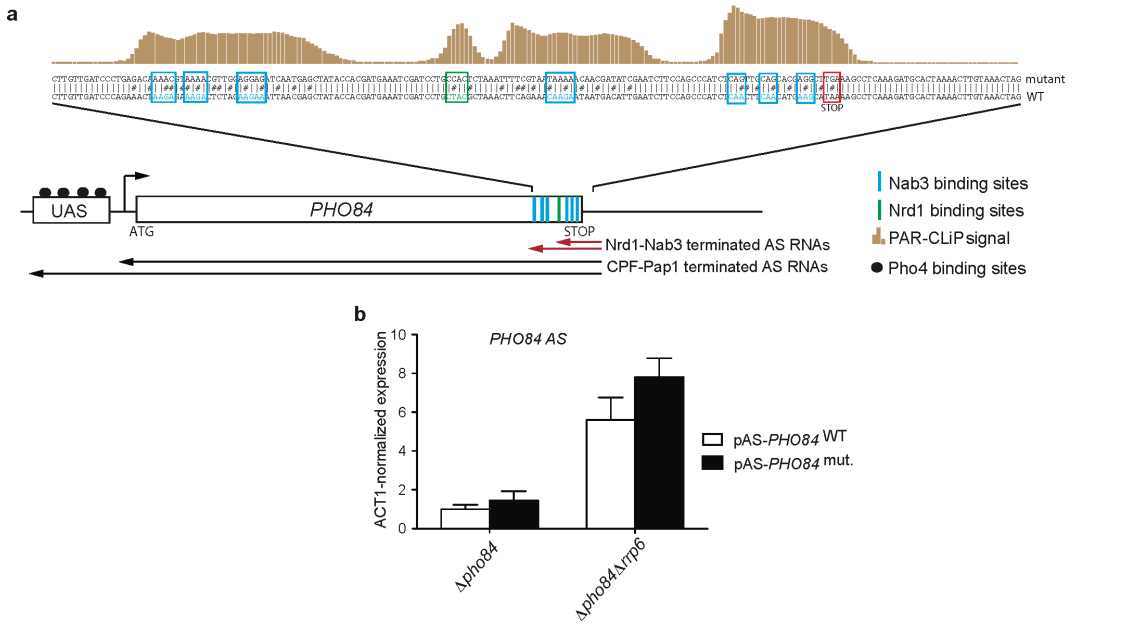


Supplementary Figure 2: *PHO84* AS RNAs are polyadenylated by Pap1.
(a) *PHO84* AS RNAs frequently accumulate in the nucleolus in a *pap1-1Δrrp6* strain after 1h at 37°C. *pap1-1Δ rrp6* cells were grown at 25°C in SD complete and shifted to 37°C for 1h prior to fixation followed by hybridization of *PHO84* AS (red) and probes against the ribosomal pre-rRNA spacer sequence *ITS2-1* (green) as a molecular marker. Nuclear DNA was stained using DAPI (blue) and cellular outlines were visualized by bright-field imaging. **(b)** In the absence of Pap1, *PHO84* AS RNAs are polyadenylated and degraded by the surveilling exosome, presumably following polyadenylation by Trf4/5. Northern blot membranes with oligo dT purified poly(A)+ RNA or polyA minus (flow-through) from the indicated strains grown at 25°C or shifted to 37°C for 1h were hybridized with *PHO84* AS specific probes as described in Figure 1. Membranes were rehybridized with *ACT1* random labelled probes to control for loading and for phenotype. In *pap1-1* at 37°C, *ACT1* mRNA is reduced in level probably due to the defect in 3' end formation.

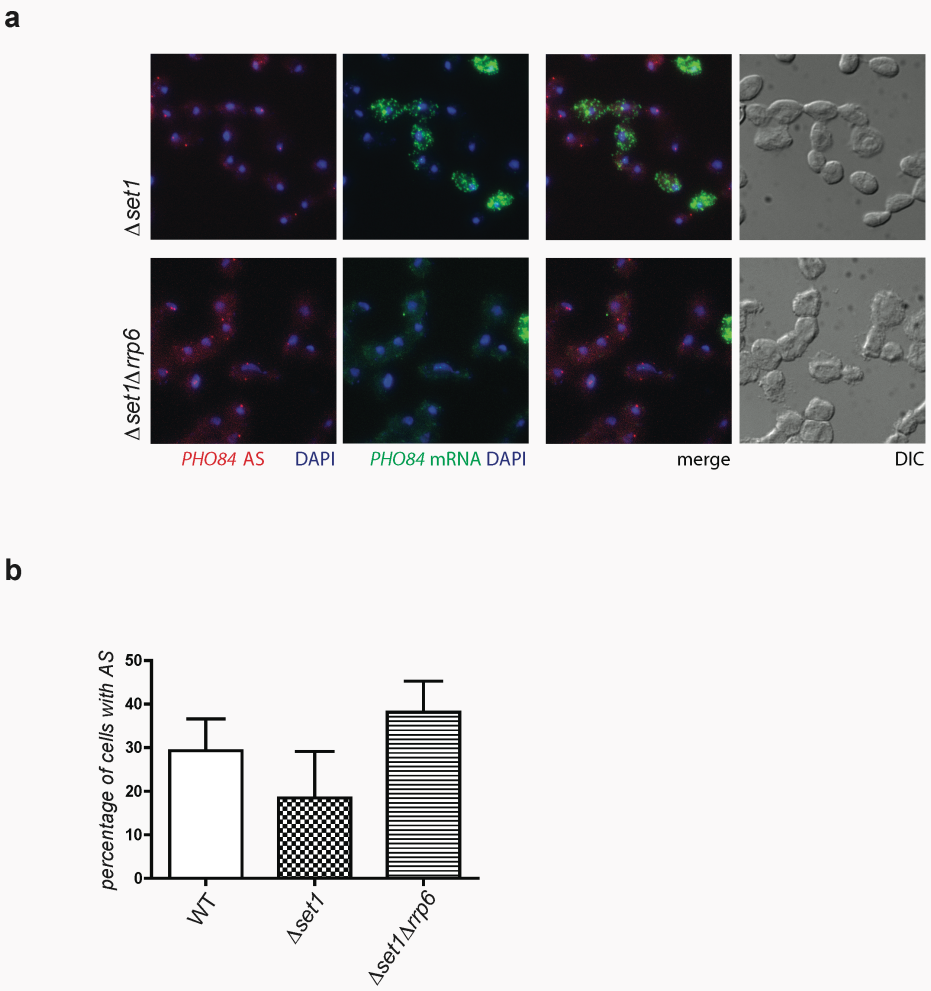
Supplementary Figure 3



Supplementary Figure 3: *PHO84* AS RNAs have mRNA-like properties.
(a) *PHO84* AS RNAs are exported to the cytoplasm by the mRNA export receptor Mex67. *mex67-5 and mex67-5 Δrrp6* cells were grown at 25°C in SD complete and shifted to 37°C for 1 hour prior to fixation followed by hybridization with *PHO84* AS probes as shown in Figure 1c. The nucleolus was stained using a *ITS2-1* probe (green), and nuclear DNA visualized by DAPI stain (blue). **(b)** Frequency disbribution of *PHO84* AS RNA expression levels after 1 hour heat shock at 37°C in *mex67-5* and *mex67-5 Δrrp6* cells. The numbers of *PHO84* AS RNAs were determined for > 100 cells.
(c) *PHO84* AS RNAs are degraded in the cytoplasm in an Xrn1 dependent manner. Δ*xrn1* cells were grown in SD complete medium at 25°C, fixed and hybridized with FISH probes complementary to *PHO84* sense and AS RNAs as shown in Figure 1c.

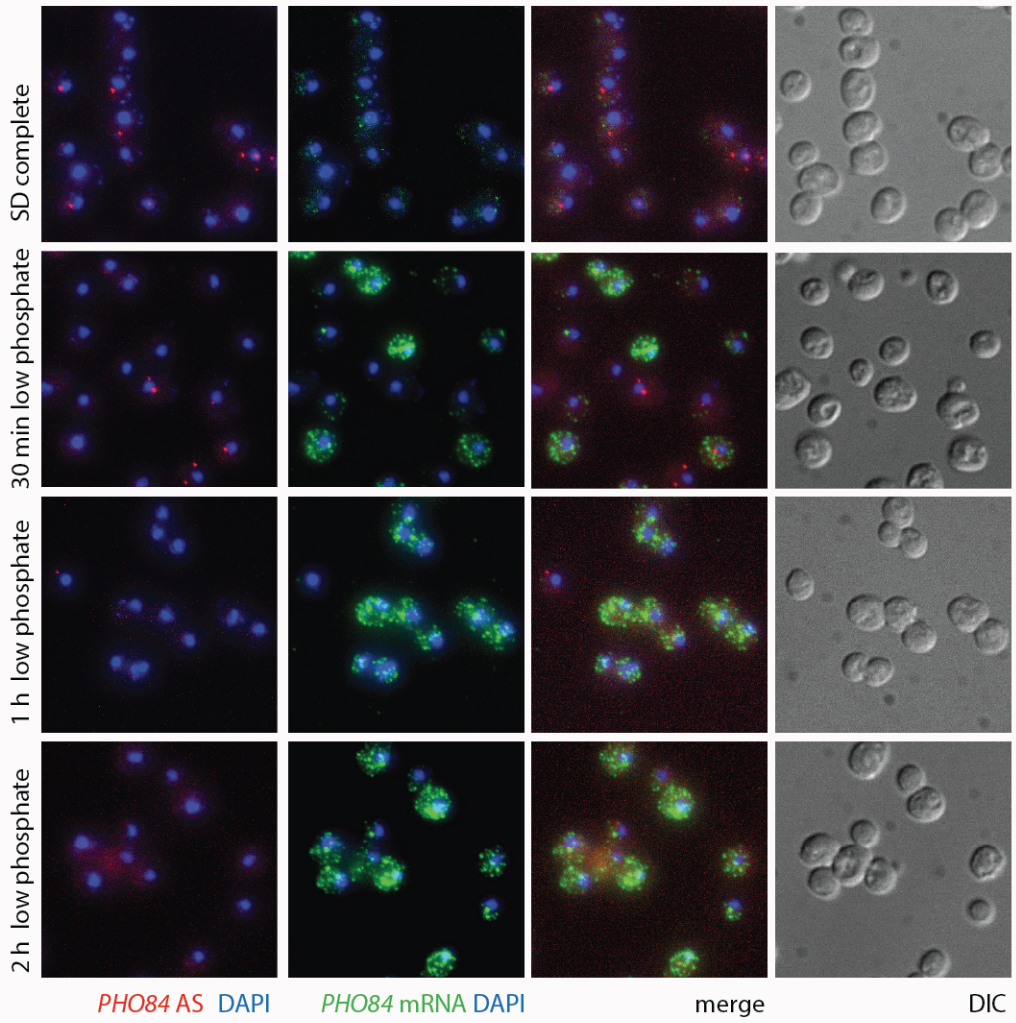


Supplementary Figure 4: Nrd1 and Nab3 binding site mutations increase AS RNA production.
(a) Schematic view of the mutagenesis of Nrd1 and Nab3 binding sites at the *PHO84* 3'end. The Nrd1 PAR-CLiP signal detected over the same region (Creamer et al., 2011) [45] is indicated above. Nrd1 and Nab3 recognition motifs were defined based on the PAR-CliP binding data and the consensus binding sites derived from these analyses (Creamer et al., 2011) [45]. The WT and mutated *PHO84* coding strands are shown. The mutations introduced maintain the *PHO84* open reading frame. The sequence of the AS RNA corresponds to the reverse and complement of the indicated sequences. The 3' ends of early terminated *PHO84* AS transcripts have been described (Neil et al., 2009) [1]. **(b)** Analysis of *PHO84* AS RNA levels in a Δ*pho84* strain transformed with a *PHO84* plasmid containing the wild-type (pAS-*PHO84* WT) or the mutant (pAS-*PHO84* mut.) sequence, in which Nrd1 and Nab3 motifs on the AS orientation are mutagenized. Plasmid encoded *PHO84* AS transcripts were measured by RT-qPCR with specific primers as described in Methods. CPF: cleavage and polyadenlytion factor.



Supplementary Figure 5: Opposite effects of Set1 and Rrp6 on *PHO84* AS RNA levels.
(a) smFISH for *PHO84* sense and AS RNA in Δ*set1* and Δ*set1*Δ*rrp6* cells. Cells were grown, fixed and hybridized to probes specific for *PHO84* sense (green) and AS RNAs (red) as shown in Figure 1c.
(b) Percentage of cells expressing *PHO84* AS RNA in individual cells of the indicated strains.

Supplementary Figure 6



Supplementary Figure 6: Low phosphate medium prevents *PHO84* AS RNA expression.
Strong induction of *PHO84* sense transcription in low phosphate medium prevents *PHO84* antisense expression. Cells were grown in SD complete medium and shifted to low phosphate medium for 0.5, 1 and 2 hours prior to fixation. smFISH is shown in green for *PHO84* sense and red for *PHO84* AS RNAs. DNA was stained using DAPI (blue) and cellular outlines were visualized using DIC optics.

	Mean number of total *PHO84* AS RNAs per cell	Half-life (min) measured by 1,10 Phenanthroline induced transcription inhibition	Transcription frequency (RNAs/hour)
Wild type	0.36±0.098	11.4±3.9	1.31
Δ*rrp6*	0.82±0.18	12±2.8	2.84

Supplementary Table 1: Transcription frequency in wild-type and *Δrrp6* cells. mRNA half-life was calculated using mean *PHO84* AS expression levels in wild-type and *Δrrp6* strains measured by smFISH and decay rates measured by qRT-PCR after transcription shutoff by 1,10 Phenanthroline. Assuming mRNA decay follows first-order kinetics, transcription frequency can be calculated using (Holstege et al. 1998; Wang et al. 2002) [56,57] :

Transcription frequency = Ln2 x (steady state RNA level)/half life (min)

Supplementary table 2: Strains, primers, and smFISH probes used in this study

STRAINS USED IN THIS STUDY

Code	Name	Genotype	Reference
W303 background			
FSY1742	WT	*MATa ade2 his3 leu2 trp1 ura3*	
FSY3117	*Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 Δrrp6::Kanr*	Camblong et al., 2007 [43]
FSY3518	*Δhda2*	*MATa ade2 his3 leu2 trp1 ura3 Δhda2::TRP1*	Camblong et al., 2007 [43]
FSY3018	*Δhda2Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 Δrrp6::Kanr Δhda2::TRP1*	this study
FSY3517	*Δset1*	*MATa ade2 his3 leu2 trp1 ura3 Δset1::TRP1*	this study
FSY3833	*Δset1Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 Δrrp6::Kanr Δset1::TRP1*	this study
FSY1982	*mex67-5*	*MATa ade2 his3 leu2 trp1 ura3 mex67-5 integrated*	Jimeno et al., 2002 [59]
FSY1985	*mex67-5 Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 mex67-5 integrated Δrrp6::Kanr*	this study
FSY2078	*Δxrn1*	*MATa ade2 his3 leu2 trp1 ura3 Δxrn1::RP1*	Jensen T.H. lab
FSY1968	*pap1-1*	*MATa ade2 his3 leu2 trp1 ura3 pap1-1*	Jensen T.H. lab
FSY1988	*pap1-1 Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 pap1-1 Δrrp6::Kanr*	Libri D. lab
FSY4838	*pap1-1 Δtrf4*	*MATa ade2 his3 leu2 trp1 ura3 pap1-1 Δtrf4::*	Libri D. lab
FSY4275	*GAL-NRD1*	*MATa ade2 his3 leu2 trp1 ura3 HisMX6-pGAL-NRD1*	Thiebaut et al., 2006 [8]
FSY4282	*GAL-NRD1 Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 HisMX6-pGAL-NRD1 Δrrp6::Kanr*	Thiebaut et al., 2006 [8]
FSY2527	*Δpho4*	*MATa ade2 his3 leu2 trp1 ura3 Δpho4::Kanr*	this study
FSY3313	*Δpho4Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 Δpho4::Kanr Δrrp6::TRP1*	this study
FSY4265	*Δpho4Δset1*	*MATa ade2 his3 leu2 trp1 ura3 Δpho4::Kanr Δset1::HIS3*	this study
FSY4266	*Δpho4Δset1Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 Δpho4::Kanr Δset1::HIS3 Δrrp6::TRP1*	this study
FSY4911	*NRD1-HA Δpho4*	*MATa ade2 his3 leu2 trp1 ura3 NRD1-HA-HIS3 Δpho4::Kanr*	this study
FSY4918	*NRD1-HA Δpho4 Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 NRD1-HA-HIS3 Δpho4::Kanr Δrrp6::TRP1*	this study
FSY4912	*NRD1-HA Δpho4 Δset1*	*MATa ade2 his3 leu2 trp1 ura3 NRD1-HA-HIS3 Δpho4::Kanr Δset1::TRP1*	this study
FSY4841	*rpb1-1*	*MATa ade2 his3 leu2 trp1 ura3 rpb1-1*	Libri D. lab
FSY4842	*rpb1-1 Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 rpb1-1 Δrrp6::URA3*	Libri D. lab
FSY3799	*Δpho84*	*MATa ade2 his3 leu2 trp1 ura3 Δpho84::Kanr*	Camblong et al., 2009 [58]
FSY3811	*Δpho84 Δrrp6*	*MATa ade2 his3 leu2 trp1 ura3 Δpho4::Kanr Δrrp6::TRP1*	Camblong et al., 2009 [58]

BY4741 background			
FSY3158	*WT*	*MATa his3 leu2 lys2 ura3*	Euroscarf
FSY4747	*Δrrp6*	*MATa his3 leu2 lys2 ura3 Δrrp6::NatMX*	Houseley J. & Tollervey D. (2006) [60]
FSY3182	*Δtrf4*	*MATa ade2 his3 leu2 ura3 Δtrf4::Kanr*	"
FSY4748	*Δtrf4Δrrp6*	*MATa ade2 his3 leu2 ura3 Δtrf4::Kanr Δrrp6::NatMX*	"
FSY4749	*Δtrf4 GAL-TRF5*	*MATa ade2 his3 leu2 ura3 Δtrf4::Kanr HisMX6-pGAL-3HA::trf5*	"

PRIMERS USED IN THIS STUDY

Code	Name	Sequence
OFS1741	*ACT1* F Mid	5'-TTCCAGCCTTCTACGTTTCCATC-3'
OFS1742	*ACT1* R Mid	5'-CGTGAGGTAGAGAGAAACCAGC-3'
OFS735	*ACT1* F 3'	5'-TACTCCGTCTGGATTGGTGGTT-3'
OFS736	*ACT1* R 3'	5'-GGTGAACGATAGATGGACCACTT-3'
OFS737	*ACT1* F 5'	5'-TGGTATGTTCTAGCGCTTGCAC-3'
OFS738	*ACT1* R 5'	5'-GTCAATATAGGAGGTTATGGGAGAGTG-3'
OFS1077	*PHO84* F 3'	5'-GAAATTAACGAGCTATACCACGATGAAATC-3'
OFS1078	*PHO84* R 3'	5'-CATGTTGAAGTTGAGATGGGCTGG-3'
OFS1158	*PHO84* F M	5'-CTGCCGCACAAGAACAAGATGG-3'
OFS1159	*PHO84* F M	5'-TTTGGAGGATGATTGTCAAGAGATTCG-3'
OFS1075	*PHO84* F 5'	5'-CCGTCAATAAAGATACTATTCATGTTGCTG-3'
OFS1076	*PHO84* F 5'	5'-AAAATCATTCAAATGGTTGTGGAAGGC-3'
OFS1717	*SCR1* F	5'-AACCGTCTTTCCTCCGTCGTAA-3'
OFS1718	*SCR1* R	5'-CTACCTTGCCGCACCAGACA-3'
OFS1249	*PHO84 -1000 SalI F*	5'-GGGGGGGGGGGTCGACCGAGAGTGATAAAGAAGAGGCGGT-3'
OFS1363	*PHO84 +355 Sal1 R*	5'-GGGGGGGGGGGTCGACGTCTCAAGTCGCTTGCTTAGTCGA-3'

smFISH probes used in this study

PHO84 mRNA, *PHO84* AS and *MDN1* probes were labeled in a single position with eigther Dylight 549 or DyLight 594 *ITS2-1* probe was labeled in 3 positions with cy5 (Zenklusen et al., 2008). Labeled positions are shown in bold.

PHO84 sense	*PHO84* AS	*ITS2-1*	*MDN1*
ccttcggtaaggtgtt cttt	cccatctcaacttcaac atg	GA**T** ATG CTT AAG TTC AGC GGG **T**AC TCC TAC CTG ATT TGA GG**T** C	cagagggaaaagca gaattg
aaatggttgtggaag gccat	ctgctacgctaaacttc aga		ttgttgctaaagtgga aggg
cttctttccagaggatc ttc	gctataccacgatgaa atcg		cggctatgtagtatag ttcc
aaaccttcgtcatcga tgga	ccttgttgatcccaga aact		gaaaacatgaccagt gatgg
ggtcttaacttgttgcc aac	gccttattcatgttgtt ggg		caacgtgccactatct ctaa
aaaccaacaccagca atgga	gttacctcacgtcatg gaaa		ggcattgcaacggga aatat
ggacatcatagtgat accca	aactgtgctagagac ggtaa		gaatccttttgtgtgga tgg
atactaccgtgccagt aaac	tttgggtactctaatcg acc		ggttgaatgaagcgt gcaaa
acaaggtttgacttgg acct	gtgccattattgcaca aacc		ccaacaatgaatcgc gtgat
ccaacagaagtggaa acctt	tttctgctgcatctggt aag		tccagaatcgccctca aaat
gcagaatggtacaga caatc	tacagatctactgctc atgg		gttcgaggaagctgt aatca
cccatgacaatacggt agaa	aacttcggtccaaaca caac		gcacatctcttagcttc gtt
ggtagtggcaaattc agaag	catcatcttgaccgctt tgt		gaacactgggtagttt agag
ttagcaaagacagca cccat	tcattacctggttactg ggt		cctgaggctcaataat gaag
tgataccaccggaga tttga	ctgctgtcggtaatct gatt		attgacgcgaccttag tact
tttgcgtattctagttc gcc	cgggttgagtttaaac agtg		tttgtcgtggatagtgt gga
aaacatgccaaccct agaac	gctggttcatggtttac ctt		ggacaaaggttaatg ggtag
ccaacttagcgttaac atcc	gtcaatggaagtacg gtaag		cgaagagaggaaac cgtttt
cactggtgtcgtgaat tttc	ccaaaggcttcgttca aaga		ccgcttttccaatgag catt
attgtcaagagattcg acgg	ggtttggaaagagctt ctac		gagtcatggcaaccc atata
cagcagtacctagca aaatc	cagtgatgaagacat ggcaa		ggcatcagtttgttcac cta
cccgtagaaagcaac atcta	cacaagaacaagatg gcgaa		caccagaggtataag tacca
ggtttgcagaataac agcac	taagttggaacttgct gctg		acaccagctctccatt caaa
tttttggaaccggcat aacc	atccttattgggttggg tac		tacccatctcccttcttt ga
atcgacagtgaagac ggata	aaggcttgtgaccaa atgtg		cgcgcttttctaaaag cgat
accgatgacacagaa caaag	gctgaatgtgatgcta gatg		tcctctggatggaatg gtta

caccaagtttatggtatgcg	ctcttatcttggttgctgct	catttgcagcctttacagtc
taaagagccaacagaccatg	aaatggagaggtgccatcat	cgctaggttcctctaattca
ccaggaacaataaaggtggt	gtgactacccactatcttct	ggttggtcaaaatgggaaac
gtaacgagttgggaaacact	cctgctattaacttcgttgc	ccccttgttcaatgaaatg
	gcaaaccactgttgctcatt	ctaacctttcgcacagctta
		caggtttgttgatgccattg
		cgtagacagaagactggatt
		tgaattctccaatagcgcca
		ctcgccgattgcttgtataa
		ccttgaggaagcaatgtcta
		ggcacatgttgggtcaaaaa
		gcgaacaatgtgctgttcat
		ccgactagtaaaacaggttc
		tgaacgactgtagttttccc
		ccaagaagatcaccagtttc
		gcatcttgtgaaacttctcg
		gtacgcttcgttccaaagtt
		cagcccatttgtcaagtaac
		cctcaaacttcttcactgag
		ccctcgacaaaattgaagac
		gccctgatagtctttaccaa
		ttcatcgagcaatagccact

3 Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription

3.1 Aims of article 2

For a second project, I was keen to expand my experimental expertise to work with higher eukaryotic systems and simultaneously immerse myself in the much more complex world of lncRNA mediated transcription regulation in a disease relevant model system. I was particularly interested in enhancer RNAs (eRNAs), a newly discovered class of lncRNAs transcribed from mammalian enhancers, which are regulatory sequences that activate gene transcription across large genomic distances in response to cell type specific signals. eRNA expression occurs genome-wide in diverse cell types and is correlated with target gene activation. Several studies had implicated eRNAs in various roles, including increasing chromatin accessibility at promoters, regulating enhancer-promoter looping interactions, and facilitating RNA pol II elongation on target loci, but no concerted model for eRNA function had emerged. To study eRNA function, we chose estrogen induced eRNAs in the MCF7 breast cancer cells as a model system. In particular, we wanted to determine whether transcription initiation from active promoters is coupled with eRNA expression on individual alleles, whether eRNAs stabilize enhancer-promoter interaction, and also characterize the pioneer factors that condition enhancers to become transcriptionally active. The results of this study were published online in *Nucleic Acids Research* in December 2016 (Rahman et al, "Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription").

3.2 Article 2

Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription

(Running Title: Single-cell analysis of eRNA expression)

Samir Rahman[1,4], Cornelia E. Zorca[1,4], Tatiana Traboulsi[1,2], Emmanuel Noutahi[1], Matthew R. Krause[3], Sylvie Mader[1,2], and Daniel Zenklusen[1]*

[1]Département de Biochimie et Médecine Moléculaire, Université de Montréal, Montréal, QC H3T 1J4, Canada.

[2]Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, QC H3T 1J4, Canada.

[3]Montreal Neurological Institute, McGill University, Montréal, QC H3A 2B4, Canada.

[4]These authors contributed equally to this work.

*Corresponding Author: Daniel Zenklusen, Ph.D.

[1]Département de Biochimie et Médecine Moléculaire

Université de Montréal

2900 Edouard-Montpetit, D-340

Montréal, QC H3T 1J4, Canada

+1 (514) 343-6327 (phone)

daniel.r.zenklusen@umontreal.ca

3.2.1 Abstract

Enhancers are intergenic DNA elements that regulate the transcription of target genes in response to signaling pathways by interacting with promoters over large genomic distances. Recent studies have revealed that enhancers are bi-directionally transcribed into enhancer RNAs (eRNAs). Using single-molecule fluorescence *in situ* hybridization (smFISH), we investigated the eRNA-mediated regulation of transcription during estrogen induction in MCF-7 cells. We demonstrate that eRNAs are localized exclusively in the nucleus and are induced with similar kinetics as target mRNAs. However, eRNAs are mostly nascent at enhancers and their steady-state levels remain lower than those of their cognate mRNAs. Surprisingly, at the single-allele level, eRNAs are rarely co-expressed with their target loci, demonstrating that active gene transcription does not require the continuous transcription of eRNAs or their accumulation at enhancers. When co-expressed, sub-diffraction distance measurements between nascent mRNA and eRNA signals reveal that co-transcription of eRNAs and mRNAs rarely occurs within closed enhancer-promoter loops. Lastly, basal eRNA transcription at enhancers, but not E2-induced transcription, is maintained upon depletion of MLL1 and ERα, suggesting some degree of chromatin accessibility prior to signal-dependent activation of transcription. Together, our findings suggest that eRNA accumulation at enhancer-promoter loops is not required to sustain target gene transcription.

Key words: transcription, enhancers, eRNAs, smRNA FISH, enhancer-promoter looping

3.2.2 Introduction

Enhancers are intergenic DNA elements that regulate the transcription of target genes by interacting with promoters over large genomic distances (1-3). They contain binding sites for transcription factors, which promote RNA polymerase II (RNAPII) recruitment and transcription activation (4). Enhancers carry unique epigenetic marks that distinguish them from promoters, including monomethylated Lysine 4 of Histone H3 (H3K4me1) and acetylated Lysine 27 of Histone H3 (H3K27ac) (5,6). In addition, these regulatory elements have an open chromatin conformation, which increases accessibility to transcription factors and to RNAPII (7). Several genome-wide studies have shown that enhancers are transcribed into long noncoding RNAs in a tissue-specific manner in response to extracellular signals. For example, stimulation of cortical neurons via membrane depolarization was shown to induce the recruitment of RNAPII to enhancers and the initiation of bi-directional transcription of noncoding RNAs, termed enhancer RNAs (eRNAs) (8). Importantly, eRNA expression positively correlated with increased mRNA expression from proximal target genes, suggesting that eRNA transcription marks active enhancers. Although genome-wide studies in other cell types have also established a correlation between eRNA and target mRNA expression, whether there is a unified mechanism by which eRNAs regulate their target genes remains unclear (8-17).

Since long non-coding RNAs are more functionally diverse than other classes of ncRNAs, such as microRNAs, rRNAs, and tRNAs, they are thought to be under lower selective constraints (18). Although enhancers possess conserved transcription factor binding sites, enhancer-derived transcripts often lack conserved motifs or secondary structures that could provide a hint for a unified mechanism of action (7). Similarly, little is known about eRNA biogenesis. Most eRNAs are capped, unspliced and non-polyadenylated with a median length of approximately 350 nucleotides (19). Recent studies showed that the RNAPII-associated integrator complex mediates transcription termination at enhancers, and that 3'cleavage of eRNAs by the integrator complex is required for their function (20). Interestingly, several experimentally validated eRNAs were detected as distinct bands in Northern blot analyses, further corroborating the idea that eRNA transcription is terminated in a uniform manner (16,21,22). In addition, ChromRNA-seq analyses have suggested that

eRNAs are enriched in the chromatin fraction (20). Collectively, these observations indicate that eRNAs act as processed transcripts *in cis* on their target loci. Supporting this idea, tethering of eRNAs to luciferase reporter plasmids enhanced promoter activity (15,16). However, it cannot be excluded that enhancer transcription rather than the eRNA transcript *per se* is required for enhancer function.

While the features that constitute a functional eRNA molecule are currently unknown, several studies have implicated eRNAs in different steps of transcription regulation, including the modulation of chromatin accessibility at promoters and the release of the negative elongation factor NELF (11,23,24). In particular, eRNAs have been suggested to mediate looping interactions between enhancers and promoters of target genes by facilitating the recruitment of specific factors known to stabilize long-range chromatin interactions, including cohesin and the mediator complex (15,21,25). However, treatment with flavopiridol to block eRNA transcription did not affect chromatin looping at the *P2RY2* and *GREB1* loci upon E2 stimulation of MCF-7 cells (14). Furthermore, eRNA inhibition affected neither enhancer-specific epigenetic modifications nor the recruitment of transcriptional regulators, suggesting that eRNA synthesis is not required for the assembly of enhancer-promoter complexes (14). Interestingly, a recent study has shown that eRNAs bind to the transcription factor Yin Yang 1 (YY1), and that this interaction can enhance the recruitment of this transcription factor to enhancers (26). Whether transcription factor trapping by eRNAs is a widespread mechanism of gene expression regulation remains to be investigated. Therefore, while several different studies suggest that eRNAs are likely to be implicated in regulating target gene transcription, the mechanistic details of eRNA function remain unclear.

One difficulty in studying induced transcriptional responses is that different cells and even different alleles within a cell may not show the same induction kinetics upon stimulation, due to the stochastic nature of many cellular processes (27,28). Different studies have shown that eRNA transcription precedes the peak of mRNA transcription (9,11). In addition, transcription often occurs in short bursts, where individual alleles switch between active and inactive states (27,29-31). Observing such behavior, therefore, requires monitoring the transcriptional response at the single-cell and single-allele level. Most previous studies

investigating eRNA function used ensemble measurements of cell populations, and were thus unable to provide spatial or allele-specific information regarding eRNA and target mRNA transcription (14,15). In particular, whether eRNAs are required for every round of transcription initiation, and whether they modulate transcription bursts is still unknown.

To investigate the role of eRNAs in transcription, we characterized the spatiotemporal expression of these noncoding transcripts in individual cells using single-molecule resolution fluorescence in situ hybridization (smFISH). Specifically, we used ERα positive MCF-7 breast cancer cells as a model system to study eRNA-mediated transcriptional programs during estrogen induction. Estrogen binds directly to the nuclear receptor ERα, which is recruited to estrogen response elements (EREs) as a homodimer and regulates target gene transcription by recruiting transcriptional cofactors (32,33). ERα binding to chromatin is facilitated by the pioneer factor FOXA1, which itself is recruited at H3K4me1 marks deposited by the histone methyltransferase MLL1 on the enhancer regions of target genes (34,35). While the mechanism whereby MLL1 recognizes its target loci is presently unknown, roles for the CpG Binding protein CGBP, the transcription factor YY1, and the core subunit of the SWI-SNF complex hSNF5 have been suggested (36-38).

Here we monitored the expression and localization of eRNAs derived from the enhancers of the *FOXC1* and *P2RY2* loci, as well as the expression of their cognate mRNAs over a time course of estrogen induction in individual MCF-7 cells by smFISH. In the uninduced state, FOXC1 and P2RY2 eRNAs were expressed at low levels, independently of MLL1 and ERα. Estrogen treatment induced the transcription of eRNAs and target mRNAs with similar kinetics, and required both chromatin modification by MLL1 and ERα recruitment. However, co-expression of eRNAs and mRNAs at individual alleles was infrequent, and did not correlate with bursting mRNA transcription. Furthermore, distance measurements between eRNA and nascent mRNA signals at co-expressing alleles, revealed infrequent co-localization within closed enhancer-promoter loops. Taken together, our data suggest that ongoing eRNA transcription is neither required to stabilize chromatin loops, nor to sustain continuous transcription from target alleles.

3.2.3 Results

eRNAs and target mRNAs are co-induced upon E2 treatment in MCF-7 cells

To examine the role of eRNAs in modulating transcription at the single-cell level, we examined their spatial expression patterns in MCF-7 breast cancer cells over a time course of 17β-estradiol (E2) induction using smFISH. This technique uniquely allows the simultaneous localization of eRNAs and target mRNAs in individual cells and at individual alleles with single-molecule resolution (29,41). Specifically, we examined the E2-induced *FOXC1* and *P2RY2* loci as model genes. Genome-wide studies have shown previously that the *FOXC1* and *P2RY2* enhancers are transcribed into eRNAs in a hormone-dependent manner (14,15). The *FOXC1* gene encodes a single exon of 3.4 kilobase-pairs (kbps) and its expression is regulated by a single E2-responsive enhancer, located 26 kb downstream of its transcription termination site. The *P2RY2* gene is 18 kb-long and its expression is regulated by an E2 responsive enhancer located 22 kb upstream of its transcription start site (TSS). The *FOXC1* enhancer is bi-directionally transcribed, producing sense and anti-sense eRNAs (Figure 3.1A). We designed oligonucleotide probes against both target mRNAs and eRNAs (based on GRO-seq data published by Li et al., 2013), which we hybridized to paraformaldehyde-fixed MCF-7 cells at different time points of E2 induction. Addition of E2 results in a transient transcriptional response that peaks at 40 minutes (13); we therefore analyzed eRNA and mRNA expression at 0, 20, 40 and 60 minutes after E2 treatment. As shown in Figure 3.1B, FOXC1 and P2RY2 mRNAs were expressed at a low level in uninduced cells. Upon E2 induction, bright spots representing active sites of transcription, consisting of multiple nascent mRNAs, were detected within the nuclei of single cells. Since MCF-7 cells are triploid for the *FOXC1* and *P2RY2* genes, 3 to 6 transcription sites were observed in interphase and dividing cells, respectively.

Analysis of the bi-directionally transcribed FOXC1 eRNAs showed similar induction kinetics as the target mRNAs, but different localization patterns. In uninduced cells, FOXC1 eRNAs were detected in the nucleus and were expressed at low levels (Figure 3.1B). Even though 38.2% of the cells expressed anti-sense and 23.5% expressed sense FOXC1 eRNAs, only a single eRNA spot was detected in the majority of these cells (Figure 3.1C). Upon E2 induction, eRNA and target mRNA expression increased in parallel, with 62.5% of the cells

expressing anti-sense and 55.8% expressing sense FOXC1 eRNAs at 40 min of treatment. Notably, the fraction of cells expressing more than one eRNA increased three fold compared to the uninduced state. Importantly, the number of eRNA spots rarely exceeded the number of alleles in either interphase or dividing cells (Figure 3.1C). A similar pattern of expression was observed for the P2RY2 sense eRNA and its cognate mRNA.

Since enhancer elements recruit many proteins, including transcriptional co-activators and RNAPII, eRNAs could be incorporated into large complexes and become inaccessible to the RNA FISH probes. To address this point, we carried out protease treatment prior to probe hybridization as described previously by Buxbaum et al., but found similar levels of eRNA expression in the presence and absence of protease digestion (42). Specifically, we observed comparable numbers of FOXC1 AS eRNAs per cell with or without pepsin treatment (Supplementary Figure S1A). Likewise, we found that the intensities of single FOXC1 AS eRNA spots did not change, suggesting that probe accessibility to the target is not obscured by eRNA incorporation into protein complexes (Supplementary Figure S1B). In addition, quantification of eRNA and mRNA fold induction levels in response to estrogen treatment by RT-qPCR was consistent with smFISH measurements, implying that eRNA and mRNA detection by smFISH is efficient (Supplementary Figure S2). Furthermore, in the presence of RNAse A, neither FOXC1 mRNAs nor AS eRNAs were detected, validating that the signals observed by smFISH are specific to RNA (Supplementary Figure S3).

In summary, our time course analysis revealed that eRNA transcripts are nuclear and induced with similar kinetics as the target mRNA. Combined with previous findings showing that eRNAs have a short half-life of about 7 minutes, and act *in cis*, these observations provide evidence that most eRNAs detected by smFISH are nascent and are mostly restricted to the enhancer from which they are transcribed (11,24).

Induction of eRNA and target mRNA transcription requires ERα and MLL1

As shown in Figure 3.1, low levels of eRNA expression were detectable at estrogen-responsive enhancers even in uninduced cells. This could be due to hormone-independent activation of ERα or to continuous basal transcription of the enhancer prior to ERα recruitment. In the latter case, low levels of eRNA transcription could be implicated in chromatin remodeling, thereby facilitating ERα accessibility once the cells are exposed to estrogen. To distinguish between these possibilities, we pre-treated the cells prior to the addition of E2 with either Tamoxifen, a drug that blocks estrogen receptor function by inhibiting E2 binding, or with ICI 182,780 (ICI), which induces ERα degradation (43-45). Both drugs effectively abolished the induction of FOXC1 and P2RY2 mRNA, as shown by the lack of bursting transcription upon E2 treatment following the addition of either inhibitor (Figure 3.2A and 3.2B, and Supplementary Figure S4A and S4B). Similarly, eRNA induction was inhibited by both Tamoxifen and ICI (Figure 3.2A and 3.2B, and Supplementary Figure S4A and S4B). However, basal eRNA expression was unaffected by ICI 182,780 treatment, suggesting that the low level of eRNA transcription observed at these enhancers in uninduced cells is independent of ERα. This observation further raises the possibility that low-level eRNA transcription primes enhancers to respond rapidly to environmental stimuli.

Since ERα inhibition did not attenuate the basal level of eRNA expression, we investigated whether upstream factors are required to activate eRNA transcription. As noted above, the histone methyltransferase MLL1 deposits the H3K4me1 mark on enhancer regions, which recruits the pioneering factor FOXA1, and subsequently, ERα (34). Thus, we examined whether siRNA-mediated MLL1 knockdown affects basal eRNA transcription. We found similar basal levels of P2RY2 eRNA transcription in the absence of E2 in MLL1-depleted cells and in the non-specific siRNA control (Figure 3.3A and 3.3B). Therefore, low basal eRNA transcription may occur prior to chromatin remodeling and participate in the recruitment of MLL1 (46,47).

In contrast to the basal level of eRNA transcription, P2RY2 eRNA induction was significantly reduced upon E2 treatment in MLL1-depleted cells compared to the non-specific siRNA control (Figure 3.3B), and paralleled a decrease in the number of transcriptionally

active *P2RY2* loci (Figure 3.3B). Consistent with these observations, ChIP analysis revealed a ten-fold decrease in ERα binding to the *P2RY2* and *FOXC1* enhancers upon E2 induction in MLL1-depleted cells compared to the control (Figure 3.3C, Supplementary Figure S5A). Furthermore, MLL1 knockdown resulted in a two-fold decrease in the level of H3K4me1 at these regulatory elements both in the presence and in the absence of E2 induction (Figure 3.3C, Supplementary Figure S5A). Taken together, these findings corroborate the RNA FISH results showing that MLL1 depletion inhibits both eRNA and target mRNA induction (Figure 3.3A and 3.3B), and suggest that MLL1-dependent deposition of H3K4me1 and the recruitment of ERα are required for the induction of eRNA transcription.

eRNA- mRNA co-expression is infrequent and is not required to maintain transcription

Previous studies have suggested that eRNAs facilitate transcription by stabilizing enhancer-promoter interactions (15,21,25). These interactions could persist for the duration of a single initiation event or for the entire length of the transcriptional response. If eRNA transcription or accumulation at a specific allele were required for enhancer-promoter communication or for maintaining transcription, frequent co-localization of eRNAs at active alleles should be observed. To test this hypothesis, we determined whether the simultaneous induction of eRNAs and target mRNAs is coordinated at individual alleles, and measured the frequency of eRNA and nascent mRNA signal co-localization. To identify *FOXC1* transcription sites (TSs), we determined the position and intensity of mRNA signals, and clustered them into low and high intensity groups, with the low intensity spots representing single mRNAs. Nuclear spots of greater than 1.5 times the intensity of the mean value of single mRNAs were scored as transcription sites, assuming that actively transcribed loci contain multiple nascent mRNAs due to transcription bursting (27,29,30,48). Since the *P2RY2* gene contains several short exons at the 5' end that are too short for smFISH probe design, we used the first intron region to detect the nascent mRNA, and applied a similar spot-clustering algorithm to locate the transcription sites (Supplementary Figure S6A). We then measured the frequency of co-localization of actively transcribing alleles with eRNAs (Fig. 3.4A) and *vice versa* (Fig. 3.4B). Specifically, co-localization was defined by the presence of an eRNA within a 400 nm radius from the center of the TS signal. This cutoff was selected to include open enhancer-promoter configurations, but to exclude signals from neighboring alleles, which are

separated by more than 1um, as determined from distance measurements between active *P2RY2* transcription sites (Supplementary Figure S6B).

As shown in Figure 3.4A, FOXC1 eRNA–TS co-localization was infrequent before E2 treatment, with only 4% of transcription sites co-localizing with an eRNA. FOXC1 eRNA–TS co-localization increased to 27% at 40 minutes of E2 treatment. While E2 induction significantly increased the proportion of FOXC1 eRNA-TS co-localization (Fisher's exact test; $p<0.0001$), the majority of transcription sites did not have an associated eRNA. As noted in Figure 3.1C, the fraction of cells expressing more than one eRNA increased three-fold at 40 min of E2 treatment compared to the uninduced state. Additionally, there was a five-fold increase in the fraction of cells with more than one active transcription site at 40 min of E2 treatment compared to the uninduced state. Despite this increase in the numbers of eRNAs and transcription sites detected per cell, the majority of TSs did not co-localize with eRNAs, indicating that mRNA transcription does not require the simultaneous expression of eRNAs. Triple co-localization of FOXC1 mRNA, antisense and sense eRNA was present at only 2% of active transcription sites at 40 min of E2 induction, further suggesting that simultaneous transcription of sense and anti-sense eRNAs is rare at active TSs.

Although the low frequency of eRNA-mRNA co-expression at individual alleles suggests that eRNAs are not required for sustained transcription, eRNAs could influence the magnitude of the transcription output. To determine whether eRNAs modulate transcription strength, we correlated the presence or absence of eRNAs at the TS with the RNAPII density on the target gene. As shown in Figure 3.4C, in uninduced cells, basal *FOXC1* expression averaged 2-3 transcripts per TS and occurred primarily from alleles that had no associated eRNAs. During peak induction, both *FOXC1* and *P2RY2* showed strong bursting transcription, independent of the presence of eRNAs; however, for the *FOXC1* gene, there was a small but significant increase in the number of nascent RNAs per transcription site at active alleles that co-localized with eRNAs compared to those that did not (Figure 3.4C). Furthermore, comparisons of mRNA signal intensities at transcription sites before and after E2 treatment showed that the RNAPII density along the gene increased over time (Supplementary Figure S7). In contrast, although the proportion of cells transcribing eRNAs increased with induction,

the eRNA signal intensities did not change (Supplementary Figure S7). This observation suggests that compared to mRNA, eRNA transcription initiation occurs at a lower frequency, and that eRNAs do not accumulate at enhancers in high numbers.

mRNA-eRNA co-transcription rarely occurs within a closed enhancer-promoter loop conformation

Chromosome conformation capture (3C) analyses conducted in the presence or absence of enhancer-derived transcripts proposed that eRNAs mediate enhancer-promoter interactions (15,21,25). Such interactions are thought to be maintained by large protein complexes, containing cohesin, the mediator complex and other co-factors, and are thought to occur within 10-100nm (49,50). Although these distances are below the 200 nm resolution limit of conventional light microscopy, signals within this range can be localized at higher spatial resolution by 2D Gaussian fitting (40). Using the coordinates of eRNA and mRNA transcription site signals as markers of enhancers and genes, respectively, we assessed whether transcribed enhancers interact with active genes. To achieve this, we first determined the minimal distance at which we could resolve co-localizing signals using a P2RY2 intron probe labeled with two different colors (Figure 3.5A). After pixel shift correction, signals identifying the same intron were co-localized within 63 nm (Figure 3.5A). We then measured pairwise eRNA-nascent mRNA distances among co-expressing alleles at 40 min of E2 induction. At this time point, 29% of anti-sense and 19% of sense FOXC1 eRNA-mRNA co-expressing alleles showed a separation of 100nm or less (Figure 3.5B). These data show that the small percentage of active transcription sites that co-express eRNAs are infrequently transcribed within a closed enhancer-promoter loop. Furthermore, these observations suggest that looping interactions are either transient or that eRNAs are preferentially transcribed before enhancer-promoter loops are established (Figure 3.5C).

3.2.4 Discussion

Our study reveals important insights into eRNA expression that are unique to the use of single-cell approaches, and complements previous investigations of eRNA function that

have largely relied on ensemble measurements. Population-level studies are limited in their ability to reveal the differential behavior of cells within a population, and cannot distinguish whether individual alleles show alternate transcriptional activity. Therefore, such studies typically report average expression profiles that are limited in their scope to accurately describe what occurs in individual cells and at individual alleles. Transcriptional responses induced by external stimuli have been shown to be particularly noisy, as gene activation depends on a cascade of events, such as signal sensing, transcription factor binding and chromatin remodeling (27,29). Our analysis of E2-mediated transcription kinetics, previously described as a strong and synchronous transcriptional response, showed highly variable induction in individual cells, both for mRNA and eRNA expression. Even at the peak of E2 induction (40 min), one third of all the cells did not exhibit strong FOXC1 and P2RY2 mRNA transcription, suggesting that only a fraction of the cells respond to the stimulus at any given time. Furthermore, individual active alleles showed variable transcriptional strength, as measured by the number of nascent mRNAs, consistent with a bursting expression pattern.

Interestingly, although the eRNA and mRNA transcription frequencies increase over the time course of E2 induction, we did not observe an increase in eRNA signal intensity at individual alleles as observed for mRNAs (Supplementary Figure S7). Typically, an increase in transcription initiation frequency results in the association of multiple nascent RNAs with a locus. The ability to detect nascent mRNAs by smFISH depends not only on the initiation frequency, but also on the length of the gene, on the elongation velocity, and on the kinetics of release of an RNA from chromatin. Although eRNA transcription units are shorter than those of mRNAs, strong transcriptional bursting at enhancers should result in an increase in eRNA signal intensity, which was not observed over the entire time course examined. Furthermore, if eRNAs were infrequently initiated and released shortly after transcription termination, smFISH probes would hybridize to partially transcribed eRNAs, yielding signals of variable intensities. Northern blot analyses have demonstrated that these transcripts are detected as bands of discrete size, which correlate well with the uniform eRNA intensities measured by smFISH and imply that most eRNAs are present as fully synthesized transcripts of defined length (16,21,22). Additionally, fractionation experiments revealed that eRNAs are mainly chromatin-associated, consistent with our findings that the number of eRNA signals rarely

exceed the number of alleles in individual nuclei. Together, these lines of evidence indicate that eRNA transcription is infrequent and that eRNA release from chromatin is slow.

Various functions have been ascribed to eRNAs during the transcription cycle. For instance, eRNAs are thought to facilitate the transition of paused RNAPII into productive elongation by mediating the release of the negative elongation factor NELF (24). If eRNAs were required for RNAPII pause release at target alleles, we would predict to observe comparable frequencies of eRNA and mRNA expression, since every cycle of mRNA transcription would require an eRNA to act as a decoy for NELF. The low transcription frequency of eRNAs observed at the *FOXC1* and *P2RY2* enhancers, however, does not favor such a model, at least for the genes investigated here.

Alternatively, several studies have proposed a role for eRNAs in facilitating or stabilizing enhancer promoter loops through interactions with cohesin and the mediator complex (15,21,25). Specifically, knockdown of different eRNAs were shown to reduce enhancer-promoter contacts in 3C assays. One limitation of 3C analysis is the inability to determine whether there is heterogeneity in enhancer-promoter interactions at individual alleles within a population. DNA FISH studies have reported both transient and stable interactions between regulatory regions and promoters, depending on the system investigated (50,51). In favor of stable looping, FRAP analysis of cohesin sub-complexes in G1 measured chromatin-bound residence times of about 24 min (52). Although it is not clear whether cohesin-stabilized enhancer-promoter interactions have a similar duration, this data suggest that such interactions could persist for many minutes (52). Our measurements of eRNA and mRNA co-expression at individual alleles revealed two interesting features: (i) that co-expressing alleles are rare, and (ii) that co-expression infrequently occurs in a closed enhancer-promoter configuration. One interpretation of this observation is that enhancer-promoter interactions are dynamic. Recent studies show compartmentalization of genes and their regulatory regions in topologically associating domains (TADs), which position enhancers and promoters in proximity. Thus, TADs are thought to facilitate contacts between regulatory regions and target gene promoters, and to circumvent the need for stable loops (49,50). Alternatively, the infrequent co-expression of eRNAs and mRNAs in a closed-loop configuration could indicate that eRNA transcription precedes looping or that eRNA

transcription is mostly inhibited in the closed loop configuration (Figure 3.5C). Positioning of ER✔, co-activators, the mediator complex and the basal transcription machinery in a closed loop configuration could indeed facilitate mRNA over eRNA transcription. Therefore, while the accumulation of eRNAs at active enhancers is unlikely to be required to stabilize enhancer-promoter loops, eRNA transcription may, nonetheless, initiate this communication, which promotes target gene activation.

Lastly, our data show that eRNA transcription at the *FOXC1* and *P2RY2* enhancers occurs at a basal level even in the absence of MLL1 and ERα. An additional role of eRNAs in mediating enhancer-promoter communication suggests that the process of transcription at enhancers modifies chromatin, thereby facilitating the binding of transcription factors and co-activators (46). Such low-level transcription at enhancers could maintain this regulatory region in an open chromatin state, and allow cells to respond rapidly to E2 stimulation. In support of this idea, we found that knockdown of MLL1 reduced both the level of H3K4me1 and the binding of ERα to *FOXC1* and *P2RY2* enhancers. Enhancer-mediated gene activation may be characterized by a feed-forward loop, whereby basal eRNA transcription facilitates recruitment of TFs and co-regulators, which then further remodel chromatin and increase the frequency of eRNA and mRNA transcription initiation. Understanding how eRNAs and TFs act either jointly or independently to stimulate transcription will require targeted deletion of different enhancer regions to alter eRNA structure or abolish eRNA expression, while maintaining the binding of ERα.

In conclusion, our smFISH analysis of eRNA and mRNA expression patterns over a time course of estrogen induction showed that the majority of eRNAs are not co-localized with active mRNA transcription sites, and *vice versa*, implying that eRNA and mRNA transcription is rarely coupled on individual alleles. The lack of co-localization of eRNAs with a TS may reflect an early stage of transcription initiation, when the enhancer is primed for activating target gene expression. This observation is consistent with previous studies showing that the onset of eRNA transcription precedes target gene activation (9,11). Reciprocally, the absence of eRNAs at TSs indicates that eRNAs are not required to sustain bursting transcription. Furthermore, the lack of accumulation of eRNAs at enhancers upon E2 induction suggests that

transcription of these noncoding RNAs is initiated at low frequency, and that eRNAs act transiently during the early stages of activation, possibly to alter the local chromatin environment, facilitate transcription factor access or to initiate enhancer-promoter communication. Notably, alleles that co-express eRNAs and mRNAs show a broad distribution of enhancer-promoter configurations, in which eRNAs are rarely transcribed within distances compatible with direct interactions. This suggests that while eRNAs may initiate enhancer-promoter communication, their transcription is mostly repressed once looping interactions are established. Elucidating the dynamics of enhancer-promoter interactions in the presence or absence of eRNAs will require live cell analyses at high spatial and temporal resolution.

3.2.5 Materials and Methods

Reagents 17β-estradiol (E2) (E1024) and the antiestrogens (Z)-4 Hydroxytamoxifen (H7904) and ICI 182,780 (I4409) were purchased from Sigma. Cells were pretreated with the antiestrogens (100 nM) for 3 hours prior to E2 induction (5 nM) for 40 min in the absence of the antiestrogens.

MCF-7 cell culture The MCF-7 cell line was maintained in α-minimal Eagle's medium (α-MEM) (Wisent, St-Bruno, QC, Canada) supplemented with 10% fetal bovine serum (FBS) (Invitrogen, 10437028) as described previously (39). The cells were attached to poly-L-Lysine (Sigma, P8920) coated coverslips in regular media. Three days prior to induction, the cells were washed twice with PBS, and the media was changed to phenol red–free Dulbecco's modified Eagle's medium (DMEM) (Wisent), supplemented with 10% charcoal-treated FBS (Invitrogen, 12676029), 1% sodium pyruvate and 1% L-glutamine (Wisent). The cells were induced with 17β-estradiol (E2) at 100 nM final concentration (except for the ERα inhibition experiments, where the E2 concentration was lowered to 5 nM) for the indicated lengths of time. Validated MLL1 and non-specific (NS) siRNAs were transfected at 50 nM final concentration using Lipofectamine RNAiMax (Invitrogen, 13778100) according to the manufacturer's instructions. Twenty-four hours after transfection, the media was replaced with supplemented phenol red-free DMEM. Estrogen treatments were initiated 48 hours later (a total of 72 hours post siRNA transfection). All experiments represent 2-3 independent biological replicates.

smRNA FISH Custom DNA probe sets were designed using Stellaris® Probe Designer, synthetized by Biosearch Technologies containing a 3' amine reactive group, and labeled with Cy5 (GEPA25001), Cy3 (GEPA23001), or Cy3.5 (GEPA23501) from Sigma or their DyLight (Thermo Scientific) equivalents DyLight 650 (62266), DyLight 550 (62263), DyLight 594 (46413). Probe sequences are shown in Supplementary Table S1. For smRNA FISH, the cells were briefly washed with 1xPBS, fixed with 4% paraformaldehyde (pH 7.2) for 10 minutes at room temperature, washed three times with 1xPBS, and stored overnight in 70% ethanol at -20^{o}C. Prior to hybridization, the cells were air-dried and rehydrated in 10% formamide/2xSSC for 10 min at room temperature. The cells were hybridized with 10-20 ng of each probe plus

40 ug ssDNA/tRNA mix resuspended in the hybridization solution (10% dextran sulfate/10% formamide/2xSSC/2 mM VRC/0.1 mg/ml BSA) for 3 hours in the dark at 37°C. Two post-hybridization washes were carried out at 37°C with 10% formamide/2xSSC for 30 min each. Samples were then rinsed with 1xPBS and mounted with ProLong Gold antifade reagent with DAPI (P36935, Invitrogen). Images were acquired with a 63x NA 1.4 oil objective on a Zeiss Axioimager Z2 equipped with an AxioCam mRm CCD camera and the following filter sets: Zeiss 488050-9901-000 (Cy5), Chroma SP102 v1 (Cy3), Chroma SP103 v2 (Cy3.5), Zeiss 488049-9901-000 (DAPI).

Protease and RNase A treatments For protease treatment, prior to smFISH probe hybridization, fixed cells were rehydrated in 1xPBS and treated with 1 mg/ml Pepsin in 10 mM HCl, at 37°C, for 1 min. Enzyme activity was inhibited with 0.1 M Glycine in 1xPBS, pH 7.2 for 10 min. For RNAse treatment, fixed cells were treated with 0.1 mg/ml RNase A for 1 hour at 37°C. The cells were washed three times with 1xPBS, pre-incubated in 10% Formamide/2xSSC, and hybridized with the indicated probes as described above.

Immunofluorescence Cells stored in 70% ethanol were air-dried and rinsed with 1xPBS for 5 min. This was followed by permeabilization with 0.5% Triton x-100/1xPBS for 10 min at RT. Cells were then washed 3 times with 1xPBS before proceeding with the smFISH protocol as described above. After smFISH hybridization, the cells were washed 3 times with 1xPBS and blocked with 4% BSA (Ambion/Life Technologies molecular biology grade BSA Catalog # AM2616) in 1xPBS for 10 min at RT. Cells were then incubated with the rabbit polyclonal anti-MLL-C (EMD Millipore, ABE240) antibody (diluted 1:100 in 1%BSA/1xPBS) for 1 hour at RT. After the primary antibody incubation, cells were washed three times with 1xPBS for 5 min. This was followed by incubation with the secondary anti-rabbit-Alexa 488 antibody (diluted 1:500 in 1%BSA/1xPBS) for 1 hr at RT in the dark. Cells were then washed 3 times with 1xPBS for 5 min and mounted on slides with Prolong Gold antifade reagent containing DAPI.

Image processing and spot detection For image analysis, 3D datasets were reduced to 2D data using maximum projections in Fiji. Spot detection was done by 2D Gaussian fitting as described previously (29,40). To correct for pixel shifts between channels, TetraSpec beads (Invitrogen T-7279) were imaged in all the channels and their position was determined by 2D Gaussian fitting. Relative pixel shifts were used to align channels after image acquisition and

spot detection using a custom script in MATLAB R2013a(8.1.0.604) (The Mathworks, Inc.). The mRNA channel was used as a reference to correct eRNA positions relative to their target mRNAs.

RNA quantification and distance measurements Nuclear masks were created in Fiji after manual segmentation of DAPI stained nuclei. Assignment of eRNA and mRNA signals within the nuclear masks was done using custom scripts in MATLAB. Detection of mRNA transcription sites was done in two steps: (i) computing I_s, and (ii) locating transcription sites by searching for nuclear spots with at least n * I_s intensity, where n=1.5. To determine the intensity of a single mRNA, mRNA spots were clustered into different classes using c*k*-means, based on intensity. For FOXCI mRNA signals, the single mRNA intensity was calculated by computing the mean intensity of spots outside the nuclear boundary that belong to the low intensity group (with cytoplasmic mRNAs corresponding to single mRNAs). For P2RY2 intron signals, the single RNA intensity was calculated by computing the mean intensity of the nuclear intron spots in the low intensity group. To measure the distance between mRNA transcription sites and eRNAs, we searched for co-localizations occurring within a 400 nm radius around each mRNA transcription site. When eRNAs were detected within that window, we measured the 2D Euclidian distance between the centroids of the eRNAs and mRNAs, as calculated above. Distances between active transcription sites were measured using Volocity 6.0 (PerkinElmer).

RTq-PCR Total RNA was harvested in TRIzol Reagent (ThermoFisher Scientific, 15596026) and cDNA synthesis was carried out with random hexamers using SuperScript II Reverse Transcriptase (ThermoFisher Scientific, 18064014) according to the manufacturer's instructions. The RTq-PCR assay was performed in duplicate on a LightCycler 480 System (Roche Life Science) using SYBR Green. The ΔC_T (delta threshold cycle) method was used for quantification and transcript levels were normalized to GAPDH. RT-qPCR primers are listed in Supplementary Table S2.

Chromatin Immunoprecipitation For ChIP, the cells were seeded at a density of $1.66x10^6$ cells per 10 cm dish in phenol red-free DMEM media supplemented with 10% charcoal-stripped serum and transfected 24 hr after seeding with 50 nM siRNAs using siLentFect Lipid Reagent (BioRad, 170-3361) according to manufacturer's instructions. The next day, media was replaced with supplemented phenol red-free DMEM. The cells were

induced for 40 min with 100 nM estradiol 72 hrs post-transfection and harvested for ChIP. Briefly, the cells were crosslinked with 1% formaldehyde for 10min at room temperature. The crosslinking reaction was quenched with 0.15 M glycine for 5 min and the cells were washed twice with ice-cold PBS and harvested. The cell pellets were resuspended in lysis buffer (5 mM PIPES pH8, 85 mM KCl, 0.5% NP40), supplemented with a mix of protease and phosphatase inhibitors. Following centrifugation, nuclei were resuspended in nuclear lysis buffer (50 mM Tris pH8.1, 10 mM EDTA pH8, 1% SDS), supplemented with a mix of protease and phosphatase inhibitors, sonicated with a Bioruptor (Diagenode) at medium power for 4 rounds of 8 min with 30 sec intervals between pulses. The DNA was quantified and 50 ug of each sample was diluted 20 times in dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7 mM Tris pH8.1, 167 mM NaCl), and incubated overnight at 4°C with 4 µg of one of the following antibodies: ERα (Santa Cruz, sc-543), H3K4me (abcam, ab8895), Rabbit IgG (Jackson ImmunoResearch, 011-000-003). The complexes were immunoprecipitated for 2 hrs at 4°C with 40 µL of a 1:1 mix of Dynabeads A and G (Life technologies). The beads were washed twice with dialysis buffer (2 mM EDTA pH8, 50 mM Tris pH8.1, 0.2% Sarkosyl) for 15 min with rotation at room temperature and 4 times with wash buffer (0.5 M LiCl, 1% NP40, 1% sodium deoxycholate, 33.2 mM Tris pH8.1). The crosslinking was reversed and the DNA was eluted by heating the beads for 30min at 65°C in elution buffer (50 mM $NaHCO_3$, 1% SDS). Proteins were digested with Proteinase K (ThermoScientific, EO0491) and the DNA was purified on EZ-10 columns (BioBasic). The abundance of the immunoprecipitated DNA fragments was quantified by real time qPCR on a LightCycler 480 System (Roche Life Science) using SYBR Green. ChIP results were analysed by the Percent Input Method. The knockdown efficiency was assessed by Western Blotting with the MLL-C antibody (Millipore, ABE240), using β-actin (Sigma, A5441) as a loading control. The qPCR primers and siRNAs used in the ChIP experiments are listed in Supplementary Table S2.

Statistical analysis K-sample permutation tests were used to analyze the number of nascent transcripts per allele over the time course of E2 induction. The Fisher's exact test was used to compare the proportions of eRNAs that were co-localized and non co-localized with active

TSs. P values were corrected for multiple comparisons using the Holm-Bonferroni method, and a threshold of $p<0.05$ was used throughout.

3.2.6 Acknowledgements

We thank F. Robert, M. Oeffinger, N. Francis, D. Scott, P. Raymond and P. Bensidoun for helpful discussions. We are grateful to J. Kulpa and S. Haidar for technical advice and to W. Li and Q. Ma for providing eRNA sequences.

3.2.7 Author contributions

SR performed the smFISH and immunofluorescence experiments, and analyzed the data. CZ did tissue culture and hormone/drug treatments. TT performed the ChIP experiments. EN wrote scripts in MATLAB. MK performed statistical analysis. SR, CZ, and DZ wrote the manuscript. DZ supervised the project.

3.2.8 References

1. Bulger, M. and Groudine, M. (2010) Enhancers: the abundance and function of regulatory sequences beyond promoters. *Developmental biology*, **339**, 250-257.
2. Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A. *et al.* (2014) Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell*, **158**, 849-860.
3. Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A. and Blobel, G.A. (2012) Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, **149**, 1233-1244.
4. Marsman, J. and Horsfield, J.A. (2012) Long distance relationships: enhancer-promoter communication and dynamic gene transcription. *Biochimica et biophysica acta*, **1819**, 1217-1227.
5. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A. *et al.* (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*, **39**, 311-318.
6. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A. and Wysocka, J. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, **470**, 279-283.
7. Natoli, G. and Andrau, J.C. (2012) Noncoding transcription at enhancers: general principles and functional models. *Annual review of genetics*, **46**, 1-19.

8. Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**, 182-187.
9. Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Ronnerblad, M., Hrydziuszko, O., Vitezic, M. *et al.* (2015) Gene regulation. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*, **347**, 1010-1014.
10. Orom, U.A. and Shiekhattar, R. (2013) Long noncoding RNAs usher in a new era in the biology of enhancers. *Cell*, **154**, 1190-1193.
11. De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L. and Natoli, G. (2010) A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS biology*, **8**, e1000384.
12. Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A. *et al.* (2011) Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*, **474**, 390-394.
13. Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T. and Kraus, W.L. (2011) A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*, **145**, 622-634.
14. Hah, N., Murakami, S., Nagari, A., Danko, C.G. and Kraus, W.L. (2013) Enhancer transcripts mark active estrogen receptor binding sites. *Genome research*, **23**, 1210-1223.
15. Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X. *et al.* (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, **498**, 516-520.
16. Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., Elkon, R., Melo, S.A., Leveille, N., Kalluri, R. *et al.* (2013) eRNAs are required for p53-dependent enhancer activity and gene transcription. *Molecular cell*, **49**, 524-535.
17. Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M. *et al.* (2013) Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, **498**, 511-515.
18. Johnsson, P., Lipovich, L., Grander, D. and Morris, K.V. (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochimica et biophysica acta*, **1840**, 1063-1071.
19. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455-461.
20. Lai, F., Gardini, A., Zhang, A. and Shiekhattar, R. (2015) Integrator mediates the biogenesis of enhancer RNAs. *Nature*, **525**, 399-403.
21. Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.S., Chen, S. *et al.* (2014) Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 7319-7324.
22. Pnueli, L., Rudnizky, S., Yosefzon, Y. and Melamed, P. (2015) RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the

gonadotropin alpha-subunit gene. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 4369-4374.
23. Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L. and Sartorelli, V. (2013) eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Molecular cell*, **51**, 606-617.
24. Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C. and Kim, T.K. (2014) Enhancer RNA facilitates NELF release from immediate early genes. *Molecular cell*, **56**, 29-42.
25. Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A. and Shiekhattar, R. (2013) Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*, **494**, 497-501.
26. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A. and Young, R.A. (2015) Transcription factor trapping by RNA in gene regulatory elements. *Science*, **350**, 978-981.
27. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. and Tyagi, S. (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS biology*, **4**, e309.
28. Sanchez, A. and Golding, I. (2013) Genetic determinants and cellular constraints in noisy gene expression. *Science*, **342**, 1188-1193.
29. Zenklusen, D., Larson, D.R. and Singer, R.H. (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature structural & molecular biology*, **15**, 1263-1271.
30. Chubb, J.R., Trcek, T., Shenoy, S.M. and Singer, R.H. (2006) Transcriptional pulsing of a developmental gene. *Current biology : CB*, **16**, 1018-1025.
31. Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U. and Naef, F. (2011) Mammalian genes are transcribed with widely different bursting kinetics. *Science*, **332**, 472-474.
32. Evans, R.M. and Mangelsdorf, D.J. (2014) Nuclear Receptors, RXR, and the Big Bang. *Cell*, **157**, 255-266.
33. Sanchez, R., Nguyen, D., Rocha, W., White, J.H. and Mader, S. (2002) Diversity in the mechanisms of gene regulation by estrogen receptors. *Bioessays*, **24**, 244-254.
34. Jeong, K.W., Andreu-Vieyra, C., You, J.S., Jones, P.A. and Stallcup, M.R. (2014) Establishment of active chromatin structure at enhancer elements by mixed-lineage leukemia 1 to initiate estrogen-dependent gene expression. *Nucleic Acids Res*, **42**, 2245-2256.
35. Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S. and Brown, M. (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*, **132**, 958-970.
36. Ansari, K.I., Mishra, B.P. and Mandal, S.S. (2008) Human CpG binding protein interacts with MLL1, MLL2 and hSet1 and regulates Hox gene expression. *Biochimica et biophysica acta*, **1779**, 66-73.
37. Rezai-Zadeh, N., Zhang, X., Namour, F., Fejer, G., Wen, Y.D., Yao, Y.L., Gyory, I., Wright, K. and Seto, E. (2003) Targeted recruitment of a histone H4-specific methyltransferase by the transcription factor YY1. *Genes & development*, **17**, 1019-1029.
38. Rozenblatt-Rosen, O., Rozovskaia, T., Burakov, D., Sedkov, Y., Tillib, S., Blechman, J., Nakamura, T., Croce, C.M., Mazo, A. and Canaani, E. (1998) The C-terminal SET

domains of ALL-1 and TRITHORAX interact with the INI1 and SNR1 proteins, components of the SWI/SNF complex. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 4152-4157.
39. Bourdeau, V., Deschenes, J., Laperriere, D., Aid, M., White, J.H. and Mader, S. (2008) Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. *Nucleic Acids Res*, **36**, 76-93.
40. Thompson, R.E., Larson, D.R. and Webb, W.W. (2002) Precise nanometer localization analysis for individual fluorescent probes. *Biophysical journal*, **82**, 2775-2783.
41. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. and Tyagi, S. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods*, **5**, 877-879.
42. Buxbaum, A.R., Wu, B. and Singer, R.H. (2014) Single beta-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science*, **343**, 419-422.
43. Deschenes, J., Bourdeau, V., White, J.H. and Mader, S. (2007) Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 kb of upstream flanking sequences. *The Journal of biological chemistry*, **282**, 17335-17339.
44. Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M.J., Song, X., Zhang, F., Ma, Q., Ohgi, K.A. *et al.* (2014) Enhancer activation requires trans-recruitment of a mega transcription factor complex. *Cell*, **159**, 358-373.
45. Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A. and Brown, M. (2000) Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell*, **103**, 843-852.
46. Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C. *et al.* (2013) Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular cell*, **51**, 310-325.
47. Milne, T.A., Dou, Y., Martin, M.E., Brock, H.W., Roeder, R.G. and Hess, J.L. (2005) MLL associates specifically with a subset of transcriptionally active target genes. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 14765-14770.
48. Bahar Halpern, K., Tanami, S., Landen, S., Chapal, M., Szlak, L., Hutzler, A., Nizhberg, A. and Itzkovitz, S. (2015) Bursty gene expression in the intact Mammalian liver. *Molecular cell*, **58**, 147-156.
49. Dekker, J., Marti-Renom, M.A. and Mirny, L.A. (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics*, **14**, 390-403.
50. Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G. and Heard, E. (2014) Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, **157**, 950-963.
51. Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W. and Furlong, E.E. (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*, **512**, 96-100.
52. Gerlich, D., Koch, B., Dupeux, F., Peters, J.M. and Ellenberg, J. (2006) Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. *Current biology : CB*, **16**, 1571-1578.

3.2.9 Figure legends

Figure 3-1. **eRNAs are low-abundance nuclear transcripts that are induced with similar kinetics as their target genes. A.** Diagram of the genomic organization of the *FOXC1* and *P2RY2* loci. **B.** smFISH showing the expression and localization of FOXC1 (left) and P2RY2 (right) eRNAs and mRNAs before and after 40 min of E2 induction in MCF-7 cells. Arrows indicate different eRNA and transcription site configurations. **C.** Frequency distributions of FOXC1 (left panels) and P2RY2 (right panels) eRNA and active mRNA transcription sites during E2 induction. Representative data of three independent experiments; n = 100 cells per time point.

Figure 3-2. Induction of FOXC1 eRNA and mRNA transcription requires ERα. A. FOXC1 eRNA and mRNA expression in the presence or absence of a 3 h pre-treatment with Tamoxifen (100 nM) or ICI (100 nM) followed by E2 induction (5 nM) for 40 minutes. **B.** Quantification of data from (A). Frequency distributions of FOXC1 antisense eRNAs and nascent FOXC1 transcripts, representative of two independent experiments; n = 70-200 cells for each condition.

Figure 3-3. MLL1 is required for E2-induced eRNA transcription.

A. smFISH analysis for P2RY2 sense eRNA and mRNA combined with MLL1 immunofluorescence in MLL1-depleted or non-specific siRNA-treated cells before and after E2 induction. **B.** Frequency distribution of P2RY2 sense eRNAs and mRNA in MLL1-depleted or non-specific siRNA-treated cells before and after E2 induction. Data are representative of 4 independent experiments; n = 160-200 cells for each condition. **C.** ChIP analysis for H3K4me1 (left) and ERα (right) presence at the *P2RY2* enhancer in MLL1-depleted or non-specific siRNA-treated cells in the presence or absence of E2 treatment; n=2 independent experiments; error bars indicate SD.

Figure 3-4. Simultaneous expression of eRNAs and mRNAs is infrequent in single cells and is not required to maintain transcription. A. Frequency of co-localization between FOXC1 (top panel) and P2RY2 (bottom panel) mRNA transcription sites and their cognate eRNAs using the transcription site as a reference. eRNA-mRNA transcription site co-

localization was scored within a 400 nm radius. **B.** Frequency of co-localization of eRNAs with mRNA transcription sites using the eRNA signal as a reference. Panels A and B indicate the mean and SD from 3 independent experiments. **C.** Quantification of the relationship between eRNA-transcription site co-localization and the RNAPII density on individual alleles. Outlines represent the relative density of data points. Data was pooled from three independent experiments; active transcription sites from n = 300 cells per time point were analyzed.

Figure 3-5. eRNA-mRNA co-expressing alleles are infrequently found in a closed enhancer-promoter loop configuration. A. Determination of the upper limit of co-localization detection (63 nm) using smFISH probes spanning the P2RY2 5'intron labeled in two different colors and 2D Gaussian fitting. **B.** Frequency distribution plots displaying the distances between mRNA transcription sites and nascent eRNAs within a 400 nm radius (red) overlaid with the localization precision plot (green) shown in (A.). The data represents the 40 min E2 induction point combined from 3 independent experiments. **C.** Cartoon illustrating the spatial organization of nascent eRNAs relative to the transcription site. Three different scenarios of eRNA and target mRNA co-expression are observed at single alleles. (i) Simultaneous eRNA and mRNA transcription consistent with a closed enhancer-promoter loop conformation; least frequent. (ii) Simultaneous eRNA and mRNA transcription in an open enhancer-promoter loop conformation and (iii) mRNA transcription from alleles that are not co-expressed with an eRNA; most prevalent. Factors involved in enhancer-promoter communication are shown schematically.

3.2.10 Figures

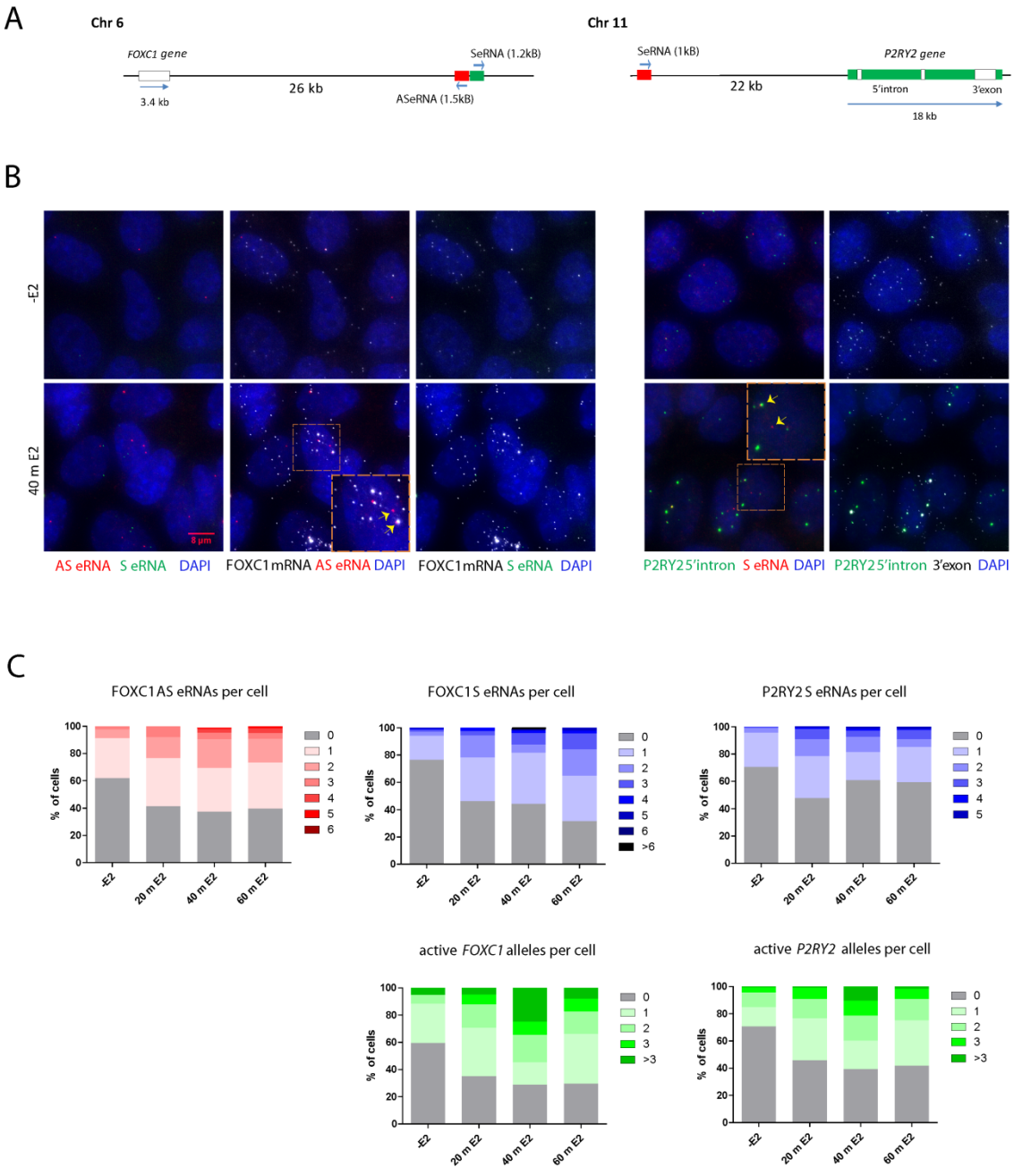


Figure 3-1: eRNAs are low-abundance nuclear transcripts that are induced with similar kinetics as their target genes

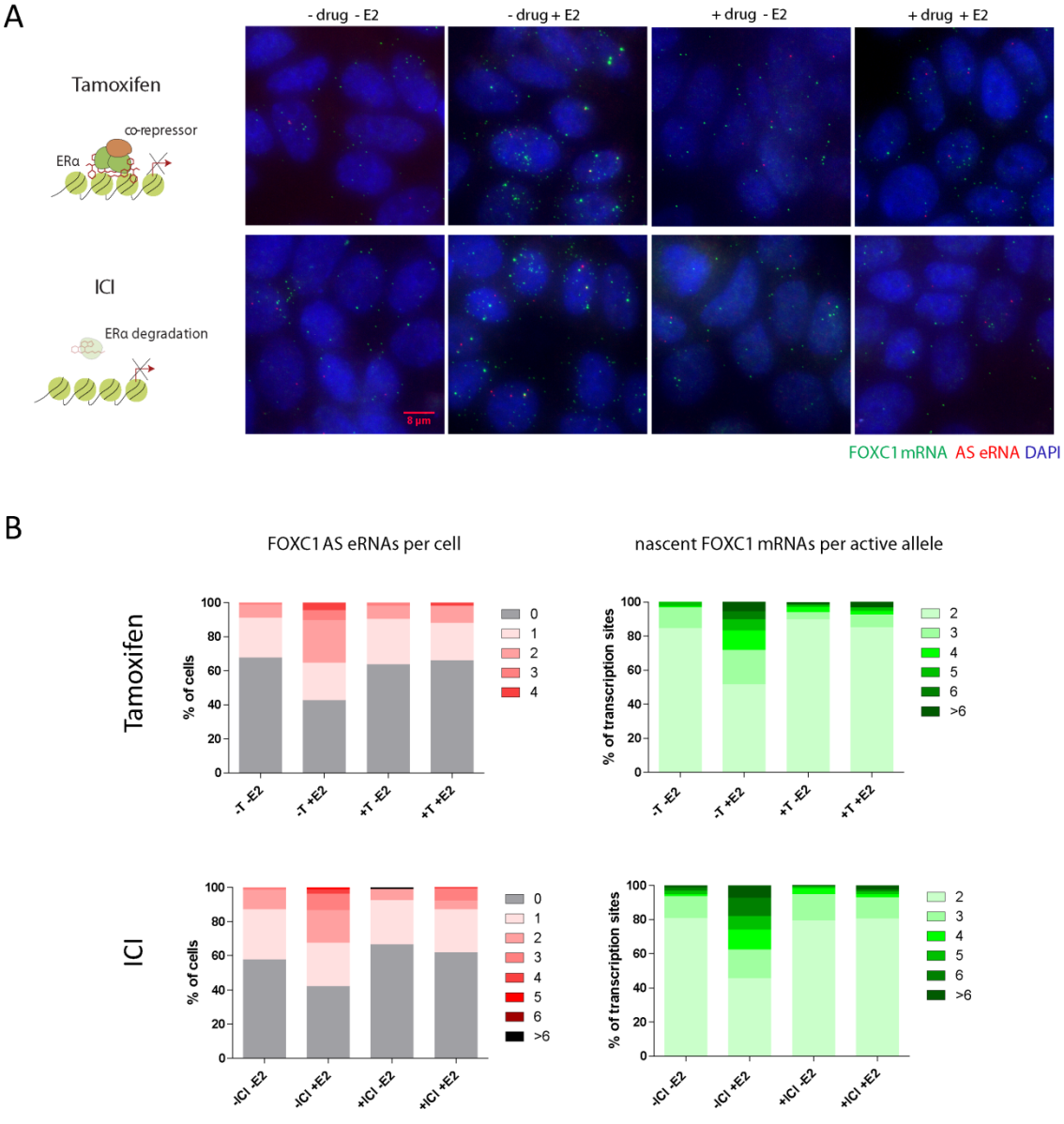


Figure 3-2: Induction of FOXC1 eRNA and mRNA transcription requires ERα

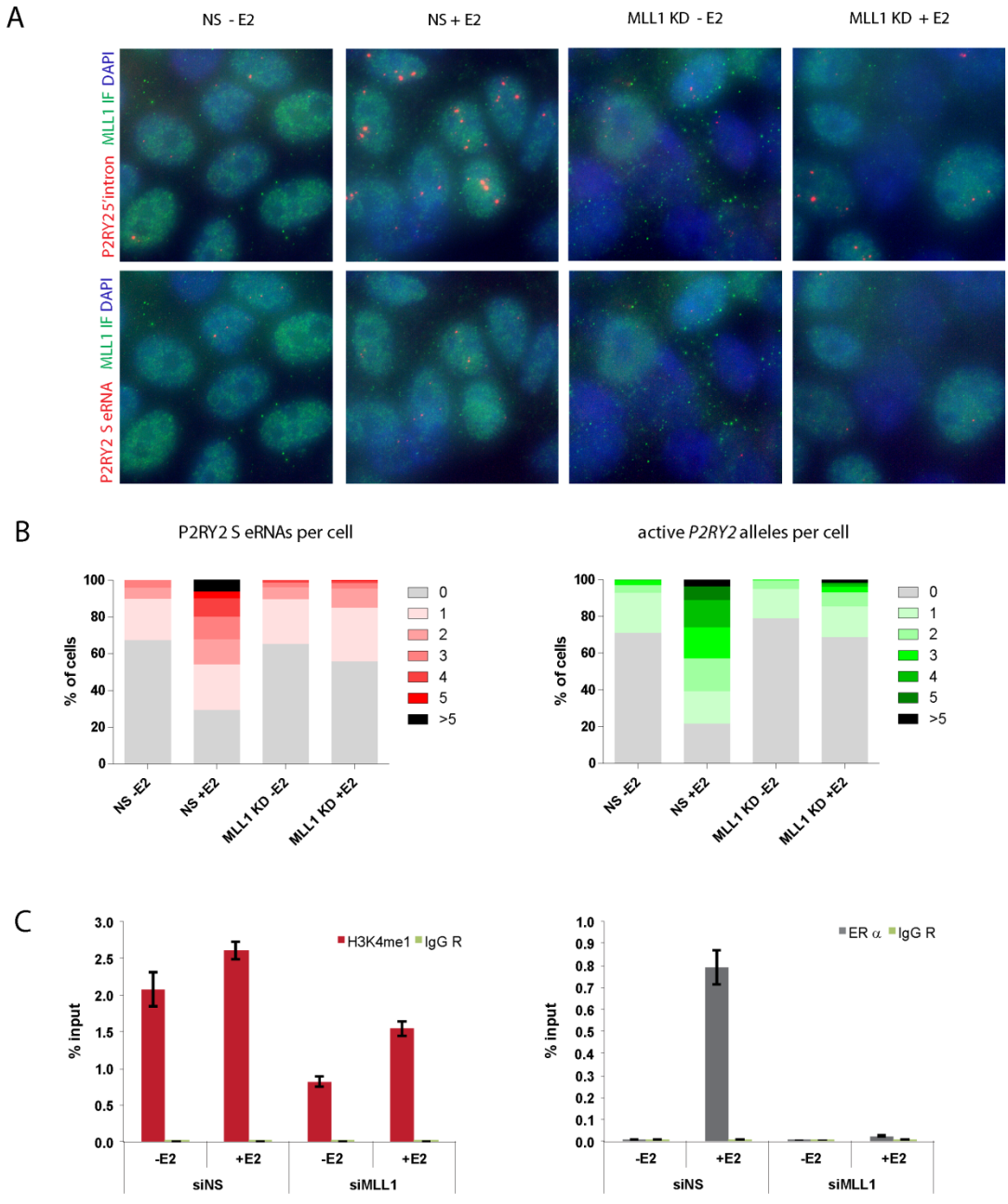


Figure 3-3: MLL1 is required for E2-induced eRNA transcription

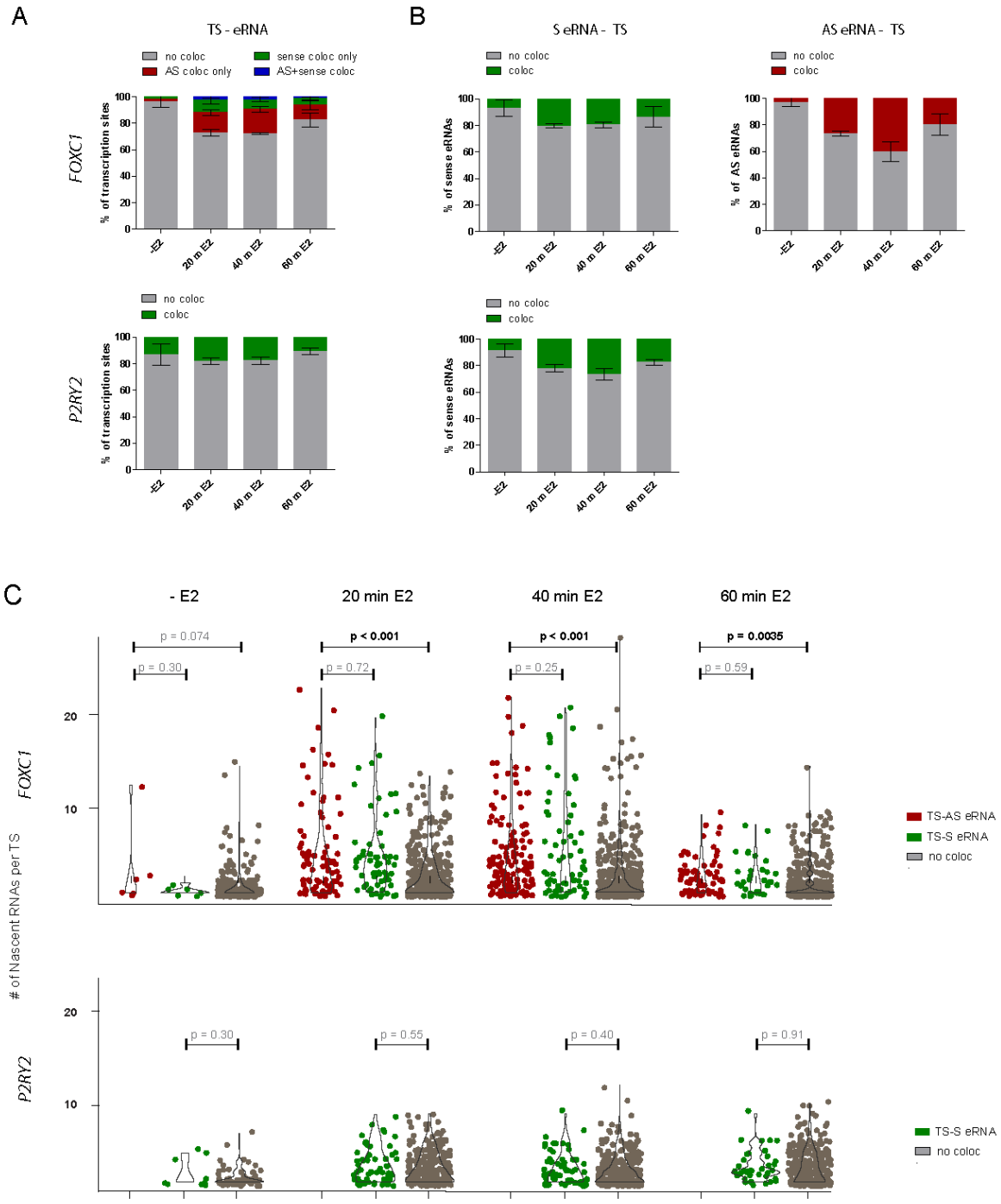


Figure 3-4: Simultaneous expression of eRNAs and mRNAs is infrequent in single cells and is not required to maintain transcription

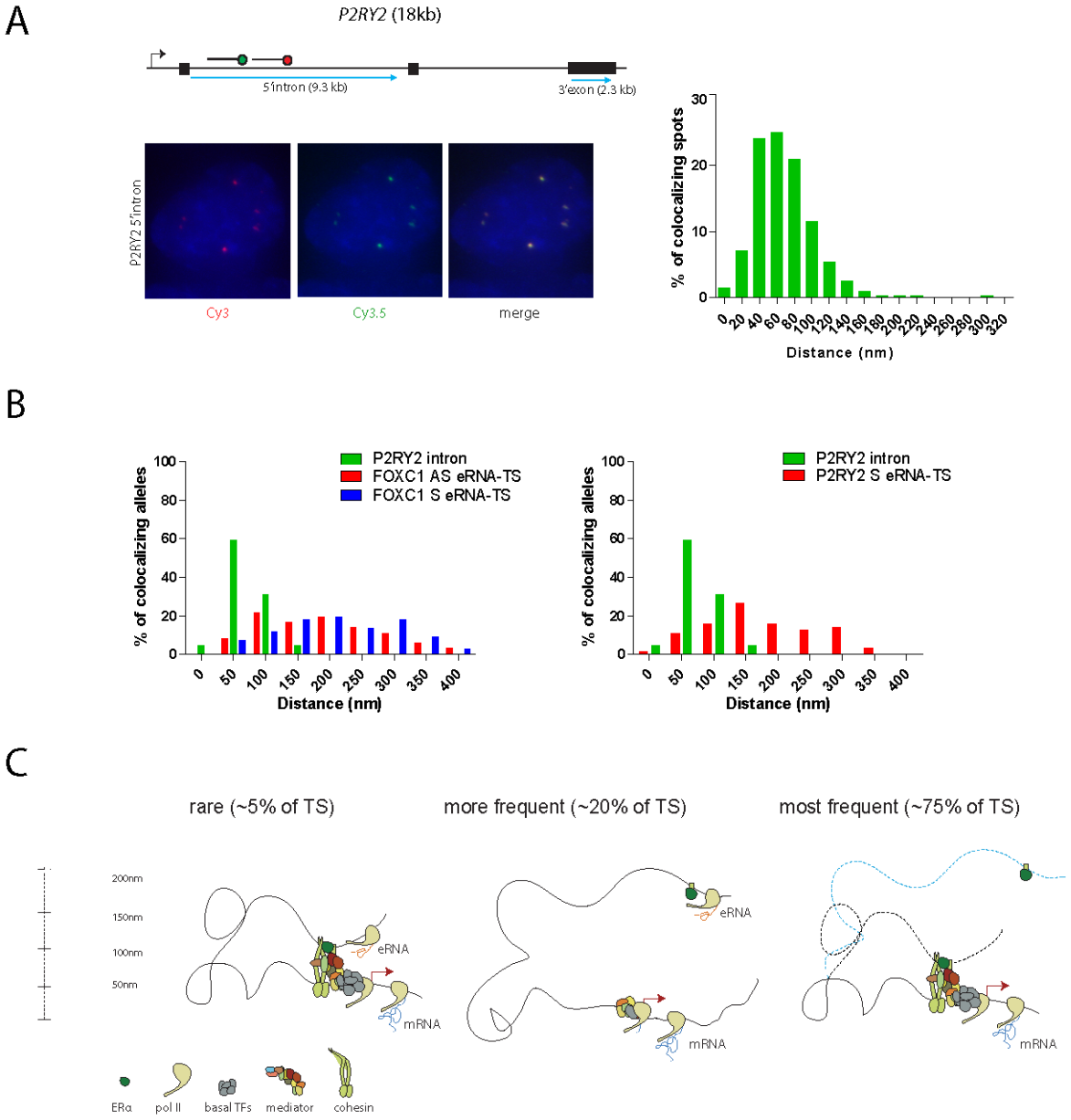
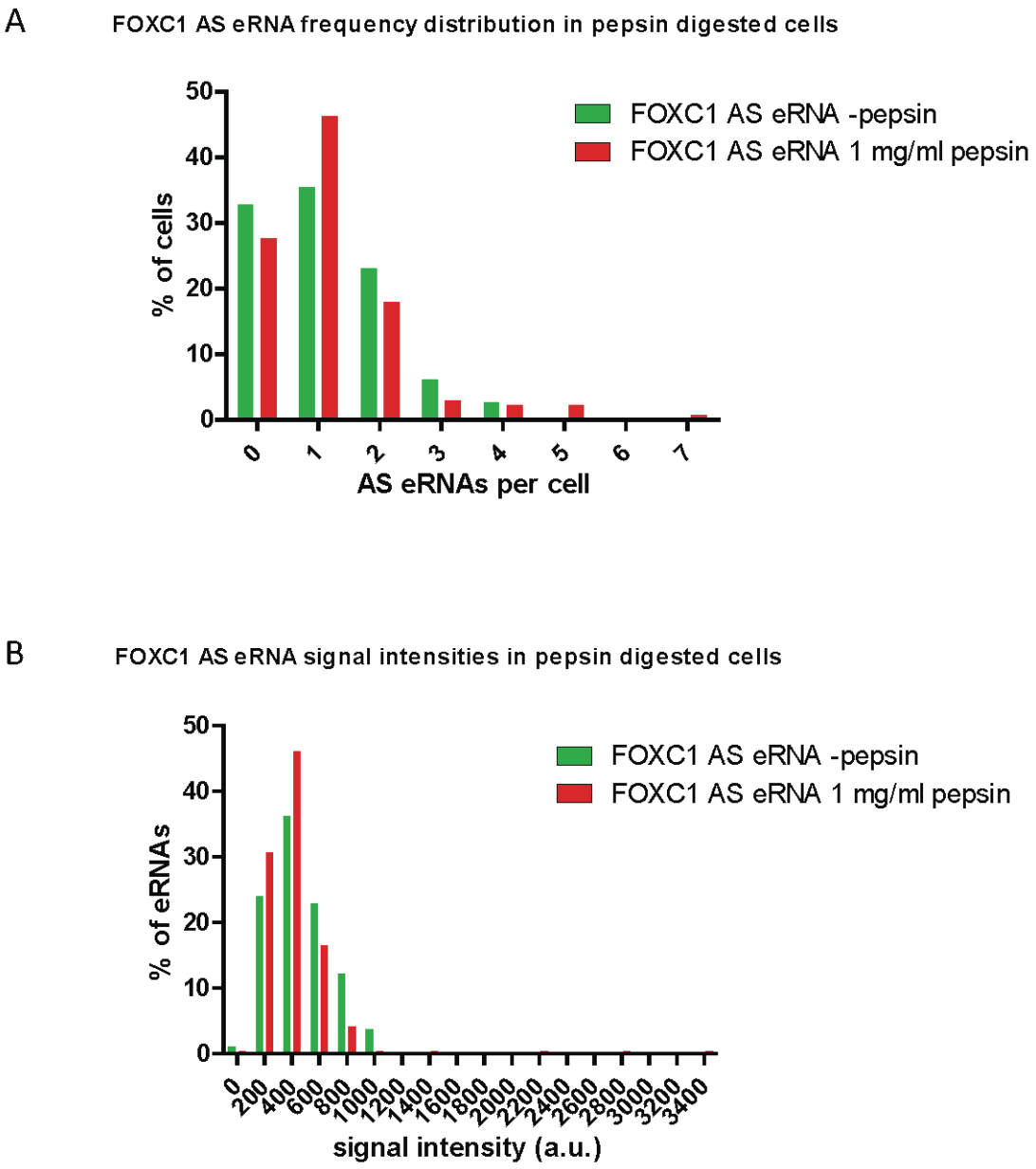


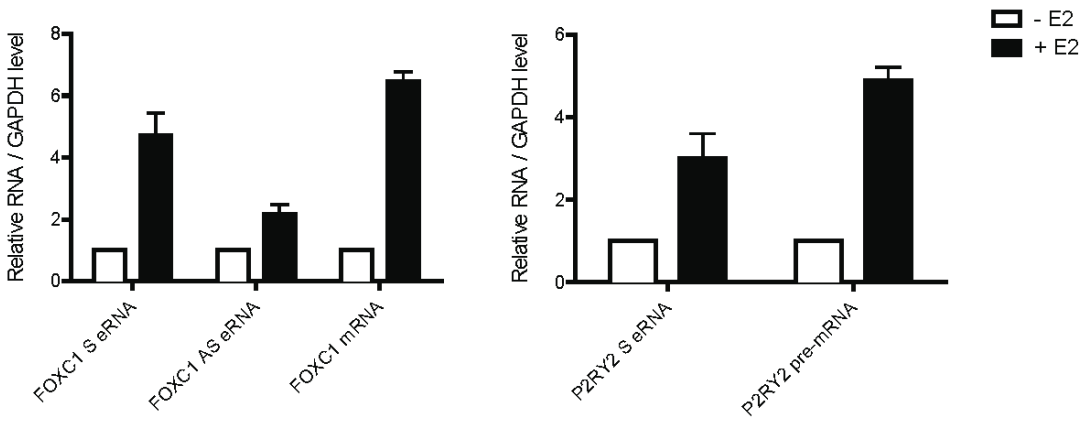
Figure 3-5: eRNA-mRNA co-expressing alleles are infrequently found in a closed enhancer-promoter loop configuration

3.2.11 Supplemental figures and tables



Supplemental Figure S1: Protease digestion to determine the efficiency of eRNA detection by smFISH

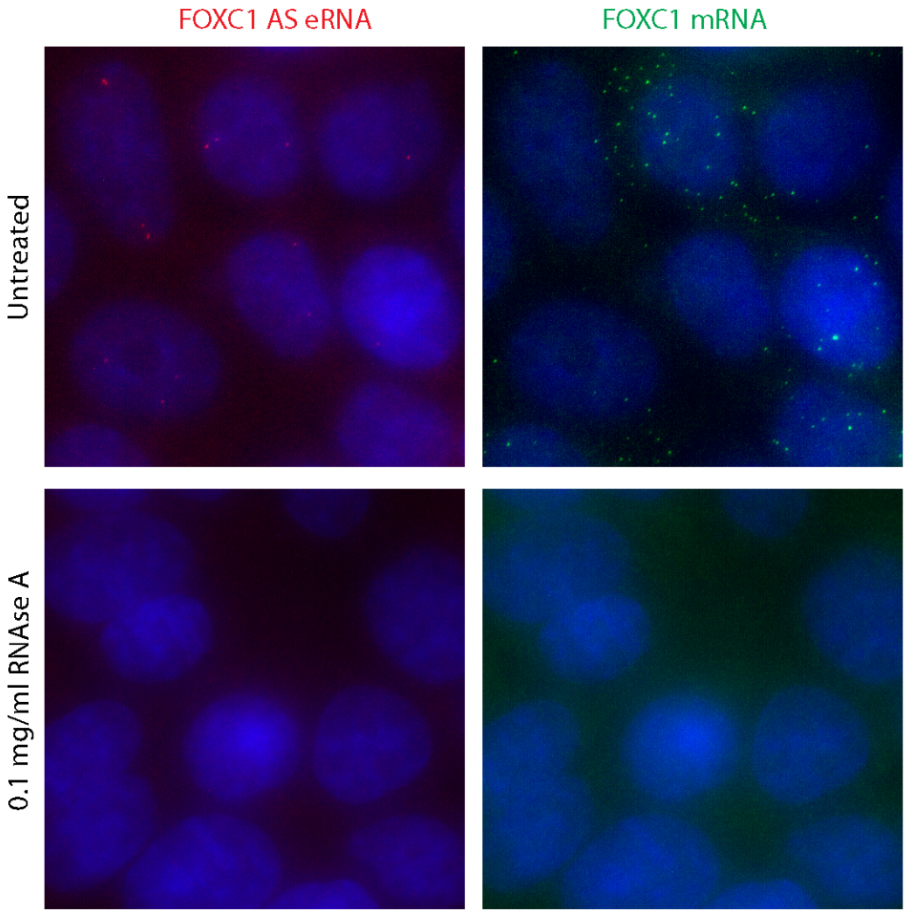
A



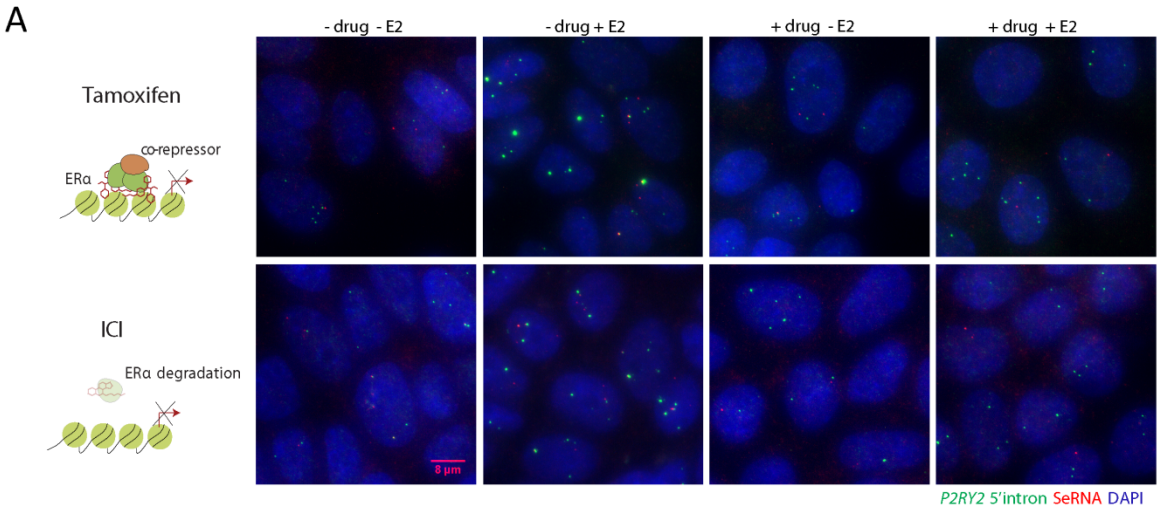
B

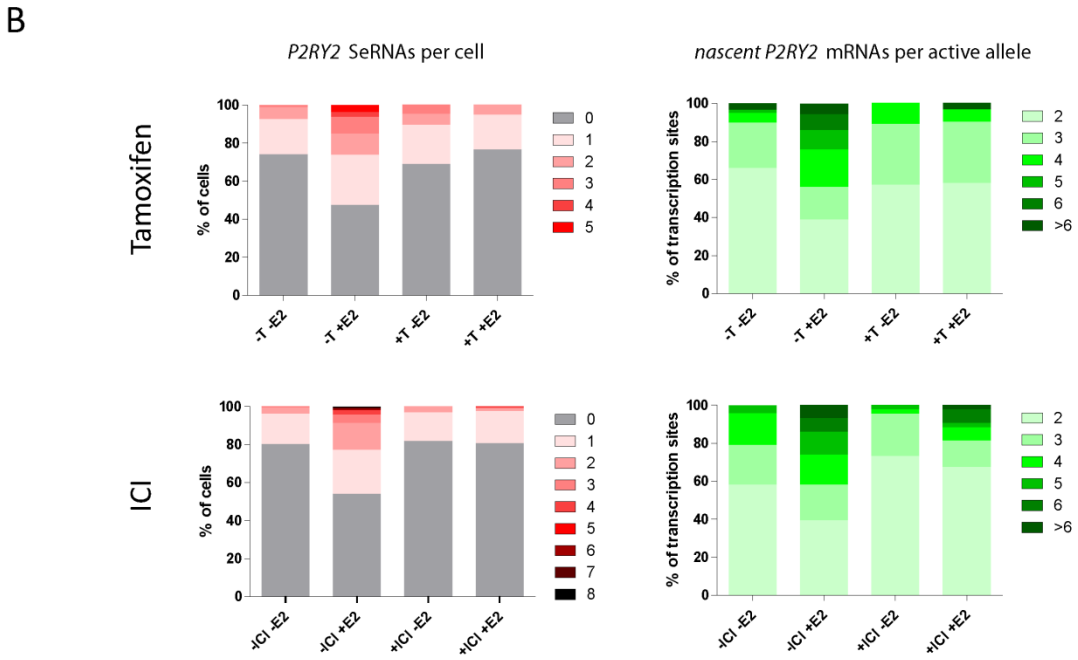
	RT-qPCR	FISH
FOXC1 S eRNA	4.70	4.73
FOXC1 AS eRNA	2.15	2.40
P2RY2 S eRNA	3.0	1.97

Supplemental Figure S2: RT-qPCR quantification of eRNA and mRNA expression in E2-stimulated MCF7 cells

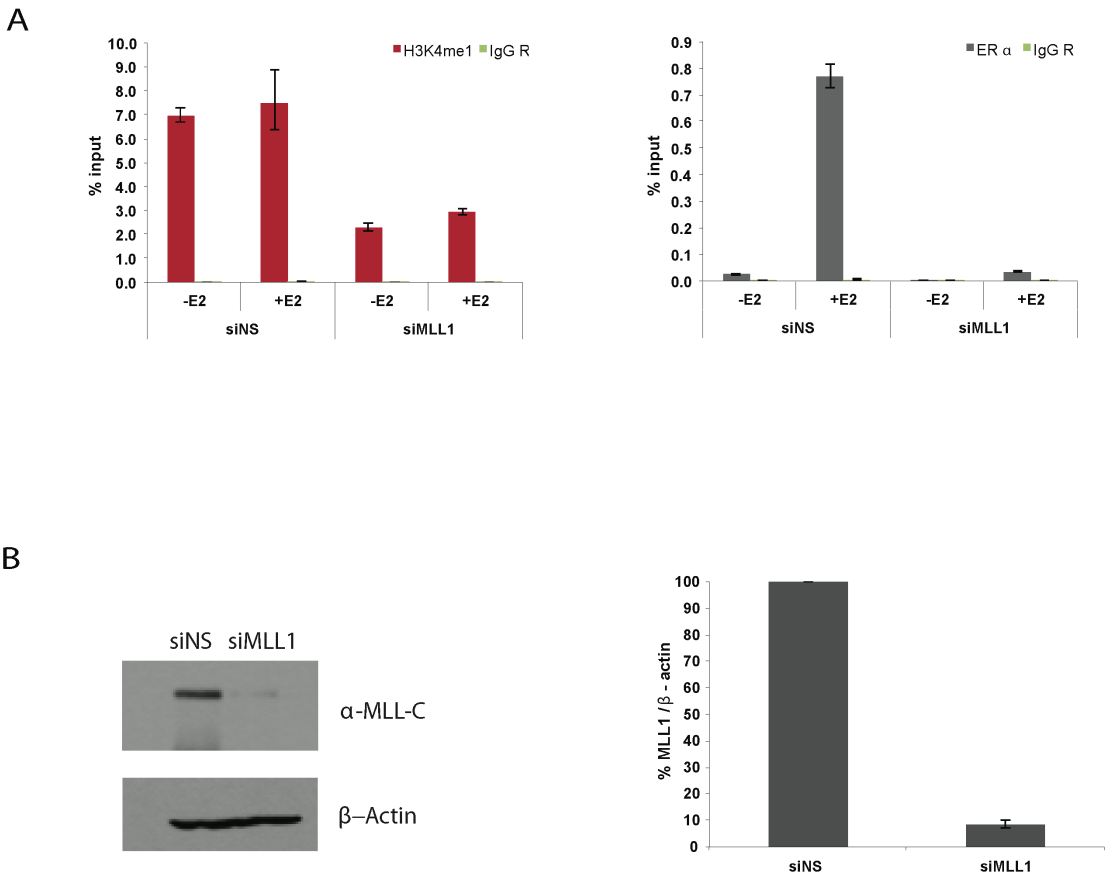


Supplemental Figure S3: RNAse A treatment to validate smFISH signal detection

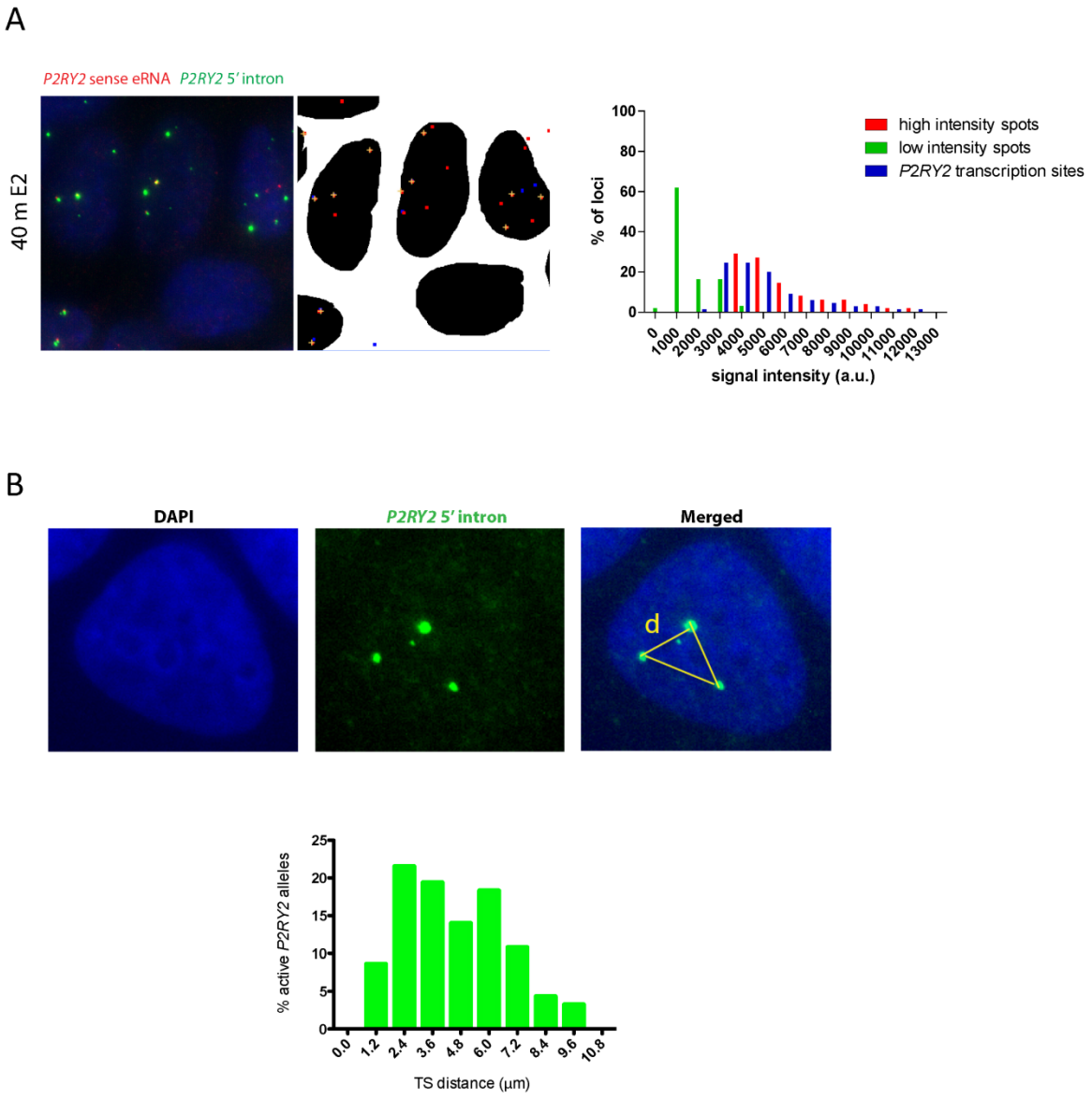




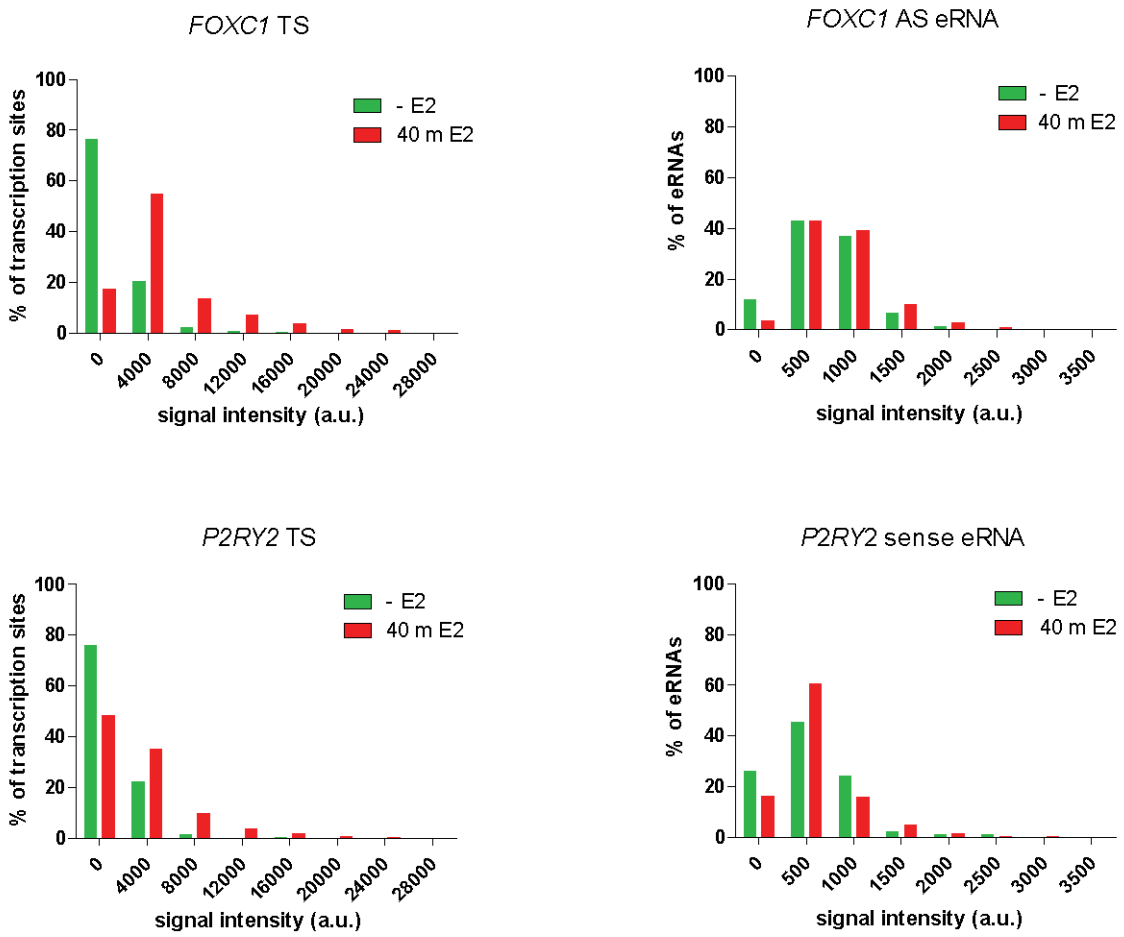
Supplemental Figure S4: Induction of P2RY2 eRNA and mRNA transcription requires ERα



Supplemental Figure S5: ChIP qPCR analysis of ERα and H3K4me1 levels on the *FOXC1* enhancer



Supplemental Figure S6: Transcription site detection



Supplemental Figure S7: Analysis of eRNA and TS intensity changes during E2 induction

Supplemental table S1: List of smFISH probes used in this study

FOXC1 probes

FOXC1 sense eRNA		**FOXC1 AS eRNA**	
FOXC1e_S_1	ttgagcagcagaatggagcg	FOXC1e_AS_1	actcagaaagtgccatggag
FOXC1e_S_2	cctgactgctggaggtaatg	FOXC1e_AS_2	tggagagaaatatgccctgc
FOXC1e_S_3	aatcatccgagagacgcgag	FOXC1e_AS_3	atgtcactgtcacgttagtg
FOXC1e_S_4	caggaggctgaccaatattt	FOXC1e_AS_4	ttgtgttcctcagggacaat
FOXC1e_S_5	cagggcagaccccaaatatc	FOXC1e_AS_5	ggatgtaattcaggggctag
FOXC1e_S_6	aagtcagagagatgcgtggg	FOXC1e_AS_6	ctgagacatgctgtcattca
FOXC1e_S_7	ccgcttcacctttcatgaag	FOXC1e_AS_7	gactggactcattttgggac
FOXC1e_S_8	aagaagtgtgccgagttgtt	FOXC1e_AS_8	aaactggtggctactcacat
FOXC1e_S_9	tagtgtgcttccactgtttg	FOXC1e_AS_9	tgcagatggtcttaggagtt
FOXC1e_S_10	ttcttcggggttctgagaac	FOXC1e_AS_10	ggtgctcattagacccttttg
FOXC1e_S_11	aggagcaggtgaaacgaggg	FOXC1e_AS_11	cctgtatatatttcttcctc
FOXC1e_S_12	taggaaaggacaggggcatc	FOXC1e_AS_12	tctggcttcctcaggaaaga
FOXC1e_S_13	aaaggagagtgctacgcagg	FOXC1e_AS_13	ctccacagctgtgagaatta
FOXC1e_S_14	caagtgagcgaggacaggat	FOXC1e_AS_14	cagggaagctgttacaggtg
FOXC1e_S_15	tagtttctgaagagcagggt	FOXC1e_AS_15	agaaaccctgaagtccagag
FOXC1e_S_16	agagaattgaggcttgctgc	FOXC1e_AS_16	taaagcttgaggtggaggga
FOXC1e_S_17	gatggaaactgcccagattt	FOXC1e_AS_17	gttatgattgacagtggggg
FOXC1e_S_18	ggggtaggtttaaagacgga	FOXC1e_AS_18	caacacggtactgtttaggc
FOXC1e_S_19	gagcaacagttatagaacct	FOXC1e_AS_19	tccagagcacatggaatgtg
FOXC1e_S_20	ttttctccctgacaaaacca	FOXC1e_AS_20	ttataaacagcagggcaggc
FOXC1e_S_21	ggtggagattttgaaggaga	FOXC1e_AS_21	cccgttttggagagaataca
FOXC1e_S_22	caggttgaccttcaccttag	FOXC1e_AS_22	cgtggaacatgctggatgtg
FOXC1e_S_23	ttctccaacctagaggacaa	FOXC1e_AS_23	tctttaagtgttggccactc
FOXC1e_S_24	taaggactcaggacgatggt	FOXC1e_AS_24	acacctgcaggtgtgttatg
FOXC1e_S_25	attactccacgtctgttatg	FOXC1e_AS_25	tagattcaccacgtttgagg
FOXC1e_S_26	agaagccatcatgaagcagg	FOXC1e_AS_26	ggtcacagagacttaggttg
FOXC1e_S_27	cacctgtgctattcacaaag	FOXC1e_AS_27	tggagaagcacatgggattt
FOXC1e_S_28	tggagagacagaccgtccag	FOXC1e_AS_28	ttcttcactgaaactgtgcc
FOXC1e_S_29	atcctgatcttcctgttgga	FOXC1e_AS_29	gacagcctggaatatgcttc
		FOXC1e_AS_30	ctcttcagttctagcaatct
		FOXC1e_AS_31	agaaggatgtaccagtccat
		FOXC1e_AS_32	acagtcagtgacggctctac
		FOXC1e_AS_33	aaatcagtgcctgatgttgg
		FOXC1e_AS_34	cctgttgaatgatatcgacc
		FOXC1e_AS_35	cctaggttccatgatgtata

FOXC1e_AS_36	tgccacttacttcctggatg
FOXC1e_AS_37	acagatccaagctgaattcc

FOXC1 mRNA

FOXC1_1	atggtgatgagcgcgatgta
FOXC1_2	taccgttctcggtcttgatg
FOXC1_3	atgatgttgtccacgctgaa
FOXC1_4	cttgcaggttgcagtggtag
FOXC1_5	tgaccggaggcagagagtag
FOXC1_6	gactcagggacgacgagctg
FOXC1_7	ctctgtgactcgaacatctc
FOXC1_8	cactggagagttgttcaagc
FOXC1_9	atttgacagctactattccc
FOXC1_10	gtacagagactggctggaag
FOXC1_11	tagacgaaagctccggacgt
FOXC1_12	gtgtgtcaaaacttgctaca
FOXC1_13	tcgatttagttcggctttga
FOXC1_14	ttagcttttcctgctttggg
FOXC1_15	gattttgccttgatgggttc
FOXC1_16	cgctggtgtggtgaatattc
FOXC1_17	gctggtgagctgaatttttg
FOXC1_18	atagagttttcttcgtgctg
FOXC1_19	gtggctctgaattaatcggt
FOXC1_20	tatctggagtaacactgtcc
FOXC1_21	atatcttacaggtgagaggc
FOXC1_22	ctcccttcaacataggataa
FOXC1_23	gcgactttcataaacgggga
FOXC1_24	ggaacactttctggcgtttg
FOXC1_25	caggcaatttaacgtcaggt
FOXC1_26	ataggtctcattcaaactga
FOXC1_27	gctctattaaagtatccaga
FOXC1_28	tagcctcaaagcaagctgac
FOXC1_29	acgtatttgtttatatgcca
FOXC1_30	ccgaatcatggactgtcatt
FOXC1_31	ggctgattcatgggcttaaa
FOXC1_32	aggcatcaccgtggtaagac
FOXC1_33	tttatctgtgcatatctggc
FOXC1_34	gcactgcaattttatatgga
FOXC1_35	tggctcacagggatgtataa
FOXC1_36	cttccttcttgttattgctt

FOXC1_37	caaaatgttctgctcctctc
FOXC1_38	gattgtcccctaggtttaac
FOXC1_39	tacaaaaggggagggccaa
FOXC1_40	tttattacatttcctgggtt
FOXC1_41	atctccccaaagtctgattg
FOXC1_42	gtgttgcctctagataggag
FOXC1_43	agcaacagcaattactgctt

P2RY2 probes

P2RY2 sense eRNA

P2RY2e_S_1	cactcaaggtcttgtgcatg
P2RY2e_S_2	tgaggagactgtcacagagg
P2RY2e_S_3	acttgactcagacagttact
P2RY2e_S_4	acaatggatttgcctatggc
P2RY2e_S_5	agcagggggagaggaaagga
P2RY2e_S_6	agagaactggctggatcttg
P2RY2e_S_7	gatacagaagccagcaaaga
P2RY2e_S_8	ccaaagccacaagtccatta
P2RY2e_S_9	ctgcagggaaatggcacctg
P2RY2e_S_10	gagctgataagaagtccaca
P2RY2e_S_11	ctcctgtcaaaactcttgga
P2RY2e_S_12	ggactttaaacagggcttca
P2RY2e_S_13	gtgaaactccactagaagac
P2RY2e_S_14	agctggcataaccagaaaat
P2RY2e_S_15	tggcaatgaccttggaacca
P2RY2e_S_16	agctctgcccatgaactaag
P2RY2e_S_17	caccaagcatgccaaaggtc
P2RY2e_S_18	gacagacctgggaacctgag
P2RY2e_S_19	atggacaggccctagaagac
P2RY2e_S_20	ttttgccactgtgaacactg
P2RY2e_S_21	gtttgcatgtgctgttgttg
P2RY2e_S_22	taaaaccctcctacctttg
P2RY2e_S_23	ataaacaacgctgttccagc
P2RY2e_S_24	ttatctaaagtccgtgttct
P2RY2e_S_25	aattctgctctcctcaaagt
P2RY2e_S_26	ccacatcatcaaccatcata
P2RY2e_S_27	actatcaagaatcactacct
P2RY2e_S_28	tcagtggtgagactccataa
P2RY2e_S_29	tccctgggtaaggaaaacta

P2RY2 5'intron

P2RY2_intron_1	acattaccaatggtaggtgg
P2RY2_intron_2	gcaccatttgccagattaaa
P2RY2_intron_3	aataaacttgcctgggtctc
P2RY2_intron_4	acccactttgataaccagat
P2RY2_intron_5	cactgttctggcaaactgtg
P2RY2_intron_6	ctgcacttggcaatacttga
P2RY2_intron_7	ttgacatcacaaatgccctg
P2RY2_intron_8	caaacaactcctgggacagt
P2RY2_intron_9	aaaagtgtttggccaagggg
P2RY2_intron_10	atgagatcaattgcaggtcc
P2RY2_intron_11	gcaggaactggtcattcaac
P2RY2_intron_12	tgcaggatcttacttgttct
P2RY2_intron_13	ttggtagtcatttctgaggg
P2RY2_intron_14	ctacaagtgtcacattcctc
P2RY2_intron_15	ccatgttctcttatagtctt
P2RY2_intron_16	taacaagcacaatcccttgc
P2RY2_intron_17	cagggcagaattcaaactcc
P2RY2_intron_18	atcagcagttgggtaatctg
P2RY2_intron_19	gatcaacaaggcatctctgg
P2RY2_intron_20	cctacagaaagcagtttggg
P2RY2_intron_21	aggacaaccaagaggaggtg
P2RY2_intron_22	tatcagggcccagtaataag
P2RY2_intron_23	ctggatgagtacagaggtca
P2RY2_intron_24	cagaaccaactatggcactc
P2RY2_intron_25	tactcatcttggaccaggaa
P2RY2_intron_26	agagggtcaaggagttgttc
P2RY2_intron_27	tgacagtgacaaggtaccag
P2RY2_intron_28	aaatcacattcccaggaacc
P2RY2_intron_29	gggtctggattcagagtaat

P2RY2e_S_30	tatccccagagagatatatc
P2RY2e_S_31	gtgacctacaaacctactgt
P2RY2e_S_32	ctcttcttgactctctttat
P2RY2e_S_33	tacccataggacatatttcc
P2RY2e_S_34	cttgtctccataatcaaact
P2RY2e_S_35	taattcttttgtgtctctga
P2RY2e_S_36	caggtcaagtgggagaactc
P2RY2e_S_37	atgattaccttgtgctagga
P2RY2e_S_38	gccatttttggggaatcatat

P2RY2_intron_30	ttttgagattctcagagcca
P2RY2_intron_31	aggctgtgtttcttcaattc
P2RY2_intron_32	tctcgcagtctaattccaag
P2RY2_intron_33	catcaaccaggaaccacatg
P2RY2_intron_34	caaggtggaaaggggggatga
P2RY2_intron_35	aaattcaggagcccatgaca
P2RY2_intron_36	ctgtcacaacagggagacac
P2RY2_intron_37	ttggcgctagcaaatcatga
P2RY2_intron_38	gctacagagacatgctacta
P2RY2_intron_39	tattgtcagtgtgggtttga
P2RY2_intron_40	taacctcagatgctatcagc
P2RY2_intron_41	cggtctgaggaaatggctaa
P2RY2_intron_42	caggctgaggagcttagaag
P2RY2_intron_43	ttcacaaaggttgccacatc
P2RY2_intron_44	acaggaaggtcacaaccaga
P2RY2_intron_45	cgcaggaattatgggaaggg
P2RY2_intron_46	tggggttatctcaggaagac
P2RY2_intron_47	gcaggtaaagaaggcatgga
P2RY2_intron_48	tcattgagggggcaagagag

P2RY2 3'exon

P2RY2_exon_1	cattgatggtgtcattccag
P2RY2_exon_2	ttgaagtcctcgttgaagcg
P2RY2_exon_3	caagaagatgtagagcgcca
P2RY2_exon_4	gacgcattccaggtcttgag
P2RY2_exon_5	gccaggtggaacatatatgt
P2RY2_exon_6	cgcatacagtgcatcagaca
P2RY2_exon_7	gtagtaatagaccagcagcg
P2RY2_exon_8	tagaagaggaagcgcaccag
P2RY2_exon_9	gatgctgcagtaaaggttgg
P2RY2_exon_10	cgctgatgcaggtgaggaag
P2RY2_exon_11	gagcgcagaggtcgtaagac
P2RY2_exon_12	gctggtggtgacaaagtaga
P2RY2_exon_13	acgaagcggctgaagagctc
P2RY2_exon_14	cagcatgactgagctgtagg
P2RY2_exon_15	aaggatgacggcaaagggca
P2RY2_exon_16	gagccatgagcacgtaacag
P2RY2_exon_17	tggaatggcaggaagcagag
P2RY2_exon_18	agcggaaggagtagtagagg
P2RY2_exon_19	taaccttgtaggccatgttg

P2RY2_exon_20	tcaaggcaactgttagcact
P2RY2_exon_21	cagccaggaagtagagcacg
P2RY2_exon_22	atctcgggcaaagcgtacga
P2RY2_exon_23	atgtcagttctgtcggatct
P2RY2_exon_24	caacacatcttctatcctct
P2RY2_exon_25	gaatgtccttagtgttctcg

Supplemental table S2: List of qPCR primers and siRNAs used in this study

RT-qPCR Primers

FOXC1 sense eRNA	5’-CATGAAAGGTGAAGCGGAAATAC-3’ (forward)
	5’- TGAAGGAGCAGGTGAAACG -3’ (reverse)
FOXC1 AS eRNA	5’-CTGAGGAACACAAGACTAGCC-3’ (forward)
	5’- ACTGGACTCATTTTGGGACATC-3’ (reverse)
FOXC1 mRNA	5’-AGTCAGCTTGCTTTGAGGCTA-3’ (forward)
	5’- AGGCATCACCGTGGTAAGAC -3’ (reverse)
P2RY2 sense eRNA	5’-AGCTTCTGGTTCCAAGGTCA-3’ (forward)
	5’- CATGTGCTGTTGTTGCTGTG-3’ (reverse)
P2RY2 intron	5’-ACTGCTCTGACCATGACCTC-3’ (forward)
	5’-TGACACTGCTTGGTAGGGAG-3’ (reverse)
GAPDH mRNA	5’-GTTTTTCTAGACGGCAGGTCA-3’ (forward)
	5- AACATCATCCCTGCCTCTACT-3’ (reverse)

ChIP-qPCR Primers

P2RY2e_ChIP	5'-CCATCAAAGCTGTTGCTTCT-3' (forward)
	5'-CCAGGATAGTGCCAGTGAAC-3' (reverse)
FOXC1e_ChIP	5'-CTGAGGAACACAAGACTAGCC-3' (forward)
	5'-ACTGGACTCATTTTGGGACATC-3' (reverse)

siRNAs

MLL1	5'-GAUUCGAACACCCAGUUAUdTdT-3' (sense)
	5'-AUAACUGGGUGUUCGAAUCdTdT-3' (anti-sense)
NS	5'-UUCUCCGAACGUGUCACGUdTdT-3' (sense)
	5'-ACGUGACACGUUCGGAGAAdTdT-3' (anti-sense)

3.2.12 Supplemental figure legends

Supplemental Figure S1: Protease digestion to determine the efficiency of eRNA detection by smFISH.

A. Frequency distributions of FOXC1 AS eRNAs in pepsin-treated and untreated MCF7 cells induced with 100 nM E2 for 40min, n=120-140 cells. **B.** Frequency distributions of FOXC1 AS eRNA signal intensities in pepsin-treated and untreated MCF7 cells induced with 100nM E2 for 40 min; n=120-140 cells.

Supplemental Figure S2: RT-qPCR quantification of eRNA and mRNA expression in E2-stimulated MCF7 cells.

A. RT-qPCR analysis of the fold change in expression of FOXC1 (left) or P2RY2 (right) eRNA and mRNA relative to GAPDH in MCF-7 cells in response to E2 treatment for 40 min; n=2; error bars indicate SEM. **B.** Comparison of RT-qPCR and smFISH measurements of FOXC1 and P2RY2 eRNAs.

Supplemental Figure S3: RNAse A treatment to validate smFISH signal detection. Representative images show FOXC1 AS eRNA and mRNA signals in MCF7 cells that were either left untreated or treated with 0.1 mg/ml of RNAse A in 2xSSC for 1 h at 37^{o}C.

Supplemental Figure S4: Induction of P2RY2 eRNA and mRNA transcription requires ERα.

A. P2RY2 eRNA and mRNA expression in the presence or absence of a 3 hour pre-treatment with Tamoxifen (100 nM) or ICI (100 nM) followed by E2 induction (5 nM) for 40 min. **B.** Quantification of data from (A). Frequency distributions of P2RY2 sense eRNAs and nascent P2RY2 transcripts, representative of 2 independent experiments, n=70-200 cells for each condition.

Supplemental Figure S5: ChIP qPCR analysis of ERα and H3K4me1 levels on the *FOXC1* enhancer.

A. ChIP analysis for H3K4me1 (left) and ERα (right) in MLL1-depleted or non-specific siRNA-treated cells in the presence or absence of E2 treatment; n=2 independent experiments; error bars indicate SD **B.** Quantification of the MLL1 knockdown efficiency by Western Blotting.

Supplemental Figure S6: Transcription site detection.

A. mRNA spots were localized by 2D Gaussian fitting and clustered into a high-intensity or a low-intensity group. The mean value of the low-intensity group was considered to correspond to the intensity of a single RNA signal, which was used to quantify the number of nascent transcripts per active allele. In the diagram, spots marked by yellow crosses correspond to *P2RY2* transcription sites (left). Frequency distributions of P2RY2 intron signal intensities in the high and low-intensity clusters from a representative field of 40 min E2-induced MCF-7 cells (right). **B.** 2D distance measurements between active transcription sites in E2 treated MCF-7 cells; n=93 alleles.

Supplemental Figure S7: Analysis of eRNA and TS intensity changes during E2 induction.

Frequency distributions of *FOXC1* and *P2RY2* eRNA and TS signal intensities (a.u.) in uninduced and in 40 min E2-induced MCF-7 cells. Signal intensity values were pooled from 3 time course experiments.

4 Discussion

4.1 General goals of thesis

Deep sequencing methods have led to the discovery of a vast number of long non-coding RNAs in eukaryotes, of which very few have been functionally characterized. The general goals of my PhD thesis were to elucidate the roles of different classes of lncRNAs in transcription regulation and to discover the key factors implicated in their biogenesis. To achieve these goals, I had focused my studies on two distinct classes of lncRNAs in yeast and mammalian cells as model systems, using single molecule quantitative microscopy in combination with genetic and biochemical approaches to glean novel mechanistic insights. In this discussion, I will elaborate on the findings of these two studies separately, discussing them within the broader context of recent literature.

4.2 Article 1: Bimodal expression of *PHO84* is modulated by early termination of antisense transcription

4.2.1 Objectives and summary of results

For my first project, I investigated a mechanism of ncRNA mediated transcription regulation in the simple eukaryote *S. cerevisiae*. *S. cerevisiae* is an important model organism for studying the fundamental mechanisms of gene regulation and has been used extensively to study different aspects of transcription. In contrast to mammalian genomes, the *S. cerevisiae* genome is very gene dense and regulatory elements are most often limited to gene promoters, suggesting less complex regulatory mechanisms than in higher eukaryotes. Despite this, *S. cerevisiae* expresses many different classes of ncRNAs and serves as an important model to study how ncRNAs participate in transcription regulation. In the first part of my thesis, I investigated the role of a class of lncRNAs that are initiated in the 3' UTR of protein coding genes and transcribed in anti-sense to coding regions. In particular, I focused on a lncRNA that is transcribed in antisense to the yeast gene *PHO84*, which encodes a high affinity phosphate transporter. This lncRNA had been studied previously and serves as a model gene to study anti-sense transcription in *S. cerevisiae*, but much of the mechanistic detail of how it functions in regulating *PHO84* transcription was still not understood. One interesting feature of the previous studies was that depleting Rrp6p, a component of the nuclear exosome,

resulted in an increased expression of the anti-sense RNA that was paralleled with a decrease in *PHO84* expression. This led to the hypothesis that *PHO84AS* RNAs stabilize in the absence of an active nuclear exosome, allowing the accumulation of these lncRNAs at the *PHO84* locus by an unknown mechanism, and resulting in the recruitment of the histone de-acetylase Hda1 to the *PHO84* promoter, leading to repression of *PHO84* transcription. Based on the model proposed in these early studies, we further elucidated the mechanisms of biogenesis and function of this AS RNA to enhance our understanding of its regulatory function, and possible mechanisms of AS mediated gene regulation in general.

Our study, published in the paper entitled "Bimodal expression of *PHO84* is modulated by early termination of antisense transcription" challenged the previous model, and revealed a more complex regulation of *PHO84* regulation than previously anticipated. We first showed that AS RNAs levels are regulated primarily at the level of transcriptional elongation by the Nrd1-Nab3-Sen1 complex, which requires Rrp6 for efficient recruitment to the 3'end of *PHO84*. The loss of Rrp6 attenuates the activity of Nrd1 and results in increased AS transcriptional read-through into the *PHO84* promoter which leads to the silencing of *PHO84* transcription. Thus, we reveal a bi-modal pattern of expression, where cells that express AS do not express *PHO84*, and vice versa. Furthermore, we could demonstrate that while *PHO84* is expressed in strong bursts in only a small fraction of cells, AS RNA is expressed continuously at low frequency maintaining the *PHO84* promoter silent in most cells until the inducing signal of phosphate starvation reaches a critical threshold. Upon induction however, sense transcription becomes dominant, as we rapidly detect a synchronous expression of *PHO84* in the entire population, and *PHO84AS* is completely repressed. In the following sections I will discuss different aspects of *PHO84* regulation and anti-sense RNA mediated gene regulation in *S. cerevisiae* in general.

4.2.2 Bi-modal switch vs dose-dependent regulation of gene transcription

Our investigation of *PHO84* regulation by AS RNA was in part motivated by specific mechanistic questions, which were not addressable by previous studies due to the technical limitations of the experiments conducted. A previous study from Camblong and colleagues

used northern blot analysis to show that deletion of the nuclear exosome component Rrp6 results in an increase in AS RNA expression, leading to complete suppression of *PHO84*. Intriguingly, wild type cells express both *PHO84* mRNA and AS RNA. We surmised that this could represent a population where sense and AS RNAs are expressed simultaneously in individual cells, or a population that consists of two discrete classes of cells, those that express sense or those that express AS. Since classical biochemical assays like northern blots produce ensemble measurements of RNAs that have been isolated from cell populations, they cannot distinguish between the two scenarios described above. If the first scenario is true, the complete suppression of *PHO84* in Δ*rrp6* cells may be a cumulative effect of the accumulation of AS RNAs, which gradually tune down the transcriptional output of the gene over time. Thus, there may be a spectrum of different expression levels of both sense and AS in individual cells. Conversely, if the second scenario holds true, in which cells either express sense or AS, *PHO84* repression by AS RNA is likely to occur in a binary switch-like manner, as opposed to the first scenario.

To address these mechanistic questions, we investigated *PHO84* regulation using smFISH, which allowed us to localize and quantify single transcripts for both *PHO84* mRNA and AS RNA simultaneously in individual cells. Our results revealed a bi-modal expression pattern, cells expressing AS RNA do not express *PHO84* mRNA and *vice versa*. Also, in contrast to previous models suggesting the accumulation of AS RNAs at sites of transcription, we did not detect *PHO84* AS RNA accumulation transcription sites, but showed that most *PHO84* AS RNAs were localized in the cytoplasm, suggesting rapid release of these RNAs after synthesis and nuclear export, similar to mRNAs. In addition, *PHO84* mRNA and AS RNA were transcribed by different modes of transcription. Cells expressing *PHO84* mRNA showed high levels of cytoplasmic *PHO84* mRNAs, as well as strong signals at the site of transcription corresponding to multiple nascent transcripts, characteristic of a bursting transcription pattern, characterized by frequent transcription re-initiation events during short active periods of promoter ON times [26,27]. Conversely, the AS RNA is expressed at lower steady-state levels, with small numbers of cytoplasmic RNAs (1-2), and transcription foci that rarely show more than one nascent transcript per allele. Therefore, *PHO84* AS RNA shows a constitutive expression pattern, that can be the result of a promoter that is always on, but

where transcription is initiated infrequently through single, uncorrelated initiation events [27]. Intriguingly, the Δ*rrp6* strain showed a 2-fold increase in the percentage of cells expressing AS RNA, with higher number of RNA per cell (2-3), and a nearly 2-fold decrease in percentage of cells expressing sense. These results revealed that a greater fraction of cells were switching off the *PHO84* promoter in response to increased numbers of AS RNA in Δ*rrp6*, which could be caused by an increase in AS RNA stability due to the role of Rrp6 in non-coding RNA degradation, or an increase in AS RNA transcription frequency.

Various studies have made use of single cell approaches to elucidate mechanisms of lncRNA mediated transcription regulation in yeast, and revealed different expression patterns of lncRNAs. Bi-modal expression, as shown for the *PHO84* mRNA and AS RNA, or a heterogeneous pattern of co-expression of lncRNA and mRNA [150,151]. smFISH has been used to investigate the regulatory circuit on the *FLO11* promoter that is mediated by a toggle switch between two lncRNAs, *PWR1* and *ICR1*, which either activate or repress *FLO11* transcription [150]. The toggle switch is modulated by competition between the transcription factors Flo8p and Sfl1p, which activate the transcription programs for *PWR1* and *ICR1,* respectively, on opposite strands. Flo8p activates transcription of *PWR1*, which interferes in *cis* with *ICR1* transcription, thereby activating *FLO11*. During FLO11 activation, Cti6p, a component of the HDAC Rpd3, maintains a repressive chromatin state at the Sfl1 binding site, inhibiting its recruitment, and thereby repressing *ICR1* transcription. *PWR1* and *ICR1* expression is mutually exclusive, and *ICR1* expression is mostly anti-correlated with *FLO11* in single cells, whereas *PWR1* is co-transcribed with *FLO11*. In wild type cells, *FLO11* shows 3 modes of expression: silent, basal (<5), and active (>5). Intriguingly, depletion of Flo8p or Cti6p, which results in de-repression of *ICR1*, does not completely silence *FLO11* in all cells, but reduces expression in most cells to a basal level. Conversely, depleting Sfl1p significantly increases the transcriptional output of *FLO11*, with 98% of cells expressing at an active, rather than basal level, containing an average of 36 transcripts per cell. Overall, this suggests that the toggle switch between *PWR1* and *ICR1* at the *FLO11* promoter regulates the transition from a basal transcriptional state to a highly active state. Such a bi-modal expression pattern between the repressive lncRNA *ICR1* and the target *FLO11* mRNA is analogous to our findings, but

whereas *FLO11* is expressed in 70% of cells, *PHO84* is only expressed in 20% of wild type cells in conditions of intermediate phosphate levels. Therefore, *PHO84* may be more efficiently silenced by AS RNA. Since *PHO84* contains a TATA box promoter, which is regulated by the SAGA complex, the mechanism of transcriptional activation would follow a 2 state bursting model, whereby the promoter remains silent most of the time and infrequently transitions to a highly active state. Our data suggests that the AS RNA expression level is negatively correlated with the frequency of *PHO84* promoter activation.

In contrast to *PHO84* and *FLO11* regulation by lncRNAs, regulation of the yeast sporulation-inducing genes *IME1* and *IME4* does not result in a bi-modal expression pattern (Van Werven et al., 2012). Sporulation is required for gametogenesis in diploid cells, but is repressed in haploid cells. The lncRNA *IRT1*, which is transcribed in sense from the promoter of *IME1*, represses *IME1* in haploid cells by recruiting the Set2 histone methyltransferase and the HDAC Set3. Therefore, regulation of *IME1* recapitulates similar epigenetic mechanisms of gene regulation as described for *PHO84*. However, the kinetics of *IME1* regulation by *IRT1* in single cells is not consistent with a bi-modal switch, but a dose-dependent graded response. When haploid cells are transferred to a sporulation inducing medium, most cells express *IME1* at low levels. Gradually, over the time course of sporulation, *IRT1* levels increase and *IME1* levels decrease, with cells harboring intermediate levels of both *IRT1* and *IME1* transcripts between 0.5-1 h post-sporulation. After 4 h in sporulation medium, haploid cells completely repress *IME1,* and accumulate large numbers of *IRT1* RNAs. *IRT1* transcription foci increase in size over time, indicating an increase in transcription initiation frequency, which gradually represses *IME1* transcription. Since haploid cells need to repress sporulation inducing genes to maintain their viability, it would not be advantageous to regulate *IME1* through a bi-modal switch, as cells would be able to rapidly re-activate *IME1*. Therefore, a gradual increase in *IRT1* transcription over time may help to establish a stable repressive chromatin state that is less easily reversible than a bi-modal promoter that allows rapid switching between an active and inactive state.

Altogether, these studies show how lncRNA mediated transcription regulation is achieved using different strategies and shows the power of single cell imaging to glean insights into the mechanisms of lncRNA mediated transcription regulation. Despite the

similarities in the epigenetic machinery that participate in the lncRNA mediated repression of target loci, such as the recruitment of histone modifying complexes, whether suppression occurs gradually or in a switch-like manner seems to depend on the function of the regulated gene. Genes that regulate processes that determine cellular identity, such as sporulation-inducing genes, which need to induce stable phenotypes, depend on gradual repression over time, which may induce the acquisition of a stable repressive chromatin state. This would prevent rapid re-activation of the gene, which might be lethal in the wrong environmental conditions. Conversely, genes that respond to fluctuations in environmental nutrient levels, such as *PHO84*, would favor a more flexible mechanism of regulation that allows rapid activation and inactivation.

4.2.3 Antisense RNA regulates *PHO84* activation by transcriptional read-through into the promoter

An important question that emerged during our study is whether regulation of the *PHO84* transcription-on switch is mediated by AS transcriptional read-through into the *PHO84* promoter rather than accumulation of AS RNAs at the locus, as previously suggested. Our smFISH data showed no AS RNA accumulation at the locus, but we detected in increase in AS RNAs in the cytoplasm upon *RRP6* deletion. Furthermore, RNA half-life measurements showed that the deletion in *RRP6* did not alter the stability of the *PHO84* AS RNA. This allowed us to postulate that Rrp6p may regulate the frequency at which the AS RNA is transcribed rather than its stability and suggested an additional role for *RRP6* to its well characterized role in RNA processing and degradation. This hypothesis was supported by previous studies showing that Rrp6p interacts with the Nrd1-Nab3-Sen1 complex in the context of transcription termination and 3'processing of small ncRNAs such as sno/snRNAs [38].

Our results also demonstrated that the previous model suggesting AS RNA accumulation at the *PHO84* locus, caused by stabilization upon Rrp6p depletion, is an unlikely mechanism by which AS RNA expression regulates *PHO84*. Several lines of evidence support this argument. Firstly, we detect most AS RNAs in the cytoplasm in Δ*rrp6* cells. Although there is an increase from 12% of cells with nascent AS RNA in wild type to 20% in Δ*rrp6*,

indicating an increase in transcription in Δ*rrp6*, the relatively low fraction of cells with an AS RNA at the site of transcription suggests that AS RNA is not retained at the locus post-transcriptionally. Furthermore, most nascent AS RNA signals detected in Δ*rrp6* correspond to a single AS transcript, similar to what is observed in wild type cells, arguing against an accumulation of AS RNAs over the *PHO84* locus. In addition, to determine if the nuclear AS RNA signal persists once transcription is abolished, we performed a transcription inhibition experiment in which the temperature sensitive RNAPII mutant *rpb1-1* was fixed at different time points after heat shock. We observed rapid disappearance of the nuclear AS RNA signal after transcription inhibition, comparable to what we observed for nascent transcripts of the protein coding *MDN1* gene (**see Figure 2-4**). Therefore, similar to an mRNA, the *PHO84* AS RNA is rapidly released from the locus post-transcription and exported to the cytoplasm. Our observations in the temperature sensitive mRNA polyA polymerase mutant *pap1-1* and the Δ*xrn1* mutant further showed that the AS RNA requires polyadenylation to be competent for export, and that once it is exported to the cytoplasm, it is degraded by the 5'-3' mRNA exonuclease Xrn1p, which requires de-capping. Therefore, the AS RNA is capped and polyadenylated, recapitulating the bona fide features of an mRNA. Interestingly, a genome wide study investigating the turnover of AS lncRNAs in *S.cerevisiae* revealed that many are stabilized in a *xrn1* deletion background, suggesting that AS lncRNAs are often metabolized like mRNAs despite the lack of conserved open reading frames (ORFs) [43]. Since our data showed that *PHO84* AS RNA is processed like an mRNA and therefore not retained at the locus, it seemed unlikely that degradation of full length AS RNAs by the nuclear exosome at their site of transcription would constitute a major part of the metabolic pathway of *PHO84* AS. Therefore, we then determined whether Rrp6p plays any role in *PHO84* AS RNA stability.

To determine the relative contributions of Rrp6p and Xrn1p to AS RNA degradation, we measured the differences in AS RNA half-lives in these mutants compared to wild type by measuring residual RNA levels at different time points after inhibiting transcription with the metal-ion chelator 1,10-Phenanthroline. Interestingly, while the AS RNA half-life is increased to 27.3 min in Δ*xrn1,* the half-lives in wild type and Δ*rrp6* had similar values of 11.4 min and 12 min, respectively. Therefore, *PHO84* AS RNA is primarily degraded in the cytoplasm by

Xrn1p, whereas Rrp6 plays no role in AS RNA turnover. This led us to hypothesize that Rrp6p may regulate the frequency of AS RNA transcription. It was already established that Rrp6 interacts with the Nrd1-Nab3-Sen1 complex to terminate transcription of sn/snoRNAs [38]. Also, PAR-CLiP analysis showed that the Nrd1-Nab3-Sen1 complex binds to AS transcripts throughout the yeast genome, including *PHO84* AS RNA [152], and we identified Nrd1 and Nab3 binding motifs at the 3'end of *PHO84*. Altogether, these findings implied a general role for Nrd1-Nab3-Sen1 in transcription termination of AS RNAs.

Investigating the role of the Nrd1 complex in *PHO84* regulation, we revealed that depletion of Nrd1 under the glucose repressible *GAL1* promoter led to an increase in AS RNA expression, which was further enhanced in Δ*rrp6*. Mutagenesis of the Nrd1 and Nab3 motifs led to a more modest increase in AS RNA expression, possibly due to only partial removal of the Nrd1-Nab3-Sen1 recognition sites. Since Rrp6 interacts with Nrd1-Nab3-Sen1, we inferred that this interaction may stabilize the occupancy of the Nrd1-Nab3-Sen1 complex on the 3'end of *PHO84*, thereby increasing the efficiency of AS RNA transcription termination. Accordingly, ChIP analysis showed a significant decrease in Nrd1 binding on the 3'end of *PHO84* upon Rrp6 deletion. Altogether, our results demonstrated that Rrp6p plays a role in regulating transcription elongation of AS RNA, possibly by facilitating the recruitment of Nrd1 to the 3' end of *PHO84*.

Intriguingly, following our work, a genome wide study showed using RNA seq and RNA pol II ChIP that Rrp6 deletion increases transcriptional read-through of RNAs that are targeted by Nrd1-Nab3-Sen1, showing a general role for Rrp6 in facilitating Nrd1-Nab3-Sen1 mediated termination [153]. Furthermore, a follow up study by Castelnuovo and colleagues demonstrated that the mechanism identified in our study is widespread in the yeast genome. Many *PHO84*-like, TATA box regulated genes express anti-sense RNAs from their 3' UTRs and those AS RNAs undergo weak termination by Nrd1-Nab3-Sen1, due to a relatively low density of Nrd1 and Nab3 motifs at the 3' ends [154]. Regulating AS transcription by Nrd1-Nab3-Sen1 might therefore be a widespread mechanism to allow cells to fine tune the expression of target genes in response to extracellular signals. Allowing transcription to initiate continuously from the accessible 3'UTR of the gene by selecting for weak transcription termination would permit AS transcription to read through the sense promoter at

a frequency high enough to keep the gene repressed unless the inducing signal reaches a critical threshold. Alternatively, Nrd1 termination at specific 3'UTRs might be actively regulated, for example by regulating the recruitment of Nrd1. Future studies will show whether such gene specific regulation of Nrd1 and/or Rrp6 activity exists.

4.2.4 Role of Set1 in antisense RNA mediated gene repression

In addition to the role for Rrp6p in the early termination of AS transcription, our work also identified the histone methyltransferase Set1p as a positive regulator of AS RNA transcription. It has long been established that Set1p deposits the H3K4me3 mark on transcriptionally active genes [155]. H3K4me3 is localized to the 5'region of active genes where Set1p is recruited co-transcriptionally by RNA pol II whose CTD is phosphorylated on Ser5, but not Ser2 [156]. Interestingly, H3K4me3 persists for a long period of time after the inactivation of transcription and dissociation of Set1, returning to near background levels after 5 h, implying that this epigenetic mark retains a molecular memory of previous transcription events. It has also been shown that deposition of H3K4me3 by Set1p is required for efficient gene activation [157,158].

Intriguingly, our data showed an increase in H3K4me3 at the 3'end of *PHO84* in a Δ*rrp6* strain, suggesting that the increase in AS RNA expression upon Rrp6p deletion could be correlated with Set1p mediated deposition of H3K4me3 on the site of AS transcription initiation. Accordingly, we hypothesized that if Set1p promotes AS transcription by depositing the H3K4me3 mark on the 3'end of *PHO84*, it will produce an antagonistic effect on *PHO84* activation relative to Rrp6p, which facilitates *PHO84* activation by preventing AS transcription elongation through the promoter. Our smFISH data in a Δ*set1* strain shows a modest 10% decrease in the fraction of cells expressing *PHO84AS*, and a decrease in AS RNAs per cell, and a 20% increase in the fraction of cells expressing *PHO84*. Conversely, deleting Rrp6 in a Δ*set1* background reverts AS expression levels to near wild type levels, and decreases the fraction of cells expressing *PHO84*. Therefore, Set1p may promote AS transcription, whereas Rrp6p acts downstream by regulating elongation via its interaction with Nrd1. Thus, Rrp6p and Set1p play opposing roles in modulating the frequency of *PHO84* activation. Interestingly, the Δ*set1* strain shows a modest increase in Nrd1 occupancy on the

3'end of *PHO84*, suggesting that Set1p may antagonize Rrp6p dependent Nrd1 recruitment. Since Set1p is known to interact with the CTD of RNAPII phosphorylated at Ser5, similar to Nrd1, it may competitively inhibit Nrd1 recruitment to the RNAPII CTD, thereby favoring AS transcription elongation. However, this mechanism does not exclude a direct role for Set1p in promoting AS transcription, mediated by the deposition of H3K4me3 on the 3'end of *PHO84*.

Interestingly, following up on our study it was shown that Set1p is required for the transcription of AS RNAs, whereas expression of most protein-coding genes is unaffected upon Set1 deletion [154]. Genes with high density of Nrd1-Nab3-Sen1 binding sites that efficiently induce Nrd1 mediated termination express AS RNAs with low transcriptional read-through into gene promoters, which therefore do not induce strong repression. These AS transcripts that are subject to efficient Nrd1 termination are more strongly dependent on Set1p for their production than AS RNAs that are weakly terminated by Nrd1, as the latter show a modest decrease in transcription upon Set1p deletion, corroborating with our results for *PHO84* AS. Therefore, AS RNAs expressed from *PHO84*- like genes may be regulated primarily by Rrp6 during transcription elongation, whereas Set1p plays a minor role in modulating the frequency of AS initiation.

Intriguingly, results from a previous study on the role of Set1p in gene repression corroborate with the findings by Castelnuovo and colleagues [159]. Margaritis et al., show that Set1p is not required for the activation of most protein-coding genes, but a subset of genes is de-repressed upon Set1 deletion. These specific genes are repressed by Set1p through two distinct mechanisms of H3K4 methylation. Set1p deposits H3K4me3 at the 3'end of these genes to promote AS transcription, but also deposits H3K4me2 within the body of the genes, which is required for full repression. Interestingly, we showed by ChIP that there is an increase in H3K4me2 within the body of *PHO84* in Δ*rrp6*. Therefore, Set1p may act either upstream of Rrp6p to promote AS transcription through H3K4me3 or downstream by depositing H3K4me2 within the gene body in a transcription elongation dependent manner. Since there are many genes in *S.Cerevisiae* similar to *PHO84* that are induced by metabolic stress, the mechanisms by which Rrp6p and Set1p regulate gene activation through the modulation of AS RNA transcription, as revealed by our study, may be broadly applicable across the yeast genome.

Recent studies also showed that AS transcription is ubiquitous across the yeast genome, and that many AS RNAs are regulated by Rrp6p and require the histone modifiers Hda1 and Set1 to induce gene repression, similar to *PHO84* [154]. Tiling array analysis of the AS transcriptome in *S.Cerevisiae* revealed that AS RNAs can be clustered into different classes based on their ability to repress sense gene expression upon accumulation in a Δ*rrp6* background, and whether this repression is HDAC (Histone de-acetylase) or histone methyltransferase dependent [154]. Genes that are suppressed by AS RNA accumulation upon depletion of Rrp6p fall into two categories, those that remain suppressed upon deletion of HDACs and the H3K4 methyltransferase Set1p, and those that are de-repressed. Therefore, the genes in the latter group require extensive histone modification coupled to AS expression to induce efficient repression. These genes are similar to *PHO84*, as they are regulated by TATA box promoters, which have a closed nucleosomal architecture in repressive conditions, and therefore require extensive chromatin remodeling to induce expression. Their analysis showed extensive AS transcriptional read-through into TATA box regulated genes. Since this class of genes requires HDACs and the histone methyltransferase activity of Set1 for efficient repression, histone modification may consolidate a repressive chromatin structure that is initially induced by AS read-through into the TATA box promoter.

4.2.5 The role of antisense transcription in nucleosome remodeling and regulation of transcriptional bursting

The nucleosome architecture of the *PHO84* upstream activating sequence (UAS) provides hints on how continuous AS RNA transcription may inhibit the *PHO84* promoter from bursting sporadically under repressive conditions, and yet allow the flexibility to respond rapidly during induction. The *PHO84* UAS contains 4 binding sites for the transcription factor Pho4p, the two internal sites being high affinity binding sites, and the 2 peripheral binding sites having low affinity for Pho4p [160]. The two peripheral Pho4 binding sites are occluded by nucleosomes in repressive conditions, but evicted by Pho4p and the SWI/SNF chromatin remodeling complex during inorganic phosphate starvation, allowing full activation of *PHO84*. It has been shown that the nucleosomes occluding the peripheral Pho4p binding sites

differ markedly in their requirements for active chromatin remodeling, due to intrinsic differences in the strength of nucleosome-DNA interactions [160]. The downstream nucleosome closer to the TATA box element is less dependent on active remodeling by the SWI/SNF complex, as it is less stably positioned on the Pho4p activation site. Intriguingly, full activation of the *PHO84* promoter requires eviction of the downstream nucleosome rather than the upstream one. Therefore, the kinetics of *PHO84* transcription is controlled by the positioning of this nucleosome, which has an inherently unstable interaction with DNA. Therefore, *PHO84* may be activated sporadically, which endows individual cells the plasticity to respond rapidly to changes in phosphate levels in the environment. Nevertheless, the sensitivity of the promoter may have to be regulated so that most cells in the population do not respond inappropriately in the absence of the inducing signal. Thus, low frequency but continuous AS transcription of may reduce the frequency of sporadic *PHO84* activation by stabilizing the nucleosome architecture on the UAS and promoter.

The repressive chromatin structure on the *PHO84* UAS established as a result of AS transcription may be short-lived due to thermodynamically unstable nucleosome-DNA interactions. Therefore, the chromatin state and position might have to be re-established frequently by de-novo AS transcription (see below). Interestingly, *PHO84* mRNA transcription is dominant over AS transcription. When cells are transferred to a phosphate depleted medium, the entire population expresses only *PHO84* sense within 1-2 hours of induction (**see Supplementary figure 6 in Chapter 2**). Therefore, under inducing conditions, single cells rapidly switch to sense expression, as the promoter becomes remodeled and accessible to Pho4p binding and efficient re-initiation by RNAPII. Therefore, while the *PHO84* promoter structure is flexible enough to respond rapidly to induction under phosphate-rich repressive conditions, continuous AS transcriptional read-through appears to lower the probability of sporadic activation of *PHO84*. After the duration of the *PHO84* burst, AS transcription may be re-initiated and RNA pol II elongation through the gene promoter may facilitate nucleosome re-assembly to reset the transcription-off switch. In Δ*rrp6* cells, Nrd1 mediated early termination of AS RNA is compromised, which allows more frequent RNA pol II elongation events through the *PHO84* UAS. If RNA pol II elongation is coupled to nucleosome positioning and HDAC recruitment, then the increase in AS transcription

frequency in Δ*rrp6* cells may stabilize the repressive chromatin structure on *PHO84,* thereby decreasing the frequency of activation of the transcription on-switch.

Genes repressed by AS RNAs in an HDAC dependent manner, similar to *PHO84*, are enriched for TATA box promoters that are strongly occluded by nucleosomes. During activation, these genes are transcribed in bursts through the co-operative action of transcription factors and chromatin remodeling complexes which evict nucleosomes to activate the transcription on-switch. A potential role of low frequency AS transcription across the promoters of these genes may consist in re-positioning nucleosomes to prevent the binding of transcription factors and therefore increase the threshold of the transcription on-switch [154]. Nucleosome re-assembly may be achieved by histone chaperones, proteins that bind histones and regulate nucleosome assembly, which interact with the RNA pol II CTD during elongation. Interestingly, it was shown that the histone chaperone Spt6, involved in the maintenance of chromatin structure, is recruited to elongating RNA pol II in an HDAC dependent manner [161]. In addition to promoting the recruitment of histone chaperones to facilitate nucleosome re-positioning during AS transcription across the promoter, HDACs may also act downstream to consolidate the repressive chromatin structure [154]. Therefore, one of the roles of AS transcription may be to fine tune the transcription on-switch of strongly bursting TATA box genes by modulating nucleosome assembly on the promoter.

It is not known what factors are required to recruit RNA polymerase to the 3' UTR of *PHO84* to initiate AS RNA transcription in phosphate rich conditions. Transcription of the *GAL10* AS RNA had previously been shown to be regulated by the transcription factor Reb1 [80]. The *PHO84* 3'UTR contains a binding site for Reb1, however deleting the Reb1 motif does not affect *PHO84AS* transcription, and multiple cryptic elements within the accessible 3'UTR were shown to promote AS transcription [81]. Therefore, whether RNAPII can initiate transcription from the 3'end of *PHO84* without a specific activator, possibly due to a exposed cryptic TATA box, or whether a specific TF is required for *PHO84* AS transcription remains to be determined.

4.2.6 How frequently do antisense RNAs need to be transcribed to modulate gene activation in response to extracellular signals?

We showed that *PHO84* AS RNA is expressed constitutively at low levels. Combining AS RNA half-life measurements with quantification of the average steady state number of AS RNAs per cell by smFISH showed that wild type cells transcribe 1.31 *PHO84* AS RNAs/hour. Δ*rrp6* cells show a two-fold increase in transcription frequency (**see Supplementary table 1 in Chapter 2**). This suggests that a two-fold increase in AS transcription frequency is sufficient to raise the threshold of *PHO84* activation by about two-fold as measured by the two-fold decrease in the percentage of cells that express *PHO84* mRNA in Δ*rrp6* (see **Figure 4-1**). Interestingly, the *GAL10AS* RNA was estimated to be transcribed at a frequency of 1 RNA/50 min, similar to *PHO84AS* in wild type [80]. Therefore, *GAL10AS* and *PHO84AS,* which both regulate their target genes via similar HDAC dependent epigenetic mechanisms, also have similar rates of synthesis. This low frequency continuous transcription may be sufficient to induce a repressive chromatin structure that is flexible, but stable enough so that it does not require a high rate of AS RNA synthesis to be maintained.

Recent studies investigating the regulation of the *GAL1-GAL10* promoter showed that transcription of *GAL10* AS RNA prevents leaky activation in repressive conditions [162]. They used a dual PP7/MS2 labeling approach to simultaneously visualize transcription of *GAL10AS* and GAL10 mRNA, respectively. A nuclease deficient version of the Cas9 protein, dCas9, was recruited specifically to the strand expressing *GAL10AS* to induce a CRISPR mediated block in AS RNA transcription. The CRISPR block of *GAL10AS* increased the sensitivity of the *GAL10* promoter, whereby it was activated more frequently in glucose-rich repressive conditions. The increase in transcriptional leakage from the *GAL1-GAL10* shared promoter resulted in a fitness defect, as cells were utilizing galactose, a less efficient source of carbon than glucose, even when glucose was abundant. Interestingly, the CRISPR block of *GAL10AS* in inducing conditions showed no effect on bursting frequency of GAL10. In galactose-rich inducing conditions, the *GAL1-GAL10* shared promoter may be continually accessible and dominant over *GAL10AS* transcription. Therefore *GAL10AS* transcription is merely spurious instead of functional in inducing conditions. These findings corroborate well with the results

from our *PHO84* study. During induction upon phosphate starvation, we detect a synchronous response in the cell population, since all cells switch to *PHO84* mRNA expression. In this context, *PHO84* AS RNA may be inconsequential, as the nuclear concentration of the transcription factor Pho4p may be sufficiently high to overcome the nucleosome barrier on the *PHO84* UAS. In addition, RNA pol II elongation on the sense strand may recruit histone chaperones to re-assemble nucleosomes on the 3'UTR, thereby repressing AS RNA transcription.

Interestingly, although the kinase Pho85p phosphorylates Pho4p to prevent nuclear import in phosphate rich media, partially phosphorylated Pho4p can localize in the nucleus under intermediate phosphate conditions and activate a subset of target genes [18]. Under these conditions, Pho4p binds differentially to phosphate-responsive promoters. Intriguingly, partially phosphorylated Pho4p binds efficiently to the *PHO84* promoter to activate its transcription, while it does not efficiently bind to or activate *PHO5*. This difference in promoter sensitivity can be attributed to the fact that the *PHO5* promoter exposes low affinity Pho4p binding sites in the repressed state, whereas the *PHO84* promoter exposes high affinity binding sites [163]. Therefore, although the nuclear concentration of Pho4p under intermediate phosphate conditions is lower than during phosphate starvation, Pho4p binds more strongly to the *PHO84* promoter than on *PHO5*. Since *PHO84* encodes a high affinity phosphate transporter that is localized in the cytoplasmic membrane, it plays an early role during the phosphate starvation response by importing phosphate into the cell when environmental phosphate levels decrease. Therefore, there may have been selective pressure on *PHO84* to maintain high affinity transcription factor binding sites, resulting in a highly sensitive promoter that is rapidly inducible. Nonetheless, since the nuclear concentration of Pho4p in non-inducing conditions is lower than during phosphate starvation, Pho4p may be less efficient in circumventing the activation threshold that is established by continuous AS transcription across the *PHO84* UAS, resulting in complete repression of the *PHO84* promoter in the majority of cells.

The live cell study of *GAL10* regulation by *GAL10AS* using the PP7/MS2 dual labeling approach showed that *GAL10AS* is present at the transcription site for only 90s. Given a RNAPII elongation rate of 2kb/min, it would take 50-100s to transcribe *GAL10AS* [164].

Therefore, *GAL10AS* is only detectable at the transcription site during synthesis, implying that GAL10AS acts by transcription, not through the post-transcriptional accumulation of RNAs at the locus. These findings corroborate with the results from our *PHO84* study, whereby inhibition of RNA pol II revealed that the detection of nuclear *PHO84AS* requires continuous transcription. Therefore, continuous low-frequency transcription of AS lncRNA may be a common regulatory mechanism in yeast to fine tune stress-inducible genes to respond appropriately to environmental signals. Future live cell imaging studies on other *PHO84*-like genes, combined with biochemical experiments to investigate changes in nucleosome architecture, will reveal further mechanistic insights on AS lncRNA mediated regulation of transcription.

4.3 Article 2: Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription

4.3.1 Objectives and summary of results

While completing the study on AS RNA mediated gene regulation in yeast, I became intrigued by the emerging field of ncRNAs in higher eukaryotes, where the noncoding genome vastly eclipses the exome in size. Similar to yeast, many different classes of ncRNAs had been identified through next generation sequencing approaches, but the role of most of these ncRNAs remains to be elucidated. In particular, I became interested in a specific class of recently discovered ncRNAs expressed from enhancer elements named enhancer RNAs (eRNAs). eRNAs are transcribed bi-directionally from enhancers, the distal transcription regulatory elements required for metazoan development and most cell-type specific transcription in higher eukaryotes, but that are absent in single celled eukaryotes. Studies in different cell types suggested eRNAs to be implicated in gene activation, but no concerted model for eRNA function seemed to emerge. Rather, different mechanistic models were suggested, including roles in chromatin accessibility, enhancer-promoter looping, and the RNA pol II transcription cycle, and these models were met with skepticism by some scientists in the field who held that eRNAs might simply be a byproduct of TF binding to enhancers and serve no active role in gene regulation [88,135,165]. My goal was to apply and expand the single cell and single molecule approaches I had established during my studies in yeast, to test

different models of how eRNAs participate in transcription regulation, with the goal to reveal general principles of eRNA function.

At the time I became interested in eRNAs in early 2013, studies investigating eRNAs mostly applied ensemble measurements from cell populations and therefore were unable to monitor transcriptional activity in a single cell context, nor on individual alleles, leaving fundamental questions of eRNA function unanswered, that we were eager to address. Notably, although GRO-seq analysis shows a genome wide correlation in the upregulation of eRNA and mRNA transcription, this does not necessarily imply that eRNA transcription is tightly coordinated with every cycle of transcription initiation from the target promoter. Furthermore, chromatin conformation capture assays that implicate a role for eRNAs in chromatin looping describe a statistical mean of different possible conformations, obscuring the dynamic behavior of individual chromatin fibers. This suggests that enhancers which transcribe eRNAs may not necessarily trigger looping in a deterministic fashion. Therefore, our primary objectives were to determine if eRNA and mRNA transcription is tightly coordinated in response to external stimuli in the context of single cells and single alleles, and if eRNAs induce and possibly stabilize enhancer-promoter loops. To achieve this, we applied the single molecule imaging tools established during the PHO84 project, expanded them to use sub-diffraction resolution distance measurements to allow visualizing and quantify transcriptional activity at a single locus, as well as to measure the proximity of transcribing enhancers and their target promoters at high spatial resolution.

In light of the objectives outlined above, we choose the estrogen induced transcription response as a model system. eRNAs were previously shown to be induced by estrogen (E2) at ERα bound enhancers in the MCF7 breast cancer cell line, which presented itself as an ideal experimental system due to its rapidly inducible, strong transcriptional response in the presence of E2 [88,133] [87]. Furthermore, results from a study by Li and colleagues suggested a putative role for some eRNAs in recruiting cohesin to maintain enhancer-promoter looping interactions [88]. Moreover, *in vitro* assays tethering eRNAs to a luciferase reporter plasmid resulted in enhanced luciferase expression, suggesting that eRNA molecules are able to activate transcription in a manner that is independent of their synthesis. Among the eRNAs whose expression most strongly correlated with target mRNA expression were eRNAs

transcribed from the enhancers of the *FOXC1* and *P2RY2* genes. Furthermore, these eRNAs were previously investigated for their roles in transcriptional activation and/or enhancer-promoter looping [88,133]. Therefore, *FOXC1* and *P2RY2* served as our model genes for our single molecule study.

Using single cell and single molecule assays, we were able to reveal important aspects of eRNA expression and function. We found that during a time course of E2 induction, enhancers and promoters activate transcription with similar kinetics in single cells, as observed by the correlated increase in nascent eRNA signals and numbers of active transcription sites for the target loci *FOXC1* and *P2RY2*. However, while the target gene shows a pattern of expression upon induction that is indicative of transcriptional bursting, as shown by the large variability in the number of nascent transcripts per active site, we do not detect an accumulation of eRNAs at enhancers, implying that eRNA transcription is initiated at a lower frequency. Moreover, most active transcription sites for both *FOXC1* and *P2RY2* (approximately 75%) do not co-express eRNAs, implying that while eRNAs may be implicated in promoter activation, they are not required for every cycle of RNA pol II transcription initiation once the promoter is active. Interestingly, we also showed through sub-diffraction-limit measurements of distances between nascent eRNAs and mRNA transcription sites, that co-expressed eRNAs are rarely found within a distance that is compatible with enhancer-promoter looping interactions, as previously predicted from polymer modeling. Therefore, while eRNAs may initiate the looping process, once an enhancer-promoter loop is established, eRNA synthesis may be mostly repressed and not required to sustain transcription from the target promoter. Furthermore, we show that depleting the estrogen receptor ERα and the histone methyltransferase MLL1, which is known to poise enhancers for subsequent activation, abrogates E2 dependent induction of eRNA transcription, with negligible effects on basal eRNA transcription. Therefore, eRNA transcription may play a very early role in the gene activation pathway by pre-programming the enhancer for activity.

4.3.2 The kinetics of transcriptional activation from enhancers and promoters in single cells during E2 induction

Our investigation of the role of E2 induced eRNAs in transcription regulation is, to our knowledge, the first to investigate eRNA-mRNA co-expression patterns at the single cell level and at a single molecule resolution. It was previously shown that E2 induction produces a transient transcriptional response that peaks at 40 min [87]. Our E2 time course shows basal transcriptional activity on the candidate loci *FOXC1* and *P2RY2* and their respective enhancers prior to E2 treatment. Prior to E2 treatment, mRNA transcription sites are rare, of low intensity and show little variability in transcriptional output. eRNAs are far less abundant, most cells showing only a single nuclear eRNA FISH signal. Since eRNAs are thought to act in *cis* and have short half-lives, most eRNAs detected are therefore likely nascent transcripts [124,135]. eRNA transcription is considered to be the most robust predictor of enhancer activity [125]. Therefore, detecting a single eRNA spot in most cells, may suggest that a small fraction of alleles harboring the enhancer locus is in an active state prior to E2 stimulation. Upon 40 min E2 induction, there is a 3-fold increase in the number of cells with more than 1 eRNA, consistent with transcriptional activation of multiple enhancer loci. The number of eRNAs detected per nucleus (3-6) is consistent with the haplotype of MCF7 cells, which are triploid, and therefore have 3-6 alleles, depending on the cell cycle stage. Concurrently, we detect an increase in the number of active mRNA transcription sites, indicative of promoter activation from multiple alleles. The mRNA transcription foci increase in size, corresponding to the accumulation of multiple nascent transcripts, and show large variability in transcriptional output, consistent with transcriptional bursting [26,166]. Interestingly, even though there is an increase in the number of eRNAs per cell during peak induction, the eRNA spot intensities remain the same as seen in untreated MCF7 cells, and show little variability across the population. This suggests that relative to the frequency of mRNA transcription initiation from the target promoter, eRNA transcription does not initiate in a bursting mode, and that there is no local accumulation of multiple eRNA transcripts at enhancers.

It is well established that enhancers and promoters share common architectural features, such as core transcription factor binding motifs and well positioned nucleosomes

flanking bi-directional initiation sites. Therefore, transcription activation from an enhancer may be mechanistically analogous to promoter activation. Thus, the frequency of ERα mediated enhancer and promoter activation may be very similar, leading to the correlated increase in the number of transcriptionally active enhancers and promoters. However, once transcription is activated, the frequency of individual cycles of RNA pol II initiation and elongation may differ between an enhancer and promoter. Protein-coding sequences are enriched in U1 snRNP 5' splice sites that promote efficient RNA pol II elongation. Therefore, if transcription is re-initiated frequently during a burst, we expect a broad range of fluorescent signal intensities at the transcription foci corresponding to nascent mRNAs of different lengths. Conversely, eRNA sequences are not enriched in U1 snRNP 5'splice sites, but often contain early polyA signals that prevent productive elongation. Furthermore, if inefficient elongation of eRNAs is coupled with a low frequency of re-initiation, it would be unlikely to detect multiple elongating polymerases on the enhancer from temporally clustered cycles of transcription. This might explain why fluorescent signal intensities of eRNAs do not vary between active alleles, even during induction.

Many of the eRNAs detected during peak induction may correspond to processed RNAs, that are retained for a short time on chromatin after transcription termination. eRNAs were shown to be terminated by the RNA pol II associated Integrator complex to produce chromatin associated processed transcripts [140].Processing of eRNAs by the Integrator complex was correlated with increased transcription on the target locus. Furthermore, several studies show that eRNAs have a defined length, indicative of uniform 3'ends [134,147,167]. Taken together with the observation that we detect eRNA signals with uniform intensities, this may imply that eRNA function in stimulating mRNA transcription does not require frequent transcription of eRNAs or the accumulation of multiple nascent eRNA transcripts, but rather retention of uniformly sized single eRNAs on chromatin.

4.3.3 The roles of eRNA transcription in regulating chromatin accessibility

Our observation of basal eRNA transcription from a fraction of alleles in the absence of E2 led us to investigate the factors that maintain enhancer transcription prior to signal-dependent activation. Initially, we examined whether basal transcription of an enhancer prior

to E2 signaling occurs independently of ERα, or if it is mediated by ERα that retains the capacity to bind to its target enhancer in the absence of E2. To distinguish between these two possibilities, we treated cells with the anti-estrogen Tamoxifen, which prevents E2 binding and recruits transcriptional co-repressors without affecting ERα binding to the ERE. Alternatively, we treated cells with ICI 182,170, an anti-estrogen that prevents dimerization of the ERα homodimer, reducing its affinity for the ERE, and inducing ERα degradation upon prolonged treatment [168]. As expected, cells pre-treated with Tamoxifen or ICI 182,170 before E2 induction do not show a change in expression upon stimulation, as measured from the number of nascent transcripts per active mRNA transcription site, as inhibition of ERα diminishes RNAPII recruitment on promoters [85]. However, the frequency of eRNA expression in cells pre-treated with ICI 182,170 does not decrease in either the percentage of cells expressing eRNAs or number of eRNAs detected per cell prior to E2 induction, even though induced eRNA transcription in response to E2 is significantly reduced. This was observed after a prolonged 3 hour treatment with ICI 182,170 prior to E2 induction, which leads to complete ERα degradation. This observation led us to postulate that other factors recruited to enhancers prior to ERα are required to activate basal eRNA transcription.

It is well established that ERα recruitment requires the pioneer chromatin remodeling factor FOXA1 [86]. It has also been shown that FOXA1 binds specifically to H3K4me1 modified enhancer chromatin throughout the genome [95]. Studies on the model E2 induced gene *TFF1* in MCF7 cells have previously shown that the histone methyltransferase MLL1 deposits the H3K4me1 mark on the *TFF1* enhancer, which is subsequently required to recruit FOXA1 and ERα to activate gene transcription [96,97]. Since H3K4me1 is a signature of an active enhancer, we speculated that MLL1 may deposit the active H3K4me1 on our candidate E2 responsive enhancers to activate basal eRNA transcription. However, our MLL1 knockdown experiments show similar results as the ERα depletion experiments. Basal eRNA transcription does not decrease upon depletion of MLL1, even though we detect a two-fold decrease in H3K4me1 enrichment and ten-fold decrease in ERα recruitment on the *P2RY2* enhancer. MLL1 depletion inhibits eRNA induction, as well as promoter activation, as we detect fewer active P2RY2 transcription sites in cells depleted of MLL1. Therefore, H3K4me1

deposition by MLL1 may condition the enhancer to respond to estrogen and activate transcription, but is not required for basal eRNA transcription.

H3K4me1 is often present on enhancers prior to nucleosome depletion, H3K27 acetylation, and enhancer activation [169]. Developmental enhancers that are pre-marked by H3K4me1 and H3K27me3 are considered to be poised, but inactive [170]. Conversely, enhancers that are enriched in H3K4me1 and H3K27ac are active, and show preferential association with RNA pol II relative to poised enhancers. Therefore, while MLL1 is not required for basal transcriptional activity on the enhancer, it may facilitate induced eRNA transcription by poising the enhancer through H3K4 monomethylation. FOXA1 may then recognize the H3K4me1 mark and recruit ERα, which activates eRNA transcription. Subsequently, induced eRNA transcription can facilitate the transition of the enhancer from a poised to an active state (see **Figure 4-2**).

Interestingly, recent studies have assigned eRNAs a role in modulating chromatin accessibility by affecting histone acetylation levels on both enhancers and promoters [147,171]. A study by Pnueli and colleagues showed that eRNA knockdown inhibits the interaction of the *Cga* gene, which encodes the α-subunit of the hormone gonadotropin, with its regulatory enhancer and decreases H3K27ac levels on both the enhancer and promoter, leading to decreased mRNA expression. Interestingly, the H3K7ac mark on the enhancer is replaced by the repressive H3K27me3 mark upon eRNA knockdown. Therefore, the eRNA molecule itself can induce H3K27ac on the enhancer to maintain enhancer accessibility and function.

Furthermore, a very recent study by Bose and colleagues that was published after our paper revealed that the histone acetyltransferase CBP interacts with eRNAs in a non-sequence specific manner. Using PAR-CLiP and ChIP-seq analysis they showed that CBP binds to eRNAs transcribed from enhancers to which CBP is recruited. Furthermore, the interaction of eRNAs with a specific regulatory region within the histone acetyltransferase (HAT) domain of CBP was shown to enhance histone acetylation *in vitro*, and knockdown of eRNAs leads to reduced H3K27ac on the respective enhancer and target promoter, resulting in decreased mRNA transcription. Interestingly, CBP recruitment is not affected by eRNA knockdown. Therefore, while the eRNA molecule itself may not be implicated in CBP recruitment, the

process of transcription on the enhancer may maintain an accessible region to help recruit CBP, which then binds to the nascent eRNA transcript to stimulate histone acetylation at the enhancer. As CBP also interacts with RNA pol II, deposition of H3K27ac can also be mediated by the process of transcription on the enhancer, whereby CBP would translocate with the elongating polymerase to extend H3K27ac across the enhancer.

Intriguingly, it was shown that in addition to inhibiting enhancer-promoter looping, knockdown of eRNAs diminishes the recruitment of RNA pol II and the histone acetyltransferase p300 on the enhancer and promoter of the antisense lncRNA that represses the *DHRS4* gene cluster [172]. Therefore, when the enhancer and promoter are brought into proximity during eRNA transcription, the eRNA may serve as a molecular bridge that delivers histone acetyltransferases, such as p300 or CBP, from the enhancer to the promoter, thereby specifically increasing chromatin accessibility at the target locus, which increases the efficiency of RNA pol II recruitment. However, basal eRNA transcription may initiate at a frequency that is too low to stimulate histone acetylation and *de novo* recruitment of RNA polymerases on the promoter. Our IF and smFISH data show that cells in which MLL1 has been depleted harbor weaker transcription foci for *P2RY2*, suggesting that MLL1 can indirectly affect the RNA pol II density on the target promoter by modulating the rate of eRNA synthesis. The role of MLL1 may consist in priming the enhancer to respond to signal dependent transcription factors to increase the frequency of eRNA transcription initiation. When eRNAs are synthesized at a higher frequency, they may enhance the histone acetylase activity of CBP, as the increased local concentration of eRNAs on chromatin may increase their affinity towards CBP. In addition, if the enhancer and promoter are rendered more accessible by histone acetylation, this may also facilitate recruitment of cohesin and mediator to stabilize enhancer-promoter looping interactions.

The observation that basal eRNA transcription is ongoing upon depletion of MLL1 suggests that there is some degree of chromatin accessibility on the enhancer prior to the downstream recruitment of chromatin remodelers. It was shown that MLL1 binds to an accessible CpG element within the *TFF1* enhancer [97]. The binding of MLL1 requires de-methylation of the CpG element. MLL1 was also shown to be recruited to CpG rich regions by the CpG binding protein [98]. Interestingly, CpG elements that are minimally accessible are also

present in Glucocorticoid receptor target enhancers, which are active in a tissue specific manner [99]. It was shown that the accessibility of CpG sequences on these enhancers relies on prior CpG demethylation, which may commission the enhancers for activation. It will be interesting for future investigation to determine if basal eRNA transcription maintains open chromatin on enhancers by regulating DNA methylation on CpG elements. This may be a plausible mechanism by which eRNAs increase enhancer sensitivity to signaling pathways. It was shown that the antisense lncRNA TARID induces DNA demethylation on a CpG element in the promoter of *TCF21*, thereby activating the gene [173] . If CpG elements are a common feature in enhancers, as they are in promoters, then eRNAs can potentially induce DNA de-methylation on CpG sequences to make them more accessible to histone methyltransferases like MLL1, which can then initiate the cascade of events that are required for gene induction.

Although our data show that H3K4me1 deposition on enhancers by MLL1 can facilitate eRNA induction in response to E2, it has also been demonstrated that eRNA transcription may be implicated in recruiting MLL1 and depositing H3K4me1/2 on enhancers [91]. The study by Kaikkonen and colleagues showed that the spread of H3K4me1/2 from the core enhancer overlaps with the spread of the eRNA GRO-seq signal, and that deposition of H3K4me1/2 depends on MLL1-4. They also show that inhibition of transcription, but not knockdown of the eRNAs, reduces H3K4me2 levels on enhancers. Therefore, deposition of active histone methylation marks by MLL1 is most likely mediated by enhancer transcription, rather than the eRNA molecule itself. This corroborates with another study which showed that MLL1 can interact with RNAPII on a subset of transcribed regions within the genome [126]. Therefore, the process of transcription on the enhancer could recruit MLL1, which then translocates with the elongating RNA pol II to deposit the H3K4me1/2 mark across the enhancer. This, in turn, increases chromatin accessibility at enhancers by recruiting FOXA1 and ERα, which increases the frequency of eRNA transcription initiation.

Interestingly, deposition of H3K4me1 by MLL1 was shown to stabilize the binding of the histone acetyltransferase TIP60, showing a tight functional link between histone methylation and histone acetylation [174]. A study by Jeong and colleagues also showed that MLL1 recruitment is enhanced by the SWI/SNF chromatin remodeling complex, which itself

is recruited by ERα. It was also shown in a recent study that the ERα co-factor FOXA1 recruits the histone methyltransferase MLL3 to direct H3K4 mono-methylation [175]. Furthermore, the study by Kaikkonen and colleagues proposes a model whereby eRNA transcription and MLL1 recruitment occur downstream of transcription factor binding and histone acetylation. Altogether, these findings demonstrate that there is no consistent linear sequence of events that defines the process of enhancer activation. Therefore, enhancer activation may rather be the cumulative effect of a network of interactions between transcription factors, histone modifiers, and eRNAs, which can associate with each other at different time points to stimulate a feed forward cycle.

4.3.4 Transcription initiation from the target promoter does not require eRNA co-expression

Observing a concurrent increase in the number of eRNAs and the number of mRNA transcription sites in single cells upon E2 induction prompted us to investigate how frequently a transcriptionally active allele would co-express eRNAs from its respective enhancer, i.e. to determine the degree of spatio-temporal co-ordination of eRNA and mRNA transcription across all active alleles. To calculate the frequency of eRNA-mRNA co-expression it was essential to assign a nascent eRNA to its target locus by considering the physical distance between enhancer and promoter. Recent studies have revealed that enhancers and their target promoters are compartmentalized within higher-order chromatin structures referred to as topologically associated domains (TADs) [105]. Furthermore, simulation of intra-TAD chromatin fiber conformations using the principles of polymer physics predicts that chromatin fibers will not form static loops, but may adopt many different conformations, resulting in a broad range of distances between regulatory elements and promoters, which was validated by DNA FISH [111]. Accordingly, we devised a program that assigns a nascent eRNA to its target promoter by searching for the most proximal eRNA within a specific radius from the center of an mRNA transcription site. Such a search should accommodate all possible chromatin fiber conformations, including direct enhancer-promoter looping interactions (10-100 nm), as well as non-closed loop configurations. We assigned a radius of 400 nm from a transcription site as our threshold for classifying eRNAs as co-expressed, which is a close approximation to the

upper limit of enhancer-promoter distances as determined from previous DNA FISH measurements [111,112].

Using these tools, we determined the frequency of eRNA-mRNA co-expression during the time course of E2 treatment. Surprisingly, our data revealed that during peak induction at 40 min E2 treatment, 75% of *FOXC1* transcription sites do not co-express an eRNA. The 25% of transcription sites that co-express eRNA mostly show co-expression of either sense or AS eRNA, but only rarely both, indicating that eRNA transcription does not initiate in both orientations simultaneously. It has been proposed that nucleosome depleted core enhancer elements may be too small to accommodate two pre-initiation complexes simultaneously [45]. Therefore, at any given time, a single PIC may initiate transcription randomly in one direction or the other. Our analysis of eRNA-mRNA co-expression for *P2RY2* showed similar patterns as *FOXC1*, whereby at least 75% of P2RY2 transcription sites during peak induction show no co-expression of eRNAs. Although we showed a statistically significant increase in co-expression during 40 min E2 treatment relative no E2 treatment, it is evident that the overwhelming majority of active alleles do not require continued eRNA synthesis and/or eRNA accumulation on enhancers to sustain transcription.

We further showed that eRNA co-expression has a negligible effect on RNAPII density at transcriptionally active alleles. For *P2RY2,* there was no statistical difference in the number of nascent transcripts per active transcription site among the alleles that co-express eRNA compared to those that do not. Although we detected a small, statistically significant, difference in the number of nascent transcripts per *FOXC1* transcription site at active alleles co-expressing eRNA during peak induction, many alleles for both *FOXC1* and *P2RY2* do not co-express eRNA but show a high RNAPII density. Therefore, our findings do not corroborate strongly with models that implicate eRNAs in the RNA pol II transcription cycle where every instance of RNA pol II pause release on the target locus requires an eRNA. A study by Zhao and colleagues on a specific class of androgen receptor (AR) induced eRNAs showed that these eRNAs recruit and activate the positive transcription elongation factor pTEFb to phosphorylate Ser2 on the RNA pol II CTD, thereby promoting productive elongation on the target gene. These eRNAs contain an HIV-1 TAR RNA like motif that forms a hairpin loop, which is requires for pTEFb activation. In this scenario, in which eRNAs possess structural

features that may be required to bind to pTEFb and stimulate its catalytic activity during every cycle of RNA pol II pause release, we would expect to observe an accumulation of eRNAs at mRNA transcription foci in much higher numbers. As stated previously, we never observe eRNAs accumulating at enhancers, irrespective of the transcriptional output from target promoters, which suggests that every cycle of RNA pol II release on the target locus is unlikely to be coupled with eRNA synthesis. Moreover, we do not detect the hairpin forming sequence that was identified for some AR induced eRNAs in the ERα induced eRNAs analyzed in our study, suggesting mechanistic differences between distinct classes of eRNAs. Furthermore, whether structural features such as hairpin loops are unique to certain kinds of eRNAs or universal is not yet known, and needs further investigation.

Another study by Schaukowitch and colleagues showed that eRNAs can bind to the negative elongation factor NELF to act as a decoy to prevent NELF association with RNA pol II [135]. Their model suggests that eRNAs accumulate to release NELF, thereby promoting RNA pol II elongation on the target gene. They also show that eRNAs have short have-lives and are downregulated prior to mRNA induction. Their results corroborate with GRO-seq analyses in many cell types which show that eRNA transcription represents an immediate early response that precedes successive waves of transcriptional activation from genes [137]. Our results also suggest that eRNA transcription may represent an immediate early response that subsides once transcription is induced from the target locus, as not only do we detect a low fraction of mRNA transcription sites that co-express eRNAs, but a similarly low fraction of eRNAs that are associated with transcriptionally active alleles. The non-mRNA co-expressing eRNAs may represent early activated enhancers that have yet to induce transcription from their target loci. However, as we do not detect accumulation of eRNAs, the model which suggests that eRNAs accumulate locally to release NELF is unlikely to explain our results, unless the accumulation is very transient and is followed by rapid destablization of eRNAs. The E2 induced eRNAs investigated in our study may act transiently during their synthesis to activate the target promoter, which can then initiate multiple cycles of transcription in the absence of continued eRNA synthesis.

4.3.5 The role of eRNAs in regulating transcriptional bursting frequency

Our data showed an increase in variability in the transcriptional response of individual alleles upon E2 induction, indicative of gene bursting as described previously [26,166]. Recently, there has been considerable interest in characterizing the specific tunable properties of bursting that are regulated by enhancers, including two single cell studies that investigated the specific parameters of bursting fine-tuned by enhancer-promoter interactions. [113,114]. A study by Fukaya and colleagues linked PP7 and MS2 live cell RNA imaging reporters to developmental enhancers to track the trajectory of transcriptional activation in single cells in developing *Drosophila* embryos [113]. They showed that different developmental enhancers change the bursting frequency of the reporter, but do not change the burst amplitude or burst duration during an individual burst. Similarly, a study by Bartman and colleagues showed using smFISH that forced enhancer-promoter looping through a Zinc finger protein based tethering system does not influence burst size, i.e, number of nascent transcripts per active allele, but increases the burst fraction, i.e. number of transcriptionally active alleles [114]. Burst fraction is an approximate measure of the frequency at which a promoter is activated. Our study corroborates well with these recent findings. We show that the increase in the number of transcriptionally active enhancers is correlated with the increase in number of mRNA transcription sites during the time course of E2 treatment, but that the RNAPII density on individual alleles is not strongly correlated with continued eRNA synthesis or accumulation of eRNAs on enhancers. Therefore, while eRNAs may activate the enhancer to stimulate a transcriptional burst from the target locus, they might play a less important role in fine tuning burst amplitude or duration.

Our data also show that gene transcription is not completely off in the uninduced state, as we detect weakly transcribing alleles in the absence of E2. Therefore, bursting in this context does not necessarily correspond to a binary switch from an OFF state to an ON state, but may represent a modulation of the rate of transcription initiation. Intriguingly, a recent live cell study demonstrates that transcriptional bursting does not necessarily imply a 2 state model of stochastic gene activation and inactivation in which the gene infrequently switches on from a completely repressed state, but may fit a continuum model, in which the promoter shows a spectrum of different levels of transcriptional activity [176]. The 2-state model presupposes that

the gene is either OFF, when the probability of transcription initiation is zero, or that the gene is ON, when transcription initiates at a constant frequency. Addressing the limitations of previous smFISH studies in their ability to elucidate the full dynamic range of transcriptional activity, Corrigan and colleagues found that their simulations of gene bursting fit a continuum model, whereby fluctuations in transcription initiation rate over short timescales can produce a spectrum of transcription states interspersed with periods of inactivity. Therefore, an active enhancer may enable its target promoter to access high amplitude transcription states by modulating changes in the rate of transcription initiation rather than activating a promoter from a completely repressed state. As discussed previously, eRNA transcription may convert an enhancer from a poised state to an active state through histone acetylation. It was also shown that histone acetylation on the enhancer induces histone acetylation on the target promoter. Therefore, increasing the frequency of enhancer activation during induction may increase the probability that the target gene will access high amplitude transcription states by increasing chromatin accessibility at the promoter. The duration or amplitude of these bursting states may be mostly independent of eRNA synthesis. Future live cell experiments using dual PP7/MS2 reporters to monitor transcription simultaneously from the enhancer and promoter will help to determine whether or not this is true.

4.3.6 eRNAs and enhancer-promoter looping

Functional studies in breast and prostate cancer cell lines have shown that eRNAs associate with cohesin or mediator, and that this association is required to maintain looping, as shown by 3C (Chromosome Conformation Capture) analysis on candidate genes [88,167]. One of key questions of our study was whether transcriptionally active enhancers form stable looping interactions with their target loci, mediated by eRNAs, or if the data previously acquired using 3C simply represents a cumulative average of variable enhancer-promoter conformations. We surmised that if eRNAs were required to stabilize looping interactions, we would frequently detect nascent eRNA signal within a distance from the active promoter which corresponds to a cross-linked enhancer-promoter interaction, as predicted from polymer modeling of chromatin fiber conformations. While smFISH does not allow the detection of all enhancer-promoter interactions within a cell, as it requires mRNA and eRNA co-expression, we can nonetheless measure the frequency of interaction between an enhancer and promoter

that are both transcriptionally active, assuming, that most detected eRNAs are nascent transcripts localized at the enhancer.

Our analysis showed, for both *FOXC1* and *P2RY2,* that eRNA-mRNA transcription site distances spans a range of distances that tails off at 300-400 nm, corroborating with distances between regulatory elements and target genes within their respective TADs, measured previously by DNA FISH [111,112]. This suggests that the eRNAs detected using smFISH are most likely nascent/enhancer associated and therefore provide a reasonable approximation of the location of the enhancer with respect to its target promoter. Simulations of intra-TAD chromatin fiber conformations by polymer modeling predict that two regions on an individual chromatin fiber would be cross-linked when closer than 80 nm, most likely by protein complexes such as mediator or cohesin. Therefore, we sought to quantify the relative fraction of eRNA-mRNA co-expressing alleles in which eRNA-TS distances would fall within this range, to determine how frequently transcriptionally active enhancers interact with their target promoters within a closed loop configuration.

Intriguingly, only about a quarter of eRNA-mRNA co-expressing alleles show eRNA-TS distances less than 100 nm. This suggests enhancer-promoter interactions to be dynamic rather than forming stable closed loops, at least for those enhancers that are detectable through eRNA transcription. If most enhancers and promoters in metazoan genomes are brought into proximity by sequestration within TADs, this may circumvent the need for stable loops, as chromatin fibers within TADs are predicted to be flexible, therefore facilitating frequent enhancer-promoter contacts in response to signaling pathways. The enhancer-promoter interaction may occur for a length of time that is sufficient to enable the target promoter to access a high activity transcription state, as described by the continuum model of bursting, but need not be maintained to sustain multiple cycles of transcription on the target locus. Furthermore, if eRNA transcription is the most robust signature of enhancer activity, we can speculate that most transcriptionally active alleles that do not co-express eRNAs are associated with enhancers that have been temporarily de-activated before the next gene burst. Thus, eRNA transcription may play an early transient role in establishing a permissive chromatin structure at the promoter to activate the burst, after which the eRNA itself becomes dispensable.

Due to the flexibility of chromatin fibers, enhancers might come into contact with promoters through stochastic collisions, a process which might not require eRNAs *per se*. However, the increase in eRNA transcription frequency will establish a more permissive chromatin environment on the enhancer, as well as the promoter, during instances when the promoter transiently contacts the enhancer. The simultaneous increase in chromatin accessibility on both the enhancer and promoter may enable complexes such as mediator and cohesin to bind more efficiently and sustain closed loop configurations for longer periods relative to the uninduced state. Perhaps, this may generate more frequent bursts, as it was shown by Bartman and colleagues that forced chromatin looping can increase the burst fraction [114]. Interestingly, Fukaya and colleagues show that inserting an insulator sequence between a reporter and enhancer reduces the frequency of bursting [113]. Thus, reducing the probability of enhancer-promoter contact may also decrease bursting frequency. Therefore, instead of sustaining the enhancer-promoter loop, eRNAs could also initiate the looping process, thereby increasing the frequency of enhancer-promoter contact during induction relative to the sporadic collisions that may occur prior to induction. This model better fits our data, as we do not detect most eRNAs within distances corresponding to cross-linked enhancer-promoter interactions, suggesting that they may not be required once the closed loop is established.

We do not know the enhancer-promoter configuration for the transcriptionally active alleles that do not co-express eRNAs, or *vice versa* for the eRNAs that do not co-express mRNAs. For future experiments, high resolution DNA FISH should be combined with smFISH to elucidate the functional link between eRNA transcription and enhancer-promoter looping. Furthermore, a CRISPR/Cas9 based strategy could be implemented to target dCas9 fused fluorescent proteins on both the core enhancer element and the promoter, while introducing PP7/MS2 repeats to monitor eRNA and mRNA transcription. This will allow determining if eRNA transcription regulates the lifetime of enhancer-promoter loops, and whether this would selectively modulate different parameters of gene bursting. However, this will be technically challenging as it requires the use of multiple fluorescent labels with non-overlapping spectra as well as imaging at high spatio-temporal resolution and sensitivity to obtain results that are interpretable in a meaningful way.

4.3.7 Perspectives for future research

One of the fundamental questions regarding the role of eRNAs in transcription regulation that has been at the forefront of many discussions in recent years is whether the eRNA molecule itself has an actual function or whether it is just a transcriptional by-product, whereby the process of RNA pol II elongation at the enhancer is what is required for enhancer function, making the eRNA itself dispensable. Various observations suggest that the eRNA molecule is important for enhancer function. First, it was shown that eRNAs may be functional in transcription activation independently of their synthesis, as tethering eRNAs to luciferase reporter plasmids increases expression [88,134]. Second, eRNAs were shown to undergo transcription termination by the RNA pol II associated Integrator complex, and disruption of this process alters enhancer function, suggesting that specific processing of eRNAs is required for enhancer function [140]. Third, RIP-qPCR assays on components of cohesin and mediator, two major complexes that are implicated in the formation of enhancer-promoter loops, show enrichment of eRNAs [88,129]. Altogether, these studies compel us to further characterize eRNAs from an RNA centric perspective, including interrogating whether eRNAs are generally terminated and/or processed in a uniform fashion, and whether they possess motifs and/or secondary structures that allow them to bind and regulate transcription co-factors. To achieve this, a next step will require mechanistic studies including proteomic analyses using approaches such as ChIRP-MS (Chromatin isolation by RNA purification with mass spectrometry) to determine if eRNAs recruit specific factors to enhancers [177]. In addition, assays such as iCLIP-seq on factors associated with active enhancers, such as mediator and cohesin or the histone acetyltransferases CBP/p300, can be used to determine if they bind to functional motifs within eRNAs [178]. Furthermore, deletion of potential motifs using CRISPR/Cas9 will validate the functional relevance of sequences within the eRNA transcripts. Moreover, chemical probing methods such as SHAPE can be used to map the secondary structure of eRNAs. SHAPE has been used recently to determine the *in vitro* secondary structure of the mouse lncRNA Braveheart, which specifies the cardiomyocyte lineage during development, demonstrating the utility of this approach [179]. Altogether, such studies will help to better understand the role of eRNA sequence in eRNA function.

Recent high-throughput studies have shown that most lincRNAs are processed co-transcriptionally in a different manner than mRNAs, expanding our knowledge of the biochemical properties that distinguish lncRNAs from mRNAs [180,181]. The primary distinguishing feature is the lack of efficient splicing in lincRNAs. Furthermore, Schlackow and colleagues show that lincRNAs are prematurely terminated and destabilized by the nuclear exosome [181]. These findings corroborate with earlier studies which have shown that although regulatory elements such as enhancers and promoters share a common architecture of transcription initiation, there are motifs downstream of transcription start sites which distinguish the fates of lncRNAs from mRNAs [45,50]. Notably, eRNAs and PROMPTs lack the U1 snRNP 5'splice sites that promote efficient splicing and elongation on protein coding sequences, but harbor early polA signals that induce premature termination. It was shown that PROMPT degradation by the nuclear exosome requires early polyA mediated termination [50]. Interestingly, PROMPTs and eRNAs are both stabilized upon depletion of the components the nuclear exosome and the nuclear exosome targeting complex (NEXT), whose RNA binding component RBM7 binds to lncRNAs. [53]. This finding suggests that the exosome is recruited to distinct classes of lncRNAs in a targeted manner. eRNAs expressed from super-enhancer regions were also shown to stabilize in exosome knockout mutants [182]. Whether eRNA degradation by the exosome is coupled to early polyA signal mediated termination, similar to PROMPTs, or occurs through a different mechanism is yet to be determined. It will be interesting to determine the role of the exosome in eRNA metabolism.

Another important question that emerges from these studies is whether the regulatory roles of eRNAs and other classes of lncRNAs depend on processes associated specifically to their production, including synthesis and co-transcriptional processing, rather than sequence specificity within the mature transcripts themselves. It was recently shown that lncRNAs that act locally on neighboring genes do not mediate their function through specific sequences within the transcripts but require their respective promoters or promoter-proximal processing mechanisms to regulate transcription of their target genes [183]. The lincRNA *blustr* in mouse, which activates the neighboring gene *Sfmbt2,* requires the U1 snRNP 5'splice site within its first intron, in addition to its promoter, to activate transcription. Conversely, inserting an early polyA signal to terminate transcription reduced target gene expression, whereas inserting

polyA signals downstream to increase the length of transcriptional read-through increased expression. Since the deletion of introns and exons within the *blustr* had no effect on target expression, the regulatory mechanism is clearly mediated by non-sequence specific promoter-proximal processing steps that affect transcription elongation. It will be interesting to see whether such principles can be applied to other classes of lincRNAs.

Intriguingly, as eRNAs are enriched in early polyA signals and not 5'splice sites, their regulatory capacity may be dependent on early termination. Supporting this idea, depleting Integrator components results in increased transcriptional read-through of eRNAs, which then fails to stimulate target gene expression [140]. Therefore, altering properties of eRNA processing/termination will be important to understand the role of processing in eRNA function. One way to achieve this would be to introduce a 5'splice site into the eRNA sequence using CRISPR/Cas9, and determine if this targeted modification affects the processing and function of the eRNA. Alternatively, the eRNA sequence could be exchanged with the target mRNA sequence so that the mRNA is under the regulation of the core enhancer and the eRNA is under the regulation of the core promoter. smFISH can then be used to determine if the altered genomic context results in the mRNA behaving like an eRNA, and *vice versa*. This would confirm that the core regulatory element takes precedence over downstream sequence specificity in defining the processing and function of the transcript. Conversely, if mRNA and eRNA behavior are unaltered in their new genomic context, this would suggest that the functionality of the transcript is defined by some element within itself rather than the sequence of the core regulatory element driving its production.

Alternatively, instead of introducing genetic perturbations, eRNA synthesis can be inhibited using CRISPR interference, blocking RNA pol II elongation by recruiting dCas9 downstream of the transcription initiation site. CRISPR interference has been used successfully in recent studies to inhibit transcription of the *GAL10* antisense RNA in *S.Cerevisiae* and the ncRNAs Rox1 and Rox2, which are involved in sex dosage compensation in *D.melanogaster* [162,184]. Furthermore, a chimeric sgRNA-eRNA construct can be tethered to the enhancer while applying CRISPR interference to determine if the eRNA can act in *trans* when its endogenous synthesis is abolished. Altogether, such studies will be essential to better understand the role of eRNAs in enhancer function.

4.4 Concluding remarks

The focus of my thesis was to elucidate the roles of different classes of long non-coding RNAs in eukaryotic transcription regulation. In the studies presented here we combined quantitative single molecule resolution microscopy with genetic and biochemical approaches to reveal novel mechanistic aspects of transcription regulation by long non-coding RNAs.

My first project demonstrated that continuous low frequency transcription of an antisense long noncoding RNA represses transcription activation of the *S.cerevisiae* phosphate transporter gene *PHO84* through a bimodal switch mechanism. My data, together with more recent studies expanding on our observations, demonstrate that regulation of non-coding RNA transcription initiating continuously from open chromatin regions, may allow cells to integrate fluctuating environmental signals to control gene activity in a switch-like manner as an adaptive response. This mechanism of gene regulation would act locally on proximal promoters in a rapid and economical manner through chromatin modifying factors associated with elongating RNA pol II, circumventing the slow and energy-intensive process of synthesizing trans-regulatory proteins and targeting them to specific gene regulatory sequences. Since antisense noncoding RNAs are also expressed in metazoans, it will be interesting for future investigation to determine if such bimodal switches can also be applied to antisense RNA regulated genes in multicellular eukaryotes.

For the second project, I investigated enhancer RNAs, a class of lncRNAs expressed from enhancers. The findings of this study have allowed us to infer potential transcription regulatory mechanisms mediated by eRNAs that challenge some of the previous models of eRNA function. Importantly, the results from our work suggest a model whereby the process of eRNA transcription may increase chromatin accessibility at the promoter to activate transcriptional bursting, but which does not require eRNA co-expression to sustain transcription on target loci. We also discovered that transcriptionally active enhancers rarely engage in closed loop interactions with their target promoters, suggesting that eRNAs are unlikely to stabilize enhancer-promoter interaction, but may initiate looping prior to promoter activation.

In summary, both of my PhD projects share a common theme where we show that low frequency noncoding RNA transcription through regulatory regions can fine-tune promoter activity *in cis,* and that this activity is highly variable within the cell population. My work has provided valuable new mechanistic insights on how lncRNAs participate in transcription regulation at the single cell level and lays the foundations for future studies that will further enhance our understanding of non-coding RNA function.

4.5 Figure legends

Figure 4-1: Transcriptional bursting of *PHO84* is regulated by continuous antisense transcription

a. Diagram showing alternate transcription between sense and AS in wild type cells grown in intermediate phosphate media. Sense transcription occurs in discrete bursts, characterized by frequent re-initation and accumulation of RNAs during the ON period. AS transcription occurs continuously as single initiation events distributed in time, occasionally transcribing through the *PHO84* UAS, due to weak termination by Nrd1-Nab3-Sen1. The nucleosome shown in red is inherently unstable and occludes Pho4p binding site D, which is close to the TATA box and the primary regulator of *PHO84* activation.

b. Diagram showing alternate transcription between sense and AS in Δ*rrp6* cells. In Δ*rrp6* there is an approximate 2-fold increase in the number of AS transcription cycles that read through the *PHO84* UAS, resulting in a 2-fold decrease in *PHO84* bursting frequency. The nucleosome shown in orange might be positioned more stably relative to wild type, due to more frequent nucleosome remodeling events coupled to HDAC recruitment.

c. Wild type cells in phosphate depleted media do not follow a 2-state ON/OFF model, but continuously transcribe *PHO84* due to complete remodeling of the UAS and promoter. AS transcription is completely abolished in inducing conditions.

Figure 4-2: The role of MLL1 and eRNA transcription in regulating enhancer accessibility and promoter activation

a. MLL1 methylates H3K4 on enhancers that are transcribed at a basal level.

b. FOXA1 binds to H3K4me1/2 modified enhancers

c. FOXA1 remodels chromatin and recruits E2 conjugated ERα to the estrogen response element (ERE), which then recruits RNA pol II to induce bi-directional eRNA transcription. The increase in eRNA transcription frequency makes chromatin more accessible on the enhancer and promoter, thereby activating bursting transcription of target mRNAs.

4.6 Figures

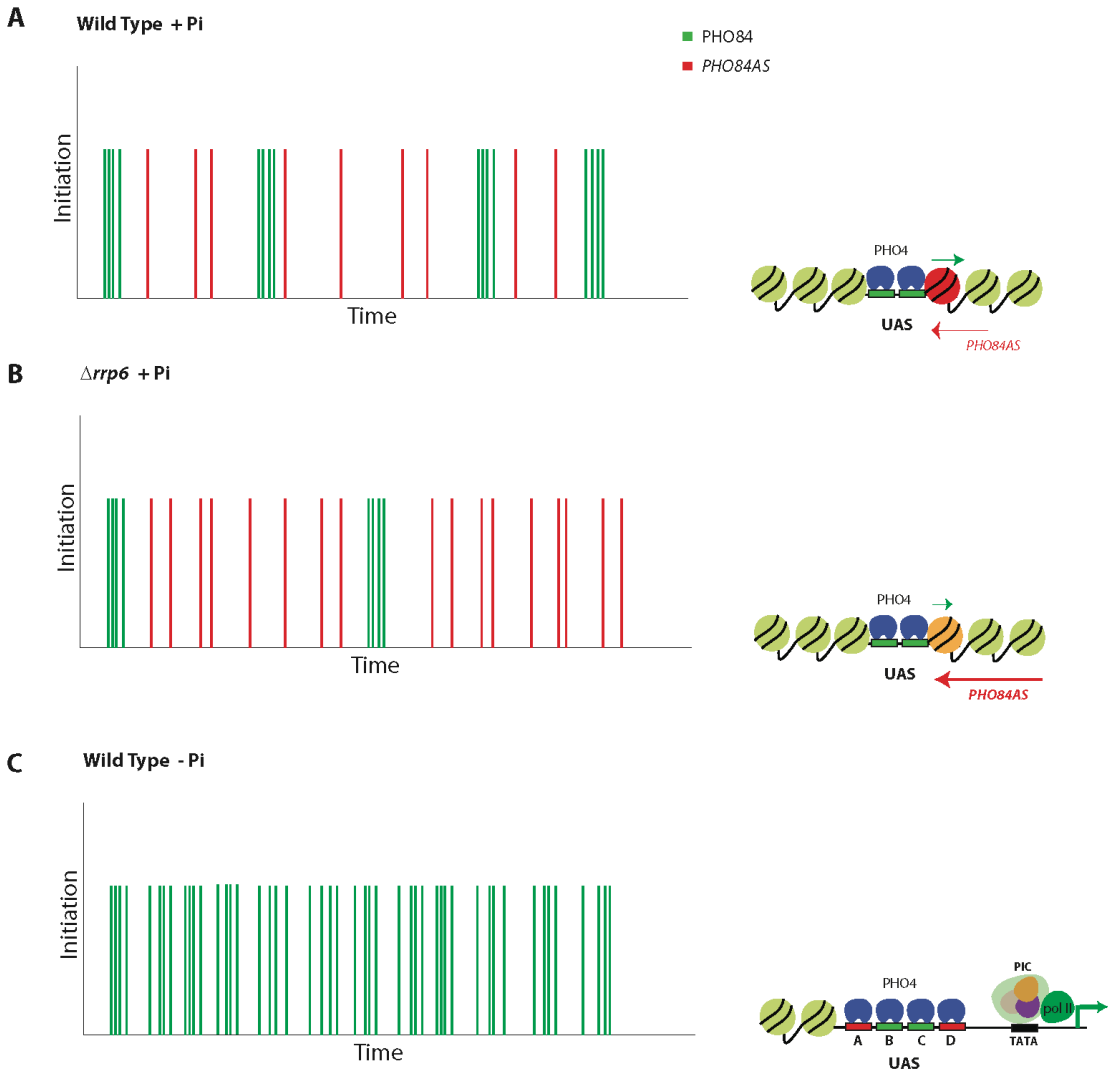


Figure 4-1: Transcriptional bursting of *PHO84* is regulated by continuous antisense transcription

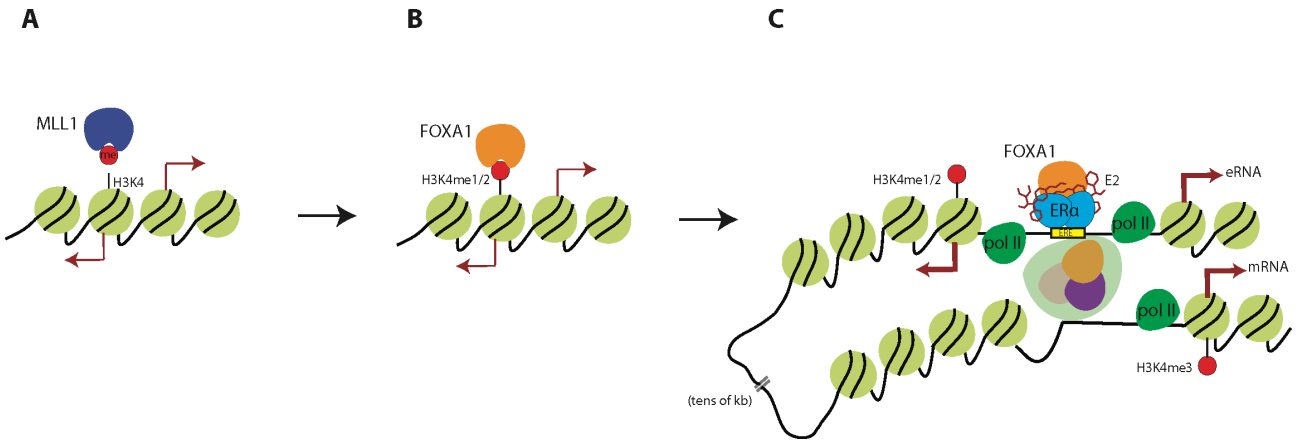


Figure 4-2: The role of MLL1 and eRNA transcription in regulating enhancer accessibility and promoter activation

Bibliography

1. CRICK, F. Central Dogma of Molecular Biology. *Nature* **227,** 561–563 (1970).
2. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–108 (2012).
3. Derrien, T. *et al.* The GENCODE v7 catalogue of human long non-coding RNAs : Analysis of their structure , evolution and expression. *Genome Res.* **22,** 1775–1789 (2012).
4. Kapranov, P. *et al.* Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.* **15,** 987–997 (2005).
5. Mattick, J. S. The genetic signatures of noncoding RNAs. *PLoS Genet.* **5,** (2009).
6. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and Functions of Long Noncoding RNAs. *Cell* **136,** 629–641 (2009).
7. Ulitsky, I. & Bartel, D. P. LincRNAs: Genomics, evolution, and mechanisms. *Cell* **154,** 26–46 (2013).
8. Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* **8,** 413–23 (2007).
9. Taft, R. J. & Mattick, J. S. Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *Genome Biol.* **5,** P1 (2003).
10. Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251–260 (1997).
11. Koster, M. J. E., Snel, B. & Timmers, H. T. M. Genesis of chromatin and transcription dynamics in the origin of species. *Cell* **161,** 724–736 (2015).
12. Almer, a, Rudolph, H., Hinnen, a & Hörz, W. Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. *EMBO J.* **5,** 2689–96 (1986).
13. Svaren, J., Schmitz, J. & Hörz, W. The transactivation domain of Pho4 is required for nucleosome disruption at the PHO5 promoter. *EMBO J.* **13,** 4856–62 (1994).
14. Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* **11,** 1587–1598 (2003).
15. Reinke, H. & Hörz, W. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol. Cell* **11,** 1599–1607 (2003).
16. Schermer, U. J., Korber, P. & Hörz, W. Histones are incorporated in trans during reassembly of the yeast PHO5 promoter. *Mol. Cell* **19,** 279–285 (2005).
17. Neill, E. M. O., Kaffman, A., Jolly, E. R. & Shea, E. K. O. Regulation of PH04 Nuclear Localization by the PH080-PH085 Cyclin-CDK Complex Author (s): Elizabeth M . O ' Neill , Arie Kaffman , Emmitt R . Jolly and Erin K . O ' Shea Published by : American Association for the Advancement of Science Stable URL : htt. **271,** 209–212 (1996).
18. Springer, M., Wykoff, D. D., Miller, N. & O'Shea, E. K. Partially phosphorylated Pho4 activates transcription of a subset of phosphate-responsive genes. *PLoS Biol.* **1,** (2003).
19. Barbaric, S., Walker, J., Schmid, A., Svejstrup, J. Q. & Hörz, W. Increasing the rate of

chromatin remodeling and gene activation - A novel role for the histone acetyltransferase Gcn5. *EMBO J.* **20,** 4944–4951 (2001).
20. Barbaric, S., Reinke, H. & Hörz, W. Multiple mechanistically distinct functions of SAGA at the PHO5 promoter. *Mol. Cell. Biol.* **23,** 3468–76 (2003).
21. Michael F. Dion, Tommy Kaplan, Minkyu Kim, Stephen Buratowski, 4 & Nir Friedman, O. J. R. Dynamics of Replication-Independent\nHistone Turnover in Budding Yeast. *Science (80-.).* 1405–1408 (2007).
22. van Werven, F. J., van Teeffelen, H. a a M., Holstege, F. C. P. & Timmers, H. T. M. Distinct promoter dynamics of the basal transcription factor TBP across the yeast genome. *Nat. Struct. Mol. Biol.* **16,** 1043–1048 (2009).
23. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304,** 1811–4 (2004).
24. Elowitz, M. B. Stochastic Gene Expression in a Single Cell. *Science (80-.).* **1183,** (2002).
25. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional Pulsing of a Developmental Gene. *Curr. Biol.* **16,** 1018–1025 (2006).
26. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* (2006). doi:10.1371/journal.pbio.0040309
27. Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* (2008). doi:10.1038/nsmb.1514
28. Blake, W. J. *et al.* Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. *Mol. Cell* **24,** 853–865 (2006).
29. Sung Rhee, H. & Franklin Pugh, B. Erratum: Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* (2012). doi:10.1038/nature11266
30. Jacquier, A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.* **10,** 833–844 (2009).
31. Berretta, J. & Morillon, A. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.* **10,** 973–982 (2009).
32. Neil, H. *et al.* Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457,** 1038–1042 (2009).
33. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457,** 1033–1037 (2009).
34. Steinmetz, E. J., Conrad, N. K., Brow, D. a & Corden, J. L. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413,** 327–331 (2001).
35. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S. & Meinhart, A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **15,** 795 – 804 (2008).
36. Webb, S., Hector, R. D., Kudla, G. & Granneman, S. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* **15,** R8 (2014).
37. Gudipati, R. K., Neil, H., Feuerbach, F., Malabat, C. & Jacquier, A. The yeast RPL9B gene is regulated by modulation between two modes of transcription termination. *EMBO J.* **31,** 2427–2437 (2012).
38. Vasiljeva, L. & Buratowski, S. Nrd1 interacts with the nuclear exosome for 3'

processing of RNA polymerase II transcripts. *Mol. Cell* **21,** 239–248 (2006).
39. Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J. & Libri, D. Transcription Termination and Nuclear Degradation of Cryptic Unstable Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome Surveillance. *Mol. Cell* (2006). doi:10.1016/j.molcel.2006.07.029
40. Arigo, J. T., Eyler, D. E., Carroll, K. L. & Corden, J. L. Termination of Cryptic Unstable Transcripts Is Directed by Yeast RNA-Binding Proteins Nrd1 and Nab3. *Mol. Cell* (2006). doi:10.1016/j.molcel.2006.07.024
41. Wyers, F. *et al.* Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* (2005). doi:10.1016/j.cell.2005.04.030
42. Geisler, S., Lojek, L., Khalil, A. M., Baker, K. E. & Coller, J. Decapping of Long Noncoding RNAs Regulates Inducible Genes. *Mol. Cell* **45,** 279–291 (2012).
43. van Dijk, E. L. *et al.* XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475,** 114–117 (2011).
44. Koch, F. *et al.* Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* **18,** 956–963 (2011).
45. Core, L. J. *et al.* Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Publ. Gr.* **46,** 3–1 (2014).
46. Andersson, R., Sandelin, A. & Danko, C. G. A unified architecture of transcriptional regulatory elements. *Trends Genet.* (2015). doi:10.1016/j.tig.2015.05.007
47. Andersson, R. *et al.* Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *bioRxiv* **5,** 005447 (2014).
48. Preker, P. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322,** 1851–1854 (2008).
49. Preker, P. *et al.* PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res.* **39,** 7179–7193 (2011).
50. Ntini, E. *et al.* Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20,** 923–8 (2013).
51. Natoli, G. & Andrau, J.-C. Noncoding Transcription at Enhancers: General Principles and Functional Models. *Annu. Rev. Genet.* **46,** 120820103026000 (2011).
52. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. (2013). doi:10.1038/nature12349
53. Lubas, M. *et al.* The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Article The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis. *CellReports* **10,** 178–192 (2015).
54. Andersen, P. R. *et al.* The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Struct. Mol. Biol.* **20,** 1367–76 (2013).
55. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507,** 455–61 (2014).
56. Cabili, M. N. *et al.* Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16,** 20 (2015).
57. Managadze, D., Rogozin, I. B., Chernikova, D., Shabalina, S. A. & Koonin, E. V.

Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. *Genome Biol. Evol.* **3,** 1390–1404 (2011).
58. Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Mol. Cell* **43,** 904–914 (2011).
59. Batista, P. J. & Chang, H. Y. Long noncoding RNAs: Cellular address codes in development and disease. *Cell* **152,** 1298–1307 (2013).
60. Sabin, L. R., Delás, M. J. & Hannon, G. J. Dogma Derailed: The Many Influences of RNA on the Genome. *Mol. Cell* **49,** 783–794 (2013).
61. Kretz, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493,** 231–5 (2013).
62. Carrieri, C. *et al.* Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* (2012). doi:10.1038/nature11508
63. Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **470,** 284–288 (2011).
64. Yoon, J. H. *et al.* LincRNA-p21 Suppresses Target mRNA Translation. *Mol. Cell* **47,** 648–655 (2012).
65. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol. Cell* **44,** 667–678 (2011).
66. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 11667–11672 (2009).
67. Zhao, J. *et al.* Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Mol. Cell* **40,** 939–953 (2010).
68. Tsai, M.-C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329,** 689–93 (2010).
69. Yang, L. *et al.* NcRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147,** 773–788 (2011).
70. Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39,** 925–938 (2010).
71. Clemson, C. M. *et al.* An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell* **33,** 717–726 (2009).
72. Elisaphenko, E. A. *et al.* A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS One* **3,** 1–11 (2008).
73. Tisseur, M., Kwapisz, M. & Morillon, A. Pervasive transcription - Lessons from yeast. *Biochimie* **93,** 1889–1896 (2011).
74. Kuehner, J. N. & Brow, D. a. Regulation of a Eukaryotic Gene by GTP-Dependent Start Site Selection and Transcription Attenuation. *Mol. Cell* **31,** 201–211 (2008).
75. Bumgarner, S. L., Dowell, R. D., Grisafi, P., Gifford, D. K. & Fink, G. R. Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 18321–18326 (2009).
76. Martens, J. a, Laprade, L. & Winston, F. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature* **429,** 571–574 (2004).
77. Martens, J. a., Wu, P. Y. J. & Winston, F. Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. *Genes Dev.* **19,**

2695–2704 (2005).
78. Hainer, S. J., Pruneski, J. a., Mitchell, R. D., Monteverde, R. M. & Martens, J. a. Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev.* **25,** 29–40 (2011).
79. Uhler, J. P., Hertel, C. & Svejstrup, J. Q. A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 8011–8016 (2007).
80. Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D. & Vogelauer, M. A ncRNA Modulates Histone Modification and mRNA Induction in the Yeast GAL Gene Cluster. *Mol. Cell* **32,** 685–695 (2008).
81. Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G. & Stutz, F. Antisense RNA Stabilization Induces Transcriptional Gene Silencing via Histone Deacetylation in S. cerevisiae. *Cell* **131,** 706–717 (2007).
82. Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors, RXR, and the big bang. *Cell* (2014). doi:10.1016/j.cell.2014.03.012
83. Shang, Y., Hu, X., DiRenzo, J., Lazar, M. a & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103,** 843–852 (2000).
84. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38,** 1289–1297 (2006).
85. Welboren, W.-J. *et al.* ChIP-Seq of ERα and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.* **28,** 1418–1428 (2009).
86. Carroll, J. S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122,** 33–43 (2005).
87. Hah, N. *et al.* A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145,** 622–634 (2011).
88. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498,** 516–20 (2013).
89. Ansari, K. I. & Mandal, S. S. Mixed lineage leukemia: Roles in gene expression, hormone signaling and mRNA processing. *FEBS Journal* (2010). doi:10.1111/j.1742-4658.2010.07606.x
90. Dillon, S. C., Zhang, X., Trievel, R. C. & Cheng, X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* **6,** 227 (2005).
91. Kaikkonen, M. U. *et al.* Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* **51,** 310–325 (2013).
92. Lee, J. E. *et al.* H3K4 mono- And di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife* (2013). doi:10.7554/eLife.01503
93. Mo, R., Rao, S. M. & Zhu, Y. J. Identification of the MLL2 complex as a coactivator for estrogen receptor ?? *J. Biol. Chem.* (2006). doi:10.1074/jbc.M513245200
94. Dreijerink, K. M. A. *et al.* Menin links estrogen receptor activation to histone H3K4 trimethylation. *Cancer Res.* (2006). doi:10.1158/0008-5472.CAN-05-4461
95. Lupien, M. *et al.* FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription. *Cell* (2008). doi:10.1016/j.cell.2008.01.018
96. Jeong, K. W. *et al.* Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. (2011). doi:10.1038/nsmb.2153
97. Jeong, K. W., Andreu-Vieyra, C., You, J. S., Jones, P. A. & Stallcup, M. R. Establishment of active chromatin structure at enhancer elements by mixed-lineage

leukemia 1 to initiate estrogen-dependent gene expression. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gkt1236
98. Ansari, K. I., Mishra, B. P. & Mandal, S. S. Human CpG binding protein interacts with MLL1, MLL2 and hSet1 and regulates Hox gene expression. *Biochim. Biophys. Acta - Gene Regul. Mech.* (2008). doi:10.1016/j.bbagrm.2007.11.006
99. Wiench, M. *et al.* DNA methylation status predicts cell type-specific enhancer activity. *EMBO J.* **30,** 3028–39 (2011).
100. Ong, C.-T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12,** 283–293 (2011).
101. Marsman, J. & Horsfield, J. a. Long distance relationships: Enhancer-promoter communication and dynamic gene transcription. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819,** 1217–1227 (2012).
102. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467,** 430–435 (2010).
103. DeMare, L. E. *et al.* The genomic landscape of cohesin-Associated chromatin interactions. *Genome Res.* **23,** 1224–1234 (2013).
104. Cremer, M. *et al.* Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosom. Res.* **9,** 541–567 (2001).
105. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
106. Dekker, J., Marti-Renom, M. a & Mirny, L. a. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* **14,** 390–403 (2013).
107. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* (2015). doi:10.1016/j.cell.2015.04.004
108. Liu, M. H. & Cheung, E. Estrogen receptor-mediated long-range chromatin interactions and transcription in breast cancer. *Mol. Cell. Endocrinol.* **382,** 624–632 (2014).
109. Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462,** 58–64 (2009).
110. Ghavi-helm, Y. *et al.* Enhancer Loops Appear Stable During Development and Are Associated With Paused Polymerase. *Nature* **512,** 96–100 (2014).
111. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157,** 950–963 (2014).
112. Williamson, I., Lettice, L. A., Hill, R. E. & Bickmore, W. A. Shh and ZRS enhancer co-localisation is specific to the zone of polarizing activity 6 key words: 5C, chromosome loop, enhancer, limb development, super-resolution microscopy. (2016). doi:10.1242/dev.139188
113. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166,** 358–368 (2016).
114. Bartman, C. R., Hsu, S. C., Hsiung, C. C. S., Raj, A. & Blobel, G. A. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol. Cell* (2016). doi:10.1016/j.molcel.2016.03.007
115. Melo, C. a, Léveillé, N. & Agami, R. eRNAs reach the heart of transcription. *Cell Res.* **23,** 1151–2 (2013).
116. Lam, M. T. Y., Li, W., Rosenfeld, M. G. & Glass, C. K. Enhancer RNAs and regulated

transcriptional programs. *Trends Biochem. Sci.* **39,** 170–182 (2014).
117. Redmond, A. M. & Carroll, J. S. Enhancer-derived RNAs: 'spicing up' transcription programs. *EMBO J.* **32,** 2096–8 (2013).
118. Lai, F. & Shiekhattar, R. Enhancer RNAs: The new molecules of transcription. *Curr. Opin. Genet. Dev.* **25,** 38–42 (2014).
119. Orom, U. A. & Shiekhattar, R. XLong noncoding RNAs usher in a new era in the biology of enhancers. *Cell* **154,** 1190–1193 (2013).
120. Stees, J. S., Varn, F., Huang, S., Strouboulis, J. & Bungert, J. Recruitment of Transcription Complexes to Enhancers and the Role of Enhancer Transcription. *Biology (Basel)*. **1,** 778–793 (2012).
121. Darrow, E. M. & Chadwick, B. P. Boosting transcription by transcription: Enhancer-associated transcripts. *Chromosom. Res.* **21,** 713–724 (2013).
122. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).
123. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474,** 390–394 (2011).
124. de Santa, F. *et al.* A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol.* **8,** (2010).
125. Zhu, Y. *et al.* Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res.* **41,** 10032–10043 (2013).
126. Milne, T. a *et al.* MLL associates specifically with a subset of transcriptionally active target genes. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 14765–14770 (2005).
127. Lam, M. T. Y. *et al.* Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498,** 511–5 (2013).
128. Mousavi, K. *et al.* ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Mol. Cell* **51,** 606–617 (2013).
129. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494,** 497–501 (2013).
130. Hsieh, C.-L. *et al.* Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 7319–24 (2014).
131. Hadjur, S. *et al.* Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460,** 410–413 (2009).
132. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20,** 578–588 (2010).
133. Hah, N., Murakami, S., Nagari, A., Danko, C. G. & Lee Kraus, W. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.* **23,** 1210–1223 (2013).
134. Melo, C. a. *et al.* ERNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. *Mol. Cell* **49,** 524–535 (2013).
135. Schaukowitch, K. *et al.* Enhancer RNA Facilitates NELF Release from Immediate Early Genes. *Mol. Cell* **56,** 29–42 (2014).
136. Zhao, Y. *et al.* Activation of P-TEFb by Androgen Receptor-Regulated Enhancer RNAs in Castration-Resistant Prostate Cancer. *Cell Rep.* (2016). doi:10.1016/j.celrep.2016.03.038
137. Weinhold, N. *et al.* Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. **347,** 1010–1015 (2015).

138. Kanno, T. *et al.* BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. *Nat. Struct. & Mol. Biol.* 1–14 (2014). doi:10.1038/nsmb.2912
139. Nagarajan, S. *et al.* Bromodomain Protein BRD4 Is Required for Estrogen Receptor-Dependent Enhancer Activation and Gene Transcription. *Cell Rep.* **8,** 460–469 (2014).
140. Lai, F., Gardini, A., Zhang, A. & Shiekhattar, R. Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525,** 399–403 (2015).
141. Baillat, D. *et al.* Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123,** 265–276 (2005).
142. Baillat, D. & Wagner, E. J. Integrator: Surprisingly diverse functions in gene expression. *Trends in Biochemical Sciences* **40,** 257–264 (2015).
143. Skaar, J. R. *et al.* The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res.* **25,** 1–18 (2015).
144. Stadelmayer, B. *et al.* ARTICLE Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* **5,** (2014).
145. Gardini, A. *et al.* Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* **56,** 128–139 (2014).
146. Yamamoto, J. *et al.* DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat. Commun.* **5,** 1–10 (2014).
147. Pnueli, L., Rudnizky, S., Yosefzon, Y. & Melamed, P. RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α-subunit gene. *Proc. Natl. Acad. Sci.* **112,** 201414841 (2015).
148. Kim, T. K. & Shiekhattar, R. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* **162,** 948–959 (2015).
149. Matharu, N. & Ahituv, N. Minor Loops in Major Folds: Enhancer–Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. *PLoS Genetics* (2015). doi:10.1371/journal.pgen.1005640
150. Bumgarner, S. L. *et al.* Single-Cell Analysis Reveals that Noncoding RNAs Contribute to Clonal Heterogeneity by Modulating Transcription Factor Recruitment. *Mol. Cell* (2012). doi:10.1016/j.molcel.2011.11.029
151. Van Werven, F. J. *et al.* Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* **150,** 1170–1181 (2012).
152. Creamer, T. J. *et al.* Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet.* **7,** (2011).
153. Fox, M. J., Gao, H., Smith-Kinnaman, W. R., Liu, Y. & Mosley, A. L. The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway. *PLoS Genet.* **11,** e1004999 (2015).
154. Castelnuovo, M. *et al.* Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res.* **42,** 4348–4362 (2014).
155. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419,** 407–411 (2002).
156. Ng, H. H., Robert, F., Young, R. A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* **11,** 709–719 (2003).

157. Boa, S., Coert, C. & Patterton, H. Saccharomyces cerevisiae Set1p is a methyltransferase specific for lysine 4 of histone H3 and is required for efficient gene expression. *Yeast* **20,** 827–835 (2003).
158. Bernstein, B. E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 8695–8700 (2002).
159. Margaritis, T. *et al.* Two Distinct Repressive Mechanisms for Histone 3 Lysine 4 Methylation through Promoting 3???-End Antisense Transcription. *PLoS Genet.* **8,** (2012).
160. Wippo, C. J. *et al.* Differential cofactor requirements for histone eviction from two nucleosomes at the yeast PHO84 promoter are determined by intrinsic nucleosome stability. *Mol. Cell. Biol.* **29,** 2960–2981 (2009).
161. Burugula, B. B. *et al.* Histone Deacetylases and Phosphorylated Polymerase II C-Terminal Domain Recruit Spt6 for Cotranscriptional Histone Reassembly. *Mol. Cell. Biol.* **34,** 4115–4129 (2014).
162. Lenstra, T. L., Coulon, A., Chow, C. C. & Larson, D. R. Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. *Mol. Cell* (2015). doi:10.1016/j.molcel.2015.09.028
163. Lam, F. H., Steger, D. J. & O'Shea, E. K. Chromatin decouples promoter threshold from dynamic range. *Nature* **453,** 246–50 (2008).
164. Mason, P. B. & Struhl, K. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol. Cell* **17,** 831–840 (2005).
165. Mousavi, K. *et al.* ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Mol. Cell* **51,** 606–617 (2013).
166. Bahar Halpern, K. *et al.* Bursty Gene Expression in the Intact Mammalian Liver. *Mol. Cell* (2015). doi:10.1016/j.molcel.2015.01.027
167. Hsieh, C.-L. *et al.* Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc. Natl. Acad. Sci. U. S. A.* (2014). doi:10.1073/pnas.1324151111
168. Ross-Innes, C. & Stark, R. Cooperative interaction between retinoic acid receptor-α and estrogen receptor in breast cancer. *Genes ...* **3,** 171–182 (2010).
169. Calo, E. & Wysocka, J. Modification of Enhancer Chromatin: What, How, and Why? *Mol. Cell* **49,** 825–837 (2013).
170. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–83 (2011).
171. Bose, D. A. *et al.* RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell* **168,** 135–149.e22 (2017).
172. Yang, Y. *et al.* Enhancer RNA-driven looping enhances the transcription of the long noncoding RNA DHRS4-AS1, a controller of the DHRS4 gene cluster. *Sci. Rep.* **6,** 20961 (2016).
173. Arab, K. *et al.* Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol. Cell* **55,** 604–614 (2014).
174. Won Jeong, K. *et al.* Recognition of enhancer element–specific histone methylation by TIP60 in transcriptional activation. *Nat. Struct. Mol. Biol.* **18,** (2011).
175. Jozwik, K. M., Chernukhin, I., Serandour, A. A., Nagarajan, S. & Carroll, J. S. FOXA1 directs H3K4 monomethylation at enhancers via recruitment of the methyltransferase MLL3. *Cell Rep.* **17,** 2715–2723 (2016).

176. Corrigan, A. M., Tunnacliffe, E., Cannon, D. & Chubb, J. R. A continuum model of transcriptional bursting. *Elife* **5,** 1–38 (2016).
177. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161,** 404–416 (2015).
178. Hauer, C. *et al.* Improved binding site assignment by high-resolution mapping of RNA–protein interactions using iCLIP. *Nat. Commun.* **6,** 7921 (2015).
179. Xue, Z. *et al.* A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. *Mol. Cell* **64,** 37–50 (2016).
180. Melé, M. *et al.* Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* 1–11 (2016). doi:10.1101/gr.214205.116
181. Schlackow, M. *et al.* Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs Article Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol. Cell* **65,** 1–14 (2017).
182. Pefanis, E. *et al.* RNA Exosome-Regulated Long Non-Coding RNA Transcription Controls Super-Enhancer Activity. *Cell* (2015). doi:10.1016/j.cell.2015.04.034
183. Engreitz, A. J. M. *et al.* Neighborhood regulation by lncRNA promoters, transcription, and splicing. *Nature* **539,** 1–15 (2016).
184. Ghosh, S., Tibbit, C. & Liu, J.-L. Effective knockdown of *Drosophila* long non-coding RNAs by CRISPR interference. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw063