Study of the activation dynamics of gene transcription by RNA polymerase 2

Sébastien Causse

▶ **To cite this version:**

Sébastien Causse. Study of the activation dynamics of gene transcription by RNA polymerase 2. Biochemistry, Molecular Biology. Université Pierre et Marie Curie - Paris VI, 2011. English. NNT : 2011PA066464 . tel-00824828

HAL Id: tel-00824828
https://theses.hal.science/tel-00824828

Submitted on 22 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité

Biologie moléculaire
Ecole doctorale : Interdisciplinaire pour le vivant

Présentée par

M. Sébastien Causse

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

Etude de la dynamique d'activation de la transcription des gènes par l'ARN Polymerase 2.

Devant le jury composé de :
M. Xavier Darzacq, directeur de thèse.

M. Edouard Bertrand et M. Domenico Libri, rapporteurs de thèse.

M. Olivier Bensaude, M. Domenico Libri, M. John T. Lis, M. Christophe Zimmer, M. Frédéric Devaux, examinateurs.

Université Pierre & Marie Curie - Paris 6
Bureau d'accueil, inscription des doctorants et base de données
Esc G, 2ème étage
15 rue de l'école de médecine
75270-PARIS CEDEX 06

Tél. Secrétariat : 01 42 34 68 35
Fax : 01 42 34 68 40
Tél. pour les étudiants de A à EL : 01 42 34 69 54
Tél. pour les étudiants de EM à MON : 01 42 34 68 41
Tél. pour les étudiants de MOO à Z : 01 42 34 68 51
E-mail : scolarite.doctorat@upmc.fr

Title :
Study of the activation dynamics of gene transcription by RNA polymerase 2.

Keywords :
transcription dynamics, fluorescence microscopy, RNA polymerase 2.

Mots clés :
Dynamiques de transcription, microscopie à fluorescence, ARN polymérase 2

Work carried out at :
Functional Imaging of Transcription
Institut de Biologie de l'Ecole Normale Supérieure (IBENS)
CNRS UMR 8197
46 rue d'Ulm 75230 Paris cedex 05

Acknowledgements

I would like to warmly thank my mother for supporting me and encouraging me ever since I got interested in Science. I also wish to thank my father, for buying me my first microscope and my first biology kit, when I was 8. I thank them both for accepting to subscribe me to junior scientific oriented magazines although they were impossible to find in South East Asia, and for allowing me to set up my first laboratory in a room in Cambodia in 1995.

I would also like to thank Natasia Causse, François Courtin, Anaïs Monge for helping me out and supporting me during these last three years.

I would especially like to thank Florine Michelin for her unyielding support.

Of course I thank my thesis supervisor, Xavier Darzacq, for accepting me into his newly created lab, and giving me both guidance and autonomy in my studies. I also thank the entire « Functional Imaging of Transcription » team for their friendship and their help during all this time.

Felix qui potuit rerum cognoscere causas.

To Toccata.

Abstract:

The precise spatial and temporal control of gene expression is an essential aspect of life. Every living organism, wether bacterial or eukaryotic, unicellular or pluricellular, requires that the genes encoded by its DNA be transcribed at the right time and according to the requirements imposed by its environment. Failure to do so can lead to dangerous situations, ranging from cell death to tumourigenesis, and including aberrant response, such as diabetes. Transcription, the first step of gene expression, has been studied in a great variety of ways, and the scientific community has accumulated vast knowledge on the particularly complex mechanisms involved. The controls and dynamics of these processes have both been studied *in vitro* and *in vivo*; however, many questions remain, notably when it comes to control of transcriptional activation. This document recapitulates my work during these last 4 years in the field of transcriptional dynamics. I have opted for an approach based almost exclusively on advanced fluorescence microscopy techniques. The reason for this choice is that microscopy gives us a very powerful temporal resolution of cellular processes. It is now the preferred way for the investigation of *in vivo* dynamics. The use of microscopy also allows the experimentator to study cells from a same sample individually, thereby avoiding so called « averaging bias » inherent to other sample based methods. This work describes the methodologies that have been developed during these past 4 years, the results these new techniques have yielded, and the questions and prospects that they in turn lead to.

Résumé:

Un des aspects essentiels du vivant est la capacité de contrôler spatialement et temporellement l'expression génétique, et ce d'une façon très précise. La portion exprimée du génome de chaque être vivant varie continuellement en fonction du temps, elle est en partie dictée par l'environnement de la cellule, qu'elle soit d'organisme unicellulaire, pluricellulaire, prokaryote ou eukaryote. Un défaut dans le contrôle de la transcription peut avoir des conséquences particulièrement dangereuses, allant de la mort cellulaire à la prolifération non contrôlée, en passant par les cas de réponses inadaptées à un signal environnant, par exemple certains diabètes. La transcription est la première étape de l'expression génétique et à ce titre ce processus a été longuement étudié par de très nombreuses approches. Notamment, la dynamique et le contrôle de la transcription a été l'objet de nombreuses recherches *in vitro* et *in vivo*. Cependant, de très nombreuses questions demeurent, en particulier en ce qui concerne le contrôle dynamique de l'activation de la transcription. Cette thèse récapitule mes 4 dernières années de travail dans ce domaine. J'ai choisi de mener une étude de la transcription axée quasi-exclusivement sur des techniques de pointe en microscopie à fluorescence. En effet, la microscopie offre la possibilité d'étudier *in vivo* un phénomène avec une résolution temporelle inaccessible par d'autres moyens. Par ailleurs, la microscopie permet également de franchir la barrière du « biais de moyenne » qui existe lorsque l'on étudie une population plutôt que des individus. La microscopie permet d'étudier chaque cellule d'un même échantillon de façon indépendante, permettant ainsi d'étudier les disparités au sein même d'un éhantillon, et notamment d'observer des phénomènes exceptionnels qui seraient passés inaperçus avec des methodes biochimiques. Cette thèse decrit les methodologies qui ont été développées durant ces 4 dernières années, les résultats que ces nouvelles techniques nous ont permis d'obtenir, et discutera des questions et des perspectives soulevées par ces résultats.

Table of Contents

Chapter 1 Introduction 7
Gene expression and RNA Polymerase 2. 7
Controling gene expression: the role of transcription factors. 14
Mechanisms of gene repression. 21
Transcription factor regulation. 24
Transcriptional dynamics. 34
The use of fluorescence microscopy. 48
Chapter 2 Results 50
Dynamic changes in RNA Polymerase 2 processivity during transcriptional activation. 50
Discussion 63
Chapter 3 Other publications 66
mRNA degradation by Xrn1 stimulates transcription. 66
Short exposure to the DNA intercalator DRAQ5 dislocates the transcription machinery and induces cell death. 75
The In Vivo Kinetics of RNA Polymerase II Elongation During Co-Transcriptional Splicing. 85
Imaging transcription in living cells. 93
Mouaikel-Libri 112
Chapter 4 Methodology Improvements 138
Live Cell Imaging. 138
FISH. 143
Hyperresolution work in progress. 145
Chapter 5 New types of data require novel analysis methodologies. 150
Chapter 6 Materials and methods. 155
Cell Culture. 155
Transfection. 155
Generation of derived cell-lines. 157
Table 1 : Generated cell lines. 157
Luciferase. 158
FISH. 158
Chemical coupling of FISH probes. 159
Immunofluorescence. 161
Fixed cell imaging. 162
Live cell imaging. 162
Chapter 7 Bibliography 163

Chapter 1 Introduction

Gene expression and RNA Polymerase 2.

Living cells all possess DNA molecules which contain a certain number of genes. These genes are sequences of DNA which encode the necessary information for a cell to produce proteins at the proper time. The expression of these genes requires the combination of multiple processes, usually grouped into two cell functions called transcription and translation. During transcription, the gene sequence is copied into molecules of messenger RNA (mRNA), which then serve as templates for the production of proteins during translation. This thesis explores the dynamics of transcription, the first step of gene expression. Transcription in itself is already a phenomenon composed of a great number of processes, and its actors are far too numerous to be all presented in my introduction. In this first chapter, I will therefore emphasize on a few of them, which I have found to be the most relevant when it comes to transcriptional dynamics. I will leave aside most of the general transcription complexes, especially those involved in initiation and chromatin remodeling, such as the mediator complex, the TBP and the TAF proteins, the SAGA and FACT complexes. I will emphasize more on the actors

controlling transcription of specific genes and the initiation to elongation transition. In particular, I will provide a detailed review of the dynamics and regulations processes of these actors, which I have found to be involved in a phenomenon I have observed during the activation of a gene array. Phenomenon which I will describe in a later chapter.

Transcription is catalyzed by RNA Polymerases. These enzymes are capable of copying one strand from a double stranded RNA molecule, into a single stranded RNA. It does so by allowing ribonucleotide triphosphates (rNTP) to align with the template DNA strand by Watson and Crick pairing. When the rNTP is aligned in the catalytic center of the enzyme, polymerization is catalyzed, adding the ribonucleotide to the nascent RNA, the energy necessary for this operation being provided by the triphosphate moiety of the rNTP. The polymerase may then move on to the next base to copy, and start over. In eukaryotes, protein encoding genes are transcribed exclusively by RNA Polymerase 2 (RNAP2). This enzyme is a 550kDa multiprotein complex of 12 subunits, the largest of which is Rpb1. This particular subunit carries the catalytic center of the complex and is of particular interest as its carboxy terminal domain (CTD) is a tandem repeat of the heptamer Tyr-Ser-Pro-Thr-Ser-Pro-Ser serving as a regulation hub where phosphorylations of the serines in position 2, 5 and 7 as well as several other modifications serve as a regulation code. There are 52 repeats of the heptamer in the human Rpb1. This

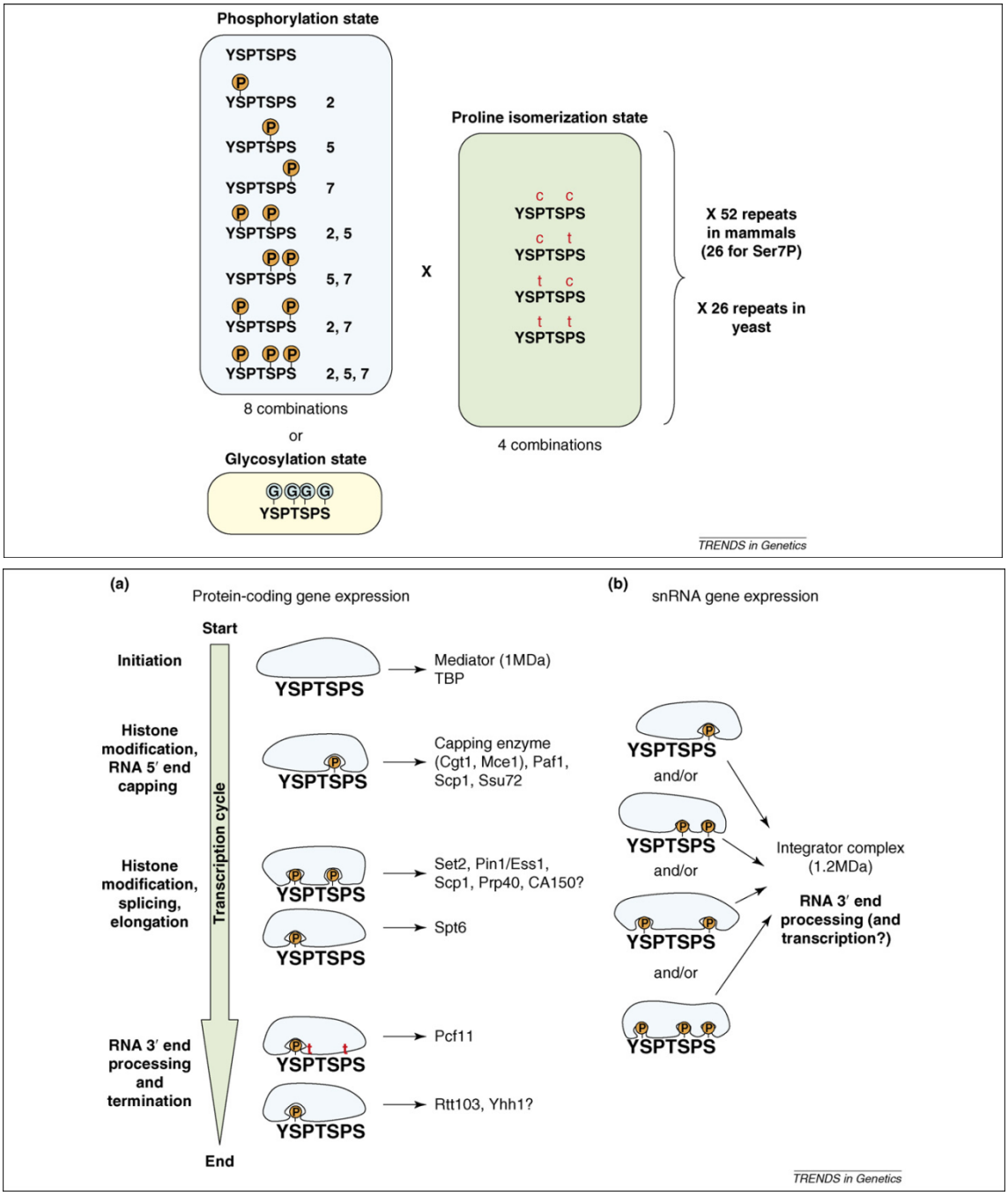


Figure 1: **Post translational modifications of the carboxy terminal domain of RNA Polymerase 2.** Upper panel: the list of all the possible post translational modifications. Each heptamer may present one combination of modifications out of the 32 possible combinations. Studies suggest that the recognized motif is however not the heptamer, but the pair of heptamer, increasing the number of combinations to 64. Furthermore, combinations might have different effects depending on whether they are presented by a proximal or a distal heptamer. Lower panel: phosphorylation patterns of the CTD may recruit different components necessary for co-transcriptional processes such as splicing or chromatin decondensation. (From Egloff and Murphy, 2008)

CTD is subject to various reversible post translational modifications and acts as a sequential anchoring point for various proteins necessary for the completion and correct maturation of mRNA, such as capping or splicing enzymes (Egloff & Murphy 2008; David L Bentley 2005; McCracken, Fong, et al. 1997). It is interesting to note that the modifications of the CTD vary as RNA Polymerase 2 transcribes the gene. Therefore, the modifications on the CTD may serve to indicate at what stage RNA Polymerase 2 is.

RNA Polymerase 2 is recruited to the promoter of genes by the Pre-Initiation Complex (PIC), composed usually of 6 general transcription factors: TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. The PIC's role is to position RNA Polymerase 2 at its proper starting point and prepare it for transcription (Kornberg 2007). As RNA Polymerase 2 is recruited, TFIIH, phosphorylates the CTD at serine 5, signaling the initiation of transcription and the recruitment by the CTD of capping enzymes (David L Bentley 2005). After transcribing a few bases, RNA Polymerase 2 enters promoter proximal pausing due to the association of the DRB sensitivity inducing factor (DSIF) and negative elongation factor (NELF). DSIF is a complex composed of the Spt4 and Spt5 proteins (Kaplan et al. 2000; Winston 2001). It was first discovered because it is essential for the sensitivity of transcribing units to an ATP analog called 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), hence its

name. When DSIF is present, transcribing RNA Polymerases 2 stall upon addition of DRB, whereas they do not in the absence of DSIF. However, it seems that DSIF is essentially a positive transcriptional factor (Morillon et al. 2003; Rondón et al. 2003). It has become apparent that in fact, NELF is the real negative elongation factor. NELF is composed of 5 subunits: NELF-A, NELF-B, NELF-C, NELF-D, NELF-E. It blocks transcription of RNA Polymerase 2 near the promoter, perhaps to ensure elongation does not occur without the 5' capping of nascent transcripts and that the CTD of RNA Polymerase 2 is properly modified to recruit co transcriptional factors.

In vivo, it seems that DSIF binds directly to the initiated RNA Polymerase 2. NELF then binds to the polymerase and DSIF complex, thus inhibiting elongation of nascent transcripts (Y Yamaguchi et al. 1999; Yuki Yamaguchi et al. 2002). NELF, it seems, does not bind to RNA Polymerase 2 or DSIF individually. Pausing caused by NELF lasts as long as NELF itself is not phosphorylated by the positive transcription elongation factor b (P-TEFb). P-TEFb is a complex composed of the CDK9 kinase and a Cyclin. In Human beings, this Cyclin can be either Cyclin T1 or Cyclin T2 (Fu et al. 1999; J Peng, Marshall, et al. 1998; J Peng, Y. Zhu, et al. 1998). The P-TEFb complex phosphorylates different substrates via its CDK9 component: its major substrate seems to be the second serines on the CTD of RNA Polymerase 2 (Isel & Karn 1999). A modification which will allow subsequent

recruitment of co-transcriptional factors by the CTD (Egloff & Murphy 2008). P-TEFb also phosphorylates NELF and DSIF. The phosphorylation of NELF detaches it from the transcribing RNA Polymerase 2, releasing it into productive elongation (Fujinaga et al. 2004). On the other hand, when the Spt5 subunit of DSIF is phosphorylated, DSIF remains with RNA Polymerase 2, and becomes a positive elongation factor (D.-K. Kim et al. 2003; J. B. Kim & P. A. Sharp 2001; Yamada et al. 2006; Ivanov et al. 2000).

P-TEFb appears as an essential factor for transcription. In fact, it seems that the vast majority of class II genes in human beings requires P-TEFb to be transcribed. It is therefore not surprising that this complex is strongly regulated inside the cell. It has been shown for example that phosphorylation of threonine 189 on Cdk9 is necessary for full enzymatic activity of P-TEFb (R. Chen et al. 2004; Q. Li et al. 2005). Also, the fact that Cdk9 has different Cyclin partners may imply a regulation of P-TEFb by these Cyclins. But the strongest evidence for a regulation is that P-TEFb can reversibly bind both a protein called Hexim and a small nuclear RNA (snRNA) called 7SK (V T Nguyen et al. 2001; Annemieke A Michels et al. 2004; Z Yang et al. 2001; Yik et al. 2003). This larger complex of P-TEFb is composed of two P-TEFb factors, bound to a dimer of Hexim proteins and a single 7SK snRNA (Q. Li et al. 2005; Blazek et al. 2005; Egloff et al. 2006; Dulac et al. 2005). When P-TEFb is assembled in this larger complex, it is unable

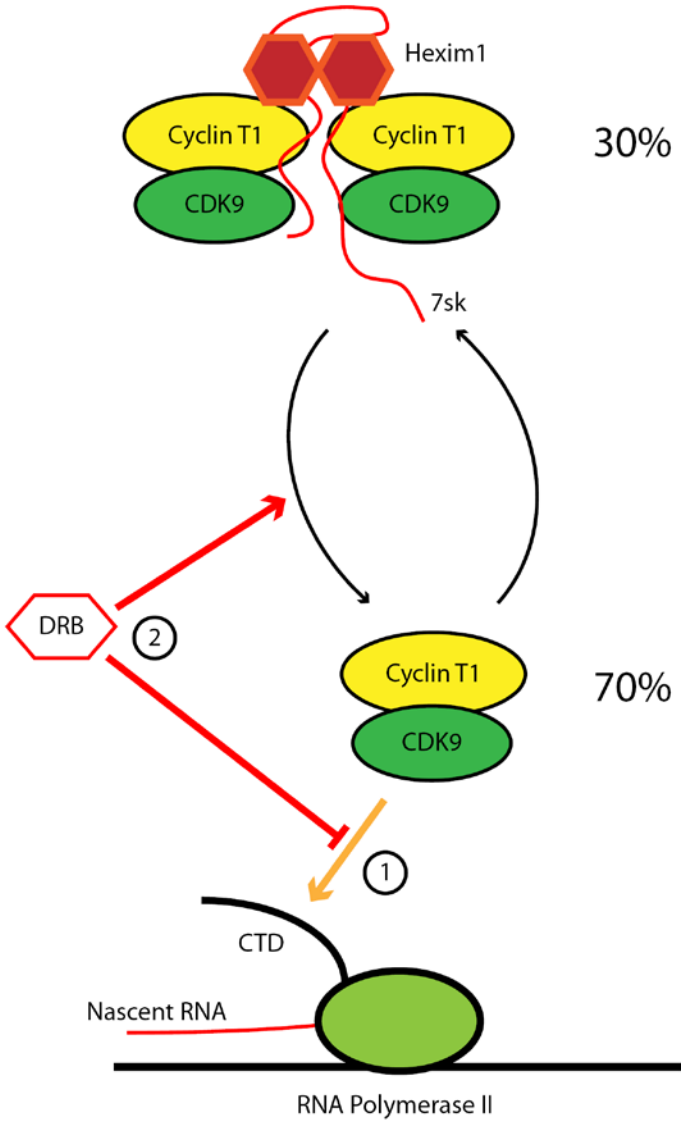


Figure 2: **P-TEFb regulation in mammalian cells.** P-TEFb is an essential general transcription factor required for the transcription of most class II genes. Correct timing of transcription in activatable genes is often carried out by controlling P-TEFb recruitment. It is present in a dynamic equilibrium of active P-TEFb and a large complex of inactive P-TEFb (percentages give the relative abundance of both complexes in a normal situation). The smaller complex of P-TEFb is the only complex capable of phosphorylating DSIF, NELF and the Serine 2 of the RNA Polymerase 2 CTD (1). Inhibiting the action of P-TEFb with DRB will unbalance the equilibrium in favor of the small complex. High levels of RNA will modify the equilibrium in favor of the large inactive complex, perhaps in a general feedback system that maintains steady RNA levels.

to phosphorylate its substrates. Very interestingly, this larger complex is predominant in many cell lines, but inhibition of transcription destabilizes the equilibrium in favour of the enzymatically active P-TEFb. This seems to indicate that this large complex is a form of regulation of transcription, and at the same time a pool of readily available P-TEFb, in case of need (V T Nguyen et al. 2001; Z Yang et al. 2001).

Controling gene expression: the role of transcription factors.

The protein content of a cell depends on the tissue it belongs to, as well as environmental conditions. Gene expression is therefore tightly regulated along the life of a cell, and at a given time, only a certain pool of the genes of a cell will be expressed. At the transcriptional level, this control is carried out by a vast number of proteins called transcription factors. Some of these factors are required for the efficient transcription of all genes, and are therefore termed general transcription factors. Others will only activate a subset of genes; these are called specific transcription factors. Some still will only be active in certain conditions, for example in the presence of a specific molecule or during certain types of stress. In human beings, more than 5% of the genes encode transcription factors; this is a very large proportion, considering that human beings are composed of a great

variety of cell types, each with its own palette of proteins (Tupler et al. 2001). This underscores the importance of a precisely controlled gene expression, and consequently the importance of transcription factors.

Specific transcription factors are usually « modular » in the sense that the binding of the transcription factor to the DNA and the activation of transcription are carried out by different regions of the protein. Transcription factors must be able to stimulate transcription of their target genes and their target genes only. They are therefore required to be able to bind specific loci. This capability is conferred by the DNA binding domain. As its name implies, the DNA binding domain is the portion of the transcription factor which is capable to bind specific DNA sequences. The structure of many of these DNA binding domains have been solved and have provided great insight into how these proteins can bind DNA in a sequence specific manner. This is usually achieved by lodging an α-helix into the major groove of the double stranded DNA. In fact, all major DNA binding domains with a known characteristic structure (both zinc-fingers, helix-turn-helix domains, leucine zipper, basic domains) were found to interact with their recognition sequence inside the major groove (Glover & Harrison 1995; M. S. Lee et al. 1989; Nardelli et al. 1991; Treisman et al. 1992). Although this interaction between the α-helix and the DNA defines the sequence specificity of the transcription factor, other elements are essential to its binding. Notably, other structures such as leucine zipper or zinc

1

2

Motif	Structure	Examples	Comments
Homeobox	Helix-turn-helix	Homeotic genes	Related to bacteriophage proteins
POU	Helix-turn-helix and Adjacent helical region	Pit-1, Oct-1 and 2, Unc86	Related to homeodomain
Paired	Helix-turn-helix	Pax, Paired	Often found in homeobox containing factors
Cys-His zinc finger	Multiple zinc fingers	Sp1, Kruppel	May form β-sheet and adjacent α-helix
Cys-cys zinc finger	Single pair of zinc fingers	Nuclear receptors	Related to E1A, Gal4
Basic domain	α-helix	c-fos, c-jun, MyoD	Associated with a leucine zipper dimerization motif or Helix-loop-helix motif
Winged helix -turn-helix	Helix-turn-helix	Fork head, c -ets, c-erg	Binds purine -rich sequences

3

Structure	Examples	Comments
Glutamine-rich	Sp1, Oct-1 and 2	Activates only from proximal domains.
Proline-rich	CTF, Oct-2, c-Jun, AP2	Weak activation from distal domains.
Acidic	GCN4, VP16, Gal4, p53, E2F1	Interacts with TFIID, the interaction may be stabilized by other amino-acids. Results in strong activation from proximal and distal domains.

4

Functional type	Example	Comments
Type I	Sp1, CTF	Stimulated initiation. Synergizes with type II A
Type II A	HIV1-Tat	Stimulates elongation Synergizes with type I
Type II B	VP16, p53, E2F1	Stimulates both initiation and elongation

Figure 3: **Transcription factor subtypes.** 1) Transcription factors are modular proteins or complexes of proteins, where the DNA binding domain and the transactivating domain are separated. 2) Different categories of DNA binding domains. These are structural types identified through sequence alignments and crystallographic studies. (From Latchman, 2008). 3) "Structural" categories of activating domains. These correspond to general amino-acid enrichment. 4) The functional classification of activating domains originally described by Blau and colleagues. It has so far not been linked to structural categories of activating domains.

finger motifs are required for the proper interaction of the transcription factor with DNA, although these motifs themselves do not directly interact with DNA. This is usually done by stabilizing the helix that will contact the binding elements on the DNA. In the case of leucine zippers, they also contribute to the sequence specificity and regulation of DNA binding. Other elements also contribute by stabilizing the transcription factor on its recognition sequences. In the case of POU for example the POU domain contacts DNA sequences adjacent to the DNA sequences bound by the homeodomain of the same transcription factor, in a helix-turn-helix motif (Klemm et al. 1994).

It is also possible to indirectly recruit an activating domain to a DNA sequence in order to achieve transcriptional activation, and there are natural occurrences of this, such as the viral factor VP16 from the herpes virus, which binds to its target sequences via the Oct-1 protein provided by the host cell (Wysocka & Winship Herr 2003). Another example is the Tat protein, which binds to the nascent RNA of the HIV long terminal repeat in order to boost the completion of the transcript (Brady & Kashanchi 2005). The DNA binding domains of transcription factors have been extensively studied ever since the discovery of the first sequence specific DNA binding protein by Dynan and Tjian in 1983 (Dynan & Tjian 1983).

When it comes to activating domains, three major structural classes have been identified, namely the acidic domains, the proline-rich domains and the glutamine-rich domains (Triezenberg 1995). In 1998, Blau *et al*. managed to identify 3 functional classes of activating domains by fusing the activation domains of various existing transcription factors to the DNA binding domain of Gal4 (Blau et al. 1996). According to that study, transcription factors may stimulate initiation of transcription of a given gene (class I), the elongation of nascent transcripts (class IIa), or both initiation and elongation (class IIb). The function of the various structural classes has been studied but so far no clear links could be made between the structural classes and the three functional classes identified by Blau *et al*.

Direct recruitment of a stretch of acidic amino acids, prolines or glutamines to the promoter of a gene is enough to activate the gene, suggesting that it is the global amino acid enrichment *per se* of the activating domain of transcription factors that is important (Ferreira et al. 2005; Gerber et al. 1994). Corroborating this hypothesis further is the fact that point mutations of various activating domains are not enough to completely abolish transcriptional activation. There are however some hydrophobic amino acids that seem to stabilize the interaction between the activation domain and general transcription factors; for example asparagines 472, phenylalanine 479 and leucine 483 of VP16 stabilize its interaction with TFIID (Uesugi et al. 1997). The mutation in VP16 of phenylalanines 442, 473, 475 and

479 to arginines also strongly affects the function of VP16, which retains its capacity to stimulate initiation, but loses its capacity to stimulate elongation (Blau et al. 1996). There are also strategies other than the use of typical activation domain structures to stimulate transcription. One example of this is the previously mentioned HIV-1 protein Tat. Tat is also a modular protein, which binds specifically to TAR, an RNA structure present at the very 5' of the HIV LTR transcripts. It also specifically binds to P-TEFb and therefore stimulates elongation of these transcripts by recruiting P-TEFb to nascent HIV transcripts (Brady & Kashanchi 2005).

In addition to the stimulation of transcriptional initiation and/or elongation, specific transcription factors may also stimulate the decondensation of chromatin. The decondensation of chromatin is essential to transcription, as the transcriptional machinery must have a fairly good access to the DNA sequence in order to transcribe genes. However, chromatin decondensation *per se* is insufficient for transcription to occur (King & Robert E. Kingston 2001; J Wong et al. 1997). The ways by which transcriptional activators can decondense chromatin is diverse. The recruitment of RNA Polymerase 2 holoenzymes in itself does remodel chromatin, as SWI/SNF is part of the holoenzyme (Greenblatt 1997; Myer & Young 1998). SWI/SNF and NURF are multiprotein complexes that can alter chromatin structure, hydrolyzing ATP in the process (Muchardt & Yaniv 1999). These factors can be

recruited to specific loci by other factors, and modify the chromatin to a transcription permissive state. A typical example of this is the recruitment of NURF by GAGA factor to heat shock genes; these genes are therefore constantly ready for transcription, which will be activated by the binding of heat shock factor (Wilkins & J T Lis 1997).

The opening of chromatin by transcriptional activators is usually not solely due to ATP dependant chromatin remodeling factors. This remodeling is usually combined to post translational modifications of histones. In fact, many transcription factors recruit co-activators such as CBP or PCAF that can acetylate lysine residues on histones. This histone acetyl-transferase (HAT) activity is sometimes found on general or specific transcription factors themselves (Doi et al. 2006; Mizzen et al. 1996). The acetylation of histones is associated to an opening of chromatin. Furthermore, acetylation of lysine 9 on histone 3 (H3K9) blocks the methylation of that lysine, a methylation that promotes condensation of chromatin. As can be expected, this last modification is also targeted by some transcriptional activators, which recruit the JHDM2A histone demethylase (Yamane et al. 2006; Whetstine et al. 2006).

Mechanisms of gene repression.

The correct expression of genes in time not only requires that transcription be activated at the correct time, but also that it stops when no longer needed, or when the protein is present in sufficient numbers. This is just as crucial as timely activation, as some proteins are part of a dynamically balanced system which must not be disturbed as it could have dramatic effects such as cell death or cancerous proliferation. It is not surprising therefore that there are ways in a cell to repress transcription, along with ways to express them. We noted in our previous chapter that transcription requires a number of events, ranging from chromatin decondensation, formation of the preinitiation complex, initiation of RNA Polymerases, and finally elongation of transcription. Transcriptional inhibitors have been shown to target these same steps in cells.

For some of the transcriptional inhibitors found so far, the way the inhibition takes place is quite clear. One of the simplest mechanism of repression is simply to occupy the binding site of a transcriptional activator. The repressor, lacking the activator domains, then simply appears as an antagonist of the transcriptional activator. The Sp3 transcriptional repressor for example functions in this way and blocks Sp1 binding sites (L Lania et al. 1997). Another obvious way to repress transcription is simply to condense chromatin. Although decondensed chromatin on

its own is not enough to activate transcription, it is absolutely necessary for chromatin to be decondensed in order for transcription to occur. Obviously, factors which would bind to genes and cause the DNA to locally condense will block transcription. And such factors are naturally present; an example is the polycomb factor, which causes chromatin condensation by methylating lysine 9 and lysine 27 on histone 3 (H3K9) protein (Czermin et al. 2002; J. Müller et al. 2002; Ringrose et al. 2004). Other factors also deacetylate histones, destabilizing the open chromatin structure of genes, and causing them to recondense. More generally, the absence of transcription on a locus will eventually lead to condensation of chromatin. This allows efficient packing of the DNA, but is an additional block that needs to be alleviated when genes are being activated.

The compaction of chromatin can be read out through a series of post-translational modifications sometimes referred to as the « histone code », not unlike the « CTD code » of RNA Polymerase II. Histones are octamers of the proteins H2A, H2B, H3 and H4, each protein being present twice. DNA can wind twice around the octamer, forming the basic block of chromatin, the nucleosome, which compacts about 150 base pairs of DNA. The N-terminal tail of histones contain a series of residues which can bear a large palette of post-translational modifications, the most well known of these being the acetylation and methylation of lysines or arginines.

Table 1 | Chromatin modifications

Mark*	Transcriptionally relevant sites†	Transcriptional role‡
DNA methylation		
Methylated cytosine (meC)	CpG islands	Repression
Histone PTMs		
Acetylated lysine (Kac)	H3 (9, 14, 18, 56), H4 (5, 8, 13, 16), H2A, H2B	Activation
Phosphorylated serine/threonine (S/Tph)	H3 (3, 10, 28), H2A, H2B	Activation
Methylated arginine (Rme)	H3 (17, 23), H4 (3)	Activation
Methylated lysine (Kme)	H3 (4, 36, 79) H3 (9, 27), H4 (20)	Activation Repression
Ubiquitylated lysine (Kub)	H2B (123§/120¶) H2A (119¶)	Activation Repression
Sumoylated lysine (Ksu)	H2B (6/7), H2A (126)	Repression
Isomerized proline (Pisom)	H3 (30–38)	Activation/repression

*The modification on either DNA or a histone.
†Well-characterized sites with regard to genomic location for DNA methylation or residues within histones for PTMs.
‡Whether the epigenetic mark is associated with activation or repression.
§Yeast (*Saccharomyces cerevisiae*).
¶Mammals.

Table 1: **The histone code and chromatin state.** The four proteins composing histones, H2A, H2B, H3 and H4 may carry various different reversible post translational modifications. Cytosine bases on the DNA may themselves be methylated. These modifications may either permit or repress transcription. (From Berger 2007)

There are other factors however which produce the specific inhibition of genes in a way yet to be discovered. These factors apparently bind to the promoter area of their target genes and disrupt the preinitiation complex, or blocks the phosphorylation of RNA Polymerases, thereby stopping these from entering elongation. For example the already mentioned NELF protein complex, which blocks transcriptional elongation and is displaced after phosphorylation by the P-TEFb complex. Such inhibiting factors have molecular mechanisms which have yet to be elucidated.

Transcription factor regulation.

In many cases, cells need to rapidly express certain genes, for example when exposed to environmental changes or certain molecules. Often these cases require a very quick transcriptional response. However, expressing these genes continuously in the absence of stimuli is very costly to the cell. Continuous expression of such genes might also hamper other essential processes of the cell. This requires a switching mechanism by which the cell will express the required gene only when the proper stimulus is present.

It has been shown that specific transcription factors are the key to this control. There are transcription factors that are present at all times in the cell, albeit in an inactive form. It is only upon stimulation of the cell that these factors gain their activity.

The types of stimuli that the cell may encounter can be very different, ranging from physical stresses such as heat shocks to a change of energy source, and comprising defensive responses to toxic molecules such as antibiotics or heavy metals. The means by which transcription factors switch from an inactive form to an active one are just as different.

Some transcription factors require the binding of a ligand in order to become active. At this time, this is by far the most common control mechanism encountered in cells, and ligands can be anything from toxic or unwanted molecules like heavy metals or antibiotics to energy sources such as galactose in yeast, or hormones in multicellular organisms. The interaction between the ligand and the transcription factor can have varying effects from one transcription factor to the next. All imply a change of conformation, but for example the binding of Ace1 to copper ions in yeasts results in binding of Ace1 to the promoter of the metallothionein gene, and its consequent transcription (Thiele 1992). In a totally different scenario, when the

thyroid hormone receptor binds its ligand, its DNA binding capability remains unchanged. The presence or absence of the ligand doesn't change the fact that this receptor remains bound to its target sequences. What the ligand does do is modify the receptor so that instead of binding co-repressors, it binds co-activators, resulting in activation of the target gene (Brzozowski et al. 1997; Wurtz et al. 1996).

This mechanism is very much related to that of other transcription factors of the nuclear receptor family. These may bind specific molecules that have little or no impact on their DNA binding capability, but modifies the receptor so that it may bind a co-activator (Brzozowski et al. 1997; Willmann & Beato 1986). Interestingly, some of these nuclear receptors, of the steroid receptor family, apparently have another control mechanism. In the absence of their ligands, these receptors are sequestered in the cytoplasm, in complex with a chaperone protein, hsp90 (Pratt 1997). When the cell is subjected to hormone treatment, the interaction between the receptor and hsp90 is disrupted, and the receptors translocate to the nucleus where they may bind their target sequences.

The regulation of a transcription factor activity by relocating it to a different compartment than the nucleus is a mechanism shared with some other transcription factors, for example Yap1, which locates to the nucleus only when levels of oxidative stress are high (Wood et al. 2003; Wood et al. 2004). However, although

Yap1 is exported out of the nucleus when oxygen levels lower, the same does not happen to steroid hormone receptors when hormone levels are depleted. In fact, when these receptors bind to their ligands and translocate to the nucleus, they remain there even after removal of the hormones, although they are in an inactive form. It is only upon the next cell division that these return to the cytoplasm.

It seems that steroid receptors are therefore regulated at two levels, but it would be very interesting to know why these receptors are sequestered in the cytoplasm in the first place, knowing that in the situation of a living organism, exposure to steroid hormones occur on a daily basis, and that the cell cycle is usually longer than a day, meaning that in physiological conditions, the steroid receptors will basically always be in the nucleus. This sequestering of the receptors in the cytoplasm by hsp90 therefore seems unnecessary, but the cost it implies for the cells does not point towards the fact that it may just be an evolutionary relic. Interestingly, the nuclear retention of steroid hormone receptors after ligand withdrawal is also mediated by hsp90, which is normally present in both the cytoplasm and the nucleus (Tago et al. 2004).

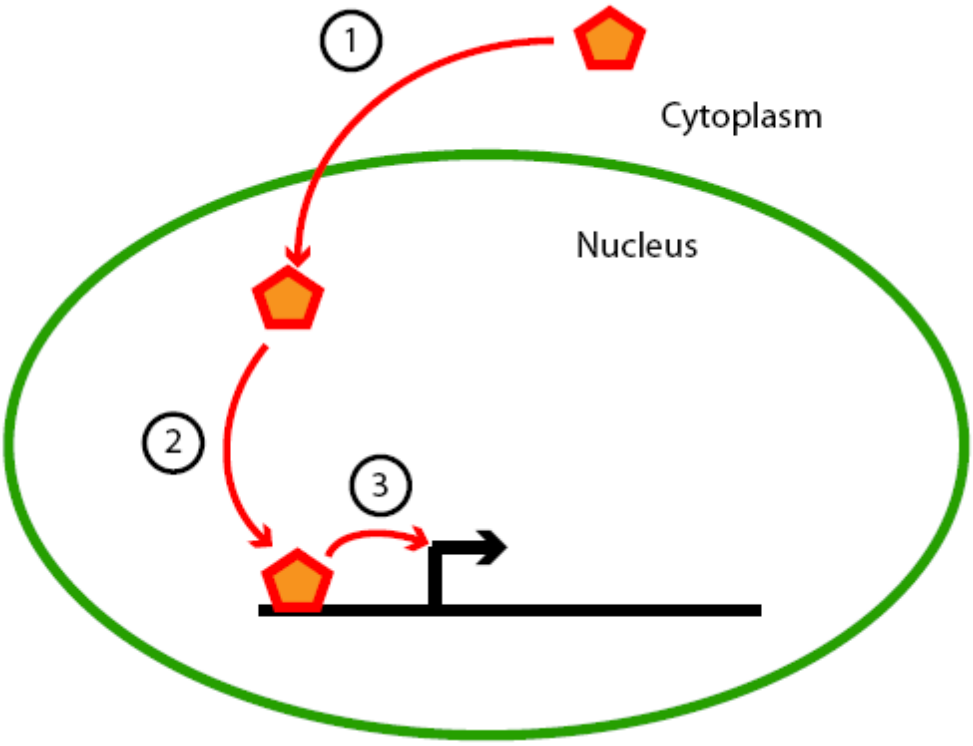


Figure 4: **Strategies of transcription factor regulation.** Transcription factors can be regulated by three main strategies: sequestering out of the nucleus (1), modification of transcription factor binding affinity (2) or modification of transcription factor transactivation potential (3). The actual method by which these regulation steps are achieved differs from one transcription factor to the next. They may imply direct binding to a ligand, binding to another cellular component or post translational modifications.

The nuclear chaperone hsp90 also has an important role in regulating another transcription factor: the Heat Shock Factor (HSF) (Zou et al. 1998). In a normal situation, HSF is also sequestered by hsp90, and is released upon heat shock, perhaps because hsp90 is required to refold other proteins that might have been denaturated by heat. The free HSF may then trimerize into its active form with the help of eEF1A and HSR1 (Kugel & Goodrich 2006; Shamovsky et al. 2006). It can then bind to Heat Shock Elements (HSE) on the DNA. The final requirement before it can stimulate transcription is a phosphorylation of serine 230 (Holmberg et al. 2001). Converting HSF from an inactive form to an active one is therefore a lengthy process, with regulations at multiple levels, but all responding to a single stimulus. Interestingly, most of these steps are actually skipped in the yeast *Saccharomyces cerevisiae*, where HSF is already in a trimer state on HSE prior to heat shock, and heat shock simply triggers its phosphorylation (Sorger et al. 1987). This seems to be a singularity as in the related yeast *Schizosaccharomyces pombe*, the heat shock regulation follows the longer pathway involving trimerization of HSF, binding to HSE and subsequent phosphorylation (Gallo et al. 1991).

Switching HSF from an inactive form to an active one does not require it's binding to a ligand, but instead involves transient interactions with a number of

other cellular components, be they proteins (eEF1A) or RNA molecules (HSR1). HSF is not the only transcription factor regulated by other cellular components. Another well known example of this is NFκB, which is also retained in the cytoplasm by an inhibitory protein called IκB (Y. Yamamoto & Richard B Gaynor 2004; A Hoffmann et al. 2006; Alexander Hoffmann & Baltimore 2006). More generally, regulation of transcription factors can be carried out not by a ligand but other proteins which can have a negative or a positive effect on the transcription factor's activity as long as they are bound to it. We have seen that this interaction may have an effect on the transactivating capabilities (as in the thyroid hormone receptor), or the binding affinity of the transcription factor (for example NFκB). These interactions may yet have another effect, this time not on binding affinity but on binding specificity. The Jun transcription factor for example may bind to different elements on the DNA. In its homodimer form, Jun binds strongly to AP1 elements, but when it is in an heterodimer form with CREB, it will lose its affinity for AP1 sites in favor of CRE sites (Jones 1990; P. Lamb & McKnight 1991).

As with RNA polymerase 2 and histones, transcription factors may be subjected to post translational modifications by various enzymes. Modifying the transcription factors in this way is yet again another way to regulate their activity, without needing a ligand or the association of another cellular component. The aforementioned CREB and HSF factors require phosphorylation in order to be

active. We have seen that in budding yeast, regulation of HSF relies exclusively on its phosphorylation, stating the importance of such post translational modifications. Likewise, phosphorylation of CREB is essential, as it allows it to associate with CBP, a co-factor necessary for transcriptional induction by CREB (Chrivia et al. 1993; Shiama 1997). Post translational modifications can not only regulate the activating potential of transcription factors, but also other regulatory steps already observed, such as DNA binding affinity, an example being the decreased binding affinity of Ets-1 as it is phosphorylated (Pufall et al. 2005), or cytoplasmic retention. This latter case is observed in the regulation of NfκB. NfκB itself is not phorphorylated, however, its inhibitor, IκB is phosphorylated when the cell is induced by phorbol ester or interleukin 1 (Y. Yamamoto & Richard B Gaynor 2004; Alexander Hoffmann & Baltimore 2006; Hayden & S. Ghosh 2004).

We have seen that phosphorylation is but one of many post translational modifications that may occur on proteins involved in transcription or transcription related processes, such as RNA Polymerase II or histones. These may also be modified through acetylation, ubiquitination and more modifications, all of which have great impact on the processes at hand. It is not surprising therefore that this is also the case in transcription factors. Acetylation of transcription factors such as p53 also have a deep impact on the function of the transcription factor. In the case of p53 for example, its acetylation on its C terminal end bestows a greater DNA

binding affinity, which in combination with phosphorylation on its N-terminus can have a major synergistic effect, as the phosphorylation of the N-terminus inhibits its binding to the inhibitory protein MDM2 (Mayo & Donner 2002; Prives 1998).

The more surprising post-translational modification of importance for transcription factors is ubiquitination. This modification consists in the addition of a 76 amino acid protein called ubiquitin, to the targeted protein. Usually, this ubiquitin tag targets the protein for degradation. It is therefore an essential component for a correct turnover of a number of proteins, including transcription factors, and RNA Polymerase 2 (Kuznetsova et al. 2003). In transcription factors however, it has another specific and unique function. In fact, it has been shown that activation of transcription by certain factors requires their ubiquitination. A study has shown that the acidic activation domain of VP16 was unable to activate transcription in yeast in which ubiquitination was deficient. However, if the VP16 activation domain was ubiquitinated artificially, transcriptional activation was rescued (S E Salghetti et al. 2001). This role for ubiquitin might be coupled with its involvement in the degradation of the transcription factor. A recent study in fact suggests that for the yeast transcription factor GCN4, the factor may bind and activate its target genes, after which the activating potential of GCN4 might be depleted. Its subsequent ubiquitination would therefore cause it to be degraded, and the binding site would be rapidly freed for the binding of a fresh GCN4 molecule

with full activating potential (Lipford et al. 2005).

The activation of preexisting transcription factor is a costly mechanism for the cell, but permits rapid activation of target genes in critical situations. In this chapter we have seen that this regulation of transcription factor activity could be carried at different levels, by different means. In effect, it requires the intervention of other molecules, which may interact stably or transiently, modifying the transcription factor in the process, through changes of conformation or post translational modifications. Although these means are quite different, they regulate transcription factor activity using the same strategies. In a previous chapter we have seen that the activity of transcription factors rely on two essential aspects: binding to DNA in a specific manner, and subsequent activation of transcription. The first aspect has an obvious prerequisite: the localization of the transcription factor in the right cellular compartment: the nucleus. It is not surprising therefore that regulation of transcription factors specifically target these crucial steps. The nuclear translocation of transcription factors for example can be impeded by binding to another extranuclear component, the binding affinity of the transcription factor may itself be changed by various means, and so can the transactivating potential of the factor. The combination of these various regulatory steps ensures a much tighter control of transcription by the cell, and also gives control over different transcription factors at the same time, by using common regulatory proteins for

different transcription factors. Notably, it may allow the cell to respond appropriately to different stimuli with the same palette of transcription factors at its disposal.

Transcriptional dynamics.

Biochemical studies have provided great insight into the roles and the interactions of various components of the cell involved in transcription. It has become clear that the dynamics of these components are of utmost importance for gene expression to function correctly. These dynamics comprise the target search dynamics of a component of the cell, *i.e.* how a general or specific transcription factor will find its target gene in the nucleus of a cell. It also involves the binding and interaction kinetics of the factor with the template DNA and other factors with which it interacts. Of course these dynamics may be quite different from a transcription factor to the next, and may also differ if the studied component is a specific or general transcription factor, a chromatin remodeler, or even a component associated with RNA polymerase 2. In this latter case, it is of interest to identify which components may affect the processivity and transcriptional kinetics of RNA Polymerase 2. Multiple studies have been carried out in the past, using both in vitro and in vivo methods, and have shed light on the dynamics of

transcription.

Historical *in vitro* experiments have shown that transcription factors, such as the aforementioned glucocorticoid receptor (GR), may bind its recognition sequence in a very stable manner (Perlmann et al. 1990). When confronted with results obtained *in vivo*, these results were contradicted. In fact, whereas Perlmann and colleagues have found a 108 minute half-life for the interaction between GR and its recognition sequence, McNally and colleagues found that in vivo, GR binds in a very transient fashion to an array of GR responsive elements (J G McNally et al. 2000). Other in vivo studies have shown similar results, where binding of a specific transcription factor, on both long and short tandem arrays, is transient (Bosisio et al. 2006; Rayasam et al. 2005; Z. D. Sharp et al. 2006; Karpova et al. 2008). Various estimates of the residence time of a transcription factor on its recognition elements show that these factors bind DNA for a few milliseconds to around 100 seconds (Hinow et al. 2006; Phair et al. 2004; Sprague et al. 2004; Farla et al. 2004). Interestingly, other factors that do not directly bind to DNA, but to DNA bound factors show the same transient binding kinetics. For some factors, this can be explained by the fact that the cofactor binds to the transcription factor and travels with it, thus having the same binding kinetics as the DNA binding factor. This is the case for example of the GRIP-1 factor, which apparently binds to GR responsive elements in complex with GR (Becker et al. 2002). On the other

hand, cofactors that bind to a transcription factor only after the binding of the said transcription factor to DNA will have slower exchange kinetics than the transcription factor itself. The exchange kinetics of the cofactor are also transient, as they transiently bind to a transcription factor which is itself transiently bound to DNA. GR also recruits such factors, for example the BRG1 protein, to its recognition elements (Johnson et al. 2008). The rapid turnover of transcription factors at the promoter of genes may be necessary for a strong transcriptional output of the gene. We have seen in a previous chapter that ubiquitination of transcription factors may in fact play a role in this sense, by marking transcription factors which have already fulfilled their role for degradation. This favors renewal of transcription factors on gene promoters, and allows rapid succession of transcriptional initiation events (Lipford et al. 2005). However, it seems that the proteasome only accounts for part of the turnover of DNA bound transcription factors. In fact, studies by FRAP in proteasome deficient samples show that in the case of GR, only 5% to 10% of the GR are immobile (Stavreva et al. 2004). A similar result was shown for NF-κB, where only 20% of the transcription factors remained on the responsive elements when the transcription factors were modified to resist the proteasome (Bosisio et al. 2006). Other processes are therefore involved in the turnover of transcription factors at gene promoters. Not very surprisingly, some of these processes involve regulators of transcription factor activity. hsp90, for example, has been shown to be essential for the turnover of GR

on tandem arrays (Stavreva et al. 2004). And an in vitro experiment has confirmed this important role of hsp90, as GR remains bound to its recognition elements in the absence of hsp90, but is essentially dissociated when hsp90 is present (Elbi et al. 2004). Other chaperones, such as p23 are also involved in the dissociation of steroid receptors from their recognition elements (Freeman & K. R. Yamamoto 2002). Finally, it appears that chromatin remodelers also impact the binding kinetics of transcription factors. The dissociation of GR from its recognition elements has been shown *in vitro* to be promoted by the activity of the Swi/Snf chromatin remodeling factor (Nagaich et al. 2004; Fletcher et al. 2002). *In vivo*, it has been shown that the Ace1 transcription factor binding kinetics is also dependent on the chromatin remodeler Rsc2, with which it interacts directly (Karpova et al. 2008). It seems therefore that transcription factors, although they have the potential to bind their recognition sequences for a long time, do not do so *in vivo*. Instead, more transient specific binding events are preferred.

Transcription factors have also been shown to bind DNA in a non specific manner. These non-specific binding events are thought to be part of an optimized strategy for finding the transcription factors' target sequence on chromatin. In effect, transcription factors essentially search for their targets by simple diffusion (Gorski et al. 2006; T Misteli 2001). Apparent diffusion kinetics were estimated by FRAP experiments on fluorescently tagged transcription factors and these have

been shown to be very fast (Phair et al. 2004; Sprague et al. 2004; Hoogstraten et al. 2002; Stenoien et al. 2001). However, these same experiments show that these factors have apparent diffusion coefficients that are much lower than what is expected for proteins of the same size (Sprague et al. 2004; Mueller et al. 2008). This slower apparent diffusion can be explained by the fact that transcription factors can interact with chromatin in a transient and non-specific manner. This interaction apparently has a rapid turnover of a few seconds (Phair et al. 2004; Gorski et al. 2006).

The privileged target search strategy of transcription factors is therefore one where transcription factors will diffuse inside the nucleus until they encounter chromatin. This interaction with chromatin is likely to be non specific, due to the scarcity of recognition elements for a given transcription factor. Therefore, the transcription factor will dissociate easily from the DNA, to once again diffuse in the nucleus. As it roams around, the transcription factor may finally bind a target element, where it will bind with a much stronger affinity, and therefore activate transcription. As we have seen in the previous chapter, the binding at recognition elements is also very transient; however, the dissociation at these elements is apparently active, whereas the dissociation of a transcription factor from a non specific DNA sequence is very likely passive (T Misteli 2001). Recent studies are in favor of this model of target search, as FRAP experiments have shown that large

fractions of transcription factors are bound to non specific DNA at any given time (Phair et al. 2004), and that single molecules of lac repressors spend the majority of their time on non specific sequences of DNA in bacteria (Elf et al. 2007). The non specific binding of transcription factors on DNA has given rise to an interesting theory in which these binding events are not just random collision events. In this theory, the non specific binding events may in fact produce active or passive sliding of the transcription factor on the DNA before the factor dissociates again. In this way, instead of simply detecting the few base pairs the factor collides with, the factor may actually scan hundreds of base pairs situated upstream and downstream of the spot it first bound to, optimizing its chances to find a specific target (Gorman & E. C. Greene 2008; Halford & Marko 2004). This theory is a direct derivate of a phenomenon that has been observed for other DNA binding proteins not involved in transcription, such as DNA repair enzymes or restriction enzymes (Gorman & E. C. Greene 2008; Halford & Marko 2004; Gowers et al. 2005). However, these were in vitro observations, and so far, there are no reports of this phenomenon happening *in vivo*.

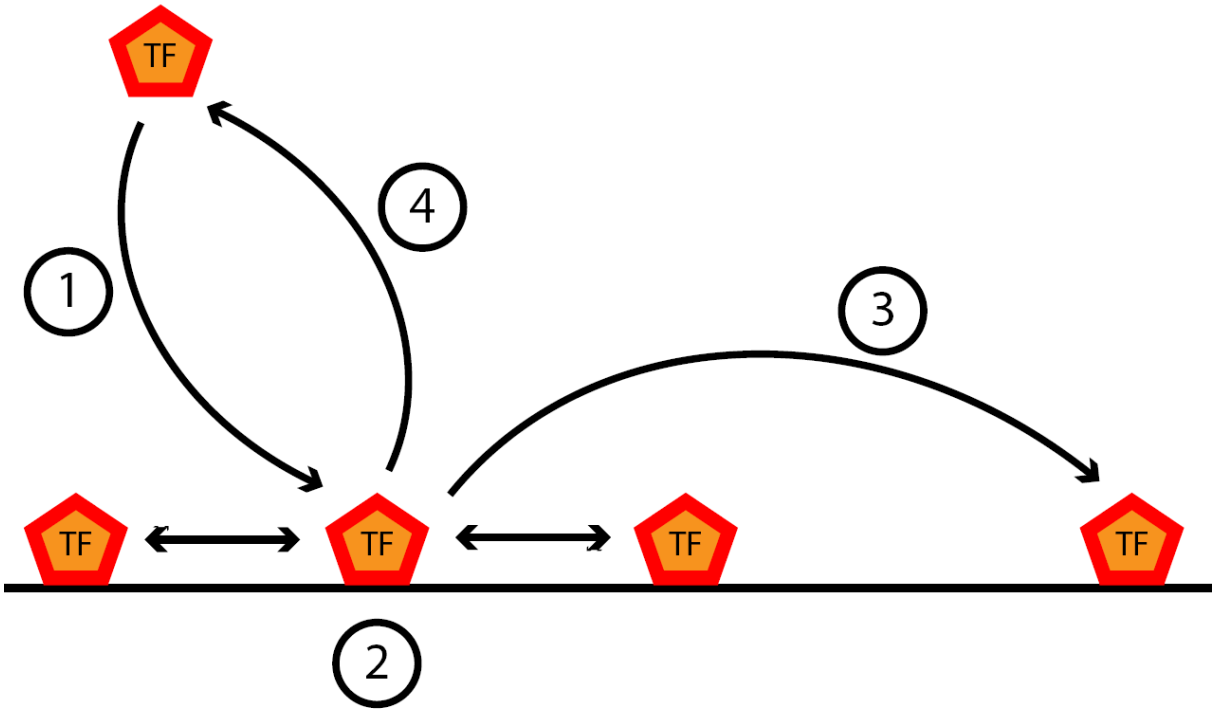


Figure 5: **Nuclear trafficking and target search strategies of transcription factors.** Transcription factors essentially diffuse in the nucleoplasm of cells. They occasionally encounter chromatin, to which they bind transiently (1). It is suggested that when bound to non-specific targets, they may have sliding motions on the chromatin (2) resulting in a local (100 bp) exploration for their targets. Alternatively, it may also explore the chromatin locally through "hops" of about 400 bp (3), or jump from a strand of DNA to another nearby strand. If no target sequences are found, the factor will dissociate passively and diffuse in the nucleoplasm (4). If it does encounter a target sequence, the transcription factor will remain bound to it and promote transcription. It will dissociate from its target mainly through the action of other proteins (After Halford and Marko, 2004).

This model of target searching by transcription factors, along with the requirements necessary for productive transcription of genes dictates that the DNA must be accessible for reading. In eukaryotic cells, DNA is tightly organized into higher order structures. This allows the DNA to be packaged into the nucleus of cells, at the expense of accessibility. One may therefore ask how a transcription factor, or any other DNA binding factor, may have access to specific DNA sequences. In order to answer these questions, the dynamics of chromatin remodeling needs to be understood. It has been long known that sites which interact with transcription factors or transcription related components are localized in nucleosome free regions, characterized by their sensitivity to DNase I and micrococcal nuclease (John et al. 2008; Wu et al. 1979). Although it was known that the displacement or the disruption of nucleosomes at these regions are common phenomena (Becker et al. 2002; Aoyagi et al. 2002; Mellor 2005), it was not clear at what level these nucleosome free regions were dynamic or static. It seems more apparent now that the remodeling of nucleosomes in the cell is in fact highly dynamic (Boeger et al. 2008). The fact that transient binding of transcription factors to chromatin is strongly mediated by chromatin remodeling factors supports a highly dynamic model. The observations of nucleosome free regions that are described in fixed cell experiments in fact do not indicate a static state, but a shift in the equilibrium of nucleosomes due to rapid transitions, themselves mediated by the constant binding of transcription factors and chromatin remodelers.

The dynamic recruitment of chromatin remodeling factors by transcription factors was recently studied by Rafalska-Metcalf and colleagues, who studied the co-recruitment of hybrid transcription factors of different activating potential and the chromatin remodeling factors GCN5, Brd2 and Brd4 (Rafalska-Metcalf et al. 2010). Their approach relied on the study of single cells with a high temporal resolution using recent live cell imaging technology. The recruitment of their transcription factor and various other proteins to a tetracycline responsive gene array could be followed by expressing proteins fused to fluorescent proteins. In their study, the transcription factor co-accumulated with GCN5, RNA Polymerase 2 and Brd4, indicating a highly dynamic binding behaviour. Brd2 however, lagged behind by 2 minutes. Although these remodelers were recruited by the transcription factor and that histones were acetylated, the presence of RNA Polymerase 2 was crucial for full chromatin decondensation. This is perhaps due to the fact that SWI/SNF is recruited by RNA Polymerase 2.

There is extensive evidence in the literature that the binding of transcription factors and the remodeling of chromatin are both very dynamic events, and to a certain extent linked to each other. Evidence also shows that RNA Polymerase 2 itself is rapidly recruited to promoters by transcription factors. It seems however that RNA Polymerase 2 recruited to a promoter is in fact extremely inefficient. In

yeast, the recruitment of TBP, TFIIB, TAF1 and Rpb1 different dynamics (Sprouse et al. 2008). This suggests that at least some of the complexes involved in transcription do not come to the promoter already preassembled. This partly explains why the formation of an initiation competent complex is inefficient. It is assumed that regulation of gene expression is essentially carried out by regulation of initiation (Juven-Gershon et al. 2008). As initiation is intrinsically inefficient, this regulation relies essentially on increasing the odds of initiation, by facilitating binding of the various transcription factors, and also increasing their residence time on the DNA. Stimuli responsive genes, which require fast transcriptional response, have another preferred step for regulation (Core & John T Lis 2008; Muse et al. 2007). In the absence of stimuli, RNA Polymerase 2 are already loaded onto the promoter, and have initiated transcription. They are however maintained in a promoter proximal paused state, and therefore no mRNA is produced. It is only upon induction by the proper stimulus that the polymerases may enter elongation. As the initiation step is already bypassed, this mechanism allows a very fast transcriptional response, all the while keeping a tight regulation.

In a recent study, Darzacq and colleagues show that on an actively transcribing gene array, 90 binding events of RNA Polymerase 2 on the promoter are required for the production of a single complete mRNA (Darzacq et al. 2007). The data obtained during their study could be fit to a three event model, which

encompassed recruitment to the promoter, initiation and elongation of transcription. On and off rates could be obtained, as well as an elongation rate and an estimated pausing time of RNA Polymerase 2 on the gene. On this gene array, under control of a VP16 activating domain, RNA Polymerase 2 was shown to have an elongation rate of 4.3kb/min. Other studies of polymerase elongation rates have produced various rates, ranging from 0.4kb/min to 4.3kb/min (Boireau et al. 2007; Kimura et al. 2002; Yao et al. 2007). The apparent discrepancy between the different elongation rates might come from the fact that FRAP experiments integrate not only elongation, but also initiation and any pausing event during transcription. In fact, by using FRAP on tagged nascent mRNA instead of RNA Polymerase 2, Darzacq and colleagues show that when pausing is taken into account, the 4.3kb/min elongation rate was reduced to a 0.4kb/min apparent elongation rate (Darzacq et al. 2007).

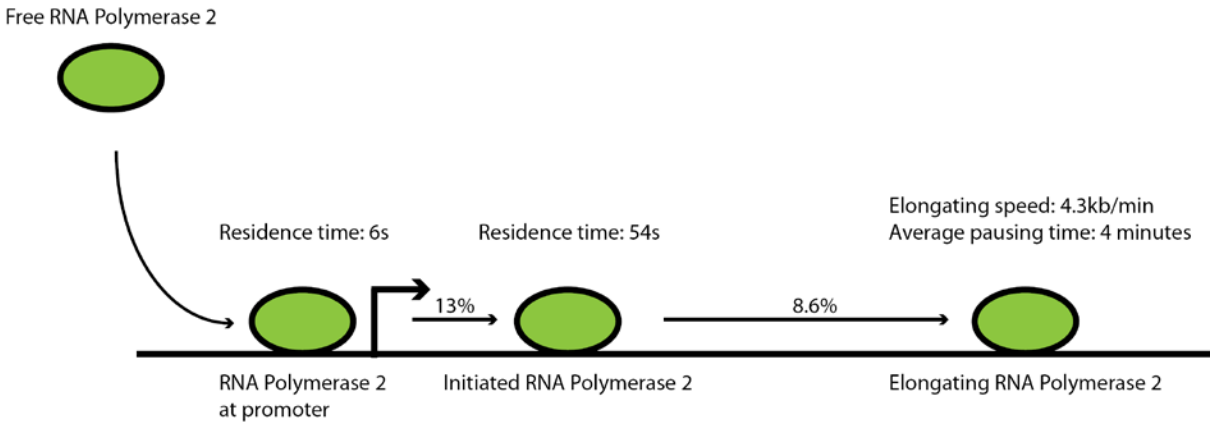


Figure 6: **Dynamics of RNA Polymerase 2 transcription.** *In vivo* dynamics of transcription on an artificial gene array as estimated by Darzacq and colleagues. This figure recapitulates some of the major findings of that study, such as the probability of a polymerase entering a specific transcription step, and the average residence time it spends in that state. In the elongation step, transcriptional kinetics and average pausing time were also estimated. (After Darzacq *et al*. 2007.)

Some of these kinetics proved to be very useful in determining the general transcriptional behavior of a specific gene along the cell cycle. Using such parameters from the literature, Zenklusen and colleagues conducted a study in which they used a markovian model to describe the genes' accessibility and the transcriptional initiation rate in yeast. By counting the number of complete mRNA of a gene and estimating the number of nascent RNAs in the nucleus using single molecule RNA Fluorescent In Situ Hybridization (RNA-FISH), they estimated which set of rates could describe their data by Monte Carlo simulation (Zenklusen et al. 2008). The model they propose is particularly interesting as it can explain both extreme transcriptional behaviours referred to as "bursts" and "constitutive" transcription, but also all the intermediates between these two extremes. In their study, they show that in yeast, most housekeeping genes have a "constitutive" mode of expression, whereas some genes such as the SAGA regulated PDR5 gene could very well be transcribed in bursts.

Gene expression requires a strong responsiveness, particularly when the cells are exposed to some forms of stress, such as a heat shock. Since transcription is the first step of gene expression, it is important that this step be controlled as accurately as possible. Knowing the dynamics at which RNA Polymerase 2 is recruited at a newly activated gene by a given transcription factor can give us valuable information on how fast a cell may start expressing a gene, and how some

deregulations in gene expression levels might occur in certain forms of diseases or cancers. The behavior and processivity of pioneering RNA polymerases 2 along a newly activated gene is also an important field of study in order to further research on the efficiency of early transcription and co-transcriptional processes. Studies of gene activation have been carried out at a single cell level, for example on the heat shock gene Hsp70, in drosophila polytene chromosomes (Yao et al. 2007; Yao et al. 2006). In this particular gene, heat shock factor (HSF1) is tightly bound on the promoter of the gene, in a singular manner compared to other transcription factors which we know bind DNA with very fast dynamics. The Hsp70 gene is one of the many genes which are poised for transcription even when heat shock is absent. It is therefore not surprising that transcription of that gene occurs rapidly after induction by heat shock. Interestingly however, the loci of heat shock responsive genes intensively recruit RNA Polymerase 2 molecules upon activation, until at some point the number of enzymes remains constant at these loci. BrdU staining confirms however that the genes are still actively transcribed at that stage, suggesting a local recycling of polymerases (Yao et al. 2007). It is not known if this is due to a reorganization of the loci into compartments, or to simple mass effect, but either case leads to the same observation that polymerases reengage into new rounds of transcription once they have ended the previous one. Other studies show a very fast activation of gene arrays that were designed in a way that they should not be poised for transcription prior to induction (Rafalska-Metcalf et al. 2010;

Janicki et al. 2004).

Following activation, it is frequently observed that transcriptional output follows oscillations. This was observed in many systems, and probably results from the complex and reversible interactions between the components necessary for transcription and the DNA template (Karpova et al. 2008; Z. Kang et al. 2002; Shang et al. 2000; Y. Liu et al. 2006; Väisänen et al. 2005; Sharma & Fondell 2002). In other systems such as the glucocorticoid receptor, this is limited to a single cycle (Becker et al. 2002; Qiu et al. 2006; John et al. 2009). The cycling of transcription demonstrates the importance of studying these processes at the single cell level. As activation and therefore cycling could be asynchronous, using a population wise study instead of a cell wise study would miss the cycling behavior, and transcription would appear as a non cycling event.

The use of fluorescence microscopy.

Recent advances in fluorescence microscopy techniques offer the opportunity to study transcription of individual genes in individual cells. The use of these techniques has provided valuable insight into the dynamics of transcription and related processes. Furthermore, whereas biochemical methods only provide

information on the situation at a certain time point, microscopy is possible in live cells, and it is therefore possible to establish techniques that will recover information from the very same cells throughout time. Live cell imaging of transcriptional activation has helped the understanding of chromatin decondensation (Janicki et al. 2004; Rafalska-Metcalf et al. 2010) and the activation dynamics of heat shock genes (Yao et al. 2007; Yao et al. 2006; Zobeck et al. 2010). Using FRAP techniques on gene arrays, various kinetics of RNA Polymerase 2 could be established (Darzacq et al. 2007). FRAP has also been the essential tool in understanding the dynamics of nuclear proteins, including all sorts of transcription factors. Fixed cell techniques such as immunostaining usually offer high sensitivity, and have also provided information on the function of various members of the splicing machinery (Spiluttini et al. 2010). Another fixed cell method, fluorescent in situ hybridization (FISH) has yielded important data essential for establishing transcriptional activity models in baker's yeast (Zenklusen et al. 2008; Gandhi et al. 2011). Latest advances in fluorescent microscopy, and notably the advent of single particle tracking methodologies and hyperresolution capable systems offer even greater opportunities in the study of transcription.

The object of my work is to study the dynamics of transcription in mammalian cells, and more precisely the dynamics of transcriptional activation, by combining various microscopy techniques in order to study a single gene array in

single cells. I will approach this study from a RNA Polymerase 2 processivity and transcript production point of view. In particular, I will study the dynamics, the processivity and the role of the very first RNA Polymerases to be loaded on the gene, the pioneering polymerases.

Chapter 2 Results

Dynamic changes in RNA Polymerase 2 processivity during transcriptional activation.

Transcription is the first step of genetic expression, and as an essential function of living cells, it is a tightly regulated phenomenon. In this study we monitored the changes in the dynamics of RNA polymerase 2 on a gene array that switches from a non transcribing state to an actively transcribing one. The purpose of such a study is to understand how the various regulatory steps are lifted as transcriptional activators bind to the promoter of a gene. The study's aim is also to provide information on the processivity of the pionneering RNA polymerases. For this purpose, we took advantage of the high degree of control the Tet-on/Tet-off system offers to the experimentator (Baron, Gossen, and Bujard 1997; Gossen and Bujard 1992). We created an artificial reporter gene array, inserted in the genome of

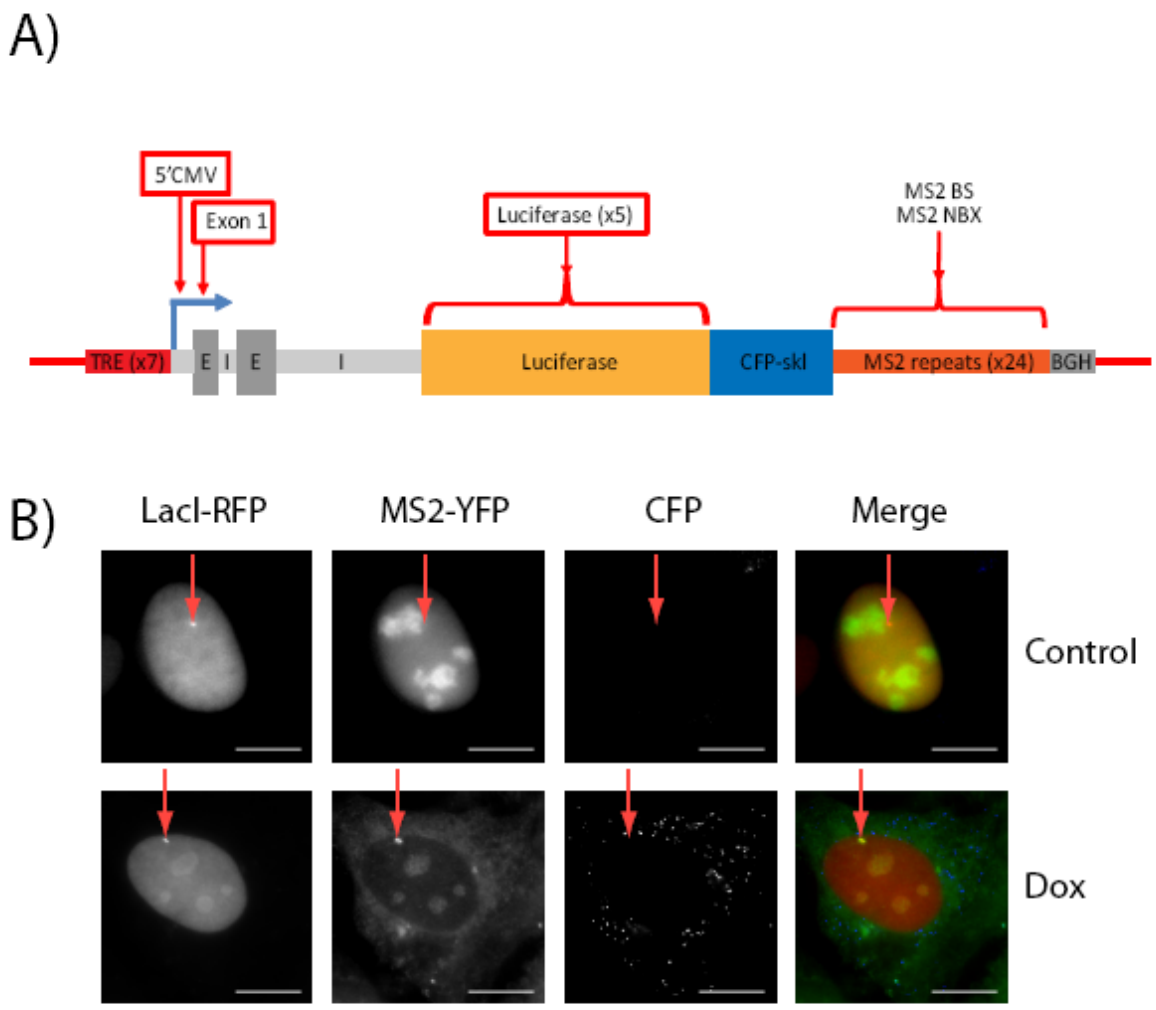


Figure 1: The reporter gene array and its signature in the 4A cell line. A) Diagram of the reporter gene inserted in the genome of U2OS cells. Red arrows indicate the regions targeted by available RNA-FISH probes, above these are the usual designations of the probes. The thickened sections of the gene are exonic regions. The thinner sections are the introns and the 5' and 3' untranslated regions. BGH stands for the Bovine Growth Hormone termination sequence. B) Images of two 4A cells, one actively transcribing after 24 hours of 2µg/mL doxycycline treatment (top row), the other inactive, before doxycycline treatment (bottom row). Red arrows show the locus of the gene array as detected using the RFP-LacI image. Scale bars represent 10µm.

human osteosarcoma cells (U2OS) [Fig 1]. This gene array was designed in such a way that the locus, its transcriptional activity and the protein product of the gene could be tracked at all time in live cells using fluorescence microscopy. Furthermore, the reporter gene was designed so that it would only respond to tetracycline-repressor based transcription factors. Its' activation or repression would therefore not interfere with the transcriptional activity of other genes, and transcription factors activating other genes would not affect our reporter gene. This high degree of control allowed us to specifically study the activation of a gene from a totally inactive (but not literally repressed) status to an actively transcribing status.

We combined this biological model with different microscopy techniques in order to study cells independently, thus avoiding averaging bias, and to identify any particular behavior during transcriptional activation that would have gone unnoticed using sample-wise techniques. We were successful in isolating one peculiar type of phenomenon, which seemingly is a transitional state between the inactive gene and the actively transcribing gene. In this state, RNA Polymerases are loaded on the gene array, but do nothing more than initiate transcription in a repetitive manner, leading to an accumulation in the cell of short incomplete transcripts. This transitional state is apparently not linked to any defect of phosphorylation of serine 2 on the CTD of Rpb1.

In order to have complete control over the transcriptional activity on our reporter gene, we decided to use an artificial gene [Fig 1]. The promoter of this gene is composed of 7 Tetracyclin operons, followed by a cytomegalovirus (CMV) minimal promoter. The open reading frame comprises a portion of the beta-globin gene (the first two exons and introns), the luciferase coding sequence, the Cyan Fluorescent Protein (CFP) coding sequence, and ends with the peroxysomal targetting signal Serine-Lysine-Leucine (SKL). In the 3' UTR, 24 sequences from the MS2 bacteriophage were inserted to allow tagging the mRNA with the MS2 coat protein (MS2cp). The total size of the reporter gene is 5.5kb. The reporter gene was inserted in a single locus of the genome of osteosarcoma (U2OS) cells, along with multiple repeats of lactose operons (LacO).

These repeats allows us to track the locus in the genome using the lactose repressor (LacI). Clonal cell lines were established, and the number of genes present in the gene array was quantified using quantitative PCR. The proper co-insertion of LacO repeats and the gene arrays was verified by colocalization of LacI and MS2cp when the gene was active. This system therefore allows us to monitor the gene array locus, its transcriptional output and the correct expression of the protein encoded. Furthermore, microscopy observations and luciferase assays confirm that the expression of the gene array is normally silent and is activated by

tet-repressor based transcription factors [Fig 1]. Amongst the multiple clonal cell-lines, we chose one (the 4A cell line) that had no background expression, had a strong luciferase response upon induction, but with a low number of copies of the reporter gene in the array (34 genes). In this cell line, newly formed accumulation of fluorescently tagged proteins of interest, Rpb1 or MS2 for example, dedicated to the transcription of the gene array are large enough to be studied using usual microscopy techniques, yet small enough so that the studied dynamics will not be biased by a mass effect.

Using video microscopy on these cell lines in which we transiently expressed pTet-on and MS2-YFP, we observed that the first MS2-YFP accumulations at the locus occurred approximately 30 minutes after induction by addition of doxycycline [Fig 2]. There were large variations from one cell to the other however, sometimes going up to 15 minutes. This is not unlike a recent report of daughter cells that resume transcription of constitutive genes in an unsynchronized manner (Ben-Ari et al. 2010). This response was far slower than usual transcriptional response to heat shock, serum treatment, heavy metal, glucocorticoid and other rtTA responsive genes (Rafalska-Metcalf et al. 2010; Yao et al. 2006; Janicki et al. 2004; Yao et al. 2007; Becker et al. 2002; Voss, John, and Hager 2006).

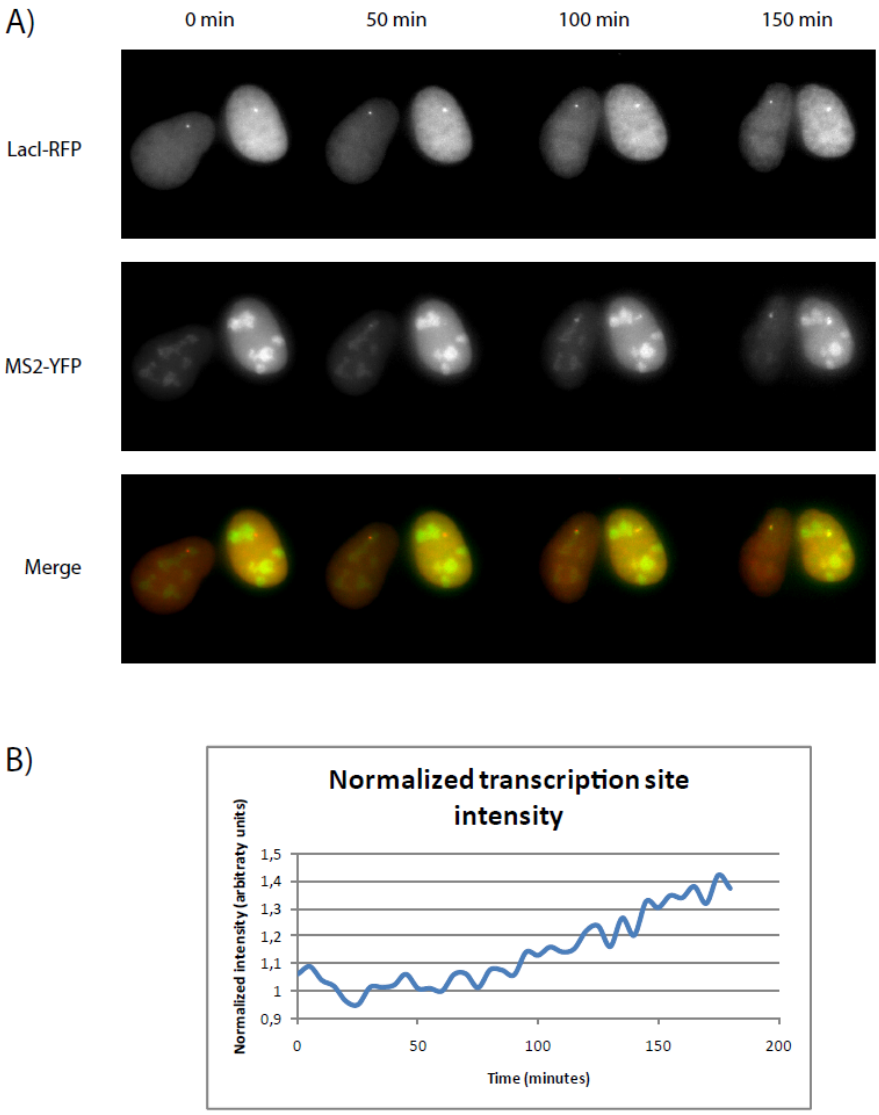


Figure 2: Transcriptional activation dynamics of the gene array. 4A cells transiently expressing Tet-on, LacI-RFP and MS2cp-YFP were imaged at 5 minutes intervals. Time above images indicates induction length with 2µg/mL doxycyclin. A) Generation of MS2cp-YFP binding transcripts at the insertion locus of the gene array. Upper row: LacI-RFP signal, the insertion locus can be clearly recognized by the local focus of LacI-RFP. Middle row: MS2cp-YFP signal, with induction time local accumulations of MS2cp-YFP appears, and colocalizes with the LacI-RFP foci. Lower row: merged images of the two upper rows. Pseudocolored red: LacI-RFP, pseudocolored green: MS2cp-YFP. B) Normalized intensity of MS2cp-YFP at the transcription site of the cell on the right in A, tracked using LacI-RFP. Cells expressed different induction dynamics, consistently with a prior study (Ben-Ari, 2011). Small scale intensity oscillations is consistent with what can be expected from z-axis drift and the used z-stack spacing (see figure n).

In fully active genes with similar promoters, RNA polymerase 2 recruitment to the genes is very fast, and RNA polymerase 2 holoenzymes have speeds of 4.3kb/min, with an average pausing time of 4 minutes (Darzacq et al. 2007). The accumulation of reverse Tetracycline repressors fused to GFP at Tetracycline responsive elements occured consistently 2 minutes after addition of doxycycline. It would therefore take approximately 8 minutes to observe accumulation of MS2-YFP at our locus, a result recently observed in another gene array under control of tetracycline responsive elements activated by the VP16 acidic activation domain (Rafalska-Metcalf et al. 2010). It is important to note two major differences in the sequence of our respective reporter genes. The first is the number of repeats in the gene arrays. The 2-6-3 cell line has approximately 200 copies of the reporter gene, whereas our cell line only has 34. Furthermore, the 2-6-3 reporter genes have promoters composed of 96 repeats of the tetracycline responsive elements. In total, the gene array in the 2-6-3 cell line has 19200 tetracycline responsive elements, against 238 in the 4A cell line. This could very possibly account for the differences in the activation dynamics of the respective gene arrays using the Tet-on system. We therefore took advantage of the slower activation dynamics of our 4A cell line to study the transcriptional dynamics and the behavior of pioneering RNA Polymerase 2 in a gene switching from an inactive to an active transcriptional status.

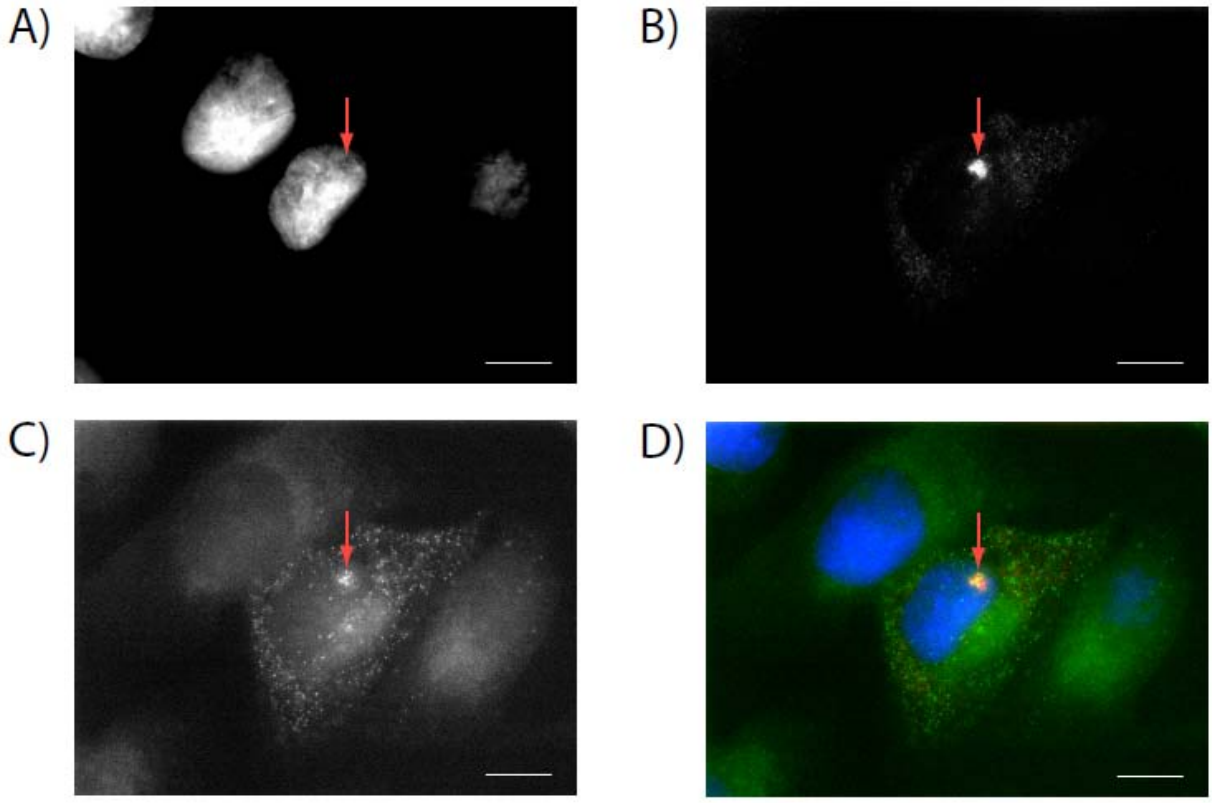
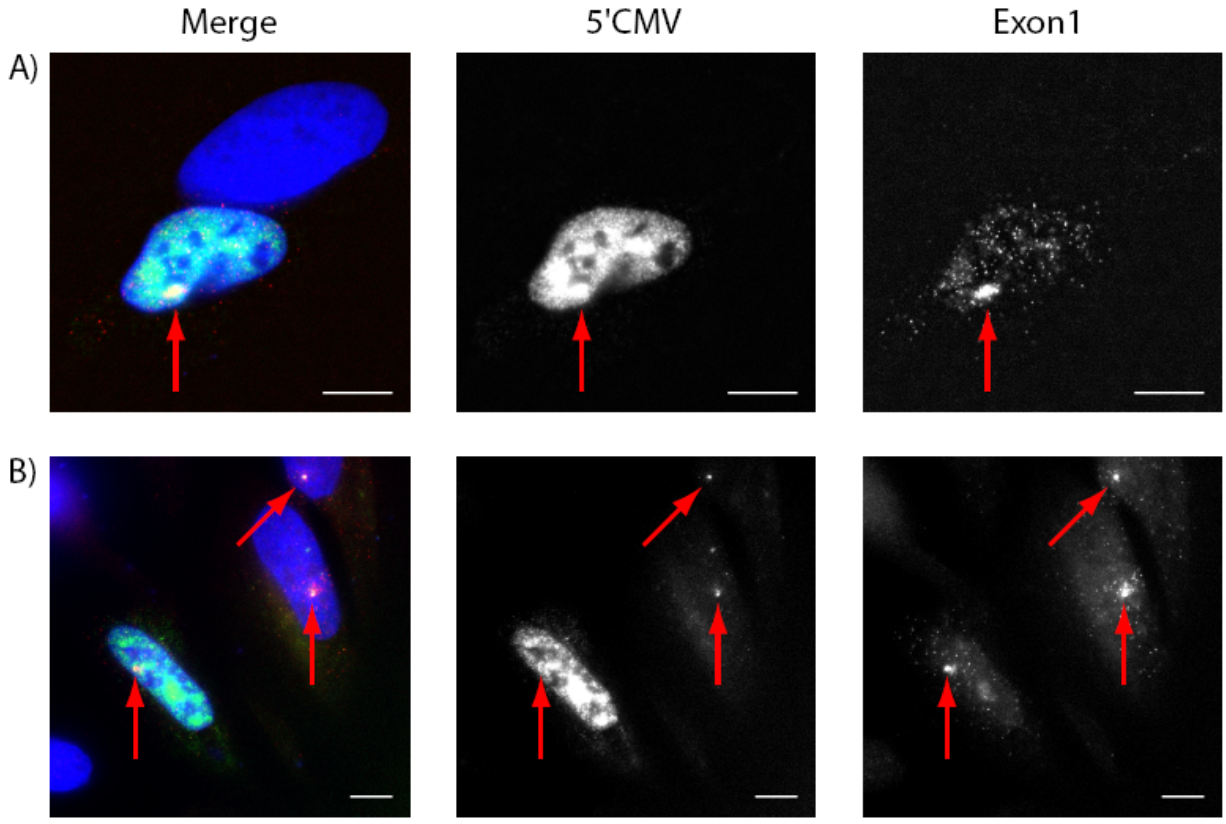


Figure 3: single probe FISH. A High stringency FISH protocol was carried out on 4A cells expressing Tet-on, induced with 2µg/mL doxycyclin for 1 hour. A) DAPI staining. B) Signal from the Exon1 probe labeled with Alexa 555. C) Signal from the 5'CMV probe labeled with Alexa 488. D) Merged image of A (pseudocolor blue), B (pseudocolored red), C (pseudocolored green). Red arrows point at the active transcription site of the lower right cell. The total absence of foci signal of either Exon1 or 5'CMV in the upper left cell indicates that it is totally inactive. Single probes of 5'CMV-Alexa488 or Exon1-Alexa555 could be identified using a 2D Gaussian fitting program, indicating that single transcripts could be identified using a single-probe high stringency FISH method.

We used Fluorescent In Situ Hybridization to verify if partial transcripts of our reporter gene are present at the locus during the first 30 minutes of induction, which would indicate the progression of RNA Polymerase 2 along the gene. Estimating the progression of RNA Polymerase 2 along the gene required the use of single 50mer DNA probes, as opposed to the combined use of multiple probes, a technique frequently used when the aim is to observe single molecules of mRNA (Zenklusen, Larson, and Singer 2008; Gandhi et al. 2011). By using high stringency in combination with bright dyes, we could detect single probes, and thus single transcripts of our reporter gene [Fig 3]. All of the probes used were capable of detecting transcripts from our reporter gene, as discrete spots could be seen both in the nucleus and cytoplasm. The active transcription site of our reporter gene could also be identified by FISH as local accumulations of our probes. Both the individual spots and the nuclear accumulations identified as transcription sites were absent when the cells did not express Tet-on or Tet-off.

A very peculiar result was observed when using the 5'CMV probe, a probe we designed to target the first 50 nucleotides of our reporter transcript. In a small proportion (9%) of our cells, we observed a very strong nuclear accumulation of signal [Fig 4]. This accumulation of signal occured regardless of the type of fluorophore we used to label the probe. Probes targeting downstream regions of the reporter gene did not accumulate in the same manner even when used in



C)

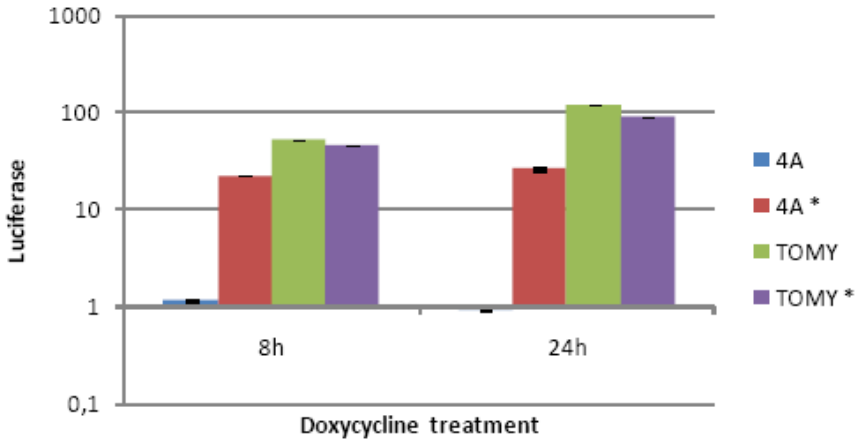
	CMV accumulation	Transcription sites	
4A	0,8	3,1	No Dox
4A TO	0,0	13,8	
4A	4,9	17,1	Dox
4A TO	0,0	35,1	

Figure 4: FISH images of the 4A cell line during activation. The DAPI image is pseudocolored in blue, the 5'CMV probe (labeled with Alexa488) targeting the first 50 bases of the transcript is pseudocolored in green, and pseudocolored red is the probe against 50 bases in the first exon of the reporter gene (labeled with Alexa555). A) A 4A cell with its nucleus saturated with the 5'CMV probe, next to an inactive cell. B) A 4A cell with a 5'CMV probe accumulation in the nucleus and cytoplasm, next to an actively transcribing cell. Scale bars represent 10 µm. Red arrows point at the transcription site, recognizable by the local accumulation of both probes. C) Percentages of cells presenting the 5'CMV probe accumulation or with visible transcription sites. The presence of a transcription site was defined by a local accumulation of both the 5'CMV probe and the Exon1 probe inside the nucleus of a cell that does not present the 5'CMV accumulation phenomenon. Doxycycline treatment corresponds to 1 hour incubation with 2µg/mL doxycycline. Two cell lines were tested: 4A (transiently expressing Tet-on), and the 4A TO, which are 4A cells stably expressing Tet-on.

combination with the 5’CMV probe and that the latter did accumulate. Probe accumulations were absent during long inductions (16h), and the proportions of cells expressing this type of accumulations were low, suggesting that this accumulation is transient. A second cell line, named 4A-To, was created from our 4A cell line, in which Tet-on was stably expressed. Although we could observe cyan fluorescence in peroxysomes in 100% of observed 4A-To cells during long inductions (16 hours), we never could observe the 5’CMV accumulation phenotype in this cell line during shorter induction times.

We performed a strong transient expression of Tet-on in the 4A-To cell line, but even in this case we could not observe the 5'CMV probe accumulation phenotype, ruling out the fact that it could be due to overexpression of Tet-on or an artifact due to the transfection procedure. We compared the luciferase activity of our 4A and 4A-To cell lines with or without induction and or transient expression of Tet-on [Fig 5]. Our assay showed that the 4A-To cell line, without doxycyclin induction, showed a higher luciferase activity than the 4A cell line. This is consistent with the previously documented leakiness of the Tet-on system. Therefore, in our 4A-To cell line, our reporter genes switch from an occasionally transcribing gene to an actively transcribing gene. Thus they might have less regulatory processes to overcome before actively transcribing, compared to a gene array with no recent transcriptional event.

A)



B)

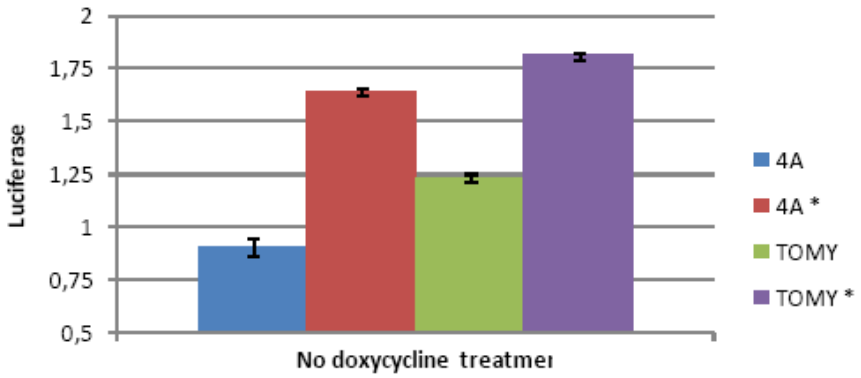
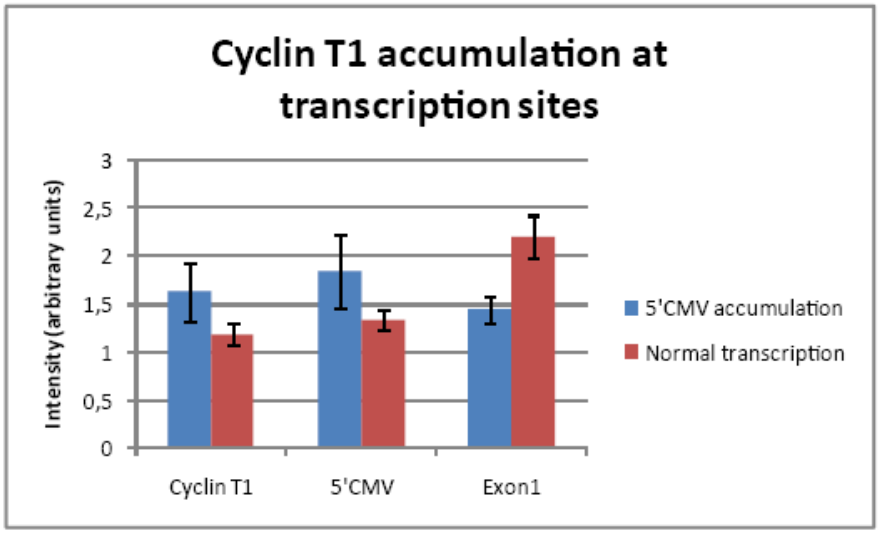


Figure 5: Luciferase assays of 4A and 4A TO cell lines. A) Luciferase activity of cell lines induced with 2µg/mL doxycycline for the indicated treatment times. B) Luciferase activity of the same samples without doxycycline induction. Values are expressed in folds of the average activity of untransfected 4A cells of all three treatment times. Asterisks indicate that Tet-on was transiently overexpressed by transfection. Error bars are standard deviations of 4 technical replicates.

A)



B)

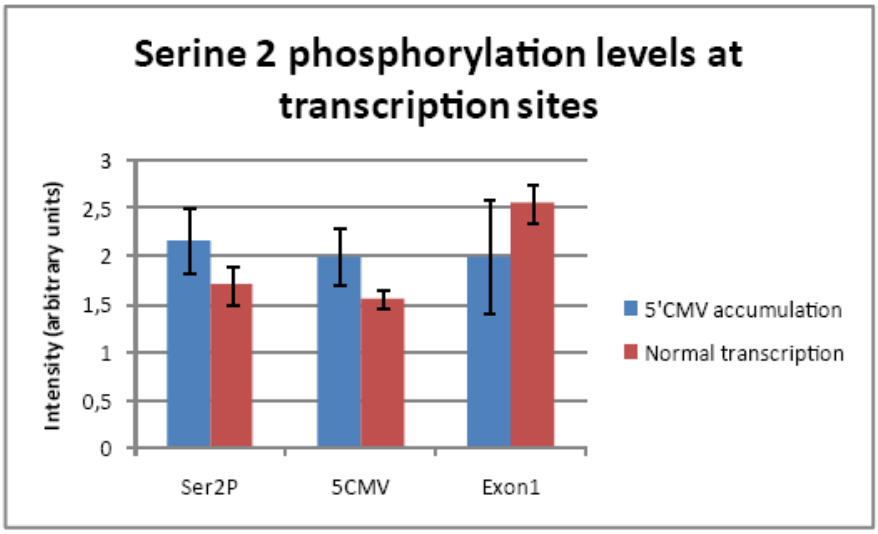


Figure 6: Immunofluorescence and FISH shows no difference in phosphorylation patterns. pTet-on was transiently expressed in 4A cells and activated with doxycycline for 15, 30 or 45 minutes. FISH was carried out, using probes against the 5'CMV and the first exon. In parallel, immunofluorescence staining against either Cyclin T1 or phosphorylation of serine 2 on the RNA Polymerase 2 CTD was carried out. Cells expressing 5'CMV probe accumulation or with transcription sites were imaged. A) Recruitment of Cyclin T1 to transcription sites of cells with or without 5'CMV accumulation. B) Serine 2 phosphorylation levels at transcription sites of cells with or without 5'CMV accumulation. Error bars represent standard errors of the mean.

We tested the phosphorylation of RNA Polymerase 2 in cells expressing 5’CMV accumulation by immunofluorescent staining of either phosphorylations of serine 2 of the Rpb1 CTD or Cyclin T1. We combined this technique with FISH using probes targeting the 5'CMV and the first exon of the beta globin gene. We selectively imaged cells expressing 5'CMV probe accumulations or with visible transcription sites (identified by colocalization of both the 5’CMV probe and the exonic probe). We analyzed the Cyclin T1 accumulation and the Serine 2 phosphorylation at the transcription site by radial analysis [Fig 6]. We observed that cells presenting 5'CMV probe accumulations did in fact have similar serine 2 phosphorylation levels compared to cells transcribing downstream regions. Cyclin T1 levels at the transcription site were on the other hand higher in cells expressing 5’CMV accumulation than in cells with expressing the gene normally. This can be explained by a much higher turnover rate of RNA Polymerases in cells with 5’CMV accumulation, requiring more P-TEFb to achieve the same global level of phosphorylation than in transcription sites transcribing downstream regions.

Discussion

Our study suggests that activation of a previously totally inactive gene, even by a strong activator, may result in a transitory accumulation of very short abortive

transcripts, which is not due to a defect in phosphorylation of serine 2 on the CTD of RNA Polymerase. Serine 2 phosphorylation is at the same levels whether there is 5'CMV accumulation or not, but Cyclin T1 accumulates more on transcription sites of cells with 5'CMV accumulation. This is consistent with a dynamic situation in which RNA Polymerases would massively start transcribing at the transcription site, but all would break off the gene. This would release shortened transcripts that would accumulate in the nucleus by saturating decay pathways. This also represents a much faster turnover of substrates for P-TEFb than in genes transcribed normally. This explains why in both situations the phosphorylation levels are equal, but why in the 5'CMV accumulating cells, Cyclin T1 is present at higher levels than in actively transcribing genes.

Gene expression is a tightly regulated set of processes at different levels, including at its first step: transcription. This study shows that a completely inactive gene can be difficult to activate, even when using a very potent activator such as VP16, which stimulates both initiation and elongation of transcription. When this activator is recruited to the promoter region of an inactive gene, cells may initially produce a very large number of shortened transcripts, and only later on produce complete transcripts. This transitional state is apparently not due to abortive initiation, as the recognized targets are larger than what has been described in the literature, and as RNA Polymerase 2 enzymes are phosphorylated on serine 2.

Furthermore, abortive initiation continues to occur even when genes are actively transcribing. The fact that only a fraction of cells express the 5'CMV accumulation phenotype and that not all cells with visible transcription sites have this phenotype indicates that it is not likely due to abortive initiation.

We have described a specific and reproducible phenomenon, but there is yet to show why these shortened transcripts appear during activation of transcription. These transcripts are not the result of abortive initiation that has been described in the literature (Core and Lis 2008)(Darzacq et al. 2007). Because the phenomenon is only observed in cases where the gene switches from an inactive state to an active state, the massive generation of shortened transcripts could be due to RNA Polymerase 2 being unable to transcribe through denser chromatin. But this does not explain why the phenomenon is only visible with the probe targeting the first 50 nucleotides of the nascent transcript.

It has been shown that very short cryptic transcripts could be produced by transcription of annotated sequences by RNA Polymerase 2 near promoter sequences, and that these transcripts may have functions (Neil et al. 2009; Proudfoot and Gullerova 2007; Xu et al. 2009). The accumulation of shortened transcripts that we have observed could be related to some sort of cryptic transcript, which may have a role during activation of transcription. It is possible that the

transcripts may not have a role *per se*, but that it is the massive repetition of initiations producing them that has a function. Combining the biophysical dynamics of the essential actors of transcription into a dynamic model of transcription may explain why the activation of a gene can be delayed in cells and why shortened transcripts may appear. This in turn can provide insight into understanding the appearance of certain phenotypes or pathologies such as cancers or viral cell cycle switching. These are the type of events that could affect an entire organism, by causing a simple dynamic deregulation of transcription in a single cell.

Chapter 3 Other publications

mRNA degradation by Xrn1 stimulates transcription.

This study stemmed from perspectives developed by Professor Mordechai Choder, from the University of Haifa. Prof. Choder's team have recently presented very strong evidence that some actors of transcription also facilitate the cytoplasmic decay of the RNA produced. They have shown that in yeast, the Rpb4 and Rpb7 heterodimer, two components of RNA Polymerase II, actively shuttle between the nucleus and the cytoplasm. Furthermore, they have shown that nuclear export of Rpb4 and Rpb7 is dependent on transcription, and that this export

facilitates the degradation of mRNAs in the cytoplasm (Goler-Baron et al. 2008; Lotan et al. 2007; Selitrennik et al. 2006). Their study thus showed that there exists at least a one-way signal from the transcription machinery that partially controls the mRNA decay machinery.

The question that arose was that there might also be a feedback message from the mRNA decay machinery to the transcriptional machinery. The implications of such a discovery would change the current vision that the control of mRNA levels in a cell is due to two uncoupled processes: transcription and mRNA decay. The study in which we participated demonstrates that the mRNA decay machinery modulates transcription to some extent, and that the very last process in the life of an mRNA stimulates the very first process in the life of an mRNA. Studies have previously shown that the Xrn1 protein, the major 3'-5' exonuclease could be inactivated in yeast without loss of viability; furthermore, microarray assays have shown that this inactivation also had no significant effect on the mRNA levels of a large number of genes (F. He et al. 2003). This is unexpected as the Xrn1 pathway is the major mRNA decay pathway in yeast. Our study has shown that in fact, yeast in which Xrn1 is mutated to an exonuclease deficient form or yeast where Xrn1 is simply deleted do have impaired mRNA degradation, and that most mRNAs have extended life-times compared to their counterparts in wild type strains. However, the mRNA levels of tested genes remained unchanged whether

Xrn1 was mutated, deleted, or present and functional.

Similar results were obtained when other members of the mRNA degradation machinery were tested. The study shows, using nuclear export deficient strains, that some components of the mRNA decay machinery shuttled between the nucleus and the cytoplasm and, by ChIP, that these components interact with genes in a fashion similar to RNA Polymerase II. Furthermore, the nuclear import of the decay factors was impaired when Xrn1 was mutated. Strangely, these decay factors did not have impaired nuclear import when Xrn1 was deleted. Finally, this study also shows that the mRNA production of genes upon activation was much slower in yeast strains where the mRNA decay machinery was deficient.

Our contribution to this study offered a better understanding of how decay factors might influence the transcriptional machinery, this time in a constitutively expressed gene. We selected the TEF4 gene as our reporter for single mRNA sensitivity Fluorescent In Situ Hybridization (FISH). The TEF4 gene is one of the few yeast cells containing an intron, allowing us to identify the transcription site. This was however not necessary, as the transcription site was identifiable by its nuclear localization alone. We could efficiently count the number of mRNAs in each cell using this technique, and quantify the intensity of each detected spot. Because TEF4 is a highly transcribed gene, we could expect that the majority of the

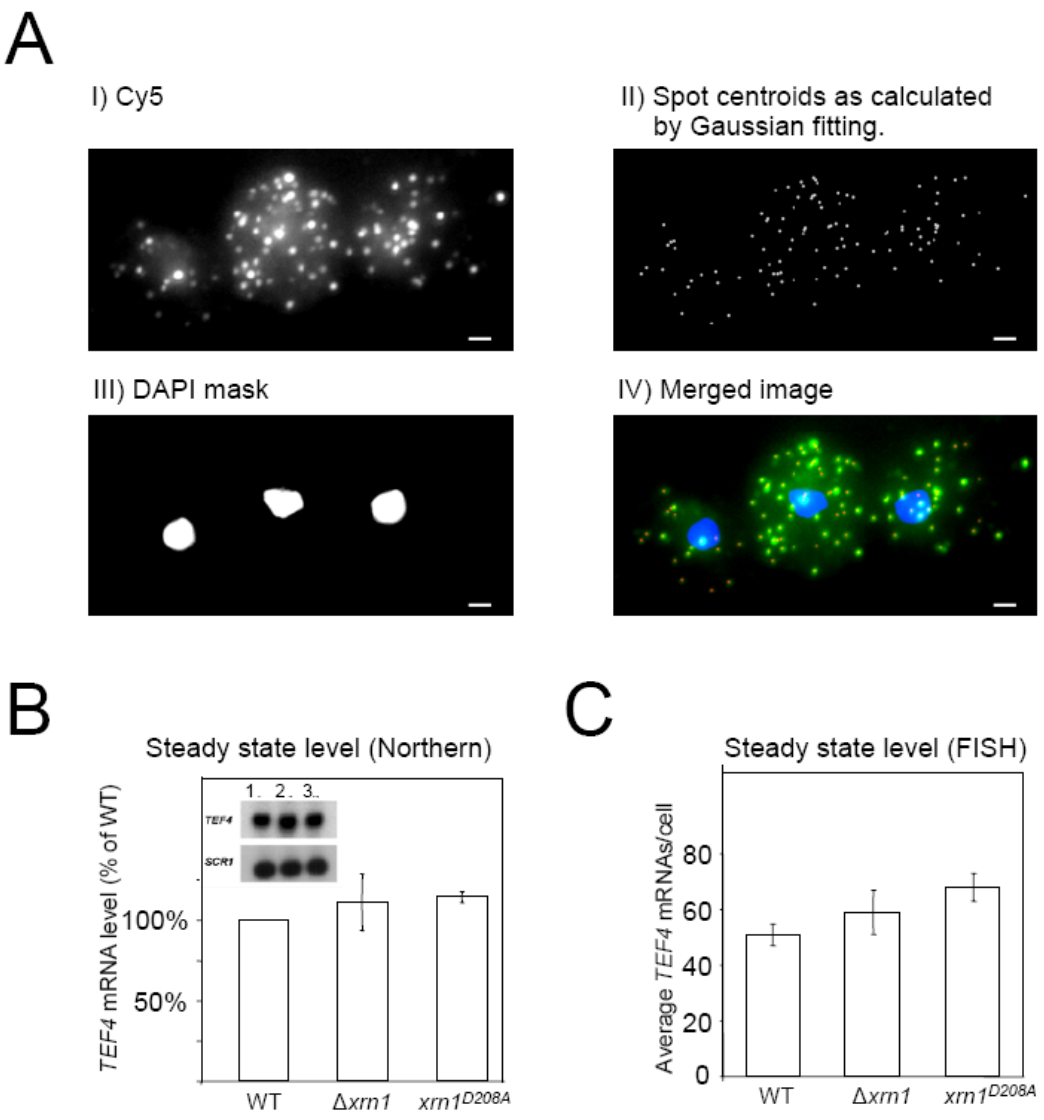


Figure: Steady state level and export of *TEF4* mRNA are not impaired in *xrn1* mutant strains. (**A**) Quantitative single molecule analysis. **(I**) Image showing the raw data of Cy5 probes, which bind to the *TEF4* mRNA. **(II**) Spot centroids. The raw image from (I) was analyzed using a 2D Gaussian fitting program, from which we obtained the coordinates of the Gaussian centroid and the integrated intensity of the corresponding spot. **(III**) Thresholded DAPI images from the cells in (I). DAPI staining was used to identify any transcription site. (**IV**) Merged image. Images (I) (pseudocolored green), (II) (pseudocolored red) and (III) (pseudocolored blue) were merged into a single image. Scale bar, 1 µm. (**B**) Northern blot quantification showing the steady state level of *TEF4* mRNA in wild type or Xrn1 impaired strains. Inset shows the Northern blot image of *TEF4* and *SCR1* RNA, as the loading control; 1 - WT, 2 – Δ*xrn1*, 3 – *xrn1*D208A. (**C**) Quantification of the average number ± SEM of *TEF4* mRNAs per cell as determined by FISH analysis. N>200 except for Δ*xrn1* (N=76).

spots detected correspond to single mRNAs. By counting these spots, we established in a manner independent from that of Prof. Choder's team that mRNA levels of the TEF4 gene remained constant, in spite of the fact that Xrn1 was mutated or deleted. We also studied the transcriptional activity of cells from different yeast strains. The three studied strains had the same genetic background, one having a functional Xrn1 gene, another having an exonuclease deficient Xrn1 and in the last strain, the Xrn1 gene was deleted.

We have developed for this study a new statistical method for comparing the transcriptional profiles of the various strains. This method pools the transcription site intensities of the various populations and redistributes them into populations of the same size as the originals. This provides us with two random profiles, from which can be calculated differences for each category of cells (not transcribing, or categorized by transcription site intensity). Repeating this step a large number of times produces a large number of differences all obtained by random resampling, from which can be extracted a probability density of observing a given difference by chance. By comparing this density to our original observed difference, we could not only say if the strains had different transcription profiles, but also where exactly the differences were.

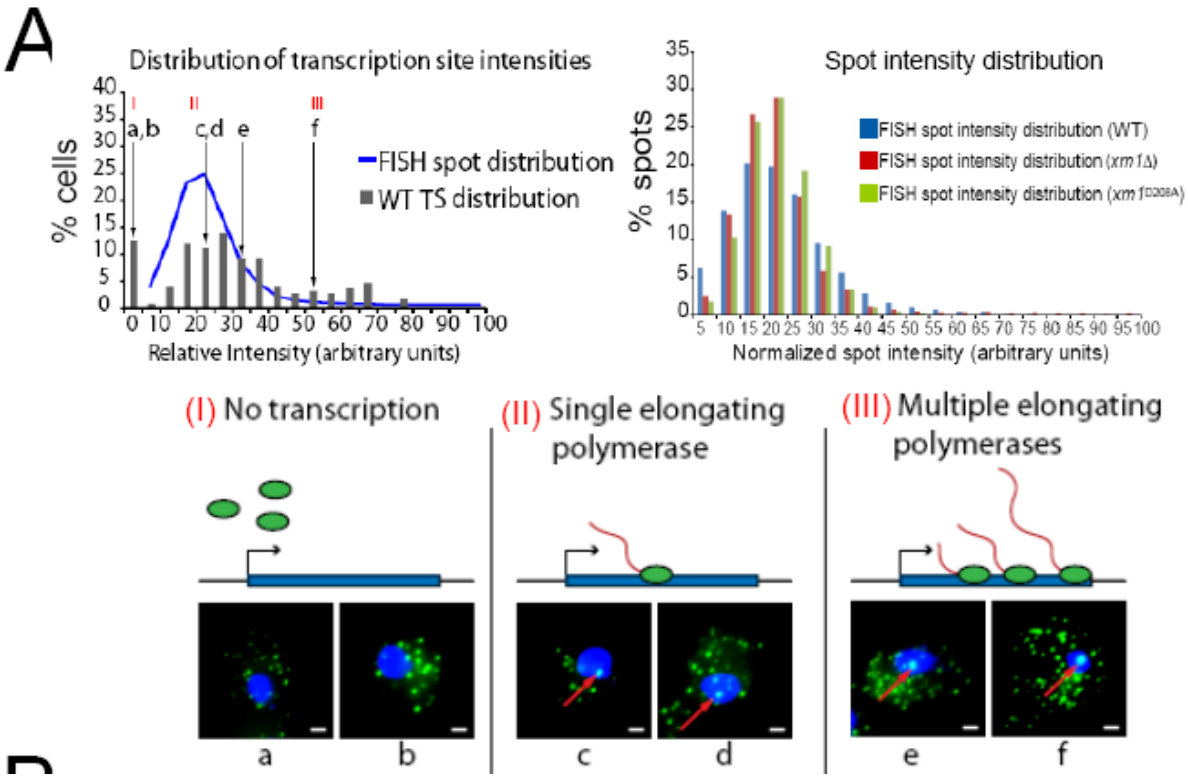


Figure: Distribution of transcription site intensities as determined by FISH analysis reveals a role for Xrn1p in transcription initiation and re-initiation. Interpretation of transcription site intensities. Upper right panel: distribution of FISH spots (most are cytoplasmic) according to their intensity. The intensity of the spots from various FISH experiments were normalized according to their median value and their distributions were plotted in percentages of total FISH spots, strain by strain. Upper left panel: example of distribution of transcription site intensity (x axis) is represented by the histogram, in percentages of WT cells (y axis). For comparison, the superimposed blue curve corresponds to the distribution of the intensities of all the FISH spots, as shown in the right panel. The first column corresponds to cells without any FISH spot in the nucleus from the same microscopy images. I, II and IIIrefer to the interpretations and cell images in the panels below. (I) Cells devoid of FISH spots in the nucleus. Cells without any FISH spots in the nucleus were interpreted as cells without any active transcription of the *TEF4* gene at the time of the fixation. Two examples of such cells are shown (a,b).(II) Cells with a low intensity FISH spot in the nucleus, which corresponds to the value of a single mRNA in the cytoplasm. This cell population is interpreted to carry a single elongating Pol II at the *TEF4* gene. Two examples of such cells are shown (c,d). In cell d, the other spots that appear inside the nucleus are either above or under it. (III) Cells with high intensity FISH spots in the nucleus. This cell population has multiple nascent *TEF4* mRNAs in the TS; hence they carry more than one elongating Pol II simultaneously at the *TEF4* gene. Two examples of such cells are shown (e,f). Arrows indicate TSs; Scale bar, 1 µm.

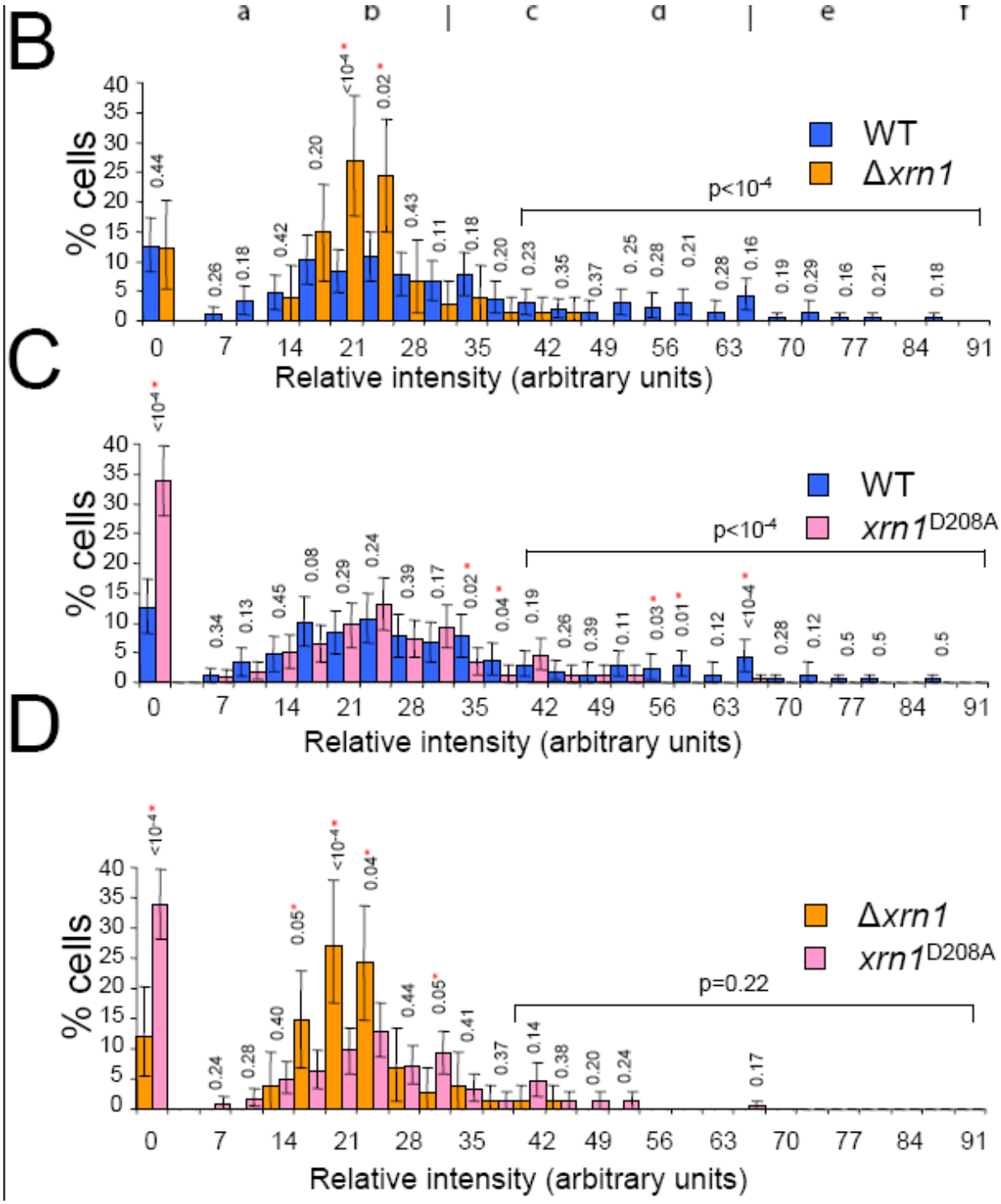


Figure: Distribution of transcription site intensities as determined by FISH analysis reveals a role for Xrn1p in transcription initiation and re-initiation. B-D) Distribution of TS intensity. Comparison between WT and Δ*xrn1* cells (B), WT and *xrn1*D208A cells (C) and Δ*xrn1* and *xrn1*D208A cells (D). Error bars represent 95% confidence intervals. Numbers above bars represent p-values of the difference between the bins of the two strains calculated by resampling. An asterisk was added when the difference is significant ($P < 0.05$). The p value was calculated for the bin 38.5-91 as well, and is shown above each histogram**.** N > 200 cells, except for Δ*xrn1* cells (N=76).

These differences would have gone undetected by classical biochemical methods due to averaging bias. We could thus add valuable insight into how decay factors could influence the transcriptional machinery, in two ways: we have shown that the presence of an exonuclease deficient Xrn1 resulted in a significantly higher proportion of cells with no active transcription of TEF4. Those cells which had an active transcription site had numbers of nascent transcripts much lower than the wild types. On the other hand, strains with no Xrn1 had about the same proportion of cells with active transcription of the TEF4 gene. However, these cells also had a lower number of nascent transcripts than the wild type strain, and very much comparable to the Xrn1 mutant strain.

All in all, this leads to two different ways in which decay factors influence the transcriptional machinery. In the case of an inactive Xrn1, it is possible that its presence retains the other decay factors out of the nucleus by binding to them and thus not allowing them to shuttle to the nucleus, a process dependent on mRNA decay. This would explain why there is a higher proportion of cells with no transcription site. On the other hand, when Xrn1 is absent, it does not hinder the nuclear translocation of the other decay factors, translocation that is perhaps facilitated by the 5’ to 3’ degradation process. In the nucleus however, the absence of Xrn1 seems to be crucial to the stimulation of transcription, as the transcriptional output is still lower than in the normal case. Again, the fact that decay factors can

shuttle freely would explain why there is a proportion of cells transcribing TEF4 similar to the wild types. And the fact that the transcriptional output is lower than the wild type strain would indicate that the Xrn1 protein is either a transcription factor as well as a decay factor, or recruits a strong activator. It also indicates that some of the other decay factors that shuttle between the cytoplasm and the nucleus are transcriptional co-activators. These hypotheses have yet to be tested, and it would be particularly interesting to find out how exactly Xrn1 and other decay factors stimulate transcription. There are obviously other questions that arise from our study, for example, are the decay factors going to activate transcription of a random gene, or are they going to specifically activate production of mRNA identical to the one they have just destroyed?

At the moment, this article was successively submitted at Science and Nature. It is now being rewritten for Cell.

Gene expression is a circular process

Gal Haimovich[1], Sebastien Causse[2], Oren Barkai[1], Xavier Darzacq[2] and Mordechai Choder[1]

[1]Department of molecular microbiology, Faculty of Medicine, Technion-Israel Institute of Technology, Bat-Galim, 31096 Haifa, Israel. [2]Functional Imaging of Transcription, Ecole Normale Supérieure, CNRS UMR 8541, 46 rue d'Ulm 75230 Paris cedex 05.

One sentence summary:

Transcription is linked to the major mRNA decay process in the cytoplasm via nucleocytoplasmic shuttling factors.

Abstract

The balance between mRNA synthesis and decay determines mRNA levels in the cell. We have recently shown that RNA polymerase II controls cytoplasmic mRNA decay. Here, we ask whether the mRNA decay machinery can, conversely, affect transcription. We show that the mRNA level in yeast is highly robust to perturbations in mRNA decay machinery. Defects in a number of cytoplasmic mRNA decay factors lead to down-regulation of transcription. Consistently, these factors shuttle between the cytoplasm and the nucleus, in a manner dependent on proper exonucleolytic degradation of mRNA. In the nucleus, the decay factors associate with chromatin and stimulate transcription initiation by RNA polymerase II. Our findings demonstrate that gene expression is a circular process in which the hitherto first and last stages are interconnected.

Cytoplasmic mRNA decay is an essential and well-controlled component of the gene expression pathway, considered to represent the end-point of this pathway. It is commonly presumed that, following the completion of mRNA degradation, the mRNA decay factors (DFs) re-enter the decay process of yet another mRNA. However, additional regulatory roles of these DFs in the gene expression pathway have never been investigated. In order to examine whether the decay process can affect transcription, we first determined the steady state level of mRNAs in mutant yeast strains lacking different mRNA DFs. The steady state level of mRNAs is determined by two opposing processes: transcription and decay. Therefore, if mRNA decay is defective, one would expect an

increase in these mRNAs. As expected, deletion of various mRNA DFs resulted in elevated levels of *EDC1* and *RPS28B* mRNAs that are degraded by non-conventional pathways (*1, 2*) (Fig. 1A and C; Fig. S1). Surprisingly, the levels of various other mRNAs did not increase (Fig. 1B and C; Fig. S1), despite their abnormally long half lives (Fig. S3 and results not shown). These results are in accordance with a previous whole genome analysis of Δ*xrn1* and Δ*dcp1* cells that revealed an unexpected low proportion of mRNAs with elevated levels (*3*). They raised the possibility that the Xrn1p and Dcp1p are not required for the decay of most mRNAs, as suggested previously (*3*). Alternatively, Dcp1p and Xrn1p can impact transcription, either directly or indirectly. The unexpected low levels of mRNAs characterize most, if not all, mutants that we investigated (Fig. 1B and C, Fig. S1), implying that any defect in the deadenylation-dependent, decapping-dependent decay machinery might affect transcription. Additional support to this notion is provided by the numerous physical and genetic interactions of various DFs with factors that are involved in the nuclear stages of gene expression found previously, mostly by high throughput approaches (Table S1).

To begin examining the possible involvement of DFs in transcription, we monitored *de novo* mRNA synthesis in response to environmental signals, such as galactose, glucose, heat shock (HS) and recovery from HS. As expected, wild type (WT) cells rapidly induced transcription of *GAL7* and *GAL10* upon galactose stimulation, and rapidly arrested transcription of these genes after switching the sugar to glucose (*4*) (Fig. 2A and B; Fig. S3A and B). Transcriptional induction of HS genes in WT cells was observed in response to HS (Fig. 2C and D; Fig. S2A and B), while transcription of non-HS genes was induced following recovery from HS (Fig. 2E and F; Fig. S2C and D). In

contrast, cells harboring an exonuclease-dead Xrn1p (Xrn1^{D208A}p) (*5*), whose mRNAs are relatively stable (Fig. S3), failed to induce efficient transcription of *GAL7* and *GAL10* upon galactose stimulation (Fig. 2A and B), HS genes in response to HS (Fig. S2A and B), as well as *TEF4* during recovery from HS (Fig. 2E and F). Similarly, cells expressing enzyme-dead Dcp2p (Dcp2-4p) (*6*) were defective in transcriptional induction of HS genes (Fig. 2C and D) and in transcriptional induction of *RPL43A* during the recovery from the HS (Fig. S2C and D). Thus, transcriptional induction is defective in the enzyme-dead strains. Surprisingly, strains lacking either *XRN1* or *DCP2* accumulated more mRNA transcripts than *xrn1*^{D208A} or *dcp2-4* strains respectively, but not more than WT (Fig. 2 and Fig. S2). Since the studied mRNA transcripts in both Δ*xrn1* and Δ*dcp2* strains are more stable than in WT (Fig. S3) (*6, 7*), we can conclude that transcription in the deletion strains is also defective. Significantly, different accumulation of mRNA transcripts in the deletion and enzyme-dead strains occurs despite identical stability of these mRNAs (Fig. S3), indicating that transcription of these genes varies in the two strains. Hence, the different transcriptional capacity is not related to the poor mRNA decay *per se*. Instead, the deletion strains support transcription in a different manner to the strains carrying enzyme dead proteins. These results strongly suggest that the presence or absence of each of these proteins directly affects transcription. Note that by 70 min following transcription induction, the *RPL34A* mRNA in WT and the *dcp2* mutants reached similar levels (Fig. S2D), reflecting the balance that starts to emerge between synthesis and decay, as the mRNA approaches its steady state level. This is in accordance with the results shown in Fig.1.

In order to better understand the transcriptional defect characterizing *xrn1* mutant cells, we employed FISH analysis designed to detect single mRNA molecules and to identify transcription sites (TSs) (*8*) (Fig. 2G and H; Fig. S4; Fig. S5). Most WT and Δ*xrn1* cells (88%) exhibited intron-containing nascent transcripts at their TSs, whereas only 66% of the $xrn1^{D208A}$ cells transcribed *TEF4* ($p=2.6x10^{-8}$) (Fig. 2I to K). This is consistent with our Northern analyses showing that the transcriptional defect of $xrn1^{D208A}$ cells is more severe than that of WT and Δ*xrn1* cells (Fig. 2A to F). However, WT cells contained TSs with more than one transcript and hence, more than one elongating RNA polymerase II (Pol II) (*8*), whereas Δ*xrn1* and $xrn1^{D208A}$ cells contained TSs with only one transcript (Fig. 2I to K; Fig. S5B to D). Since multiple molecules per TS are indicative of frequent transcription initiation or re-initiation events (*8*), we propose that Xrn1p is required for initiation and/or re-initiation. Note that the three strains carry a similar number of cytoplasmic fluorescent foci per cell, each representing one *TEF4* mRNA (Fig. S4B and C). This is consistent with *TEF4* mRNA levels obtained by Northern analyses (Fig. S4B, inset, Fig. S2D, 70 min). Importantly, since all mature *TEF4* mRNAs are found in the cytoplasm of all the three strains (confocal microscopic results - not shown), export of *TEF4* mRNA out of the nucleus does not require Xrn1p or its enzymatic activity.

The studied DFs are detected by and large in the cytoplasm (e.g. ref. (*9*)). To examine if they also reside in the nucleus, as our results imply, we utilized the *xpo1-1*, *mex67-5* strain, used to assay the nucleocytoplasmic shuttling activity of several proteins, including Pab1p (*10*). Similarly to Pab1p-GFP, several DFs accumulated in the nucleus upon heat inactivation of the exportins Xpo1-1p and Mex67-5p. Nuclear accumulation

was insensitive to translation inhibition by cycloheximide (CHX) (Fig. 3A; Fig. S6). It is worth noting that nuclear accumulation of Pat1p and Dhh1p in the Δ*lsm1* strain was reported previously, suggesting that they are exported in complex with Lsm1p (*9*). These observations and ours led us to conclude that all the examined DFs normally shuttle between the two compartments. We suspect that these DFs are usually visualized in the cytoplasm because their export rate exceeds that of their import. Interestingly, the equilibrium between export and import kinetics could be altered in response to environmental cues such as starvation (Fig. S7). Furthermore, nuclear accumulation of some of the examined DFs could be detected in several WT strains under optimal conditions (Fig. S7A and B). Interestingly, both Xrn1^{D208A}p-GFP and Dcp2p-RFP were poorly imported in *xrn1^{D208A} xpo1-1*, *mex67-5* cells. In contrast to the DFs, Pab1p-GFP was imported in *xrn1^{D208A} xpo1-1*, *mex67-5* cells as efficiently as in *XRN1 xpo1-1*, *mex67-5* cells (Fig. 3B), ruling out the possibility of a general defect in protein import in the triple mutant strain. Furthermore, during starvation, *xrn1^{D208A}* cells poorly imported Dcp2p-RFP, Ccr4p-RFP, Pat1p-RFP and Dcp1p-RFP (Fig. S7E). These results suggest that import of the DFs requires proper mRNA degradation. Interestingly, in the absence of Xrn1p, Dcp2p-RFP was imported efficiently (Fig. 3B). Since Xrn1^{D208A}p binds decapped RNA without degrading it (*5*), it is possible that the association of Xrn1^{D208A}p with the RNA blocks Dcp2p import. These import features explain why *xrn1^{D208A}* cells are more defective in transcription than cells lacking *XRN1*, reinforcing the dependence of transcription on the proper execution of the major cytoplasmic mRNA decay pathway.

Next, we examined whether DFs are capable of binding chromatin, using chromatin immunoprecipitation (ChIP) analysis. As shown in Fig. 4A, Xrn1p-TAP

associated with chromatin along the *PMA1* gene. Its association pattern partially parallels that of RNA polymerase II (Pol II). Xrn1p-TAP also associated with the *TEF4* promoter (Fig 4B), as well as with promoters of *RPL30*, *ACT1* and *YEF3* (Fig S8) but not with rDNA (Fig 4B). Several other tagged factors, excluding Dcp1p, associated with *PMA1*and *TEF4* promoters (Fig. S8) as well as with other *PMA1* amplicons, *ACT1*, *RPL30* and *YEF3* (results not shown). Collectively, our ChIP results are consistent with a direct role of these DFs in transcription.

Coupling of two processes requires that the activity of a certain protein in the first process is a prerequisite for its function in the subsequent step. Following this criterion, it was found that mRNA decay is coupled to translation, which, in turn, is coupled to mRNA export, maturation, and transcription (recently reviewed in (*11*)). Gene expression was therefore considered a linear system. Here, we show that factors of the hitherto last stage are shuttling proteins that also stimulate the first stage. A positive correlation is found between transcriptional efficiency and the capacity of DFs to shuttle. Specifically, cells lacking *XRN1*, which can import DFs efficiently, transcribe better than those harboring $xrn1^{D208A}$, which are defective in importing $Xrn1^{D208A}$ p as well as other DFs. Taken together, we propose that the exonuclease activity of Xrn1p is required for both DF import and, subsequently, for transcription.

The involvement of the studied DFs in the transcription reveals the circular nature of the gene expression pathway (Fig. 4C). Cyclic processes are inherently robust, because defects in one stage affect the overall pace of the entire process, thereby maintaining the essential balance between the stages. The maintenance of mRNA levels is one manifestation of this principle.

References

1. D. Muhlrad, R. Parker, *Embo J* **24**, 1033 (2005).
2. G. Badis, C. Saveanu, M. Fromont-Racine, A. Jacquier, *Mol Cell* **15**, 5 (2004).
3. F. He *et al.*, *Mol Cell* **12**, 1439 (2003).
4. D. Lohr, P. Venkov, J. Zlatanova, *Faseb J* **9**, 777 (1995).
5. J. A. Solinger, D. Pascolini, W. D. Heyer, *Mol Cell Biol* **19**, 5930 (1999).
6. T. Dunckley, R. Parker, *Embo J* **18**, 5411 (Oct 1, 1999).
7. F. W. Larimer, C. L. Hsu, M. K. Maupin, A. Stevens, *Gene* **120**, 51 (1992).
8. D. Zenklusen, D. R. Larson, R. H. Singer, *Nat Struct Mol Biol* **15**, 1263 (2008).
9. D. Teixeira, R. Parker, *Mol Biol Cell* **18**, 2274 (2007).
10. C. Brune, S. E. Munchel, N. Fischer, A. V. Podtelejnikov, K. Weis, *Rna* **11**, 517 (2005).
11. S. Komili, P. A. Silver, *Nat Rev Genet* **9**, 38 (2008).
12. . Materials and methods are available as supporting material on Science online.
13. R. Lotan *et al.*, *Genes Dev* **19**, 3004 (2005).
14. R. Lotan, V. Goler-Baron, L. Duek, G. Haimovich, M. Choder, *J Cell Biol* **178**, 1133 (2007).
15. S. A. Johnson, G. Cubberley, D. L. Bentley, *Mol Cell* **33**, 215 (2009).

Short exposure to the DNA intercalator DRAQ5 dislocates the transcription machinery and induces cell death.

Novel fluorescent dyes for scientific purposes appear regularly in the catalogs of suppliers. Some of these dyes serve as specific markers *per se* for subcellular compartments such as mitochondria or the nucleus. Suppliers however do not have the interest or the time to test the sometimes adverse effects of their products on the entire spectrum of biological models and processes. In the end, it is up to users to test for these side effects and to communicate them to the rest of the scientific community. Although tedious, this is a necessary step for the advancement of Science.

Here, we tested a novel DNA intercalator named DRAQ5 (Deep red fluorescing Anthraquinone number 5). DRAQ5 is highly permeant and easily intercalates with high affinity into DNA, producing very accurate images of the DNA localization and density in living cells. This study demonstrates that DRAQ5 severely affects transcription even after short exposures, and in the long run induces cell death. Bernard Vandenbunder's team has showed that exposure to DRAQ5 causes relocation of several factors involved in transcription, such as Hexim1, CBX8, Cyclin T1 and even Rpb1, the largest subunit of RNA Polymerase II. The dislocation of RNA Polymerase II foci was also observed in our hands.

These changes could not be observed when cells were stained with Hoechst 33342, another DNA intercalating dye. Moreover, when cells with active genes bearing MS2 stem loop repeats were detected using MS2cp-YFP, foci corresponding to the transcription sites rapidly recessed and were no longer detected after 5 minutes of exposure to DRAQ5. It also appeared that nucleoli, where MS2cp-YFP tends to accumulate, were disrupted by exposure to DRAQ5.

In order to test biochemically for a negative effect of DRAQ5 on the transcriptional machinery, we quantified the mRNA levels of the cMyc and the GAPDH genes. Whereas the GAPDH has a very long life time, the cMyc mRNA has a very half-life, in the order of 30 minutes. Our data suggests that a 30 minute exposure to DRAQ5 blocks transcription in a fashion very similar to Actinomycin D, a well documented transcriptional inhibitor. In fact, at early times (1 hour), levels of cMyc mRNA were lower in cells treated during the first 30 minutes with DRAQ5 than in cells continuously treated with Actinomycin D for 1 hour. It seems that some transcription could reappear at later stages after the removal of DRAQ5, but this was not tested in detail.

We demonstrated in this study that the novel DRAQ5 fluorescent dye has a strong negative effect in transcription, and that this strong effect must be taken into account by researchers that wish to use it in their studies.

Photochemistry and Photobiology, 2011, 87: 256–261

Research Note

Short Exposure to the DNA Intercalator DRAQ5 Dislocates the Transcription Machinery and Induces Cell Death

Elodie Richard[†1,2], Sébastien Causse[†3], Corentin Spriet[1], Nicolas Fourré[1], Dave Trinel[1], Xavier Darzacq[3], Bernard Vandenbunder[1] and Laurent Heliot[*1,2]

[1]Interdisciplinary Research Institute, USR 3078 CNRS and Lille 1 University, Bioophotonics Laboratory, Parc de la Haute Borne, Villeneuve d'Ascq Cedex, France
[2]Bioimaging Core Facility of Lille (BICFaL), IFR 147, USR 3078 CNRS and Lille1 University, Parc de la Haute Borne, Villeneuve d'Ascq Cedex, France
[3]Ecole Normale Supérieure, Institut de Biologie de l'ENS (IBENS) CNRS UMR 8197, Paris Cedex, France

Received 8 April 2010, accepted 22 October 2010, DOI: 10.1111/j.1751-1097.2010.00852.x

ABSTRACT

The fluorescent probe DRAQ5 which rapidly permeates cells and binds to DNA is potentially useful for functional studies of molecular dynamics and interactions in living nuclei. Within minutes after the incubation of human osteosarcoma U2OS cells with 5 μM DRAQ5, the distributions of RNA polymerase II and some of its associated regulatory proteins HEXIM and cyclin T1 in the nucleus are severely impaired, and transcription is inhibited. Furthermore, 30 min exposure to DRAQ5 induces death of U2OS cells 24 h later. Incubation with Hoechst 33342 under similar conditions does not induce these effects. These results emphasize the importance of carefully examining the functional consequences of labeling DNA with intercalating fluorescent dyes before use.

INTRODUCTION

Imaging approaches in living cells provide essential information about the kinetics of the formation of transcriptional complexes and their localization in the nucleus (1). Functional imaging studies require the simultaneous observation of the proteins of interest involved in these complexes and of the chromatin structure which plays a crucial role in transcription control.

Thanks to its high affinity mode of DNA intercalation, the permeant far red fluorescent dye DRAQ5 (deep red fluorescing anthraquinone Nr. 5) has been shown to reflect accurately the spatial concentration of DNA in live cells. DRAQ5 is easy to use, shows low photobleaching compared with GFP and other DNA dyes such as propidium iodine or TOTO-3, allows simultaneous imaging of GFP labeled proteins and does not require UV exposure for excitation (2). It has been used for blood cell discrimination and for analysis of the cell cycle by flow cytometry (3,4). However incubation with DRAQ5 induces chromatin aggregation and the detachment of eGFP tagged histones H1 and H2 from DNA in HeLa cells (5).

The epigenetic information carried by chromatin is read by effector proteins such as the Polycomb protein CBX8 which can discriminate histone H3 methylation marks and implement transcriptional silencing (6). RNA polymerase II (RNA Pol II), chromatin remodeling complexes and these effector proteins are the main targets of the regulatory networks controlling gene expression. RNA Pol II does not initiate alone the transcription of mRNAs. Essential elements of the transcription machinery have been identified by biochemical studies. Amongst them the positive transcription elongation factor b (P-TEFb) is an heterodimer consisting of the kinase CDK9 and one of the C-cyclins which controls the transition from the preinitiation to the elongation complex by phosphorylating the carboxyterminal domain of RNA Pol II (7). P-TEFb also forms a large inactive complex with the small nuclear RNA 7SK and other partners such as HEXIM (8).

Using human osteosarcoma U2OS cells expressing tagged proteins we show that incubation with DRAQ5 rapidly induces changes in the location of the transcriptional repressor CBX8, of RNA Pol II and of members of RNA-Pol II regulatory complex cyclin T1 and HEXIM. Furthermore, a short exposure to commonly used doses of DRAQ5 yields a blockade in cell cycle progression and cell death.

MATERIAL AND METHODS

Cell culture and transfection. Human osteosarcoma U2OS cells obtained from ATCC (Rockville) were grown in standard Dulbecco's Modified Medium (Invitrogen) supplemented with 10% fetal calf serum (FCS, Invitrogen) 16 h before transfection, 2.10^4 cells cm^{-2} were seeded on 32 mm diameter glass coverslips put in 6 wells plates (Falcon). Transient transfection was carried out with 1 μg plasmid using FuGENE (Roche) according to the manufacturer recommendations. Plasmids were encoding RNA-Pol II large subunit fused with eYFP (RPB1-eYFP), MS2-GFP, cyclin T1-Cerulean, HEXIM1-YFP or CBX8-eGFP.

†These two authors contributed equally to this work.
*Corresponding author email: laurent.heliot@iri.univ-lille1.fr (Laurent Heliot)
© 2010 The Authors
Photochemistry and Photobiology © 2010 The American Society of Photobiology 0031-8655/11

Staining procedure. DRAQ5 supplied by Biostatus (Shepshed, UK) as an acidified aqueous solution at 10 mM was directly added to the cells at a 5 μM final concentration for microscopy experiments and cells were observed without further washing according to the supplier protocol. Unless otherwise stated, Hoechst dye 33342 (Invitrogen) was used at a 2 μM final concentration according to the supplier protocol.

Fluorescence microscopy. 24 h after transfection, cells were incubated either with DRAQ5 or Hoechst 33342 in L15 medium without phenol red supplemented with 10% FCS. Live cell imaging was carried out on a TE2000 microscope (Nikon) equipped with a 37°C incubator and a QuantEM 512SC (Photometrics) EmCCD camera. Images were taken every 15 min without changing the gain of the camera and the exposure time. Image processing and analysis were performed with the ImageJ program (9).

FRAP experiments. Fluorescence recovery after photobleaching (FRAP) experiments were performed on a Leica SP5 XMT confocal microscope (Leica Microsystems, Gmbh) using a 63× NA 1.4 objective. Half of the cell nuclei were bleached for 3 s at 100% power of the 488, 491, 514 laser lines. Fluorescence recovery was monitored with 10% power of the 514 nm argon laser line every 1.5 s during 1 min; unspecific photobleaching was found to be negligible during acquisitions on control nuclei. Recovery curves were extracted from regions of interest, normalized on prebleach and postbleach values and averaged over all measured regions. At least 15 cells were acquired for each condition and a representative experiment was chosen for each condition. Our analysis focuses on the slow part of the recovery curves because freely diffusive species have been bleached during the experiment.

RT-QPCR analysis of GAPDH and cMyc expression. 5.10^5 U2OS cells were seeded in 6 well plates. The next day, cells were treated with either a 30 min pulse of 2 μM Hoechst33342 or 5 μM DRAQ5, or a continuous treatment of 5 μM actinomycin D in growth medium. In control cells, growth medium was renewed at treatment times. Total RNA was extracted using an Illustra RNAspin mini kit (GE Healthcare). Reverse transcription of 200 ng of total RNA was carried out using superscript III first strand synthesis supermix (Invitrogen) with oligo-dT primers. Maxima SYBR Green qPCR Master Mix (Fermentas) was used for the quantitative PCR, which was carried out on a LightCycler 480 (Roche). Analysis was done using the LightCycler 480 program (Roche). Each sample was assayed in triplicate.

Primer pairs for quantitative PCR:

GAPDH forward	5′-CACTTTGTCAAGCTCATTTC-3′
GAPDH reverse	5′-AGGGTCTCTCTCTTCCTCTT-3′
cMyc forward	5′-TGCTCCATGAGGAGACAC-3′
cMyc reverse	5′-TTTTCCACAGAAACAACATC-3′

Cell viability assay. 2.10^4 cells per well were seeded on 24 wells plates (Falcon) in standard culture conditions (DMEM with 10% FCS). 24 h later, the number of cells was approximately 4.10^4 in each well. DNA intercalating dyes were added at concentrations recommended in the Biostatus and Invitrogen data sheets for staining live animal cells : 1–20 μM for DRAQ5 and 0.3–8 μM for Hoechst 33342. After 30 min incubation, the medium was removed, three PBS washes were performed and fresh medium added. 24 h after this incubation with intercalating dyes, cells were harvested, stained with the vital dye Trypan blue (Invitrogen) according to the supplier recommendation and living cells counted in a Thoma chamber (Glasstic® Slide, Hycor Biomedical Inc.).

RESULTS

In order to evaluate the functional effects of nuclear architecture changes induced by the incubation of cells with DRAQ5 (5), we first investigated the localization of CBX8, a member of the Polycomb group proteins which binds to methylated histone H3 and is recruited to chromatin (10). CBX8-GFP fluorescence in U2OS transfected cells was distributed throughout the nuclei, with bright patches corresponding to the nucleoli (Fig. 1A). Within minutes after its addition into the culture medium, DRAQ5 fluorescence was observed in U2OS nuclei, a plateau being reached before 15 min (Fig. 1B). Simultaneously the bright fluorescence of CBX8-GFP in the nucleoli decreased. CBX8-GFP fluorescence accumulated into more intense spots whereas the fluorescence background in the nuclei decreased, suggesting a condensation of free CBX8-GFP (Fig. 1A). In contrast, Hoechst 33342 induced no change in the distribution of CBX8-GFP (Fig. 1C) whereas its binding to DNA was evidenced by its fluorescence in the nucleus, still increasing after 15 min (Fig. 1D).

In the absence of intercalating dye, RNA-Pol II large subunit fused with eYFP (RPB1-eYFP) showed a punctuate pattern of fluorescence in the nucleus (Fig. 2A,D). Upon incubation with DRAQ5 (Fig. 2C), RPB1-eYFP fluorescence became more uniform with the disappearance of the more intense dots (Fig. 2A). No change in RPB1-eYFP fluorescence was observed (Fig. 2D) upon DNA staining by Hoechst 33342 (Fig. 2F).

We next investigated the distribution of two proteins involved in the regulation of RNA-Pol II activity: cyclin T1, which is a subunit of the P-TEFb positive transcription elongation complex, and HEXIM which binds and inactivates P-TEFb. Cyclin T1-cerulean fluorescence that was distributed homogeneously throughout the nuclei, condensed within minutes after the addition of 5 μM DRAQ5 (Fig. 2B). A similar condensation was observed with HEXIM1-YFP (Fig. 3A). We wondered what would be the effects of lower DRAQ5 concentrations on nuclear organization. 0.1 μM

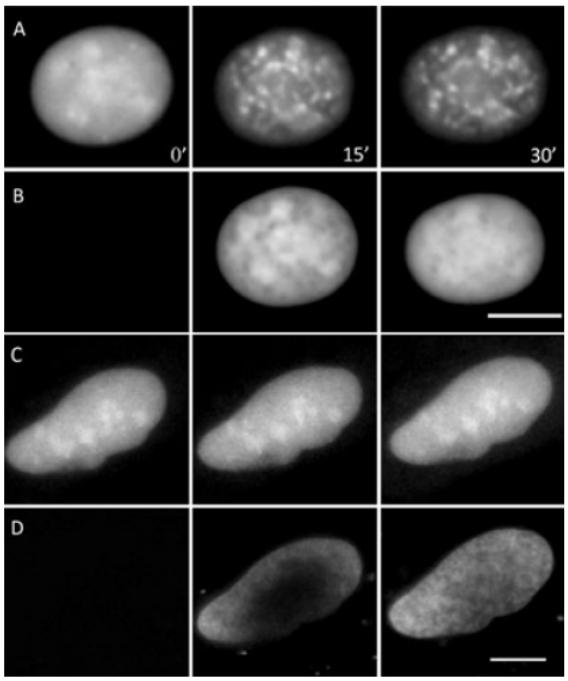


Figure 1. (A) Changes in the spatial distribution of CBX8-GFP in U2OS cells before and after 15 min and 30 min incubation with 5 μM DRAQ5. (B) Simultaneous DRAQ5 staining changes. (C, D) Same experiment with 2 μM Hoechst 33342. Bars: 10 μm.

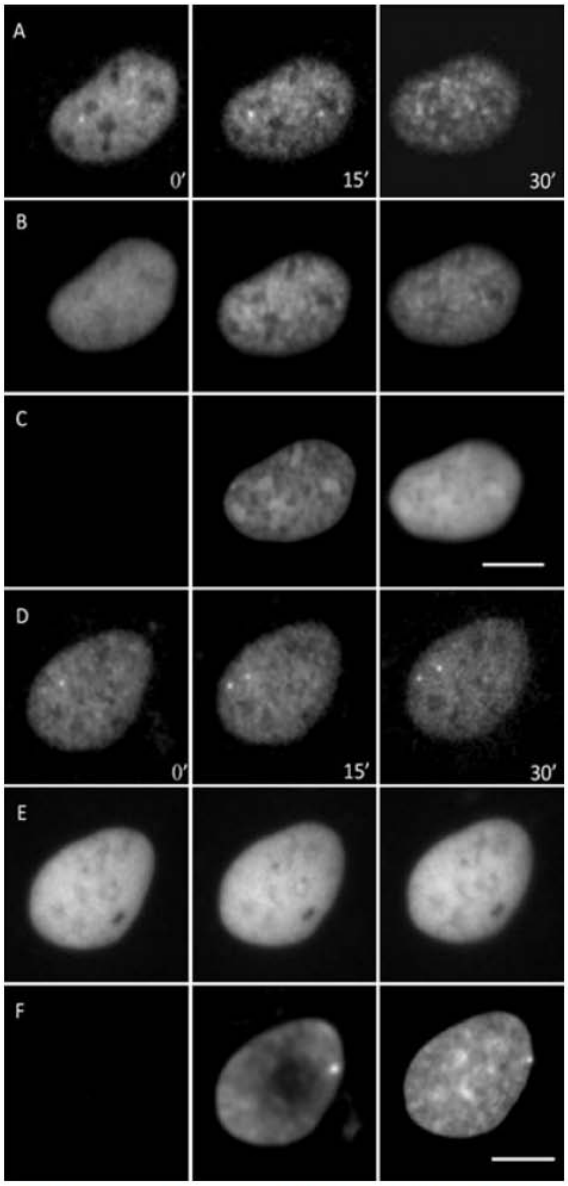


Figure 2. Changes in the spatial distribution of RPB1-eYFP (A) and cyclin T1 cerulean (B) in U2OS cells before and after 15 min and 30 min incubation with 5 μM DRAQ5, with simultaneous DRAQ5 staining changes (C). (D, E, F) Same experiment with 2 μM Hoechst 33342. Bars: 10 μm.

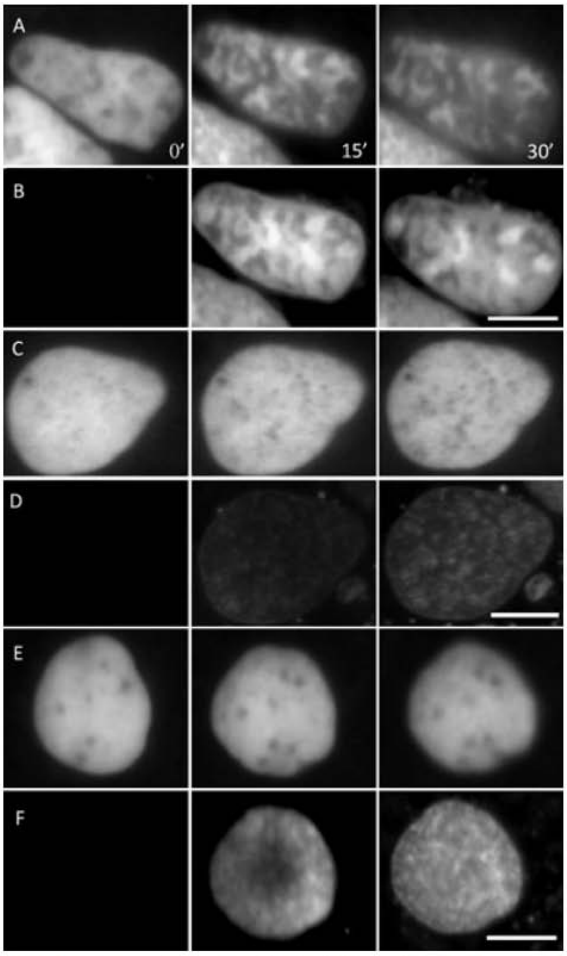


Figure 3. Changes in the spatial distribution of HEXIM1-YFP in U2OS cells before and after 15 min and 30 min incubation with 5 μM (A) or 0.1 μM (C) DRAQ5, with simultaneous DRAQ5 staining changes (B, D). (E, F) Same experiment with 2 μM Hoechst 33342. Bars: 10 μm.

DRAQ5 staining whose low intensity required the use of an EMCCD camera for its visualization was also associated with HEXIM1-YFP condensation (Fig. 3C,D). In contrast, the distribution of cyclin T1 (Fig. 2E) or HEXIM (Fig. 3E) in U2OS cells was not disturbed upon addition of Hoechst 33342.

We next examined what would be the effects of DRAQ5 on RNA-Pol II activity. FRAP experiments have revealed the existence of an immobile RNA-Pol II fraction which has been shown to be transcriptionally active (11). As DRAQ5 induced major changes in RNA-Pol II distribution, we performed a half-nucleus FRAP rather than a spot FRAP in order to get a more integrated view of RNA-Pol II kinetics. In the absence of DRAQ5, similar recovery curves were obtained inside and outside the intense foci; 15 min after DRAQ5 incubation, these intense foci have disappeared and we found that the immobile RNA-Pol II fraction was significantly reduced in regions where it was more condensed (Fig. 4A,B). In order to visualize directly the effects of DRAQ5 on mRNA accumulation, we next used an U2OS cell line carrying 64 repeats of an inducible transgene encoding a mRNA with 24 MS2 repeats in its 3′ untranslated region (1). Transcripts encoded by the transgene appear as bright spots stained by MS2-GFP in the nuclei, which disappeared within minutes after DRAQ5 incubation (Fig. 4C). We also compared the effects of DRAQ5, Hoechst 33342 and actinomycin D on the accumulation of the short lived cMyc transcripts (Fig. 4D); GAPDH was used as an internal control (Fig. 4D1). A 12 ($\pm$1)-fold

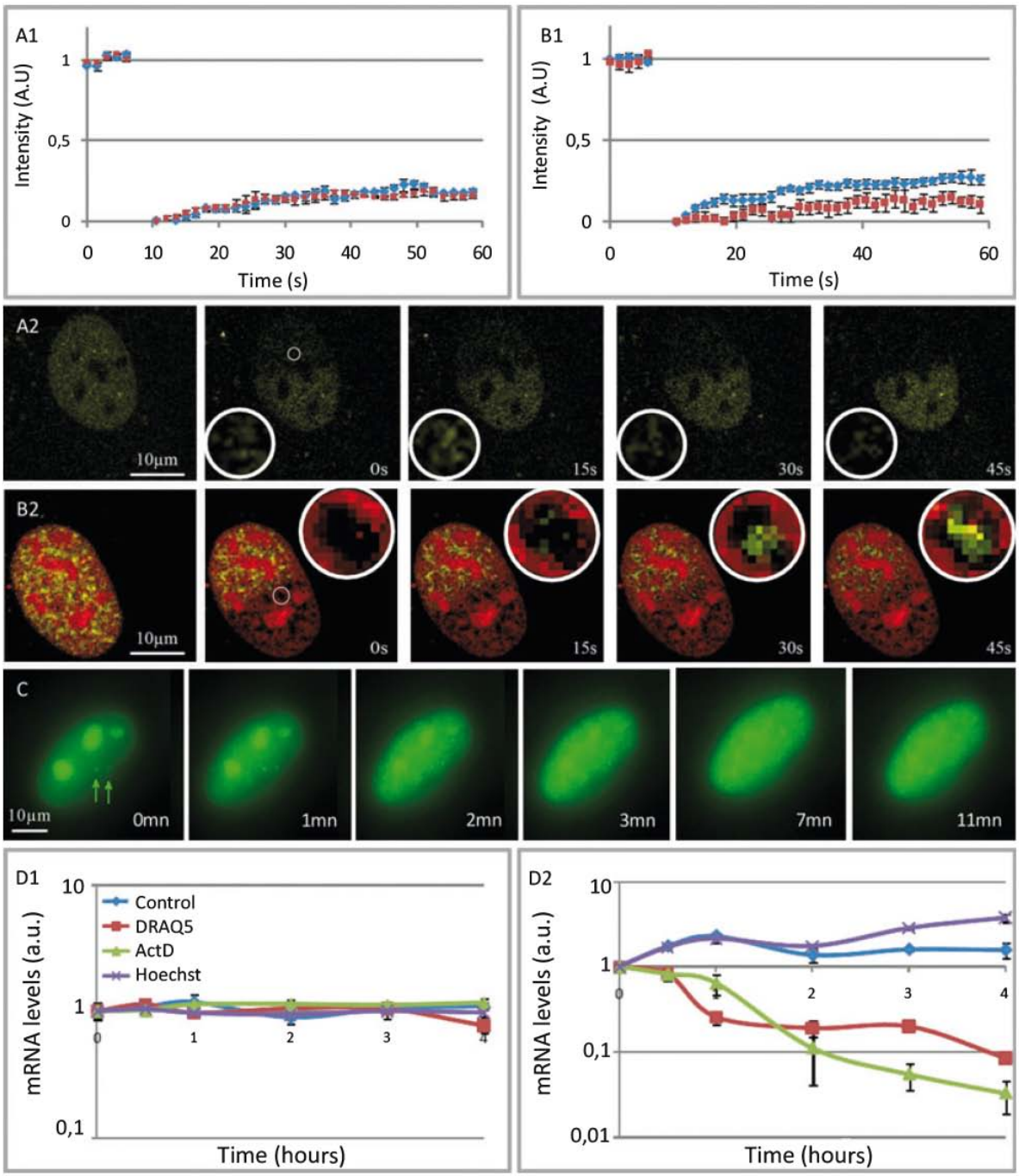


Figure 4. The effects of DRAQ5 on RNA-Pol II transcriptional activity. FRAP experiments were performed on half U2OS cell nuclei expressing RPB1-YFP in the absence (A) or 15 min after 5 µM Draq5 incubation (B). (A1, B1) RPB1 recovery curves measured in condensed (blue) or uncondensed (red) regions. (A2, B2) Associated images with RPB1-YFP in yellow and DRAQ5 in red. (C) Time-lapse imaging of an U2OS cell carrying repeats of an induced transgene and expressing MS2-GFP, following the addition of 5 µM DRAQ5. Green arrows correspond to induced transgene localization. MS2-GFP fluorescence on transcription sites and on nucleoli is rapidly disturbed upon DRAQ5 incubation. Bars: 10 µm. RT-QPCR analysis of the accumulation of GAPDH (D1) and cMyc (D2) transcripts in U2OS cells following incubation with 5 µM DRAQ5, 2 µM Hoechst 33342 or 5 µM actinomycin D. At time 30 min after the incubation with DRAQ5 or Hoechst 33342, the medium was removed, cells were rinsed and fresh medium added; for control cells medium was replaced by fresh medium.

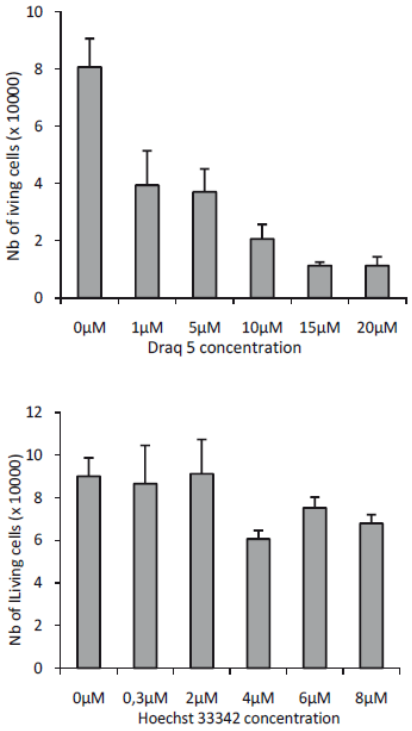


Figure 5. Cytotoxic effects of the intercalating dyes DRAQ5 and Hoechst 33342 for U2OS cells. The experiment was performed as described in Material and methods and run in triplicate.

decrease in the amount of cMyc transcripts is measured by RT-QPCR analysis 4 h following incubation with 5 μM DRAQ5 (Fig. 4D2). 2 μM Hoechst 33342 does not induce such an effect. At time 1 h, DRAQ5 inhibits transcription more efficiently than actinomycin D does. The plateau observed between times 2 and 3 h in the cMyc decay curve may be attributed to a release in RNA-pol II inhibition following DRAQ5 removal at time 30 min.

The observation of these rapid changes induced by DRAQ5 on the localization of key components and on the activity of the transcription machinery prompted us to study the effects of short exposures with this intercalating dye on U2OS cell growth. Cells were incubated during 30 min with a range of DRAQ5 concentrations currently used in cytometry studies, subsequently washed three times with PBS and fed with fresh culture medium. 24 h later, cell growth was blocked by 5 μM DRAQ5. Cell death was observed at higher concentrations. In contrast, the effect of Hoechst 33342 on U2OS cell growth was limited (Fig. 5).

DISCUSSION

A recent review emphasizes the use of DRAQ5 in the drug discovery process for screening out toxicants from compounds engineered by manufacturers (12). We present here evidence that DRAQ5 by itself can induce cell death.

It has been shown that long term (18 h) exposure of human lung adenocarcinoma epithelial cells A549 to DRAQ5 induces a DNA damage response and a perturbation of cell cycle progression (13). Our experiments with U2OS cells indicate that 30 min exposure to DRAQ5 can induce cell death. DRAQ5 belongs to the family of antraquinone synthetic DNA binding agents. These agents are structurally related to the anticancer drug mitoxantrone and studies on the cytotoxicity of this drug have been concentrated on topoisomerase II. It has been also shown that DRAQ5 intercalation into DNA displaces bound histone H1 that continuously exchange from chromatin and subsequently causes chromatin aggregation (5). We show that the change of chromatin organization induced upon DRAQ5 binding to DNA is associated with rapid and drastic changes of the fluorescence signals associated with YFP tagged RNA-Pol II large subunit, cerulean tagged cyclin T1 and eGFP tagged CBX8. These changes cannot result from a Fluorescence Resonance Energy Transfer (FRET) mechanism between DRAQ5, YFP, GFP and cerulean because the DRAQ5 absorption spectrum and the GFP derived reporter emission spectra are well separated (12,14), Therefore fluorescence changes observed in our studies reflect changes in RNA Pol II and associated PTEF-b complex localization. We also show that the transcriptionally active fraction of RNA Pol II that stays immobile during FRAP experiments is reduced in regions where it condenses upon DRAQ5 incubation. In addition, a 12-fold decrease in cMyc transcripts accumulation is observed 4 h after DRAQ5 incubation with U2OS cells. In human fibroblasts (MRC5SV) after 1 h incubation DRAQ5 inhibits [H3] incorporation (15). In U2OS cells carrying a gene array, within minutes after DRAQ5 addition, we show that transcripts disappear from the transcription site. These results suggest that highly dynamic molecular complexes (16) carrying transcription dissociate upon DRAQ5 binding to DNA, which may be sufficient to account for the cytostatic/cytotoxic effects of this drug. This dissociation could result from direct competition between DRAQ5 and transcription factories for DNA binding and from long range effects mediated by DRAQ5 DNA unwinding.

DRAQ5 and Hoechst 33342 are both DNA intercalating agents. Molecular modeling and X-ray crystallographic studies suggest the existence of a partial overlap of their binding into a minor groove of the DNA (14). DRAQ5 is able to quench the fluorescence of bound Hoechst 33342 whereas Hoechst 33342 was unable to induce any detectable displacement of DRAQ5 from the nuclei of live cells (14). In sharp contrast with DRAQ5, short exposures with Hoechst 33342 do not change chromatin functional organization and does not affect significantly U2OS cell growth. We suggest that differences in the affinity of these dyes evidenced by different nuclear uptake kinetics, in their efficiency for unwinding double strand DNA or differences in the number of labeled DNA sites may account for their differential cytostatic/cytotoxic effects.

Fluorescent dyes or tags are routinely used for the observation of molecular events in living cells. We show here that DRAQ5 which has been shown to be valuable for the visualization of DNA in fixed cells (12,17) or the identification of cells by flow cytometry (3,4,18) can be ultimately cytotoxic. Before using DRAQ5 as an intravital dye, one should therefore carefully examine to what extent it disturbs the physiological events studied. Alternatively, this dye can be useful to transiently open DNA structure and to decipher the role of this structure on gene expression (19).

Acknowledgements—This work was performed on the platform of the Bio Imaging Core Facility of Lille 1 (BICFaL/BICeL). It was supported by the CNRS, the Lille 1 University, the Région Nord-Pas de Calais, the European Regional Developmental Funds, and the Agence Nationale de la Recherche (ANR 2007 PFTV 011 01). ER is recipient of a grant from the Conseil Regional. We thank the "groupe de travail Noyau" of the GDR2588 and the French Network for Microscopy (RTmfm) for scientific discussions and help.

REFERENCES

1. Darzacq, X., Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair and R. H. Singer (2007) In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806.
2. Martin, R. M., H. Leonhardt and M. C. Cardoso (2005) DNA labeling in living cells. *Cytometry A* **67**, 45–52.
3. Smith, P. J., N. Blunt, M. Wiltshire, T. Hoy, P. Teesdale-Spittle, M. R. Craven, J. V. Watson, W. B. Amos, R. J. Errington and L. H. Patterson (2000) Characteristics of a novel deep red/infrared fluorescent cell-permeant DNA probe, DRAQ5, in intact human cells analyzed by flow cytometry, confocal and multiphoton microscopy. *Cytometry* **40**, 280–291.
4. Smith, P. J., M. Wiltshire, S. Davies, L. H. Patterson and T. Hoy (1999) A novel cell permeant and far red-fluorescing DNA probe, DRAQ5, for blood cell discrimination by flow cytometry. *J. Immunol. Methods* **229**, 131–139.
5. Wojcik, K. and J. W. Dobrucki (2008) Interaction of a DNA intercalator DRAQ5, and a minor groove binder SYTO17, with chromatin in live cells–influence on chromatin organization and histone-DNA interactions. *Cytometry A* **73**, 555–562.
6. Simon, J. A. and R. E. Kingston (2009) Mechanisms of polycomb gene silencing: Knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **10**, 697–708.
7. Kohoutek, J. (2009) P-TEFb- the final frontier. *Cell Div.* **4**, 19.
8. Diribarne, G. and O. Bensaude (2009) 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. *RNA Biol.* **6**, 122–128.
9. Rasband, W. (1997–2007) *ImageJ*. U.S. National Institutes of Health, Bethesda, MD. Available at: http://rsb.info.nih.gov/ij/.
10. Vincenz, C. and T. K. Kerppola (2008) Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proc. Natl. Acad. Sci. U S A* **105**, 16572–16577.
11. Kimura, H., K. Sugaya and P. R. Cook (2002) The transcription cycle of RNA polymerase II in living cells. *J. Cell Biol.* **159**, 777–782.
12. Edward, R. (2009) Use of DNA-specific anthraquinone dyes to directly reveal cytoplasmic and nuclear boundaries in live and fixed cells. *Mol. Cells* **27**, 391–396.
13. Zhao, H., F. Traganos, J. Dobrucki, D. Wlodkowic and Z. Darzynkiewicz (2009) Induction of DNA damage response by the supravital probes of nucleic acids. *Cytometry A* **75**, 510–519.
14. Njoh, K. L., L. H. Patterson, M. Zloh, M. Wiltshire, J. Fisher, S. Chappell, S. Ameer-Beg, Y. Bai, D. Matthews, R. J. Errington and P. J. Smith (2006) Spectral analysis of the DNA targeting bisalkylaminoanthraquinone DRAQ5 in intact living cells. *Cytometry A* **69**, 805–814.
15. Mari, P. O., V. Verbiest, S. Sabbioneda, A. M. Gourdin, N. Wijgers, C. Dinant, A. R. Lehmann, W. Vermeulen and G. Giglia-Mari (2010) Influence of the live cell DNA marker DRAQ5 on chromatin-associated processes. *DNA Repair (Amst.)* **9**, 848–855.
16. Soutoglou, E. and T. Misteli (2007) Mobility and immobility of chromatin in transcription and genome stability. *Curr. Opin. Genet. Dev.* **17**, 435–442.
17. Berger, M., A. Farcas, M. Geertz, P. Zhelyazkova, K. Brix, A. Travers and G. Muskhelishvili (2010) Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU. *EMBO Rep.* **11**, 59–64.
18. de Haas, A. H., H. W. Boddeke, N. Brouwer and K. Biber (2007) Optimized isolation enables ex vivo analysis of microglia from various central nervous system regions. *Glia* **55**, 1374–1384.
19. Sabbioneda, S., A. M. Gourdin, C. M. Green, A. Zotter, G. Giglia-Mari, A. Houtsmuller, W. Vermeulen and A. R. Lehmann (2008) Effect of proliferating cell nuclear antigen ubiquitination and chromatin structure on the dynamic properties of the Y-family DNA polymerases. *Mol. Biol. Cell* **19**, 5193–5202.

The In Vivo Kinetics of RNA Polymerase II Elongation During Co-Transcriptional Splicing.

One of the most important co-transcriptional processes in higher eukaryotes is splicing. It is called a co-transcriptional process because proteins involved in the splicing of RNA are known to be recruited by the phosphorylated CTD of RNA Polymerase 2, which only occurs during transcription (Mortillaro et al. 1996; T Misteli & D L Spector 1999; E. Kim et al. 1997; Vincent et al. 1996). Furthermore, the splicing of introns starts before the complete mRNA leaves the site of transcription (Audibert et al. 2002; Singh & Padgett 2009). Recently, it has been demonstrated that the speed at which RNA Polymerase 2 transcribes a gene can influence alternative splicing sites (A. R. Kornblihtt et al. 2004; de la Mata et al. 2003; Muñoz et al. 2010). In this study, we studied the possibility of the opposite, *ie.* that splicing might slow down transcription by RNA Polymerase 2. This required generating multiple cell lines from the same genetic background, and containing the same reporter gene, but with varying intron amounts and lengths.

This study demonstrates that increasing the number of introns in a gene will not decrease the speed of RNA Polymerase 2. More importantly, through the combined results of FRAP and videomicroscopy on MS2-YFP, GFP-Rpb1 and FISH, the study demonstrates that pre-mRNA which have completed transcription

but are not yet entirely spliced are retained at the transcription site until they are fully spliced. This study relied strongly on two 4 dimensional microscopy techniques, live cell imaging and Fluorescence Recovery After Photobleaching (4D FRAP). Our know-how in this field and the imaging systems at our disposal made it possible to collect valuable data in our laboratory, where I have set up part of the system and the automatic imaging protocols for its correct use.

The In Vivo Kinetics of RNA Polymerase II Elongation during Co-Transcriptional Splicing

Yehuda Brody[1], Noa Neufeld[1], Nicole Bieberstein[2], Sebastien Z. Causse[3], Eva-Maria Böhnlein[2], Karla M. Neugebauer[2], Xavier Darzacq[3], Yaron Shav-Tal[1]*

1 The Mina & Everard Goodman Faculty of Life Sciences & Institute of Nanotechnology, Bar-Ilan University, Ramat Gan, Israel, 2 Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany, 3 Functional Imaging of Transcription, Ecole Normale Supérieure, Institut de Biologie de l'ENS, IBENS, CNRS, UMR8197, Paris, France

Abstract

RNA processing events that take place on the transcribed pre-mRNA include capping, splicing, editing, 3′ processing, and polyadenylation. Most of these processes occur co-transcriptionally while the RNA polymerase II (Pol II) enzyme is engaged in transcriptional elongation. How Pol II elongation rates are influenced by splicing is not well understood. We generated a family of inducible gene constructs containing increasing numbers of introns and exons, which were stably integrated in human cells to serve as actively transcribing gene loci. By monitoring the association of the transcription and splicing machineries on these genes in vivo, we showed that only U1 snRNP localized to the intronless gene, consistent with a splicing-independent role for U1 snRNP in transcription. In contrast, all snRNPs accumulated on intron-containing genes, and increasing the number of introns increased the amount of spliceosome components recruited. This indicates that nascent RNA can assemble multiple spliceosomes simultaneously. Kinetic measurements of Pol II elongation in vivo, Pol II ChIP, as well as use of Spliceostatin and Meayamycin splicing inhibitors showed that polymerase elongation rates were uncoupled from ongoing splicing. This study shows that transcription elongation kinetics proceed independently of splicing at the model genes studied here. Surprisingly, retention of polyadenylated mRNA was detected at the transcription site after transcription termination. This suggests that the polymerase is released from chromatin prior to the completion of splicing, and the pre-mRNA is post-transcriptionally processed while still tethered to chromatin near the gene end.

Citation: Brody Y, Neufeld N, Bieberstein N, Causse SZ, Böhnlein E-M, et al. (2011) The In Vivo Kinetics of RNA Polymerase II Elongation during Co-Transcriptional Splicing. PLoS Biol 9(1): e1000573. doi:10.1371/journal.pbio.1000573

Academic Editor: Tom Misteli, National Cancer Institute, United States of America

Received May 14, 2010; **Accepted** November 19, 2010; **Published** January 11, 2011

Copyright: © 2011 Brody et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a Germany Israeli Foundation grant (929-192.13/2006) to YST and KM; the Ministry of Science Israel-France grants to YST and XD; and ANR-06-JC-0118 to XD. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: lacI, lac repressor protein; lacO, lac operator repeats.

* E-mail: shavtaly@mail.biu.ac.il

Introduction

The processes of transcription and RNA processing are co-transcriptionally integrated [1–7]. A variety of studies have demonstrated that pre-mRNA splicing is coupled to transcription and that intron removal can occur at the site of transcription while the polymerase is still engaged in active transcription. Other mRNA processing events such as 5′ capping and 3′-end formation are also coupled to transcription.

A unique feature of the eukaryotic RNA Pol II is the presence of a long and highly conserved tail at the C-terminus of the large subunit (Rbp1) of the polymerase, termed the CTD (carboxy-terminal repeat domain). It is proposed that the CTD serves as a "jumping board" for protein factors that assemble on the nascent transcript as it emerges from the polymerase [8–10]. For instance, certain splicing factors from the SR protein family were found associated with the phosphorylated CTD [11–14]. SR proteins are most probably not associated with the transcriptional initiation complex but join in only after the elongation related CTD-phosphorylation has begun. This was demonstrated by showing that RNA Pol II poised at the *Fos* gene promoter is not associated with SR proteins when the gene is in an uninduced state, and only after transcription had begun were SR proteins recruited [15]. Chromatin immunoprecipitation (ChIP) experiments have also demonstrated the association of RNA processing factors with transcribing genes [16–19].

Co-transcriptional splicing implies that splicing, at least in part, should occur at the site of transcription. This is supported by kinetic studies showing that splicing is completed within a short ~5–10 min time-frame from the time of transcription, and these times are not influenced by intron length [20–24]. Splicing factors are dynamically recruited to active gene loci [25–29]. The kinetics of transcription can influence alternative splice site selection. Polymerase slowdown and pausing affected the pattern of alternative splicing in vivo as seen with a mutant polymerase form that was intrinsically slow-elongating and favored the inclusion of an exon [30]. The current model is that slowing down of the polymerase or polymerase pausing leads to the selection of a weak splice site [1,31].

In order to examine the opposite relationship, namely whether the kinetics of the elongating RNA Pol II can be influenced by splicing events occurring co-transcriptionally on the associated pre-mRNA, we generated a series of comparable cell lines harboring inducible gene constructs that contain increasing numbers of introns

Author Summary

The pre-mRNA emerging from RNA polymerase II during eukaryotic transcription undergoes a series of processing events. These include 5′-capping, intron excision and exon ligation during splicing, 3′-end processing, and polyadenylation. Processing events occur co-transcriptionally, meaning that a variety of enzymes assemble on the pre-mRNA while the polymerase is still engaged in transcription. The concept of co-transcriptional mRNA processing raises questions about the possible coupling between the transcribing polymerase and the processing machineries. Here we examine how the co-transcriptional assembly of the splicing machinery (the spliceosome) might affect the elongation kinetics of the RNA polymerase. Using live-cell microscopy, we followed the kinetics of transcription of genes containing increasing numbers of introns and measured the recruitment of transcription and splicing factors. Surprisingly, a sub-set of splicing factors was recruited to an intronless gene, implying that there is a polymerase-coupled scanning mechanism for intronic sequences. There was no difference in polymerase elongation rates on genes with or without introns, suggesting that the spliceosome does not modulate elongation kinetics. Experiments including inhibition of splicing or transcription, together with stochastic computational simulation, demonstrated that pre-mRNAs can be retained on the gene when polymerase termination precedes completion of splicing. Altogether we show that polymerase elongation kinetics are not affected by splicing events on the emerging pre-mRNA, that increased splicing leads to more splicing factors being recruited to the mRNA, and that post-transcriptional splicing can proceed at the site of transcription in the absence of the polymerase.

and exons. These genes were stably integrated into a human cell line and formed tandem arrays containing multiple copies of the integrated gene, thereby creating a genomic locus that upon transcriptional activation is easily monitored using live-cell imaging. Such transcription sites have been used previously to analyze the kinetics of recruitment of a variety of nuclear factors to actively transcribing genes (reviewed in [32,33]). Previous work has demonstrated that the real-time kinetics of RNA Pol II elongation on an active gene can be faster than measured by many biochemical measurements, and is coupled to stochastic pauses of the enzyme [34,35]. Recent studies have confirmed that Pol II elongation can proceed at rates of 3–4 kb/min [21,22,36].

In this study we used the above in vivo approach to examine Pol II elongation rates either on genes that do not contain an intron and therefore are not expected to recruit the spliceosome to the nascent transcript or on genes containing varying numbers of intron/exons and that undergo co-transcriptional splicing. We show that polymerase elongation kinetics are not modulated by splicing events taking place on the emerging pre-mRNA, that increased splicing leads to an increase in the splicing factors recruited to the mRNA, and that post-transcriptional splicing can proceed at the site of transcription without the presence of the polymerase.

Results

Generation of Cell Lines with Integrated Intron-Containing Genes

In order to examine the elongation kinetics of RNA Pol II during splicing, we generated the following gene constructs. The first construct consisted of the human *β-globin* mini-gene with three exons and two introns and was therefore designated E3 (Figure 1A) [37,38]. Exon 3 was truncated and fused in-frame to a cyan fluorescent protein (CFP) coding region containing the peroxisomal targeting tripeptide Ser-Lys-Leu (SKL) in its C-terminus [39], thereby generating cyan fluorescing peroxisomes throughout the cytoplasm, indicative of productive gene expression. Immediately downstream were a series of 18 MS2 sequence-repeats. When transcribed, these repeats form 18 stem-loops in the mRNA and provide high-affinity binding sites for the MS2 coat protein [40,41]. The detection of these mRNAs in vivo is accomplished by the binding of dimers of the YFP-MS2 coat-protein to each stem-loop, thereby forming fluorescently labeled mRNAs [42]. This occurs simultaneously during active transcription [34,41].

β-globin was chosen as a model gene since it has been extensively used for in vitro splicing assays [43]. In this manner we could use a system that is close to the endogenous state of a gene. The first intron of *β-globin* is short (131 nt), while most typical genes have a long first intron [44]. The length of the second *β-globin* intron is 992 nt, which is close to the median size of internal introns (~1,400 nt) [44]. Thereby, we could use repeats of the second intron to generate additional genes containing increasing numbers of introns.

In order to compare between genes, we generated additional gene constructs (Figure 1A): (a) E1 – an intronless gene in which exons 1, 2 and their introns were deleted, thereby leaving part of exon 3 + CFP-SKL to serve as a single exon; and (b) E4, in which intron 2, flanked by the splice sites and part of exons 2 and 3, was duplicated, thereby forming a gene construct with 4 exons and 3 introns. The E6 construct (containing another two repeats of intron 2) will be described later on. All the constructs were under the inducible control of the Tet-On system. In the presence of the rtTA (Tet-On) transactivator and the addition of doxycycline (dox) to the medium, transcription was induced.

Each of the constructs was stably integrated into human U2OS Tet-On stable cells. The different sized gene constructs produced the correctly spliced mRNAs as seen by RT-PCR on mRNA extracted from the three cell lines following transcriptional activation by dox: E3 = 2,279 nt (unspliced 3,402 nt), E4 = 2,321 nt (unspliced 4,436 nt) (Figure 1B). The E1 mRNA (1,941 nt) did not undergo splicing. The genes were co-integrated with a plasmid containing 256 lac operator repeats (*lacO*) in order to enable the detection of the genomic locus of integration [29,45]. Using a fusion protein of the lac repressor protein (LacI), e.g. RFP-LacI, the genomic locus was tagged, allowing the detection of the integration locus independent of the transcriptional status of the gene. Selection of positive stable clones was performed after induction with dox and identification of cytoplasmic CFP-peroxisomes in all cell clones (Figures 1C and S1), altogether indicating accurate transcription, splicing, export, and translation. Co-integration of the 256 *lacO* repeats was verified by the expression of RFP-lacI and the detection of single genomic integration sites (Figure 1C). YFP-MS2-labeled mRNAs were detected at the transcription site, as well as in the nucleoplasm and the cytoplasm as expected (Figure 1C). Several stable clones were kept for each gene, and we proceeded to work with three cell lines designated E1, E3, and E4, in accordance with the integrated gene.

Stable transfection results in the integration of multiple copies of the gene construct at one genomic locus that is termed a "tandem gene array," which is useful for live-cell imaging experiments of active transcription sites [32]. We confirmed that RNA Pol II in both active phosphorylated forms was recruited to the transcription sites using the H14 (CTD phosphorylated on serine 5), H5

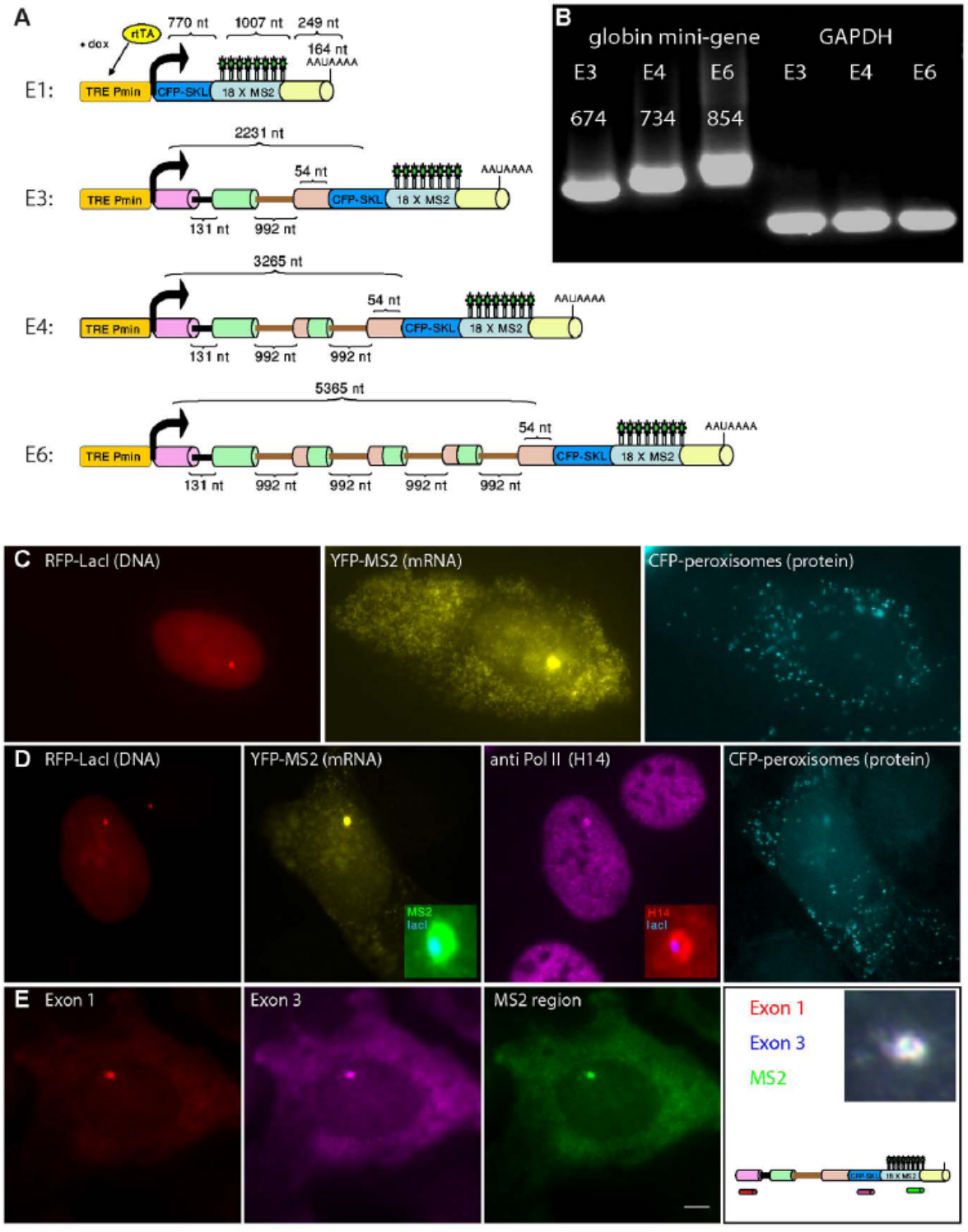


Figure 1. Generating a cell system with comparable genes containing increasing numbers of introns. (A) Gene constructs containing increasing numbers of introns and exons of the human *β-globin* mini-gene were generated and termed E1, E3, E4, and E6. Common to all the genes are: Tet-inducible promoter (orange; Tet responsive element (TRE), minimal promoter (Pmin)) that is induced in the presence of rtTA (yellow) and

doxycycline (dox); the genes consist of the *β-globin* intron and exon sequences (exon 1, purple; exon 2, green; exon 3, pink) in frame with a CFP-SKL peroxisomal protein (blue), and a series of 18 MS2 repeats in the 3′UTR (cyan) for the in vivo labeling of the mRNA via binding of YFP-MS2 proteins (green stars). The 3′-end of the gene contains a polyA site. Intron sizes, the total size of the gene, the MS2 repeat region, and the 3′UTR are marked. (B) RT-PCR showing the production of the correctly spliced E3, E4, and E6 mRNAs. GAPDH was used as a control. Sizes(nt) are marked above the bands. (C) E3 cell induced with dox for 12 h (activation times are similar in all experiments unless indicated otherwise) showing the gene locus marked by RFP-LacI (red), the transcribed mRNA at the transcription site and throughout the cell labeled with YFP-MS2 (yellow), and the CFP-SKL protein product in cytoplasmic peroxisomes (cyan). (D) Actively transcribing transcription sites (yellow, YFP-MS2 tagged mRNA) and gene locus (red, RFP-LacI) show the recruitment of endogenous RNA Pol II (purple) by immunofluorescence with the H14 Ab that detects initiating polymerases. Enlargements show the overlay at the transcription site. CFP-peroxisomes are seen in cyan. (E) RNA-FISH showing mRNAs detected throughout an E3 cell with fluorescent probes to the first exon (red), the CFP exon (exon 3, purple), and the MS2 repeats (green). Scheme shows the probe binding sites and the enlargement of the transcription site (bar, 5 μm).
doi:10.1371/journal.pbio.1000573.g001

(CTD phosphorylated on serine 2), and 8WG16 antibodies (specific to the CTD repeats) (Figures 1D and S2A,B). RNA FISH with probes to different regions in the pre-mRNA and mRNA were used to detect the distribution of the mRNAs in the cells following dox induction. Probes to exon 1, the CFP coding region (exon 3), or the MS2 region (3′UTR) of the E3 gene, detected the active transcription sites as well as cytoplasmic mRNAs in induced cells, and some nuclear signal, suggesting robust nucleo-cytoplasmic transport of the mRNAs to the cytoplasm (Figure 1E). No RNA FISH signal was seen in un-induced cells. Video S1 shows the activation of transcription, nucleo-cytoplasmatic transport of the tagged mRNA, and the appearance of the CFP peroxisomes in the cytoplasm. Altogether, we generated a series of cell lines in which the kinetics of transcription could be monitored on integrated genes in real-time.

Co-Transcriptional Splicing Occurs at the Site of Transcription

Probes to the first or second intron regions displayed signal only at the sites of transcription without additional signal in the nucleoplasm or cytoplasm (Figure 2A,B), showing that the splicing of introns occurred at the site of transcription. In order for co-transcriptional splicing to occur, splicing factors must be recruited to the pre-mRNA either by diffusing from the nucleoplasm or possibly via the CTD of RNA Pol II. A previous proteomic study immunopurified more than 100 proteins associated with RNA Pol II in a non-transcribing extract [46], among them U1 snRNP components and factors of the SR family. In order to assay the recruitment of splicing factors in vivo we examined whether different factors were enriched on the active E3 and E4 gene arrays, using immunofluorescence to endogenous proteins or expression of fluorescently-tagged versions of these proteins. snRNPs, splicing factors, and other RNA processing factors were detected (Figures S2C–J and S3). These data are summarized in Table 1. In addition, using RNA FISH probes specific to the different snRNAs we could detect U1, U2, U4, U5, and U6 snRNAs at the site (Figures 2C–E, S2D–E, and Table 1). Not all molecules expected to be present at the transcription sites were detected, e.g. PSF, $p54^{nrb}$, and 7SK-RNA (Figure S3B,C, and F). We conclude that co-transcriptional splicing is occurring on the pre-mRNAs associated with the active transcription sites, and that the mRNA processing factors are recruited either via association with the transcribing polymerase or directly with the nascent mRNAs.

The U1 snRNP Is Recruited to an Intronless Gene

We then tested for the recruitment of the spliceosomal U snRNPs to the intronless E1 gene. Using an antibody that recognizes the 2,2,7-trimethylguanosine (m_3G) cap of snRNA (anti-TMG) we could detect TMG staining on the transcription site of E1 (Figure S4A). RNA FISH showed that U1 snRNA was present on the site, as well as the U1A protein (Figures 2F and S4B). However, all other snRNAs or U2AF65 were not found on the E1 sites (Figure 2G–H, S4C–E and Table 1).

The recruitment of U1 snRNA to the E1 and E3 genes was measured and compared by correlating the intensity of the U1 and U2 signals on the transcription site to the same two signals within the nucleoplasm. In the nucleoplasm we expected to find co-localization of U1 and U2 snRNA signals, since spliced genes are predominant. In the case of the E3 and E4 genes, the U1/U2 correlation at the transcription site was the same as in the nucleoplasm (Figure 2I). However, in the case of the E1 gene a different correlation was observed between U1 and U2 at the transcription site compared to the U1/U2 nucleoplasmic ratio (Figure 2J). We then quantified this recruitment. The measured ratio of U1 and U2 snRNA signals on the different genes demonstrated a constant ratio between U1 and U2 on the intron-containing genes (E3, E4, and E6), while on the E1 gene the presence of U1 snRNA was dominant and U2 was not present (Figure 2K). Interestingly in a recent study, the association of U1 components on a gene construct in which splicing did not take place due to mutations in the 5′- and 3′-splice sites was demonstrated [47]. We therefore conclude that U1 snRNPs are associated with the transcriptional machinery on all transcribing genes, no matter whether the pre-mRNA contains introns or not, probably through interactions with the CTD of the active polymerase [46,47].

The Effect of Co-Transcriptional Splicing on Pol II Elongation Kinetics

After showing that pre-mRNA splicing is taking place at the site of transcription in the E3 and E4 genes, we examined whether polymerase elongation kinetics differed between the intronless and intron-containing genes. We used the previously described in vivo elongation assay [34,35,41] to measure the kinetics of RNA Pol II on the three genes. In this assay, active transcription sites are detected with the YFP-MS2 coat protein. This fluorescent signal at the transcription site consists of the steady state kinetics of mRNA synthesis as well as mRNA release from the site. Using a fluorescence recovery after photobleaching (FRAP) approach, the YFP-MS2 signal at the transcription site is photobleached and the recovery of the YFP-MS2 signal is monitored over time. Fluorescence recovery signifies the generation of new MS2 stem-loops in the nascent transcripts and their binding by fluorescent YFP-MS2 molecules entering from the surrounding nucleoplasm (Figure 3A and 3B).

This same approach was implemented on the E1, E3, and E4 genes. The cell lines were co-transfected with RFP-LacI and YFP-MS2, and transcription was induced by dox for at least 6 h until transcription reached steady state. Active transcription sites at steady state did not show any significant difference in YFP-MS2 intensity for at least 20 min. Using a 3D-FRAP microscope system we could photobleach the transcription site (YFP-MS2 signal) and continue to detect the gene locus using the RFP-LacI DNA label. The advantage of using the 3D-FRAP microscope is that rapid

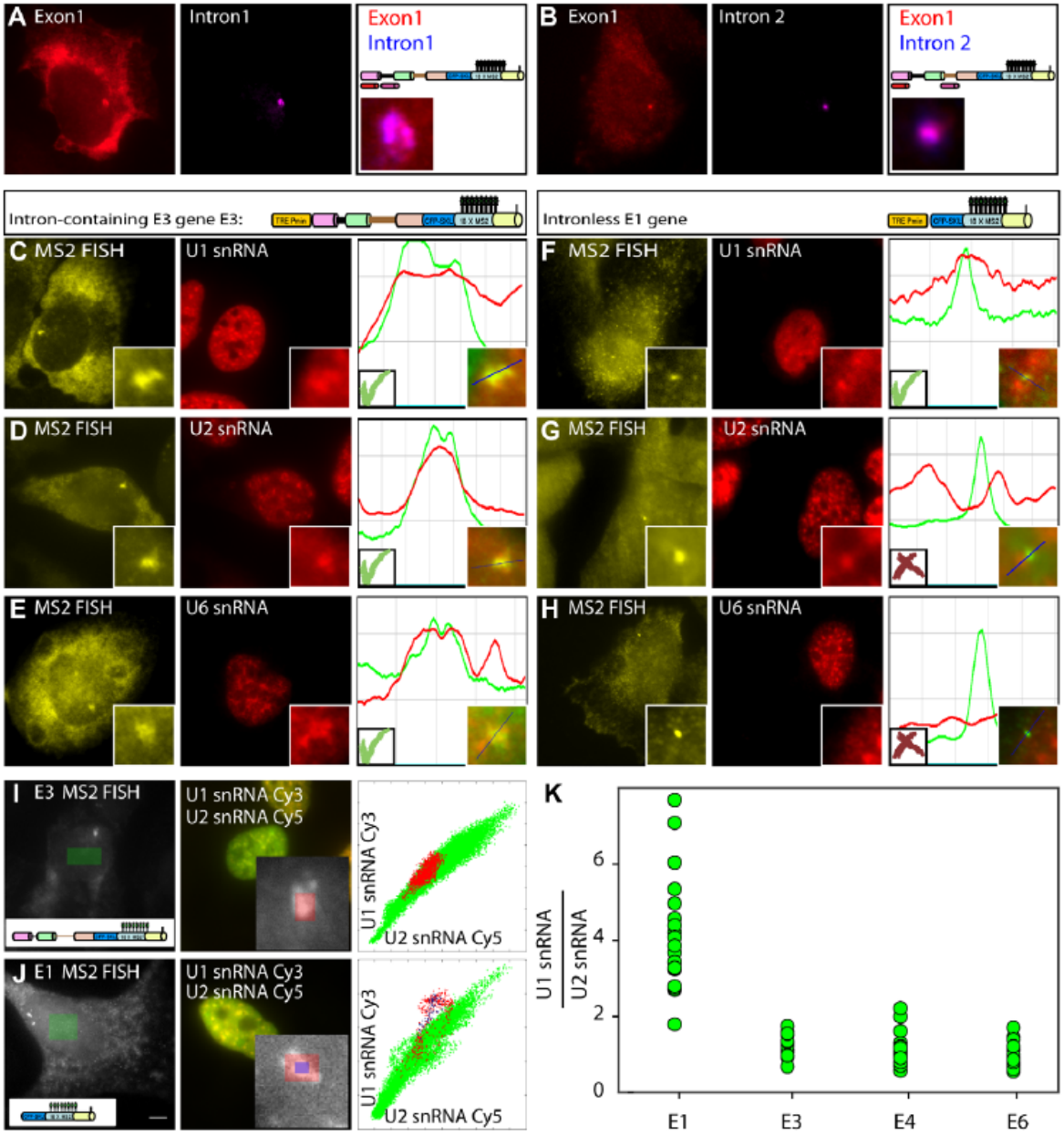


Figure 2. Co-transcriptional splicing occurs at the site of transcription. (A) A probe to the first intron or (B) the second intron (purple) was detected in an E3 cell only at the transcription site while the mature mRNA was detected throughout the cell (probe to exon 1, red). Schemes on the right show the probe binding regions and an enlargement of the colocalized signals at the transcription site. (C–H) The recruitment of the core pre-mRNA splicing machinery to active transcription sites was examined in E3 (left column) and E1 (right column) cells. RNA-FISH with an MS2 probe (yellow) shows the active transcription sites together with snRNA probes in red: (C and F) U1 snRNA, (D and G) U2 snRNA, and (E and H) U6 snRNA. Enlargements show the signals at the transcription sites. Plots on the right depict the degree of colocalization of the signals at the transcription site across the blue line. (I) RNA-FISH signal of U1 and U2 snRNAs (middle) at the active transcription site (seen with the MS2 probe, left) of an E3 cell and (J) an E1 cell were measured. The green box shows a random nucleoplasmic area, and the red and blue boxes show the area of the transcription site. The plot shows the correlation of the U1 and U2 snRNA signals at the transcription site (red and blue dots) in comparison to the nucleoplasm (green dots) (bar, 5 μm). (K) The ratio between U1 and U2 snRNA signals measured by RNA-FISH, at the transcription sites of all genes. The data in (I), (J), and (K) show a constant correlation in the intron-containing genes, and no correlation in the intronless E1 gene, meaning that U1 and not U2 is recruited even to an intronless gene.

doi:10.1371/journal.pbio.1000573.g002

Table 1. Examination of the proteins and RNAs that are recruited to intron-containing and intronless genes.

Recruited to Intron-Containing Transcription Sites (E3/E4/E6)	Not Recruited to Intron-Containing Transcription Sites (E3/E4/E6)	Recruited to an Intronless Gene (E1)	Not Recruited to an Intronless Gene (E1)
Active polymerase II (H14)	PSF	Active polymerase II (H14)	U2 snRNA
Active polymerase II (H5)	$p54^{nrb}$	Active polymerase II (H5)	U4 snRNA
Polymerase II CTD repeats (8WG16)	7SK RNA	Polymerase II CTD repeats (8WG16)	U5 snRNA
2,2,7-trimethylguanosine (m_3G) (anti-TMG)		2,2,7-trimethylguanosine (m_3G) (anti-TMG)	U6 snRNA
U1 snRNA		U1 snRNA	U2AF65
U2 snRNA		U1A	
U4 snRNA			
U5 snRNA			
U6 snRNA			
U1A			
U2AF65			
SF2/ASF			
SC35			
9G8			
TLS			
PTB			
Cyclin-T1			

doi:10.1371/journal.pbio.1000573.t001

3 dimensional (3D)-imaging over time (4D) could be performed. Thereby, the signal of the whole nuclear volume could be collected with no effect on the measurements, and motion of the transcription site was easily tracked (Figure 3A and Video S2). This is in comparison to most confocal microscopes that collect the information only in 2 dimensions, and therefore measurements would be sensitive to any DNA motion in the *z*-axis.

Transcription sites were followed for extended periods (up to 30 min). The YFP-MS2 signal intensity on the transcription sites was quantified using a tracking algorithm. We measured recovery kinetics on the transcription sites of all the three gene-types in order to examine whether different elongation kinetics could be identified. We compared a polymerase passing through the MS2 region of an intronless E1 gene to a polymerase moving through the MS2 region of the intron-containing E3/E4 genes that undergo co-transcriptional splicing (after the exons and introns are transcribed). The FRAP recovery curves measured on the E1, E3, and E4 genes all showed similar recovery kinetics (Figure 3C), indicating that the polymerase transcribing the MS2 region was not affected by the presence of active splicing from the upstream sequences. We conclude that upstream splicing events do not necessarily affect the elongation kinetics on our model genes. However, it is possible that other differences such as in intron sequence or intron size (e.g. the *β-globin* first intron is shorter than average) might cause some effect on elongation. These experiments were performed on several microscope stations to verify reproducibility (Figure S5). Altogether, these measurements indicate that polymerase transcription rates are not modulated by the co-transcriptionally assembled spliceosomal machinery.

Increasing Intron Numbers Leads to an Increase in Spliceosome Component Recruitment

We examined a similar gene construct in which additional introns and exons were very closely spaced. Thereby, we expected transcription to end before the completion of splicing of all introns, which would allow the further examination of elongation kinetics. The E6 gene was generated using the E4 construct and consisted of 6 exons and 5 introns by using duplications of intron 2 (Figure 1A). In this manner, we generated a gene with 6 exons (close to the 8 exon median size of genes [44]). The E6 mRNA (2,437 nt) was correctly spliced (unspliced 6,536 nt, Figure 1B), the CFP-SKL peroxisomal protein was expressed normally (Figure S1 and Video S1), and similar recruitment of splicing factors was observed (unpublished data).

First, we verified that there was an increase in accumulation of introns at the transcription site of the E6 gene compared to the other genes, as expected. 3D volumes of RNA-FISH labeled cells were collected, deconvolved, and analyzed. We calculated the fluorescence ratio of intron 2 in comparison to the last exon (CFP exon), in all the cell lines. Higher levels of introns generated were detected on the E6 gene (Figure 4A) compared to the other genes. As expected, there was a gradual accumulation of intron 2 at the active transcription sites in dependence to the number of introns in each gene.

We next asked whether there would be increased accumulation of the splicing machinery on genes containing more introns. We therefore quantified the relative amounts of the splicing machinery (U snRNAs) that accumulated on each transcript using 3D volumes of RNA-FISH labeled cells (Figure 4B and 4C). Quantification of the fluorescence ratio of U5 or U6 snRNAs in comparison to the first exon present in each nascent transcript demonstrated a higher accumulation of snRNAs that was dependent on the number of introns in the gene. This indicated that the addition of exons resulted in an increase in the recruitment of the splicing machinery per transcript. These data also imply that the elongation machinery does not wait for each splicing event to complete before moving forward on the gene, but rather continues to elongate while more spliceosomal units accumulate on the transcript.

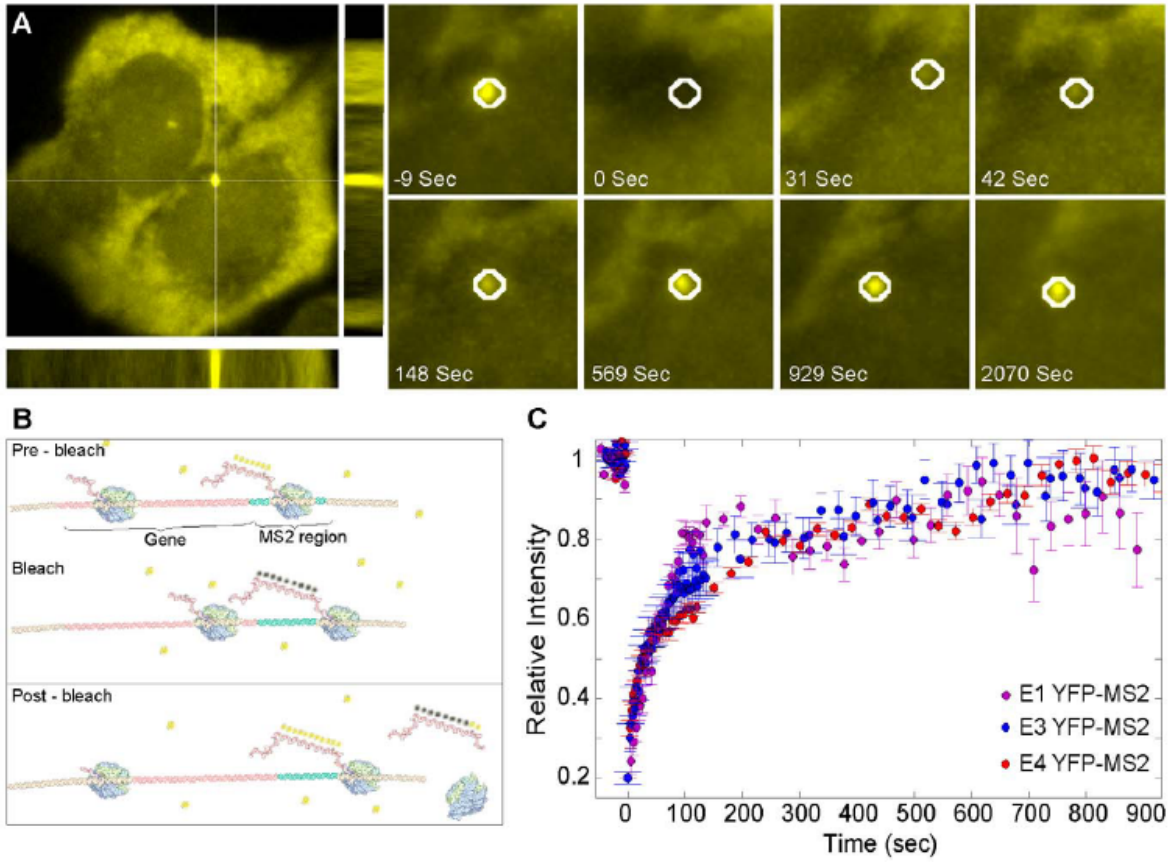


Figure 3. Comparison of transcriptional kinetics on active E1, E3, and E4 transcription sites. (A) An active transcription site (YFP-MS2, bottom cell) was photobleached and the fluorescence recovery was tracked over time (frames on right). These experiments were performed on a 3D FRAP system (XZ and YZ representations are seen under and to the right of the original image, respectively). (B) Scheme of the elongation assay showing bleaching of the YFP-MS2 proteins (yellow) bound to the mRNA and the recovery of fluorescence as new polymerases elongate through the MS2 region. (C) Recovery curves of the YFP-MS2 FRAP measurements performed on E1, E3, and E4 transcription sites. Average of $n>10$ experiments for each cell type with SDE.
doi:10.1371/journal.pbio.1000573.g003

The E6 Gene Exhibits Slower Recovery Kinetics

Next, polymerase kinetics on the E6 gene were measured. Interestingly YFP-MS2 FRAP curves showed significantly slower recovery kinetics compared to the previous measurements on the E1, E3, and E4 genes (Figure 4D). In order to verify that this was not a clone-related difference, and that the measurements were not affected by the number of gene units in the tandem array or from the DNA integration site, we compared two different E6 clones, E6-22 (40 gene copies) to clone E6-3 (9 gene copies). The copy number of gene integration in each cell line was quantified by real-time PCR on genomic DNA. We found that both E6 clones recovered with identical and relatively slower kinetics (Figure S6A and S6B).

To identify the cause for the kinetic difference between the genes, we analyzed the measured FRAP recovery curves. Two previous studies have used this MS2 FRAP approach to measure Pol II transcription rates on specific genes [34,41]. The assay measures the recovery of fluorescence as the polymerase transcribes the MS2 region right after the exons and introns are transcribed (Figure 3B). The recovery signal shows bi-phasic kinetics (Figure S7A), which best fit to two exponentials. This implies that the measurements are detecting two processes that are occurring in parallel, which require the interpretation of the significance of each phase in the curves. The abovementioned studies have used extensive modeling to study such FRAP curves and demonstrated that the fast phase of fluorescence recovery is generated by the actively elongating polymerases. The slow phase of recovery was explained as either a population of polymerases in a paused state along the gene [34] or a termination delay [41]. The slow kinetics on the E6 gene could therefore be caused either by polymerase slowdown or by retention of the mRNA after termination. This was further tested.

We first examined whether the kinetic effect was due to a change in polymerase activity. FRAP of the active polymerase transcribing on the E3 and E6 genes was analyzed using a transfected GFP-Pol II that was recruited to the active genes. Similar FRAP recovery kinetics were measured on both genes (Figure 5A). This measurement differs from the above measurement of YFP-MS2 FRAP recovery. While the latter measures elongation through the MS2 region only, Pol II FRAP encompasses the whole polymerase transcriptional process on the gene, ranging from initiation through elongation and

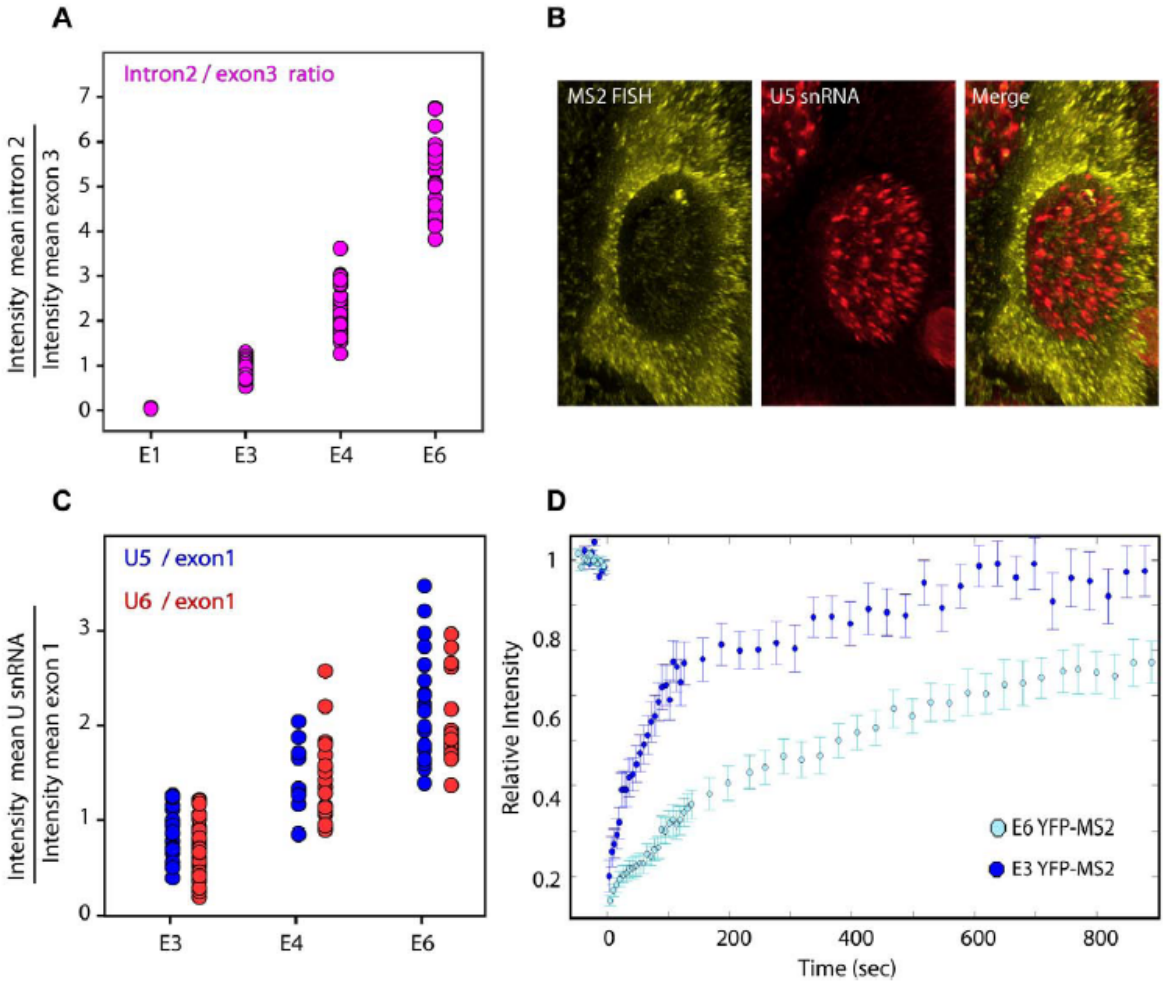


Figure 4. Increased spliceosome recruitment and change in transcriptional kinetics in relation to increasing intron numbers. (A) Quantification of the RNA-FISH signals showing the ratio of the second intron to the last exon (CFP region) on the different transcription sites (all normalized to the E3 ratio). (B) A 3D representation of the deconvolved images of an RNA-FISH experiment in which MS2 mRNA (in yellow) and U5 snRNA (in red) were detected. The signals at the transcription site were measured for both channels. The elongated form of the signal is due to the 3D projection of the images. (C) Quantification of the RNA-FISH signal showing the ratio between transcription site associated U5 (blue dots) or U6 (red dots) snRNAs in comparison to exon 1 on the E3, E4, and E6 genes. Each dot represents a transcription site ratio. (D) FRAP recovery curves of YFP-MS2 on E3 versus E6 genes with standard error bars (n>10).
doi:10.1371/journal.pbio.1000573.g004

termination [34]. The Pol II FRAP measurements (Figure 5A) indicated that the polymerases on E3 and E6 genes were functioning with similar kinetics with regard to the general transcriptional flow. However, since only the very slow recovery phase is attributed to the elongating phase of the polymerase (consists of only <30% of the signal) [34], it is difficult to resolve between the polymerase kinetics on E3 and E6. Therefore, we obtained chromatin immunoprecipitation (ChIP) information to compare the distribution of Pol II along the genes, which might detect regions of paused polymerases. Using pull-down of the polymerase with an anti-Pol II antibody followed by real-time PCR on 7 different gene regions, we corroborated the polymerase FRAP findings and demonstrated that there was no statistical differences in the density of Pol II along the two genes (Figure 5B). Additionally, from these data we can learn that the highest Pol II density is found at the promoter and the polyA site, which is in agreement with other studies [16]. Altogether, these results showed that the detected kinetic change in E6 was not directly connected to the polymerase.

Splicing Inhibition Relieves the Slow Kinetics on the E6 Gene

To examine whether the E6 kinetic delay was due to ongoing splicing, we used two splicing inhibitors, Spliceostatin A (SSA) [48] and Meayamycin [49]. SSA and Meayamycin are inhibitors of pre-mRNA splicing via SF3b binding, which binds to the intron branch-point and acts as a subcomplex in the U2 snRNP. SSA treatment caused the accumulation and retention of unspliced pre-mRNA in the nucleoplasm, which continued to accumulate in nuclear speckles, as detected by RNA-FISH with an intron probe (Figure 5C), as previously described [48]. We now show this same

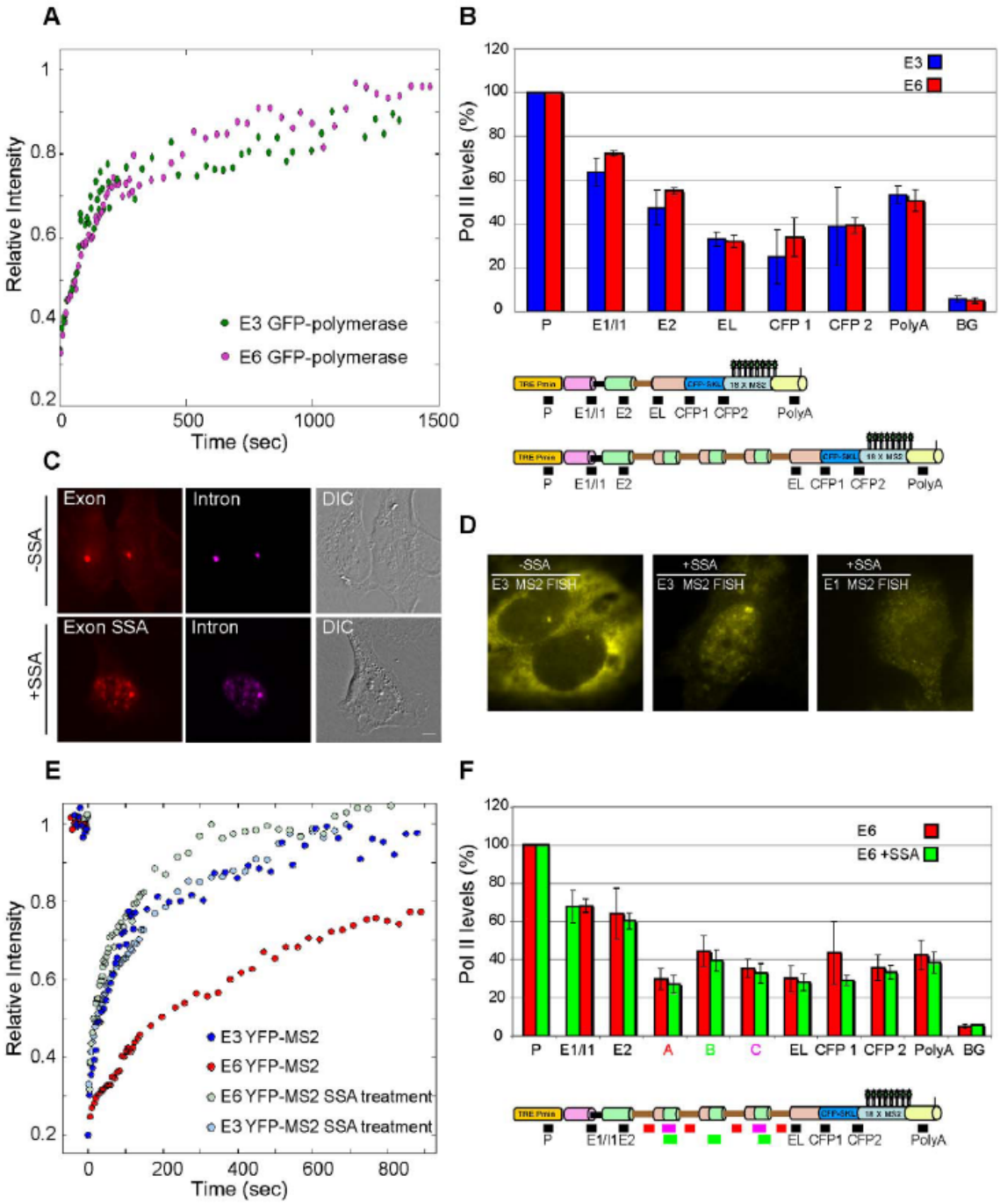


Figure 5. Dynamics of E6 mRNA are affected by active splicing and not by polymerase kinetics. (A) FRAP recovery curves of transfected GFP-Pol II recruited to the E3 and E6 genes. (B) ChIP analysis of Pol II distribution along the E3 (blue bars) and E6 (red bars) genes. Black boxes under the schemes indicate primer positions. The data were normalized to the promoter by setting the promoter value at 100%. A *t*-test for each primer set showed no significant differences ($p>0.05$). (C) Inhibition of splicing by Spliceostatin (SSA) for 9 h showed that unspliced pre-mRNA (intron probe, magenta) was distributed throughout the cell and in speckles (RNA-FISH on E6 cells) and was retained in the nucleus (bottom), whereas in untreated cells pre-mRNA was detected only at the site of transcription (intron probe, magenta; exon probe, red) (bar, 5 μm). (D) SSA treatment (6 h) caused the nuclear retention of intron-containing E3 mRNA (middle) versus the regular nucleo-cytoplasmatic dispersal of E1 intronless mRNA (right). Left,

untreated E3 cell. RNA-FISH was performed to the MS2 region of the mRNAs (yellow). (E) FRAP recovery curves of YFP-MS2 at the transcription sites of untreated and SSA-treated E3 and E6 cells. (F) ChIP analysis of Pol II distribution along the E6 gene in untreated (red bars) and SSA-treated (green bars) cells. Boxes under the schemes indicate primer positions. The data were normalized to the promoter by setting the promoter value at 100%. A *t*-test for each primer set showed no significant differences ($p>0.05$).
doi:10.1371/journal.pbio.1000573.g005

effect also for Meayamycin (Figure S8A). Interestingly, the dispersal of the intronless E1 mRNA was not affected, and it did not accumulate in speckles (Figure 5D). Then we measured YFP-MS2 recoveries by FRAP on the E6 gene in untreated cells as described above and incubated the same transcribing cells with either SSA or with Meayamycin. Under both treatments the slow kinetics of E6 were transformed to the regular kinetics of E3/E4/E1 genes (Figure 5E and S8B), whereas splicing inhibition did not cause any effect on the kinetics of the E3 gene (Figure 5E). On the other hand, FRAP of GFP-Pol II in the presence of SSA on the E6 gene (Figure S8C), and ChIP experiments in the presence of SSA (Figure 5F) did not show any change in kinetics. These results demonstrate that splicing inhibition did not modify Pol II distribution or kinetics along the genes. This emphasized that the changes in kinetics between E6 and E3 genes were due to active splicing events and were not connected to the elongation activity of the polymerase.

RNA Splicing Can Lead to Retention of the Pre-mRNA on the Transcription Site

The above experiments suggested that the slow recovery FRAP curves of the YFP-MS2 experiments were due to a delay of the nascent transcripts on the E6 sites caused by splicing but without connection to the polymerase itself. First, we verified that the full transcripts were 3′-end processed and were present at the site of transcription of all the genes prior to release. This was shown by the recruitment of GFP-PABP2 to the active transcription sites and by the presence of polyA tails as seen by RNA-FISH (Figure 6A and 6B). RNA-FISH measurements of the ratio of polyA signal to the last exon (CFP) on the sites showed that a constant ratio was retained on all the transcription sites (Figure 6C), meaning that there was no accumulation of mRNAs lacking polyA on the different genes and no change in polyadenylation levels. This suggested that there was an accumulation of transcripts that were not attached to Pol II, at the transcription sites.

To characterize the delay time of the mRNAs on the E6 transcription sites, we next examined the kinetics of mRNA release from the gene. Actinomycin D (ActD) was added to actively transcribing cells containing YFP-MS2 tagged transcription sites. The "shut-down" of the transcription sites was monitored over time under the microscope. Indeed, slower clearing of the mRNAs from the gene was observed in E6 cells ($t_{1/2}$ = 17 min) compared to E3 cells ($t_{1/2}$ = 7 min) and E4 ($t_{1/2}$ = 8.9 min) (Figure 6D, Video S3). This experiment correlates with the YFP-MS2 FRAP measurements (Figure 4D) showing slower mRNA kinetics on the E6 gene.

Finally, to demonstrate that only the mRNAs disconnected from the polymerase were stalled on the E6 gene, we quantified the intensity ratio of the polymerase signal and the mRNA signal on the transcription sites using anti-Pol II immunofluorescence and RNA-FISH to the first exon. If indeed each mRNA is attached to one polymerase, we would expect a constant ratio for all cell lines. However, we found that the proportion of mRNA to Pol II on E6 transcription sites was three times higher than on E3 and E4 sites (Figure 6E). When SSA was added and splicing was inhibited, the mRNA/Pol ratio on the E6 sites was reduced to the levels of E3 and E4 sites, meaning that under conditions of splicing inhibition, the unspliced E6 pre-mRNAs were being released without stalling similarly to the E3 and E4 mRNAs. These experiments implied that in a situation like the E6 gene which contains closely spaced introns, and in which the polymerase has finished transcribing yet the pre-mRNA is not fully spliced, the mRNA is retained at the transcription site until the completion of splicing.

To better understand the given observed results we simulated the transcriptional process using a stochastic Monte-Carlo simulation (Figure S9A and Video S4). The simulation mimicked the transcriptional process based on the transcriptional models described in the previous studies [34,41]. Kinetic parameters were obtained by fitting the simulations to the experimental FRAP curves of E3 (Figures 6F). See Material and Methods for explanation of parameter scanning. The parameters that gave the best fit were: an elongation speed of 3.6 kb/min, agreeing with previous measurements [21,22,36]; ~42% of the polymerases had entered a paused state for an average time of 6 min [34]; and an average retention time (after end of elongation) of 50 s before release from the transcription site (Figures S9 and S10) [41]. We found that the number of genes in the array and the initiation rate had no effect on the measured kinetic results, due to the normalization of the data to the steady-state pre-bleach signal. This observation was confirmed experimentally by measuring FRAP experiments with low amounts of DRB to abrogate elongation. Indeed, no kinetic effect on E3 or E6 YFP-MS2 FRAP recoveries was observed (Figure S11), while the addition of SSA even in the presence of DRB transformed E6 kinetics to E3 kinetics, signifying the effect of splicing on the retention of mRNA at the site of transcription.

We then searched for a reliable fit to the E6 data by keeping the kinetic parameters of E3 and changing only the time-scale of transcript release. We found that the average time for a transcript to remain at the E6 site of transcription was on average ~10 min (Figure 6F). This is in agreement with the actinomycin D experiment that showed a 10 min difference in transcript clearance times between E3 and E6 (Figure 6D). We also used the simulations to calculate the ratio of polymerases to nascent transcripts on the active genes, and compared to the experimental data. The simulation showed a 3-fold increase in this ratio on the E6 gene, similar to the result obtained with quantitative FISH (Figure 6E).

We then used this simulation to examine the intron/exon ratio on the different genes. In this experiment (Figure 4A) we found a gradual increase in intron signal at the transcription sites that correlated to the number of introns in each of the genes. The simulation output was compared to these RNA-FISH measurements. The simulation provided a ratio between the number of polymerases/transcripts present on the duplicated intron 2 region and the number of transcripts that had already passed through the CFP region. Simulations were run for the different genes with the abovementioned kinetic parameters assigned to E3 (short retention time). Since this experiment follows the polymerase moving along the genes (= elongation), we expected an increase in the intron/exon ratio from E1 to E6 genes. Indeed, the simulation showed the same increment in intron/transcript ratio as seen in the experimental results (compared in Figure 6G). However, when the calculated long retention kinetics at the end point of the gene was included in the simulation of E6 kinetics, we obtained an output that showed a low intron/exon ratio, which did not agree

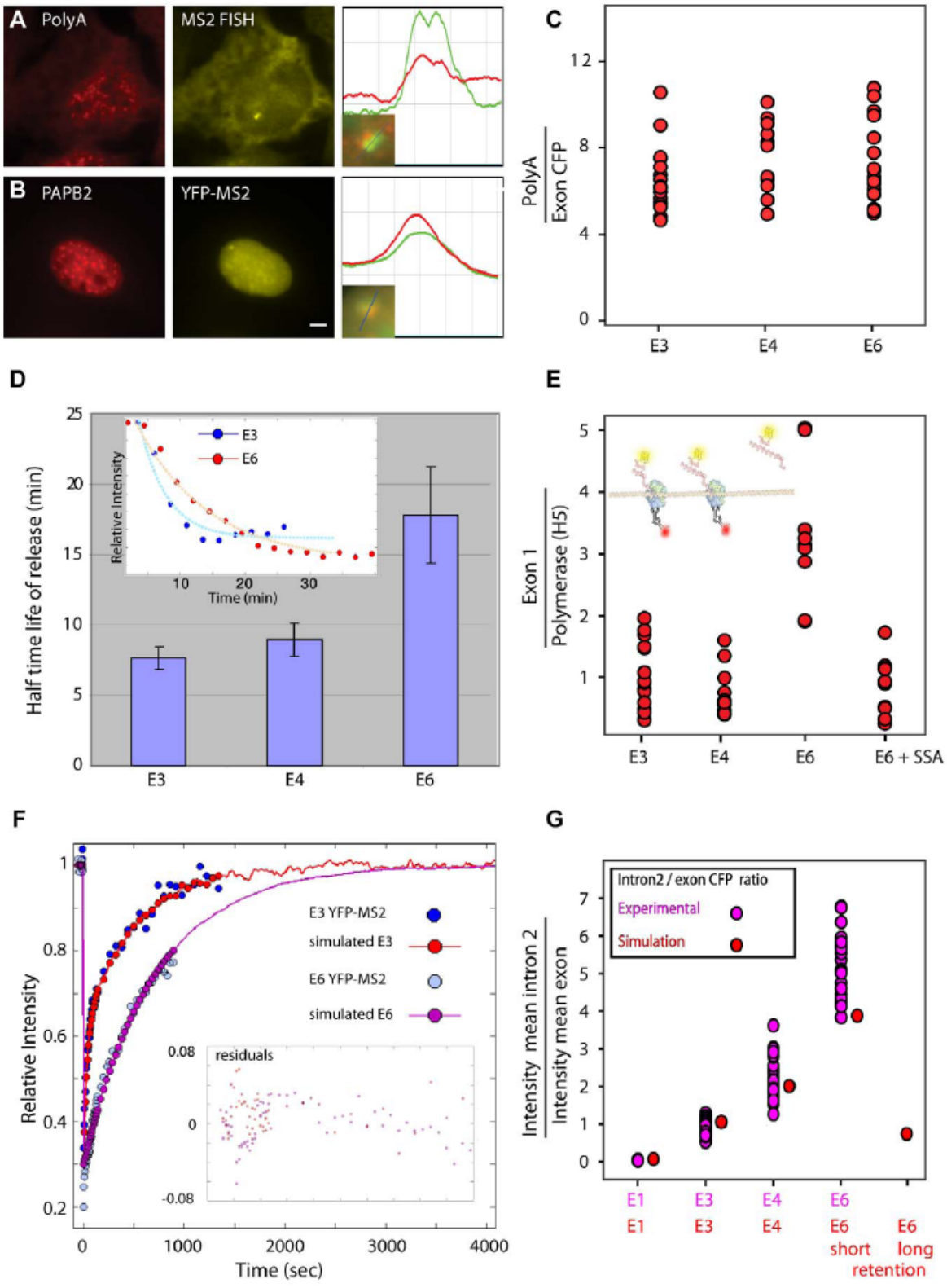

A
PolyA
MS2 FISH
B
PAPB2
YFP-MS2
C
PolyA / Exon CFP
E3
E4
E6
D
Half time life of release (min)
Relative Intensity
Time (min)
E3
E4
E6
E
Exon 1 / Polymerase (H5)
E3
E4
E6
E6 + SSA
F
Relative Intensity
Time (sec)
E3 YFP-MS2
simulated E3
E6 YFP-MS2
simulated E6
residuals
G
Intensity mean intron 2 / Intensity mean exon
Intron2 / exon CFP ratio
Experimental
Simulation
E1
E1
E3
E3
E4
E4
E6
E6 short retention
E6 long retention

Figure 6. RNA splicing leads to retention of the pre-mRNA on the transcription site. (A) RNA FISH in an E3 cell with a polyT probe (red) and an MS2 probe (yellow) showing that polyadenylated mRNAs were formed and were present on the transcription sites. (B) Recruitment of GFP-PABP2 (pseudocolored red) to the transcription site labeled with YFP-MS2 (yellow). Plots on the right in (A) and (B) depict the degree of colocalization of the signals at the transcription site across the line (bar, 5 μm). (C) The ratio of the RNA-FISH signal of polyA to the last exon on the E3, E4, and E6 active transcription sites, showing no aberrant accumulation of transcripts without polyA signal. (D) Half-time of release of mRNA from transcription sites following actinomycin D treatment (time measured from ActD addition) in E3 (n = 7) and E4 (n = 7) cells compared to E6 cells (n = 8). Inset shows two sample time-series of transcription site inactivation and the exponential fit from which half-times of clearance were calculated (blue, E3; red, E6). (E) The ratio between the polymerase signal and the first exon on the E3, E4, E6, and SSA-treated E6 genes. Each red dot represents the ratio of intensities measured on a single transcription site. The scheme at the top shows the polymerase antibody (red) and the mRNA (FISH exon 1, yellow). (F) YFP-MS2 FRAP data and the fitted simulation of the E3 and E6 genes. The calculated residuals of the fit are presented in the inset. (G) RNA-FISH ratio of intron 2 to the last exon at the transcription site in E1, E3, E4, and E6 cells. The experimental data points (pink, from Figure 4A) were compared to the simulation output (red) of the same experiment. Simulations were performed with either short (50 sec, E3 kinetics; E1, E3, E4, E6) or long (10 min, E6 kinetics; E6 rightmost point) transcript retention times.
doi:10.1371/journal.pbio.1000573.g006

with the experimental data (Figure 6G). Therefore, the continuous flow of intron generation observed with the simulation, in contrast with the slow kinetics of the transcript, imply that the release time of the transcript from the gene site depends on splicing and not on the polymerase kinetics.

Discussion

Gene expression profiles are influenced by the presence or lack of an intron. It has been known for many years that the addition of an intron into the cDNA of a gene sequence enhances its expression [50]. Indeed, the act of intron removal can have effects on the initiation of transcription, mRNA polyadenylation, mRNA export, and the cytoplasmic fate of the transcript such as translation, mRNA localization, and mRNA turnover [51]. For instance, several studies have demonstrated that the presence of an intron in the pre-mRNA and subsequent splicing can elevate the rates of mRNA export and thereby the accumulation of message in the cytoplasm [52–54]. Hence, the presence of an intron exerts an overall positive influence on gene expression.

How does intron removal affect the levels of transcription and mRNA production? Some level of regulation can be found on the DNA itself, for instance by intronic sequences that modulate nucleosome assembly in the vicinity of the promoter [55]. Splicing signals in the 5′ promoter-proximal intron can stimulate polymerase II initiation, processivity, and recruitment of transcription factors [56–59], and complexes containing transcription and splicing factors have been identified [60]. Nonetheless, the increase of Pol II activity on intron-containing genes versus intronless genes was modest [56]. Recently, the depletion of the SR protein SC35 in cells was shown to attenuate the elongation of Pol II within the gene sequence [61].

In this study we wished to examine the effects of splicing on Pol II elongation in vivo, independent of promoter influences and transcriptional initiation events. Therefore, all the gene constructs were controlled by the same inducible promoter, and elongation kinetics were examined within the gene as far as possible from the initiation site. Elongation rates were measured immediately downstream of the region transcribing the exon-intron-exon elements, or in other words we could monitor the polymerase speed as it moved along the DNA after the splicing codes in the pre-mRNA had already been synthesized. This meant that polymerase movement towards the 3′-end of the gene was occurring simultaneously with splicing events accompanying the newly synthesized pre-mRNA. We measured intron accumulation on the active transcription sites that was dependent on intron numbers.

The presence of various splicing machinery factors at the transcription sites was detected only when the gene loci were actively transcribing. Furthermore, RNA-FISH to intronic and exonic sequences demonstrated the presence of the introns only at the sites of transcription and not in the nucleoplasm, signifying that this is the location of the removal of these introns. However, when splicing was inhibited with Spliceostatin or Meayamycin, intron-containing pre-mRNA could be detected also throughout the nucleus, in particular within nuclear speckles, as previously described for SSA [48]. The intronless E1 mRNA was not affected by these treatments. Splicing factors associated with the actively transcribing genes belonged to the U snRNP family, including snRNAs (U1, U2, U4, U5, U6) and protein components (U1A, U1-70K, U2AF65). Also SR proteins such as SC35, SF2/ASF, and 9G8 were recruited to the active transcription sites upon activation. Other factors detected at the sites were the transcription/splicing-related protein TLS/FUS, splicing regulator PTB, the elongation regulator cyclin T1, and polyA-binding protein 2 (PABP2). However, factors like PSF and $p54^{nrb}$ [62] previously demonstrated to be associated with the Pol II CTD [63] were not detected at the transcribing genes, nor were they pulled down with a purified polymerase [46]. Distinguishing between the two approaches might hint to transcription factor dynamics. While purified protein complexes can give an idea of the "presence" of a factor in a transcription complex, the imaging assays can indicate the longevity of "residence time" at the active gene. The fact that these factors were not identified accumulating on the transcribing gene arrays using antibodies to the endogenous proteins could indicate short-lived or low-affinity interactions with other factors that do associate with the actively transcribing gene.

In this study we examined spliceosomal recruitment to comparable intron-containing genes as well as an intronless gene. As anticipated, spliceosomal factors were not found on the active transcription site of an intronless gene, except for components of the U1 snRNP. These data are in agreement with a recent report that quantified the recruitment of splicing factors to an intron-containing gene in which the splice sites were mutated and therefore did not undergo splicing [64]. In that study the recruitment of U1 snRNA and U1-70K were detected on the gene. In our analysis the intronless E1 gene does not contain any splice sites and therefore strengthens the notion that U1 snRNP is associated with the CTD of Pol II by default and presumably accompanies the polymerase during transcription in order to scan for the presence of a 5′ splice site. Only when such positive identification of splicing signals has occurred will the rest of the splicing machinery be recruited to the pre-mRNA.

We quantified an increased accumulation of spliceosomal components on the pre-mRNA that was dependent on the number of introns within the gene sequence. These observations assist in understanding the dynamics of spliceosome assembly on multiple introns and would be consistent with the scenario of a step-wise model of spliceosome assembly on each intron [65], as well as exon-tethering by the polymerase [66]. Our live-cell analysis suggests that the polymerase does not wait for splicing to

take place but rather runs ahead such that more introns accumulate. Together with the exon-tethering model it would mean that exons are being held through the CTD and thereby recruiting the spliceosome for each splice site.

The kinetic approach we used to measure Pol II transcription kinetics enables the quantification of elongation that is free of influences of initiation at the promoter [34]. The region in which elongation is measured is the MS2 portion of the mRNA in which the generation of each additional MS2 stem-loop results in an increase of YFP-MS2 signals on the transcription site. We inserted the MS2 region downstream of the exon-intron-exon modules and at the 3′-end of the mRNA, thereby monitoring the elongating polymerase when associated with the splicing machinery. When we compared the elongation kinetics of the polymerase moving through the MS2 region after transcribing an E3 or E4 mRNA, we found that the kinetics were similar and in fact did not differ than when transcribing an intronless E1 mRNA. A recent study examining transcription of a variety of endogenous genes along intronic and exonic regions using quantitative RT-PCR has shown that polymerase rates were similar on all genes and did not differ when transcribing intron or exon regions [21].

By generating a gene with many and closely spaced introns and exons (E6), we created a situation in which the transcription machinery should have reached the end of the gene before full splicing of the transcript was completed. Although the generation of the E6 and E4 genes required the duplication of intron 2 and the flanking splice sites, we did not find that this had an effect on the splicing outcome (RT-PCR) of E6 and E4 pre-mRNAs, nor on the expression levels of the CFP-peroxisome protein product. Also, there was no indication of aberrant localization of the E6 mRNA on the transcription site or in nuclear speckles, which could be indicative of aberrant mRNA production.

The FRAP recovery kinetics of the E6 gene were significantly slower than the E3 and E4 genes. In order to examine whether the polymerase was the cause of these slow kinetics, or if the splicing machinery was responsible, we conducted a series of experiments. Using splicing inhibitors we saw that the impediment of the slow kinetics was released and returned to the levels of E3/E4/E1 genes, demonstrating that the slow-down was due to an active splicing environment. On the other hand, GFP-Pol II recovery kinetics and ChIP distribution patterns demonstrated that the polymerase was not paused or hindered during elongation. Use of actinomycin D that inhibits transcription by intercalating into the DNA and obstructing the path of the polymerase demonstrated that the clearance of the mRNAs from the E6 gene was slower. These transcripts were polyadenylated as seen by recruitment of PABP2 and the presence of polyA tails. The four different transcripts showed polyA accumulation on the transcription sites implying that the transcripts remained attached to the gene until 3′-end processing occurred. The quantification of the ratio of polymerase to nascent transcripts showed that there was accumulation of pre-mRNAs on the E6 gene, but not on the other genes. This accumulation was alleviated when splicing was inhibited. Together with parameter estimation by the simulation we demonstrate that unspliced E6 transcripts are retained at the site of transcription for up to 10 min before being released.

The plausibility of a co-transcriptional splicing mechanism can be imagined just by the fact that human gene exons tend to be short compared to extremely larger intronic sequences [67], and genes contain an average of 5–12 exons [68]. This means that if co-transcriptional splicing did not take place, then the splicing machinery would have to hold back until transcription has completed. This would not make much biological sense, especially when envisaging the transcriptional timeframe of very large genes such as the dystrophin gene, which lies in the range of 16 h [69]. In fact it has been proposed that the immediate association of U1 snRNP and splicing factors with the emerging pre-mRNA is a means to block the association with hnRNPs, which are splicing inhibitory proteins [46]. We demonstrated that the polymerase activity is uncoupled from splicing once transcription has terminated. In a scenario in which transcription has ended and splicing continues, the pre-mRNA and splicing machinery are retained at the site of transcription until the completion of splicing.

Materials and Methods

Plasmid Construction

E4 generation: The E3 construct (termed E3-minus in [37]) was digested with BstXI. Re-circularization was prevented by Shrimp Alkaline Phosphatase (Fermentas). An adaptor composed of the hybridized A1 & A2 sequences was ligated using T4 DNA ligase (5u/µl, Fermentas, Ontario, Canada). A1:TTGATATCACC-GGTGGGCCCGGAAAAGAATTCGTCCATCACT

A2:ATGGACGAATTCTTTTCCGGGCCCACCGGTGAT-ATCAAAGTG

The adaptor was EcoRV/BstXI digested, and a PmlI/BstXI intron2-exon3 insert that was separately removed from E3 was then ligated into the adaptor to generate E4.

E6 was constructed using the same strategy but on the E4 construct. E1 was generated by first digesting the E3 plasmid BstXI/SacII, followed by ligation with an adaptor composed of A9 & A10.

A9:AAGCTTTGATCAGCGGCCGCATCACT

A10:GGTTCGAAACTAGTCGCCGGCGTA

RNA extractions were carried out with the RNeasy mini kit (Qiagen). cDNA was synthesized using the RevertAid™ First Strand cDNA Synthesis Kit and an oligo(dT) primer (Fermentas). RT-PCR and sequencing of the correctly spliced mRNAs were performed with primers to the first exon and last exon regions.

A3: AGCAACCTCAAACAGACACC

A4: GGTCTTGTAGTTGCCGTCGT

Other plasmids used in this study were: RFP-LacI, CFP-LacI, YFP-MS2(-NLS), GFP-Pol II [34], (NES-)YFP-MS2(-NLS) [36], MS2-mCherry, GFP-PABP2, GFP-PTB, GFP-TLS, and GFP-cyclin T1.

Cell Culture and Transfections

Human U2OS osteosarcoma Tet-On cells (Clontech) were maintained in low glucose Dulbecco's modified Eagle's medium (DMEM, Biological Industries, Israel) containing 10% fetal bovine serum (FBS, HyClone, Logan, UT). For transient transfections, cells were transfected with 1–5 µg of plasmid DNA and 40 µg of sheared salmon sperm DNA (Sigma) using electroporation (Gene Pulser Xcell, Bio-Rad, Hercules, CA) or with FuGENE 6 transfection reagent (Roche). Stable co-transfections were performed by electroporation of the E genes together with a 256 *lacO* repeat plasmid and a puromycin resistance plasmid using a 1:10:10 ratio (puromycin res., *lacO*, E gene). Cells were cultured for 1–2 d without selection and then selected for 2–3 wk with 100 µg/ml puromycin (AG scientific, San Diego, CA). Single colonies were picked and positive clones were selected after the addition of doxycycline (1 µg/ml) and identification of CFP-peroxisomes and by transient transfection of RFP-LacI and YFP-MS2. For drug treatments cells were imaged before and right after the addition of 5 mg/ml actinomycin D (Sigma). For DRB treatment, cells were treated with 20 µg/µl for 2 h before imaging. For splicing inhibition, cells were photobleached before and after incubation with 100 ng/ml of Spliceostatin for 6 h, kindly

provided by Minoru Yoshida (RIKEN Advanced Science Institute, Japan), or Meayamycin (300 nM) for 6 h before imaging, kindly provided by Kazunori Koide (University of Pittsburgh, USA).

qPCR

gDNA was isolated using the PUREGENE™ DNA Purification System (Gentra). For the generation of standards a 10-fold dilution was prepared. Primers used for the amplification of CFP region were GGATCACTCTCGGCATGGAC (forward) and TGCACATACCGGAGCCATTG (reverse). Primers used for the human β-globin reference gene were CAGTGCAGGCTGCCTATCAGA (forward) and GAATCCAGATGCTCAAGGCCCTT (reverse). The reaction (20 µl total volume) contained 1× SYBR green mix (Thermo scientific, Waltham, MA), 1 µl of gDNA dilution, and 0.5 mM of each primer and water. The reaction was performed using the Chromo4™ Real-Time Detector (Biorad) following the protocol: 15 min at 95°C for enzyme activation, followed by 40 cycles consisting of 15 s denaturation at 92°C, 10 s annealing at 64°C, and 20 s extension at 72°C. Fluorescent signal was measured at the end of each cycle. A final dissociation stage was performed to generate a melting curve for verification of amplification product specificity. For calculating the gene copy number, we first calculated the reaction efficiency for each set of primers by using triplicates of 5× dilutions of gDNA, using the provided software Opticon2 (Biorad). We measured the average cycle number for each gDNA sample using both primer sets and calculated the concentration (Qc) of each reaction by: $Q_c = efficiancy^{-cycle_time}$. Then we could calculate the gene copy number of the samples: $\frac{Q_c(CFP_primers)}{\frac{Q_c(beta_globin)}{2}}$

We verified that the U2OS cell line has only two alleles of the *β-globin* gene by comparing to a non-cancerous cell line.

Chromatin Immunoprecipitation and Quantitative PCR

ChIP and quantitative PCR were performed as previously described [18]. After an overnight induction with 1 µg/ml doxycycline, cells were harvested and lysed in SDS lysis buffer. Protein concentration was determined via Amido Black and 1 mg total protein was used per IP. For SSA treatment, cells were treated with 100 ng/ml SSA prior to lysis. Antibodies used for the ChIP were 5 µg/IP CTD4H8 against RNA polymerase II (Santa Cruz Biotechnology, SC-47701) and 10 µg/IP non-immune mouse IgG (Sigma, I5381) for the mock ChIPs. Uncrosslinking and Proteinase K treatment (100 µg/IP) were performed for 6 h at 65°C. DNA was extracted using the Qiagen PCR purification kit and used as a template for quantitative PCR. Primer sequences used in the qPCR analysis can be obtained upon request. For each amplicon, ChIP signal relative to input was calculated based on the Ct values. The values obtained for control antibody were then subtracted from those of specific antibodies. Thus, the following equation was used: 2(Ct(Input)−Ct(spIP))−2(Ct(Input)−Ct(cIP)), where: spIP−IP with specific antibody and cIP−IP with a control antibody. Values obtained for amplicons within gene regions were further normalized to the promotor region.

Immunofluorescence

Cells grown on coverslips were fixed for 20 min in 4% PFA and then permeabilized in 0.5% Triton X-100 for 3 min. After blocking with 5% BSA, cells were immunostained for 1 h with: mouse anti-RNA Pol II antibody (clones H14, H5, 8WG16 (Covance)), mouse anti-SF2/ASF (Zymed), mouse anti-PSF (B92, Sigma), mouse anti-2,2,7-trimethylguanosine (TMG) (Mercury), rabbit anti-U1A, rabbit anti-U2AF65 (Abcam), mouse anti-SC35 (Sigma), mouse anti-9G8, and mouse anti-$p54^{nrb}$ (BD Bioscience). After subsequent washes the cells were incubated for 1 h with a secondary Ab: Cy5-labeled goat anti-mouse IgG, Cy3-labeled goat anti-mouse IgG, FITC-labeled anti-mouse IgM (Jackson ImmunoResearch, West Grove, PA), Alexa488-labeled goat anti-mouse IgG, Mouse anti-rabbit Alexa488, and rabbit anti-mouse Alexa594 (Molecular Probes). Coverslips were mounted in mounting medium.

Fluorescence In Situ Hybridization

RNA FISH was performed on dox activated cells as previously described [70] using 10 ng (probe per slide) of 5′-Cy3 or Cy5 labeled fluorescent probes, as follows:

CFP: ATATAGACGTTGTGGCTGATGTAGTTGTACTCCAGCTTGTGCCCCA

GGATA

Exon1: CATGAATTCTTTGCCAAAGTGATGGGCCAGCACACAGACCAGCA

CGTTGC

Intron1: TCTCGTGACAGAGAAGGGTGTAAAAGCTTCTAGCCTTTTCTCTTA

CCTTA.

Intron2: TAGCAAAAGGGCCTAGCTTGGACTCAGAATAATCCAGCCTTATC

CCAACC

MS2: CTAGGCAATTAGGTACCTTAGGATCTAATGAACCCGGGAATACT

GCAGAC

PolyA: an oligo(dT) probe

The U snRNA set of probes containing 5-amino-allyl thymidines (Eurogentec) were labeled with Cy3 or Cy5 monoreactive dyes (GE Healthcare).

U1-1: CtGGGAAAACCACCtTCGTGATCAtGGTATCTCCCCtGCCAGGTAAGtAT

U1-2: CGAACGCAGtCCCCCACtACCACAAATTAtGCAGTCGAGtTTCCCACAtT

U2: AGtGGACGGAGCAAGCtCCTATTCCAtCTCCCTGCtCCAAAAATCCATtT

U4-1: tAGCAATAAtCGCGCCTCGGAtAAACCTCATtGGCTACGATACtGCCACT

U4-2: AGACtGTCAAAAATtGCCAATGCCGACtATATTGCAAGtCGTCACGGCGGtA

U5: GGCAAGGCTCAAAAAATTGGGTTAAGACTCAGAGTTGTTCCTCTCCACGGTA

U6: CGtGTCATCCTtGCGCAGGGGCCAtGCTAATCTTCtCTGTATCGTtCCAA

7SK snRNA: TtGGATGTGTCTGGAGTCTtGGAAGCTTGACTACtT

In some cases, an immunostaining was performed after FISH using the standard protocol.

Fluorescence Microscopy, Live-Cell Imaging, and Data Analysis

Wide-field fluorescence images were obtained using the Cell^R system based on an Olympus IX81 fully motorized inverted microscope (63× Plan-Apo, 1.42NA and 100× Plan-Apo, 1.4NA) fitted with an Orca-AG CCD camera (Hamamatsu), rapid wavelength switching, active focus control (ZDC), a motorized stage (Scan IM, Märzhäuser, Wetzlar-Steindorf, Germany), and driven by the Cell^R software. The microscope is equipped with an on-scope incubator which includes temperature and CO_2 control (Life Imaging Services, Reinach, Switzerland).

For time-lapse imaging of ActD treatment, several cell positions under non-treated conditions were chosen and recorded. Then 5 mg/ml of actinomycin D was added and cells were imaged every 5 min for 1.5 h.

Images and videos were analyzed using self-written ImageJ software macros (National Institutes of Health, Bethesda, MD). The co-localization intensity profile was generated with an ImageJ macro. The channel intensity correlation was generated by Matlab. The 4D image sequences were sum Z-projections. The transcription site was tracked and the mean intensity was measured at every time-point by finding the local maxima of the red (RFP-LacI) for each time point in a specific region of interest (ROI). The background was taken from an ROI outside of the cell and subtracted from all other measurements. Photobleach correction was obtained using another area in the nucleus or in another nucleus. Images acquired before drug addition were used as the initial conditions for normalization of the data. The half-time of site inactivation was calculated for each transcription site using Matlab. Each transcription site measured intensity dataset was fitted to a bi-exponential equation f(x) = a*exp(b*x)+c*exp(d*x). We then searched for the $t_{\frac{1}{2}}$, which brings the equation to the value of $f(t_{\frac{1}{2}})=\frac{f(t_0)-f(t_{ss})}{2}+f(t_{ss})$. For each E cell type we collected at least 7 cell transcription site measurements, and the averages and standard error were calculated.

FRAP

FRAP experiments were performed using a 3D-FRAP system (Photometrics) built on an Olympus IX81 microscope (63× Plan-Apo, 1.4 NA) equipped with an EM-CCD (Quant-EM, Roper), 405 & 491nm lasers, Lambda DG-4 light source (Sutter), XY&Z stages (Prior), and driven by MetaMorph (Molecular Devices). Experiments were performed at 37°C with 5% CO_2 using a live-cell chamber system (Tokai). Before and after the bleaching, cells were imaged in the YFP channel for the detection of YFP-MS2 (mRNPs) and in the CFP channel for the detection of CFP-lacI (genomic locus). For each acquisition, 7 Z-slices were taken every 350 nm. The active transcription site was bleached using the 491 nm laser. Six pre-bleach images were acquired. Post-bleach images were acquired in a sequence of 3 time frequencies: 15 images every 3 s, 15 images every 6 s, and 26 (or 45 for long experiments) images every 30 s.

Experiments were performed on two additional systems to verify their consistency: (1) a similar 3D FRAP system in the Darzacq laboratory under the same conditions; (2) Zeiss LSM 510 Meta laser scanning confocal microscope (Plan-Apochromat 63×, 1.4 NA oil objective (Jena, Germany)). Cells were scanned at 488 nm for GFP-MS2 or 514 nm for YFP-MS2, together with a 543 nm laser for the detection of RFP-lacI at the genomic locus. Post-bleach recovery images were acquired every 1 s.

The experiments were analyzed using self-written ImageJ macros that sum the projected Z-stacks at every time-point, and track and measure the mean intensity of the transcription site at every time-point by finding the local maxima in a specific region of interest (ROI). For each time-point, the background taken from a ROI outside of the cell was subtracted from all other measurements. T(t) and I(t) were measured for each time-point as the average intensity of the nucleus and the average intensity in the bleached ROI, respectively. The average of the pre-bleach images used as the initial conditions are referred to as Ti = nuclear intensity and Ii = intensity in ROI before bleaching. Ic(t) is the corrected intensity of the bleached ROI at time t [42,71]:

$$Ic_{(t)}=\frac{I_{(t)}T_i}{I_iT_{(t)}}$$

Data from at least 10 experiments for each cell-line were collected and the averaged FRAP measurements were fitted to the simulation.

The diffusion of the YFP-MS2 protein during the FRAP analysis was disregarded since the diffusion rate of free nucleoplasmic YFP-MS2 is very rapid, while the bound YFP-MS2 is associated with high affinity to the mRNA, and does not detach, diffuse, and bind again at the transcription site [34,41,42].

Quantitative RNA FISH and Immunofluorescence Analysis

Quantitative RNA FISH was used to compare the ratio of FISH signals from two different probes in two different channels. Although each cell line had a different number of gene-copies integrated, the ratio per transcript is comparable. Z stacks (91 slices, 200 nm steps) of the immunostained or RNA FISH samples were collected. Images were acquired for the quantification of the exons, introns, U snRNAs, or Pol II present on the transcription sites. Each experiment in which all E cell types were analyzed were acquired under the same conditions and on the same day. Images were deconvolved using a point spread function (PSF) based deconvolution algorithm Huygens Essential II (Scientific Volume Imaging, Hilversum, The Netherlands). The sum of pixel values at the transcription site in each channel were measured using Imaris (Bitplane, Saint Paul, MN), and the ratio between the channels is presented.

Computational Simulation

Monte Carlo simulations (Matlab) of the transcriptional process were based on the mechanistic models of [34,41]. The simulation was used to simulate the FRAP experiments, and by fitting the simulation to the experimental data we could retrieve the kinetic parameters of the process. The simulation performs stochastic decisions by using random numbers obtained from the simulation and checks whether they are smaller than the kinetic parameters. For each gene (E1, E3, E4, and E6), we generated a set of identical arrays where the length of each array corresponded to the length of the gene. RNA polymerases were implemented as counters that "slide" over these arrays. "Transcriptional initiation" occurred stochastically at a given "initiation rate," after which a constant speed of elongation was retained. Whenever a polymerase moves through the "MS2 region" of the gene, the polymerase/transcript accumulates a "fluorescent signal" that was maintained until the end of the gene. Polymerases were stochastically released at a given termination rate [41]. Polymerases could also randomly enter and exit a paused state [34]. The simulation reached steady state after long times, as expected.

To simulate the FRAP experiments, we tracked the total "fluorescent" signal. After the system entered steady state, we "bleached" (deleted) the fluorescent molecules and measured the buildup of "fluorescent" signal during the recovery. The signal was normalized as in the FRAP experimental procedure. The kinetic parameters, namely, the elongation speed, the termination rate, and the rates of entering and exiting the paused state, were chosen as those that best fit the experimental FRAP data. We found that the number of gene arrays and the initiation rate had no effect on the measured kinetic results, due to the normalization to the pre-bleach signal.

To find the above four kinetic parameters that best fit the experimental data we had to explore a four-dimensional parameter space. For each configuration, the mean square difference (MSD) between the simulated FRAP curve and the experimental curve was calculated (Figure S10). For the E3 gene, the kinetic parameters that gave the best fit (lowest MSD) were: 3.6 kb/min, ~42% of the polymerases entering a paused state for

an average of 6 min, and transcript retention of 50 s after the end of elongation. To obtain the kinetic parameters for the E6 gene, we used the best fit values of the E3 elongation speed and pausing transition rates, and changed only the termination rate.

Supporting Information

Figure S1 Transcriptional induction of the different cell lines. The cell lines containing the stably integrated E1, E3, E4, and E6 genes were transcriptionally induced by dox for 12 h. RNA-FISH with a probe to the MS2 region shows the transcription sites (yellow). The CFP protein product is targeted to peroxisomes (cyan). DIC on the right (bar, 5 μm).
Found at: doi:10.1371/journal.pbio.1000573.s001 (3.64 MB TIF)

Figure S2 Recruitment of RNA processing factors to active transcription sites of E3 cells. The gene locus was identified with RFP-LacI (red) and the recruitment of endogenous Pol II (green) was identified by (A) 8WG16 Ab (specific to the CTD repeats) and (B) H5 Ab (CTD phosphorylated on serine 2). Active transcription sites (MS2, yellow) can be seen together with immunofluorescence (red) with (C) anti-TMG identifying the cap structure of snRNAs. (D) RNA-FISH to U4 snRNA and (E) U5 snRNA, together with a probe to the MS2 region (yellow). (F) Immunofluorescence with anti-U1A, (G) anti-U2AF65, (H) anti-SF2, (I) anti-9G8, and (J) anti-SC-35; together with the YFP-MS2 protein. Enlargements show the merged signals at the transcription sites. Plots depict the degree of co-localization of the signals at the transcription site across the depicted line (bar, 5 μm).
Found at: doi:10.1371/journal.pbio.1000573.s002 (3.54 MB TIF)

Figure S3 Recruitment of transcription-related factors to active transcription sites of E3 cells. (A) Transiently expressed GFP-TLS, (B) GFP-PSF, (C) GFP-p54^{nrb}, (D) GFP-PTB, (E) GFP cyclin T1 (pseudocolored red), imaged together with YFP-MS2. (F) RNA-FISH to 7SK RNA (red) and MS2 repeats (yellow). Enlargements show the merged signals at the transcription sites. Plots depict the degree of colocalization of the signals at the transcription site across the depicted line (bar, 5 μm).
Found at: doi:10.1371/journal.pbio.1000573.s003 (3.78 MB TIF)

Figure S4 Recruitment of RNA processing factors to transcription sites of the intronless E1 gene. Immunofluorescence (red) with (A) anti-TMG that identifies the cap of snRNAs, (B) anti-U1A, and (C) anti-U2AF65, together with the mRNA seen by transfected YFP-MS2 labeling (yellow). (D) RNA-FISH to U4 snRNA and (E) U5 snRNA, together with a probe to the MS2 region (yellow). Enlargements show the signals at the transcription sites. Plots depict the degree of co-localization of the signals at the transcription site across the depicted line (bar, 5 μm).
Found at: doi:10.1371/journal.pbio.1000573.s004 (1.67 MB TIF)

Figure S5 Comparison of FRAP results on E3 cells performed on different microscopes. The recovery of photobleached YFP-MS2 on actively transcribing sites was monitored in E3 cells using: (a) a Zeiss confocal microscope; (b) a 3D FRAP system in our laboratory (system 1); and (c) a 3D FRAP system in the Darzacq laboratory (system 2) (average from at least $n>10$ in each experiment).
Found at: doi:10.1371/journal.pbio.1000573.s005 (0.14 MB TIF)

Figure S6 Genomic positioning and gene copy number have no effect on the measured kinetics. (A) FRAP results of different E6 clones (E6-3 and E6-22-22) performed on different microscopes. (B) Real-time PCR on genomic DNA was used to quantify the number of genes integrated into each gene array gene repeats: E3 = 24, E4 = 5, E6-22-22 = 40, E6-3 = 9 with standard deviation.
Found at: doi:10.1371/journal.pbio.1000573.s006 (0.46 MB TIF)

Figure S7 Analysis of FRAP curves. FRAP recovery curves of the E3 transcription site show bi-phasic kinetics, which points to two parallel processes. The red line is the curve fit for the sum of the two-exponential equation and shows a better fit than the green line that was fitted to one exponential. The yellow line presents the fast phase (70%) and the pink line is the slow phase (30%).
Found at: doi:10.1371/journal.pbio.1000573.s007 (0.15 MB TIF)

Figure S8 The effects of splicing inhibitors on the elongation kinetics. (A) Inhibition of splicing by Meayaymycin (6 h) showed that the unspliced pre-mRNA was distributed throughout the cell and in speckles (RNA-FISH on E6 cells) and was retained in the nucleus (bottom), whereas in untreated cells pre-mRNA was detected only at the site of transcription. (B) FRAP recovery curves of untreated and Meayamycin-treated E6 cells shows similar kinetics as E3 cells. (C) The treatment of cells with SSA for 9 h followed by FRAP analysis of GFP-Pol II recruited to the transcription sites shows that splicing inhibition did not affect polymerase kinetics.
Found at: doi:10.1371/journal.pbio.1000573.s008 (1.51 MB TIF)

Figure S9 Searching the kinetic parameter space for the best fit to experimental data. (A) Simulation plots of an elongating (red) or paused (blue) polymerase population. Top part shows an example of an experiment in which the MS2 fluorescence accumulates (arrow) and is then released. Bottom part shows the accumulation of fluorescence in the MS2 region (left arrow) as well as a stochastically paused polymerase (right arrow). (B,C) The simulations rely on four kinetic parameters: the elongation speed, the termination time, and entering and exiting rates of the paused state (K_{in}, K_{out}). To find the best fit to the experimental data, the kinetic parameters were varied in a systematic way. The parameters were plotted two at a time because the parameter space is four dimensional. (A) Plot of the mean square difference (MSD) between the experimental FRAP curve and the simulated curve for each pair of elongation speed and termination time. (B) Similar analysis with the pausing transition rates. The MSD is color coded: the dark blue corresponds to parameter combinations that result in the best fit to the experimental data.
Found at: doi:10.1371/journal.pbio.1000573.s009 (1.25 MB TIF)

Figure S10 Exploring how the simulation parameters affect the kinetics. (A) Changes in the elongation speed. (B) Changes in the probability to enter or exit a paused state changed the pausing time during transcription, and the average percentage of polymerases that paused in steady state. (C) Changing only the termination time.
Found at: doi:10.1371/journal.pbio.1000573.s010 (0.64 MB TIF)

Figure S11 Low DRB does not affect kinetics on E3 and E6 genes. No change in the FRAP recovery kinetics of YFP-MS2 on cells treated with DRB (20 μg/μl for 2 h before imaging): E3 cells (brown), E6 cells (dark green), or E6 cells with SSA (cyan) SSA; compared to untreated E3 (blue) and E6 (light green) cells.
Found at: doi:10.1371/journal.pbio.1000573.s011 (0.14 MB TIF)

Video S1 Transcriptional activation of E3 (top) and E6 (bottom) cells. The DNA locus is tagged with CFP-LacI and the cytoplasmic peroxisomes are also seen in CFP. The mRNA is tagged with YFP-MS2. As transcriptional induction proceeds, the transcription sites appear and then mRNAs disperse into the

nucleoplasm. CFP-peroxisomes appear at later times. Since the videos were acquired in 3D and the z-stacks covered the cell volumes, we could show the nuclear CFP-LacI locus in the beginning of the videos and then later shift to the cytoplasmic plane that best shows the CFP-peroxisomes. Cells were imaged 15 min after dox addition and the time in the videos is measured from dox addition. E3, imaged every 15 min; E6, imaged every 20 min.
Found at: doi:10.1371/journal.pbio.1000573.s012 (1.59 MB MOV)

Video S2 FRAP experiment of an E4 cell. mRNA was labeled with YFP-MS2. 2D image over time of a summed projection of a 4D video. Right: Enlarged tracking of the transcription site. Time zero indicates the bleach time.
Found at: doi:10.1371/journal.pbio.1000573.s013 (0.28 MB MOV)

Video S3 Addition of actinomycin D to transcribing E6 cells and recording transcription site inactivation over time. The mRNA was tagged with YFP-MS2 (green) and the transcription site locus with RFP-LacI (red). Time zero indicates ActD addition time.
Found at: doi:10.1371/journal.pbio.1000573.s014 (0.94 MB MOV)

Video S4 Example of a simulation run. Each dot denotes a polymerase moving along the gene: elongating = red dot, and stochastically pausing = dot becomes green. When the polymerase terminates and the transcript still remains on the transcription site, the dot becomes yellow. The bleached polymerase becomes blue. The *x*-axis = gene length of E3; in the *y*-axis each line represents a gene within the transcription site. Top: a single gene view. Bottom: an array of 1,000 genes running in parallel. The last frame contains a graph showing the MS2 intensity measured from the simulation.
Found at: doi:10.1371/journal.pbio.1000573.s015 (4.30 MB MOV)

Acknowledgments

We thank Minoru Yoshida (RIKEN Advanced Science Institute, Japan) for Spliceostatin A, Kazunori Koide (University of Pittsburgh) for Meayamycin, and Shai Carmi (BIU) for critical reading of the manuscript. YST is the Jane Stern Lebell Family Fellow in Life Sciences at BIU. YB is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship.

Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: YB YST. Performed the experiments: YB NN NB EWB. Analyzed the data: YB. Contributed reagents/materials/analysis tools: SZC KMN XD. Wrote the paper: YST.

References

1. Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G (2004) Multiple links between transcription and splicing. RNA 10: 1489–1498.
2. Bentley DL (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol 17: 251–256.
3. Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. Nature 416: 499–506.
4. Moore MJ, Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 136: 688–700.
5. Pandit S, Wang D, Fu XD (2008) Functional integration of transcriptional and RNA processing machineries. Curr Opin Cell Biol 20: 260–265.
6. Neugebauer KM (2002) On the importance of being co-transcriptional. J Cell Sci 115: 3865–3871.
7. Perales R, Bentley D (2009) "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol Cell 36: 178–191.
8. Lewis JD, Tollervey D (2000) Like attracts like: getting RNA processing together in the nucleus. Science 288: 1385–1389.
9. de Almeida SF, Carmo-Fonseca M (2008) The CTD role in cotranscriptional RNA processing and surveillance. FEBS Lett 582: 1971–1976.
10. Lacadie SA, Rosbash M (2005) Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5′ss base pairing in yeast. Mol Cell 19: 65–75.
11. Misteli T, Spector DL (1999) RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. Mol Cell 3: 697–705.
12. Vincent M, Lauriault P, Dubois MF, Lavoie S, Bensaude O, et al. (1996) The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. Nucleic Acids Res 24: 4649–4652.
13. Mortillaro MJ, Blencowe BJ, Wei X, Nakayasu H, Du L, et al. (1996) A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. Proc Natl Acad Sci U S A 93: 8253–8257.
14. Kim E, Du L, Bregman DB, Warren SL (1997) Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. J Cell Biol 136: 19–28.
15. Sapra AK, Anko ML, Grishina I, Lorenz M, Pabis M, et al. (2009) SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. Mol Cell 34: 179–190.
16. Glover-Cutter K, Kim S, Espinosa J, Bentley DL (2008) RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat Struct Mol Biol 15: 71–78.
17. Moore MJ, Schwartzfarb EM, Silver PA, Yu MC (2006) Differential recruitment of the splicing machinery during transcription predicts genome-wide patterns of mRNA splicing. Mol Cell 24: 903–915.
18. Listerman I, Sapra AK, Neugebauer KM (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat Struct Mol Biol 13: 815–822.
19. Swinburne IA, Meyer CA, Liu XS, Silver PA, Brodsky AS (2006) Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. Genome Res 16: 912–921.
20. Audibert A, Weil D, Dautry F (2002) In vivo kinetics of mRNA splicing and transport in mammalian cells. Mol Cell Biol 22: 6706–6718.
21. Singh J, Padgett RA (2009) Rates of in situ transcription and splicing in large human genes. Nat Struct Mol Biol 16: 1128–1133.
22. Wada Y, Ohta Y, Xu M, Tsutsumi S, Minami T, et al. (2009) A wave of nascent transcription on activated human genes. Proc Natl Acad Sci U S A 106: 18357–18361.
23. Kessler O, Jiang Y, Chasin LA (1993) Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. Mol Cell Biol 13: 6211–6222.
24. Beyer AL, Osheim YN (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev 2: 754–765.
25. Neugebauer KM, Roth MB (1997) Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription. Genes Dev 11: 1148–1159.
26. Huang S, Spector DL (1991) Nascent pre-mRNA transcripts are associated with nuclear regions enriched in splicing factors. Genes Dev 5: 2288–2302.
27. Zhang G, Taneja KL, Singer RH, Green MR (1994) Localization of pre-mRNA splicing in mammalian nuclei. Nature 372: 809–812.
28. Misteli T, Caceres JF, Spector DL (1997) The dynamics of a pre-mRNA splicing factor in living cells. Nature 387: 523–527.
29. Janicki SM, Tsukamoto T, Salghetti SE, Tansey WP, Sachidanandam R, et al. (2004) From silencing to gene expression: real-time analysis in single cells. Cell 116: 683–698.
30. de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, et al. (2003) A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell 12: 525–532.
31. Munoz MJ, de la Mata M, Kornblihtt AR (2010) The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends Biochem Sci 35: 497–504.
32. Darzacq X, Singer RH, Shav-Tal Y (2005) Dynamics of transcription and mRNA export. Curr Opin Cell Biol 17: 332–339.
33. Shav-Tal Y, Darzacq X, Singer RH (2006) Gene expression within a dynamic nuclear landscape. EMBO J 25: 3469–3479.
34. Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, et al. (2007) In vivo dynamics of RNA polymerase II transcription. Nat Struct Mol Biol 14: 796–806.
35. Lamond AI, Swedlow JR (2007) RNA polymerase II transcription in living color. Nat Struct Mol Biol 14: 788–790.
36. Ben-Ari Y, Brody Y, Kinor N, Mor A, Tsukamoto T, et al. (2010) The life of an mRNA in space and time. J Cell Sci 123: 1761–1774.
37. Darzacq X, Kittur N, Roy S, Shav-Tal Y, Singer RH, et al. (2006) Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. J Cell Biol 173: 207–218.

38. Mor A, Suliman S, Ben-Yishay R, Yunger S, Brody Y, et al. (2010) Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat Cell Biol 12: 543–552.
39. Tsukamoto T, Hashiguchi N, Janicki SM, Tumbar T, Belmont AS, et al. (2000) Visualization of gene activity in living cells. Nat Cell Biol 2: 871–878.
40. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, et al. (1998) Localization of ASH1 mRNA particles in living yeast. Mol Cell 2: 437–445.
41. Boireau S, Maiuri P, Basyuk E, de la Mata M, Knezevich A, et al. (2007) The transcriptional cycle of HIV-1 in real-time and live cells. J Cell Biol 179: 291–304.
42. Shav-Tal Y, Darzacq X, Shenoy SM, Fusco D, Janicki SM, et al. (2004) Dynamics of single mRNPs in nuclei of living cells. Science 304: 1797–1800.
43. Kole R, Weissman SM (1982) Accurate in vitro splicing of human beta-globin RNA. Nucleic Acids Res 10: 5429–5445.
44. Scherer S (2008) A short guide to the human genome. Cold Spring HarborNY: Cold Spring Harbor Laboratory Press. 173 p.
45. Tumbar T, Sudlow G, Belmont AS (1999) Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. J Cell Biol 145: 1341–1354.
46. Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, et al. (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. Mol Cell 26: 867–881.
47. Spiluttini B, Gu B, Belagal P, Smirnova AS, Nguyen VT, et al. (2010) Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. J Cell Sci 123: 2085–2093.
48. Kaida D, Motoyoshi H, Tashiro E, Nojima T, Hagiwara M, et al. (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol 3: 576–583.
49. Albert BJ, McPherson PA, O'Brien K, Czaicki NL, Destefino V, et al. (2009) Meayamycin inhibits pre-messenger RNA splicing and exhibits picomolar activity against multidrug-resistant cells. Mol Cancer Ther 8: 2308–2318.
50. Hamer DH, Smith KD, Boyer SH, Leder P (1979) SV40 recombinants carrying rabbit beta-globin gene coding sequences. Cell 17: 725–735.
51. Le Hir H, Nott A, Moore MJ (2003) How introns influence and enhance eukaryotic gene expression. Trends Biochem Sci 28: 215–220.
52. Valencia P, Dias AP, Reed R (2008) Splicing promotes rapid and efficient mRNA export in mammalian cells. Proc Natl Acad Sci U S A 105: 3386–3391.
53. Luo MJ, Reed R (1999) Splicing is required for rapid and efficient mRNA export in metazoans. Proc Natl Acad Sci U S A 96: 14937–14942.
54. Mor A, Suliman S, Ben-Yishay R, Yunger S, Brody Y, et al. (2010) Dynamics of single mRNP nucleo-cytoplasmic transport and export through the nuclear pore in living cells. Nat Cell Biol 12: 543–552.
55. Liu K, Sandgren EP, Palmiter RD, Stein A (1995) Rat growth hormone gene introns stimulate nucleosome alignment in vitro and in transgenic mice. Proc Natl Acad Sci U S A 92: 7724–7728.
56. Furger A, O'Sullivan JM, Binnie A, Lee BA, Proudfoot NJ (2002) Promoter proximal splice sites enhance transcription. Genes Dev 16: 2792–2799.
57. Kwek KY, Murphy S, Furger A, Thomas B, O'Gorman W, et al. (2002) U1 snRNA associates with TFIIH and regulates transcriptional initiation. Nat Struct Biol 9: 800–805.
58. Damgaard CK, Kahns S, Lykke-Andersen S, Nielsen AL, Jensen TH, et al. (2008) A 5′ splice site enhances the recruitment of basal transcription initiation factors in vivo. Mol Cell 29: 271–278.
59. Fong YW, Zhou Q (2001) Stimulatory effect of splicing factors on transcriptional elongation. Nature 414: 929–933.
60. Kameoka S, Duque P, Konarska MM (2004) p54(nrb) associates with the 5′ splice site within large transcription/splicing complexes. EMBO J 23: 1782–1791.
61. Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD (2008) The splicing factor SC35 has an active role in transcriptional elongation. Nat Struct Mol Biol 15: 819–826.
62. Shav-Tal Y, Zipori D (2002) PSF and p54(nrb)/NonO–multi-functional nuclear proteins. FEBS Lett 531: 109–114.
63. Emili A, Shales M, McCracken S, Xie W, Tucker PW, et al. (2002) Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. RNA 8: 1102–1111.
64. Spiluttini B, Gu B, Belagal P, Smirnova A, Nguyen VT, et al. (2010) Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. J Cell Sci.
65. Rino J, Carmo-Fonseca M (2009) The spliceosome: a self-organized macromolecular machine in the nucleus? Trends Cell Biol 19: 375–384.
66. Dye MJ, Gromak N, Proudfoot NJ (2006) Exon tethering in transcription by RNA polymerase II. Mol Cell 21: 849–859.
67. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921.
68. Sakharkar M, Passetti F, de Souza JE, Long M, de Souza SJ (2002) ExInt: an Exon Intron Database. Nucleic Acids Res 30: 191–194.
69. Tennyson CN, Klamut HJ, Worton RG (1995) The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nat Genet 9: 184–190.
70. Chartrand P, Bertrand E, Singer RH, Long RM (2000) Sensitive and high-resolution detection of RNA in situ. Methods Enzymol 318: 493–506.
71. Phair RD, Misteli T (2000) High mobility of proteins in the mammalian cell nucleus. Nature 404: 604–609.

Imaging transcription in living cells.

This annual review of biophysics article was written jointly by members of the Janelia Farm Research Consortium on Imaging Transcription. The review describes the state of the art in live-cell imaging of transcription. The recent advances in the field of live cell fluorescence microscopy have permitted a large number of studies to be carried out at the single cell level. These studies have given the scientific community formidable insight into the dynamics of transcription in-vivo. However, these still nascent techniques require taking certain precautions. This review points out the possibilities and limits of such methods in detail, and concludes on the progress that has yet to be done in order to perform more powerful and precise studies in the field of transcription. I have personally contributed to the writing of this review, in particular the natural and artificial gene array chapters. I have also assembled most of the figures of the review.

REVIEWS IN ADVANCE

Imaging Transcription in Living Cells

Xavier Darzacq,[1,2] Jie Yao,[1,3] Daniel Larson,[1,4] Sébastien Z. Causse,[1,2] Lana Bosanac,[1,2] Valeria deTurris,[1,4] Vera M. Ruda,[1,2] Timothee Lionnet,[1,4] Daniel Zenklusen,[1,4] Benjamin Guglielmi,[1,3] Robert Tjian,[1,3] and Robert H. Singer[1,4]

[1]Janelia Farm Research Consortium on Imaging Transcription, Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147

[2]Imagerie Fonctionelle de la Transcription, Ecole Normale Supérieure CNRS UMR 8541, 75230 Paris cedex 05, France; email: darzacq@ens.fr

[3]Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720; email: tjcal@berkeley.edu

[4]Department of Anatomy and Structural Biology, Cell Biology and Neuroscience, Gruss-Lipper Biophotonics Center, Albert Einstein College of Medicine of Yeshiva University, Bronx, New York 10461; email: rhsinger@aecom.yu.edu

Annu. Rev. Biophys. 2009. 38:173–96

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

This article's doi: 10.1146/annurev.biophys.050708.133728

Copyright © 2009 by Annual Reviews. All rights reserved

1936-122X/09/0609-0173$20.00

Key Words

modeling, RNA synthesis in live cells, polymerase kinetics

Abstract

The advent of new technologies for the imaging of living cells has made it possible to determine the properties of transcription, the kinetics of polymerase movement, the association of transcription factors, and the progression of the polymerase on the gene. We report here the current state of the field and the progress necessary to achieve a more complete understanding of the various steps in transcription. Our Consortium is dedicated to developing and implementing the technology to further this understanding.

Contents

INTRODUCTION 174
IMAGING GENE ARRAYS 174
Drosophila melanogaster Polytene Chromosomes 175
Natural Tandem Repeated Genes ... 178
Artificial Gene Arrays 179
Viruses 180
NUCLEAR ORGANIZATION OF TRANSCRIPTION 182
Transcription Factories 182
Gene Positioning 183
IMAGING ENDOGENOUS GENES 185
Steroid Receptors 185
Imaging a Single Gene 186
ANALYSIS OF KINETICS 190
Mobility of Transcription Factors ... 190
Dynamics 190
Modeling 190
CONCLUSION 191
THE FUTURE 191

INTRODUCTION

With the uncovering of the ever-growing fraction of the animal genome that is transcribed, transcription is more than ever the centerpiece of cell metabolism. Through biochemical analysis and genetics, many if not most of the proteins implicated in transcription have been identified. Decades of in vitro studies determined that the transcription process could be separated in three steps: preinitiation complex formation, initiation, and elongation. Each one of these steps may be subjected to regulation, accounting for the fine-tuning of gene expression. But biochemistry tells us only what is possible, not what actually happens, in the very specific milieu of a living cell.

It has become feasible, using the new advances in microscopy, to interrogate the processes that make up transcription and break them down into their component parts. Accurate quantification is possible due to technology that has evolved over the years to detect and measure photons. The components of the transcription reaction can then be assigned rate constants describing their forward and reverse rates. As a result of these analyses, new models have arisen to fit these data. Among those are the observations that for some models the transcriptional complex can be transient, existing only for a few seconds, and that the entire process is inefficient, yet other factors can be stably associated for hours. Understanding the kinetic components that give rise to these disparate time constants will be an important function of the new technologies.

This review is dedicated to exploring the work that is contributing to the real-time analysis of transcription, an aspect that contributes data not addressed by chromatin immunoprecipitation, microarray studies, or other bulk assays that cannot resolve the events occurring in single cells. Many cell and tissue models are described including bacteria, polytene chromosomes, various reporter genes, and cell lines. Different approaches include fluorescence recovery after photobleaching (FRAP), fluorescence correlation spectroscopy (FCS), fluorescence resonance energy transfer (FRET), and multiphoton microscopy (MPM) (see sidebar, Methods Used to Analyze Transcription in Living Cells). This is a new field that is rapidly emerging, and these initial forays represent the beginning of a new territory in the area of gene expression research.

The review is organized by the model systems that have contributed to studies in live cell imaging: naturally occurring and artificial gene arrays, viral genes, steroid receptor responsive genes, and single copy endogenous genes. There are four sections: imaging gene arrays, imaging the nuclear organization of transcription, imaging single copy genes, and the analysis of imaging data.

FRAP: fluorescence recovery after photobleaching

FCS: fluorescence correlation spectroscopy

FRET: fluorescence resonance energy transfer

MPM: multiphoton microscopy

IMAGING GENE ARRAYS

Some gene families or duplicated genes in vertebrate genomes have naturally regrouped into tandem arrayed genes. These genes therefore lie as neighbors on the same locus

in the genome and are often under the control of the same transcriptional regulators. In addition, the fruit fly (*Drosophila melanogaster*) contains giant polytene chromosomes in certain larval cell types where the entire genome is multiplied. This spatial clustering of genes offers researchers many opportunities to study transcription, especially when using fluorescence microscopy, in which accumulation of fluorescence means a better signal-to-noise ratio. Another advantage of such arrays is that when studying a fluorescent molecule that interacts directly or indirectly with the arrays, the majority of the fluorescence will be due to interacting molecules (in contrast to freely diffusing molecules, which might be preponderant at a nonamplified locus). This increases the resolution on binding events that reflect catalytic activities. So far three cases of amplified genes have been used to study transcription as illustrated in **Figure 1**: polytene salivary gland chromosomes in *Drosophila* (95), the ribosomal DNA nucleolar clusters (36), and artificially developed mammalian gene arrays (68). These examples are covered more fully below.

Drosophila melanogaster Polytene Chromosomes

Drosophila polytene chromosomes are found in many larval cell types formed by endoreduplication during development, i.e., these cells undergo DNA replication without cell division. Polytene chromosomes from salivary gland cells contain approximately 1000 copies of DNA. Condensed and decondensed chromatin form unique band and interband structures that can be distinguished with a light microscope. The chromosome banding patterns were categorized and named by Bridges (13) and have been used as a marker for cytogenetic localization of individual genes along the polytene chromosomes (**Figure 2*a***). Examining the unique cytogenetic pattern has allowed early genetic mapping such as gene deletion, gene duplication, chromosome translocation, and inversion. Furthermore, localizing protein factors on polytene chromosomes with antibody-based immunostaining techniques has provided a means to study protein-DNA interaction in vivo.

The naturally amplified chromatin template in *Drosophila* polytene chromosomes provides an opportunity to overcome the sensitivity

METHODS USED TO ANALYZE TRANSCRIPTION IN LIVING CELLS

Image Sampling

Cells imaged under wide-field fluorescence microscopy using long exposure times can yield images of transcription factors specifically bound to the relatively stationary DNA in contrast to the unresolved undersampled free transcription factors. Tracking of individual transcription factors was possible using stroboscopic laser excitation.

FRAP

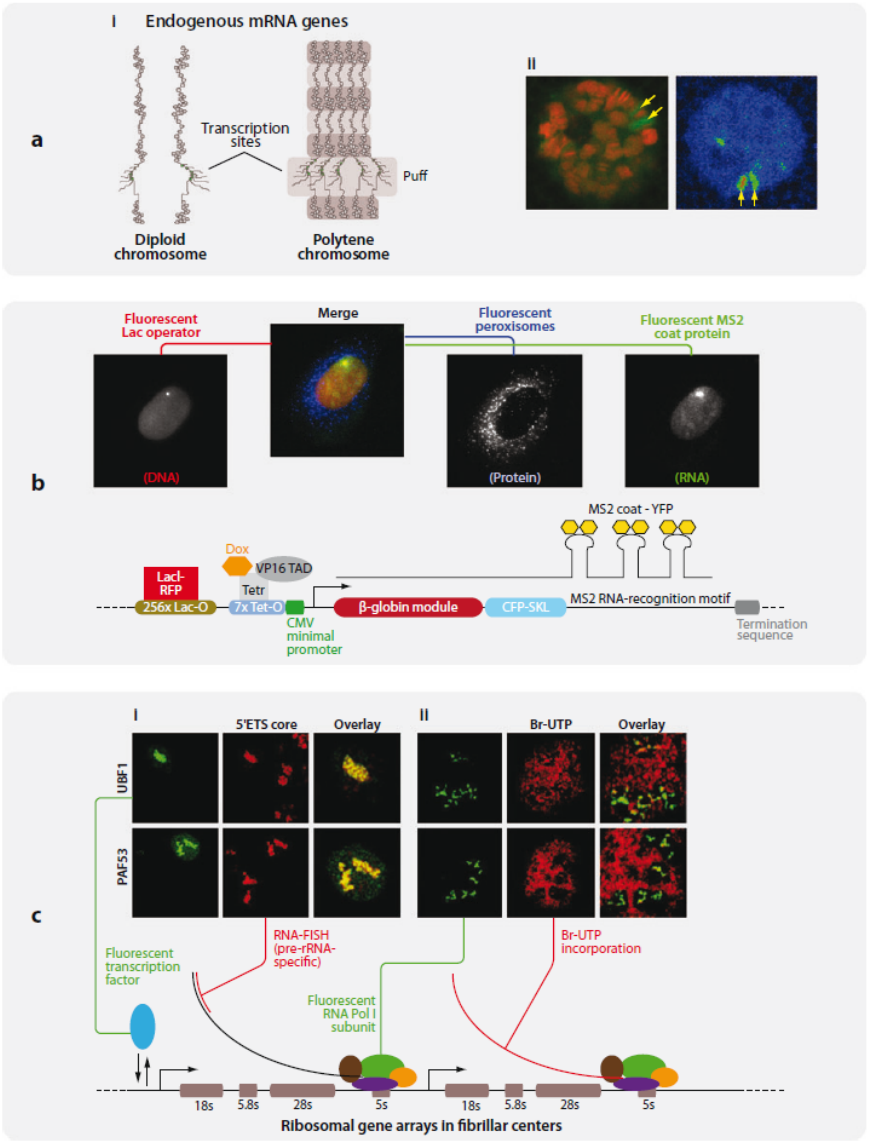
FRAP involves the irreversible photobleaching of a specified area in the nucleus by a focused laser of biologically nonabsorptive wavelength. The diffusion of surrounding fluorescent molecules into the bleached spot portrays a recovery characteristic of the movement of tagged fluorophores.

FLIP

By the method of FLIP, a specified cell area is repeatedly bleached and the loss of fluorescence of surrounding areas are monitored. This gives a more exact way to analyze the mobility of a protein which is likewise involved in binding. Dissociation kinetics of proteins from compartments can thus be determined more precisely.

FCS

FCS is a tool for binding measurements. A laser beam is focused in the cell within a femtoliter range volume. Fluctuations in fluorescence signals are measured over short periods of time, the output reflecting movement of labeled proteins through the volume. This gives a direct measurement of concentration, diffusion constants and binding constants. Generally, FCS is well suited to fast processes on a scale of milliseconds, whereas FRAP is better suited to slower processes on scale of seconds. Cross-correlation FCS can likewise be used to determine physical interaction between two species by simultaneous comparison of their fluctuation traces.



176 Darzacq et al.

limitation in visualizing transcription factors associated with endogenous gene loci in living cells. However, imaging transcription factors in living salivary gland tissues had been challenged by the thickness and optical properties of the tissue samples. Recently, Webb, Lis, and colleagues (94–96) have reported that MPM provides the experimental capability of resolving individual genetic loci (**Figure 2*b–e***) and studying dynamic interactions of a green fluorescent protein (GFP)-fused heat shock factor (HSF) or RNA polymerase II (Pol II) with active *hsp*70 gene loci in living tissues and in real time.

This is an elegant system for studying the dynamics of transcription regulation in vivo. The power of this approach lies in the combination of naturally amplified templates, *Drosophila* transgenic techniques, and MPM imaging, which provides optical sectioning deep within living tissues (99). The rapid and robust heat shock gene activation has allowed unambiguous localization of endogenous *hsp*70 loci and further assisted the visualization of the associated factors (**Figure 2*f,g***). Furthermore, multicolor fluorescent proteins and mutant gene alleles can be introduced by simple genetic crosses.

Some novel insights of transcription dynamics have arisen from the application of this method. For instance, the transient association (a few seconds) of a transcription activator with a gene promoter, the so called hit-and-run model (54), has been thought to be universal for all activators. However, in *Drosophila* polytene chromosomes, HSF is transiently associated with *hsp*83 gene loci (half life about 10 s) before heat shock but becomes stably associated with *hsp*70 gene loci (half-life > 5 min) during heat shock (**Figure 2*b,i***) (95). Therefore, some transcription factors, such as HSF, are stably associated with their target genes and can provide a stable platform that supports multiple rounds of transcription. This may be a relatively common property of strong promoters, as suggested by other reports. (*a*) A stable association with DNA in vivo has been reported for yeast Gal4 activator (60); (*b*) higher transcription levels are correlated with a more stable association of Pol I subunits with rRNA genes (36). In addition, Pol II exhibits efficient recruitment to the locus and enters into elongation during early heat shock, and it is locally recycled during late heat shock (94). Furthermore, this MPM-based FRAP assay has been coupled with chromatin immunoprecipitation assays to study the dynamics of distinct populations of Pol II molecules: During early heat shock, FRAP was performed on *hsp*70 genes when the productively elongating forms of Pol II were eliminated by inhibiting P-TEFb kinase. This study shows that the remaining transcriptionally engaged/paused Pol II molecules near the transcription start site of the genes are stably associated with *hsp*70 genes (61). Although this experimental system will continue to provide unique insights into the dynamic mechanisms of gene regulation in *Drosophila*, it would be of interest to determine to what extent the transcription kinetics at micron-scale gene structures in polytene cells resembles the kinetics at nanoscale gene structures in diploid cells.

GFP: green fluorescent protein

HSF: heat shock factor

Pol I (II): RNA polymerase I (II)

Figure 1

Three models for visualizing transcription in vivo. (*a*) *Drosophila* polytene chromosome. (*i*) diagram of diploid chromosome and polytene chromosome. Active transcription site of heat shock gene *hsp*70 loci at 87A and 87C exhibits "chromosome puffs" (*yellow arrows*). (*ii*) *hsp*70 loci are enriched with Pol II (*red*: Hoechst33342 stain; *green*: EGFP-Rpb3). (Reproduced with permission from Reference 95.) (*Right*) Upon heat shock, EGFP-Rpb3 exhibits a strong doublet at 87A and 87C loci that can be recognized in polytene nucleus (*arrows*). Image in pseudocolor. (*b*) An artificial reporter gene was stably inserted as multiple copies into a single locus in the genome of U2OS cells. The locus of integration was tracked with the use of a Lac-repressor-fused RFP. The MS2 stem loop structures on the RNA, which bind stably to MS2 coat proteins, fused with YFP, allowing the investigator to localize single molecules of mRNA in vivo (78). (*c*) Ribosomal RNA transcription occurring inside fibrillar centers in CMT3 cells. (*i*) PAF53, a Pol I subunit, and UBF1, a rDNA transcription factor, colocalize with nascent rRNA in fibrillar centers (*green*: PAF53-GFP or UBF1-GFP; *red*: fluorescent in situ hybridization probe directed to rRNA 5′ETS core). (*ii*) GFP-tagged PAF53 and UBF1 accumulate in active rRNA transcription sites (*green*: PAF53-GFP or UBF1-GFP; *red*: Bromo-UTP stain) (Adapted with permission from Reference 24). Abbreviations: EGFP, enhanced green fluorescent protein; ETS, external transcribed spacer; Pol I (II), polymerase I (II).

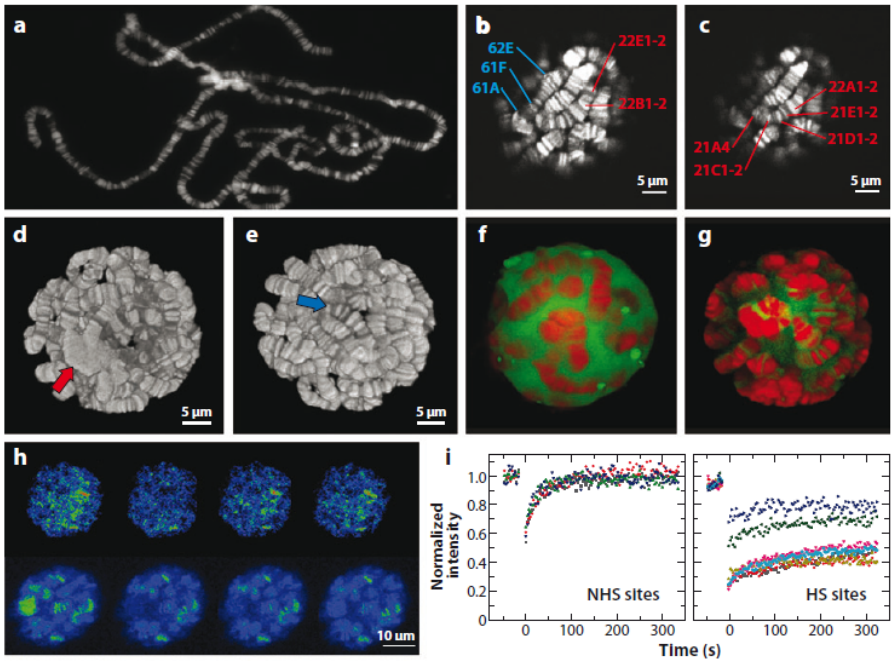


Figure 2

Imaging transcription kinetics at *Drosophila* polytene chromosomes. (*a*) An image of *Drosophila* polytene chromosomes spread onto a glass slide and stained by Hoechst33342. (*b*, *c*) Optical sections of polytene chromosomes in a live salivary gland stained with Hoechst33342. The *z*-distance is 0.5 µm. Labels identify specific bands on chromosome arms 2L (*red*) and 3L (*blue*). (*d*, *e*) Three-dimensional reconstructions of a polytene nucleus. Red and blue arrows indicate the centromeric region and telomere, respectively. Panels *b*–*e* are adapted with permission from Reference 95. (*f*, *g*) HSF expression and localization in salivary glands at NHS (*f*) and after HS (*g*). Shown are the three-dimensional reconstructions of the two-photon optical sections of polytene nuclei expressing HSF-EGFP (*green*) stained with Hoechst33342 (*red*). (*h*, *i*) FRAP of HSF-EGFP at endogenous gene loci in salivary glands. (*h*) Intensity images of a nucleus of a heat-shocked salivary gland in pseudocolor during FRAP. (*i*) FRAP curves at NHS and HS sites. (Adapted with permission from References 95 and 96). Abbreviations: EGFP, enhanced green fluorescent protein; FRAP, fluorescence recovery after photobleaching; HS, heat shock; HSF, heat shock factor; NHS, non-heat shock.

Natural Tandem Repeated Genes

The Misteli laboratory used the ribosomal genes' organization into such arrays to study Pol I transcription. Their work offered the first attempt to measure kinetic rates for a RNA polymerase in living cells. They proposed an elongation rate of 5.7 kb per minute. They also observed that different subunits of the Pol I complex were loaded onto the genes with different kinetic rates, which can be interpreted as a pioneering subunit or subcomplex interacting with the DNA, serving as a docking platform for the other subunits (24). More recently, the same group analyzed the order of assembly of Pol I components and the regulation of this assembly during G_1/G_0 and S phases. According to their findings, Pol I is assembled in an on-the-spot stepwise process that reflects

transcriptional efficiency (36). This is in contrast to in vitro studies that show that a preassembled RNA polymerase holoenzyme can be recruited to a promoter site and efficiently transcribe an RNA molecule (22), or that there is no exchange between Pol I subunits in yeast (Pol I remains intact without subunit exchange through multiple rounds of transcription in *Saccharomyces cerevisiae*) (22, 75). Similar studies have yet to be done with Pol II, whose assembly might also be influenced by the promoter sequence of the studied gene and the various specific transcriptional regulators involved.

Another example of natural array is the *CUP1* locus in the baker's yeast *S. cerevisiae* (47). *CUP1* exists as a small natural tandem array of 10 copies, transcribed by the Pol II. The transcriptional activator Ace1 binds to *CUP1* promoter in the presence of copper and activates transcription. Fusion of three copies of GFP to Ace1 enabled visualization of the *CUP1* array in live cells. The behavior of Ace1-GFP on the *CUP1* promoter has been monitored by FRAP. A complete fluorescence recovery occurred within 2 min after photobleaching. It is similar but somewhat slower than that observed for other transcriptional activators. Furthermore, at longer timescales a slow cycling of Ace1-GFP binding to *CUP1* can be detected. After computational analysis, Ace1 behavior is compatible with a model in which the slow cycle reflects the number of accessible binding sites at promoters and each accessible site can be bound by fast cycling molecules. It is suggested that the oscillation of histone occupancy at the locus accounts for cycling accessibility. At this promoter the fast cycle is responsible for transcription initiation and the slow cycle for adjusting the amount of mRNA synthesis. This simple natural model can be combined with powerful yeast genetics to explore further the implication of transcription factors and chromatin remodeling in the kinetics of transcription.

Gene arrays therefore offer a huge number of possibilities when it comes to studying transcription and related processes. A recent estimate suggested that such tandem gene arrays represent 14% of all genes in vertebrate genomes, although most are made of only two genes (62). That means that these gene arrays are available for investigators who wish to study transcription in a natural genomic context. However, these natural gene arrays offer low control over the fundamental mechanisms involved in transcription. A number of teams have therefore developed artificial gene arrays in which the reporter genes were tinkered with to study specific core mechanisms. These modified arrays represent a good compromise between natural conditions and control by the investigator.

MMTV: mouse mammary tumor virus

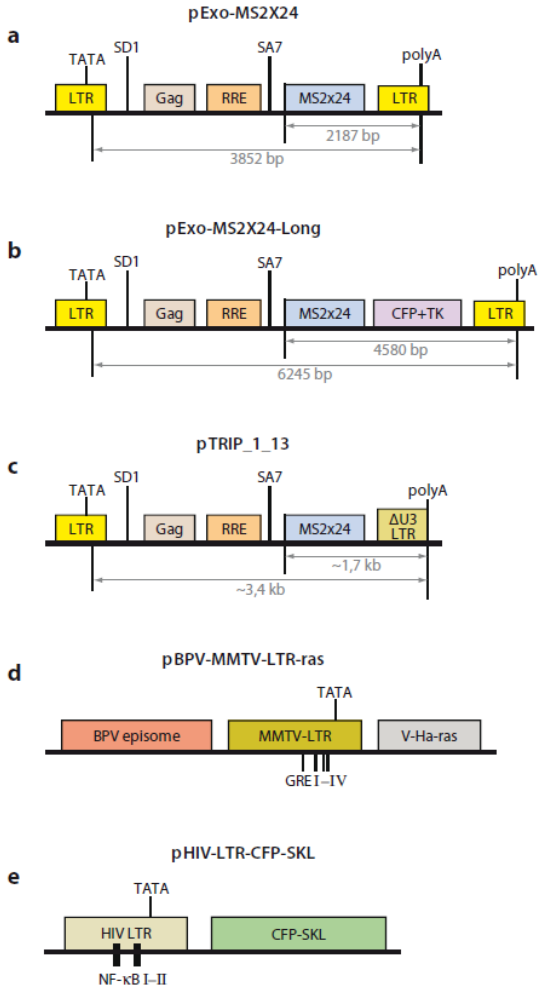
LTR: long terminal repeat

GR: glucocorticoid receptor

Artificial Gene Arrays

The first purpose of an artificial array created by Tsukamoto et al. (89) was to study chromatin remodeling during transcriptional activation. By inserting this Tet-inducible reporter gene containing Lac operator sites into large arrays, they observed changes in the chromatin upon activation and verified that the fluorescent protein encoded was correctly expressed. This system was then improved by Janicki et al. (43), who inserted 24 repeats of the MS2 bacteriophage replicase translational operator, which allowed them to visualize the mRNA transcribed from the genes. This work has permitted a real-time parallel analysis of transcription and modifications in the chromatin, notably by following histone 3.3 depositions and HP1α depletion (43). The same system was then used to estimate various kinetic steps of transcription (21). Other studies on chromatin remodeling made use of a gene array first described in 2000 by McNally (54), made up of mouse mammary tumor virus long terminal repeats (MMTV-LTRs), which can be activated by glucocorticoid receptors (GRs) (**Figure 3*d***) (45, 54, 58, 80). These studies have given great insight into the importance of various chromatin remodelers such as BRM and BRG1 during transcription. Finally, whether natural or artificial, gene arrays could be of great use in the study of many other nuclear processes, and the use of this powerful tool has only begun to reveal new and interesting observations.

One of the most surprising results concerning transcription obtained with gene arrays was that transcription in vivo is an inefficient process. The studies concerning Pol I by Dundr and colleagues (24) show that only about 10% of polymerases in the fibrillar centers are actually engaged in elongation. Similarly, Darzacq and colleagues (21) have shown that transcription in a Pol II array of genes was also inefficient, with only 1% of binding events resulting in the production of a complete mRNA. In the same study, they have determined the kinetics of different steps of transcription using FRAP experiments on an artificial gene array. They estimated that the elongation rate of Pol II is around 75 nucleotides per second, slightly slower than the 90 nucleotides per second published for Pol I (21, 24). These estimates are in accordance with previous findings of 50 and 100 nucleotides per second, respectively, using an independent method (76).



Viruses

Among viral species there exists a great variety of genomic structures and, consequently,

Figure 3

Diagrams of vectors containing viral sequences and kinetic parameters derived from the analyses used to produce tandem arrays for real-time live cell imaging results. (*a*) Vector used by Molle and colleagues to determine the residency times of Tat and Cdk9 on the HIV-1-LTR-driven transgene. (*b*, *c*) Modifications of the vector presented in panel *a* used in Reference 8 to measure transcription elongation rate. (*d*) MMTV/LTR containing glucocorticoid receptor (GR) binding sites developed by McNally and colleagues was used to determine various kinetic parameters of gene activation by hormones (45, 54, 58, 80). Similarly, the vector presented in panel *e* has been used to study NF-κB interaction with its binding sites (9). The original vector designations and references in which they were first described are given. The kinetic parameters of transcription obtained on tandem arrays in each of these vectors. Systems are discussed in the text. Diagrams are not drawn to scale. (For a compilation of other existing gene arrays, see Reference 68.) Abbreviations: BPV: bovine papilloma virus; Gag: retroviral gene coding for internal structural proteins; GRE: glucocorticoid-responsive element; NF-κB: nuclear factor κB; RRE: Rev-responsive element; SA7: splice acceptor 7; SD1: splice donor 1; SKL: Ser-Lys-Leu peroxisome-targeting peptide; v-Ha-ras: Harvey viral ras.

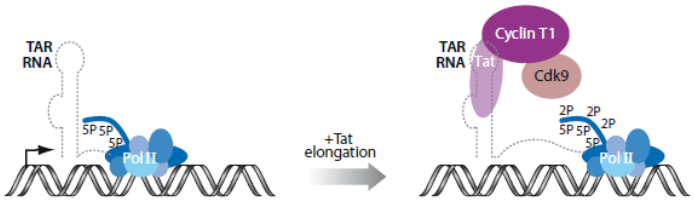


Figure 4

The role of the TAR:Tat:P-TEFb complex in HIV transcription. RNA polymerase II (Pol II) is stalled soon after initiation of transcription of the HIV provirus. The viral protein Tat recruits host cellular P-TEFb to the nascent stem-bulge-loop leader RNA, TAR (*trans*-activation responsive). The elongation factor P-TEFb composed of the cyclin T and the kinase Cdk9 phosphorylates the C-terminal domain repeats of the large subunit of Pol II at the Ser2 positions to stimulate processive elongation.

mechanisms of replication and gene expression. Only a small part of all known viruses depends on host cell polymerases in both replication and transcription.

For retroviruses, such as HIV-1 and MMTV, replication is the process whereby genome-sized RNA, which also functions as mRNA, is produced by host cell Pol II from the provirus integrated in the host genome. Thus, transcription is a means of their replication, as well as gene expression, and its tight regulation is important for the viral life cycle.

Some viruses use their own proteins to modify and redirect the activity of host cell's transcriptional machinery. In the case of HIV-1, transcription is activated by the viral protein Tat, which recruits the elongation factor P-TEFb [consisting of cyclin T and Cdk9, which phosphorylates the C-terminal domain of the large subunit of RNAP II] to the nascent stem-bulge-loop leader RNA, TAR (*trans*-activation responsive) (**Figure 4**) (44). Recently, cell lines have been created harboring tandem arrays of a reporter that carries the elements required for HIV-1 RNA production (**Figure 3*a–c***) (8, 57). In these cells the dynamics of the TAR:Tat:P-TEFb complex components has been analyzed by FRAP at the transcription sites visualized by expressing a nuclear MS2 phage coat protein (MS2cp) fusion with a fluorescent protein. Comparison of Cdk9-GFP dynamics at sites activated by Tat or phorbol 12-myristate 13-acetate/ionomycin showed that Cdk9 residency time at the HIV-1 transcription site was several times longer in the presence of Tat than in the absence of Tat (71 s and 11 s, respectively) and that it was similar to the residency time measured for Tat-GFP itself (55 s), suggesting that significant fractions of Tat and Cdk9 are present at the site as parts of the same complex, likely interacting with elongating Pol II (57). The transcription elongation rate measured by FRAP on MS2-GFP on the same HIV-1 tandem array and its variants with a longer transcribed region or without the U3 region in the 3′ LTR (which is required for efficient transcript 3′-end formation) (**Figure 3*a–c***) was estimated to be approximately 1.9 kb min^{-1} (8). The use of this HIV-MS2 tandem array also allowed the estimation of the Pol II residency time at the transcription site and its comparison to RNA production rates. The authors calculated a total polymerase residency time of 333 s, of which 114 s were attributed to elongation, 63 s to 3′-end processing and/or transcript release, and 156 s to polymerase remaining on the gene after RNA release (8).

Unlike P-TEFb, which stays at the transcription site induced by Tat for approximately 1 min (**Figure 4**), other transcription factors interacting with viral promoters interact transiently with the arrays. For example, mRFP-tagged NF-κB proteins interact with a multicopy array of transgenes containing the HIV 5′ LTR for only a few seconds (**Figure 3*e***) (9). FRAP of GFP-tagged GRs at the tandem array

MS2cp: coat protein of MS2 phage

containing MMTV-LTR promoters showed even shorter residency times (**Figure 3*d***) (54, 80).

NUCLEAR ORGANIZATION OF TRANSCRIPTION

Transcription Factories

Pol II is a multisubunit enzyme responsible for the transcription of most eukaryotic genes. The composite holoenzyme generated by the association of Pol II with other large complexes involved in related functions such as capping, splicing, and polyadenylation ensures the efficient production of mature transcripts. The key element necessary for coupling transcription with all the maturation steps is the large subunit of Pol II and in particular its C-terminal domain, essential for tethering the different machineries and regulating them temporally. The discovery of this intricate network gave rise to the idea that a specialized molecular machine is assembled at the site of transcription, nucleating from the promoter of an active gene. Because several of these factories may cluster together to ensure high local concentrations and therefore efficient interactions with all the partners involved, it is important to understand how different transcription units are transcribed and how their identity, nuclear surroundings, and positions could affect their expression.

Over the past fifteen years Cook and collaborators have put forward the concept of a superstructure called transcription factories, an assemblage of transcription and RNA-processing enzymes containing multiple genes. Before the advent of new visualizing methodologies, the transcription sites in mammalian cells were marked by elongation of nascent RNA in the presence of [^{3}H]uridine, [^{32}P]uridine, or Br-UTP, and subsequent observation at the fluorescent or electron microscope (39, 41, 42, 66). With these techniques they were able to see multiple nuclear foci sensitive to α-amanitin and containing splicing components (41). Those foci remained visible also after nucleolytic removal of most of the chromatin, highlighting the presence of an underlining structure responsible for the clustering of transcription units in which transcripts are both synthesized and processed (39). Moreover, these results are consistent with polymerases confined by the nucleoskeleton into factories and transcription occurring as templates slide past attached polymerases (40). Quantitative analysis (42) also showed that a typical factory contains approximately 30 engaged polymerases. Because two-thirds or more transcription units are associated with one polymerase at any time, each factory could contain at least 20 different transcription units.

A recent work on transcription factories (27) increased the resolution obtained with the electron microscope by coupling this technique with electron spectroscopic imaging. Electron spectroscopic imaging is a high-resolution and potent ultrastructural method that can be used to map atomic distribution in unstained preparations. Combining immunolabeling of the newly synthesized BrU-RNA with the distribution maps of nitrogen (N) and phosphorus (P) enabled specific atomic signature marking these nucleoplasmic sites to be identified. Template and nascent RNAs were attached to the surface of enormous protein-rich structures 87 nm in diameter and with a mass of 10 MDa. These structures appear porous, large enough to contain all the different protein complexes required for the complete maturation of the transcript. This finding suggested the idea that the polymerase was anchored, probably at the surface of the core, and that the DNA diffuses or loops to come in contact with a specific factory. Eskiw et al. (27) suggest that only a minority of all the machinery in the site is active, but that the high local concentrations will guarantee robust and efficient processivity.

Other questions concern how many transcription factors exist in a cell and how they should be classified. The first level of organization is the division of the three polymerases into different factories (67, 93). A more complicated issue is determining the influence of the genes' characteristics (promoter or presence of introns) on their arrangement within the

nucleus. Using replicating minichromosomes from Cos7 cells analyzed by FISH (fluorescent in situ hybridization) and 3C (chromosome conformation capture), Xu & Cook (93) examined whether the factories were specialized and the importance of the genes' distinctive characteristics. Their results confirmed that plasmids were concentrated in transcribing foci and that those being copied by different polymerases were not transcribed by the same foci. Moreover, units transcribed by Pol II, with different promoters (CMV and U2) or with the same promoter but with or without an intron in the coding sequence, are seen in nonoverlapping foci.

Even more intriguing are the results from live cells where the fluorescent tagging system of Pol II large subunit has been exploited (87). The GFP-tagged version was stably expressed in a Chinese hamster ovary cell line bearing a temperature-sensitive Pol II mutant, tsTM4, and it was observed that the fluorescent version was functional and normally assembled in the complex and rescued the phenotype. Because each factory contains only a few polymerases, it would be difficult to image those foci in live cells; however, significant results could be obtained by the study of polymerase kinetics in the nucleus of living cells. FRAP and FLIP (fluorescence loss in photobleaching) experiments in the nucleus gave important information on fractions of the enzyme in different states (48). They revealed two kinetic components in the Pol II population: A fast mobile component showed that ~75% of the molecules were diffusing freely and the immobile component showed that ~25% of the molecules were transiently immobile with a $t_{1/2}$ of ~20 min. This latter fraction was likely the active one, since incubation with DRB (5,6-dichloro-1-β-D-ribofuranosyl-benzimidazole), a potent inhibitor of elongation, eliminated it. Their model of the transcription cycle supports the idea that the enzyme spends most of the time diffusing and exchanging between the nucleoplasm and a promoter or a transcription factory. Once bound, a third of the time is mainly dedicated to elongation. With the improved use of this photobleaching technique, a more detailed analysis also resolved a third component resistant to DRB but sensitive to heat shock, representing the bound but not yet engaged fraction (37).

Whether these factories exist in most cells is a question that needs to be addressed with more sensitive technologies. Given that resolution problems pervade the experiments concerning testing of this concept, it will fall to the more quantitative, high-resolution methods to determine whether there simply exist gene-rich regional concentrations of transcription, or whether the factories are truly higher-order structures.

FISH: fluorescent in situ hybridization

Gene Positioning

The influence of the position of a gene with respect to the nuclear periphery on transcriptional competence has been extensively studied in recent years. Historically, the nuclear periphery has been seen as a nuclear substructure enriched in heterochromatin and thereby an area of transcriptional repression. However, data from yeast showing that active genes are often found in the nuclear periphery and in association with the nuclear pore complex led to a series of studies investigating the influence of nuclear positioning on transcriptional competence (3, 17).

In an early study, Cabal et al. (14) showed that the yeast *GAL1* gene changed its position from a mostly internal position to a preferential location at the nuclear periphery when the gene was activated, supporting the idea that the nuclear periphery harbors active genes. To do this, they used a fluorescently labeled *GAL1* locus in living cells by inserting an array of 112 TetO operators downstream of the *GAL1* gene, which upon coexpression of GFP-tagged TetR turns fluorescent (14). Nuclear positioning and movement of a locus were then followed using 4D live cell microscopy. Importantly, they found that the movement of the locus was not fully constrained but restricted to a 2D sliding movement at the nuclear envelope and was suggested to act as a gating mechanism to allow efficient mRNA processing and export.

Genes in yeast have been analyzed using this technique and were shown to move to the nuclear periphery upon activation (2, 12, 14, 23, 88). The requirements for the translocations, however, were often gene specific. In addition to components of the nuclear pore complex, promoters or elements in the 3′UTR, the SAGA complex of transcription factors, and components of the mRNA export machinery are involved (2, 12, 14, 23, 74, 88). Similarly, gene movement to the periphery has been suggested to occur before transcription starts for some genes, but it has also been suggested to occur as a result of transcription for other genes (12, 14, 23, 74, 88). It still remains to be shown if general principles exist that mediate the perinuclear localization of active genes in yeast and what fraction of genes use this mechanism to regulate their expression. Peripheral localization has also been suggested to mediate epigenetic memory over many generations (11).

These data from yeast led to the question whether such a mechanism may exist in higher eukaryotes. In yeast and in higher eukaryotes, chromatin loci in general are not statically positioned within the nucleus. In yeast as well as in higher eukaryotes, chromatin during interphase is mobile but mostly constrained within a radius of approximately 0.5–1 μm. That is less than 1% of the volume of a typical 10-μm spherical mammalian nucleus but half of the diameter of a yeast nucleus (52). In yeast, if a locus is located at the nuclear periphery, diffusion and the accessibility of binding sites at the nuclear periphery might be sufficient to allow tethering, as most genes likely encounter the nuclear periphery at least occasionally. In higher eukaryotes, however, if such events existed, it might require a more active movement, as a locus would have to move several microns to attach to the nuclear periphery or to nuclear pores. Peripheral heterochromatin is often interrupted at nuclear pores, indicating the presence of euchromatin in the vicinity of nuclear pores and making it possible that, like in yeast, active genes might get tethered to nuclear pore complexes to stimulate expression.

Recent live cell studies suggested that repositioning of genes from or to the nuclear periphery might have some influence on gene expression in higher eukaryotes, but that it might not be a major factor mediating gene expression. Imaging the naturally amplified *Drosophila* polytene nuclei in living salivary gland tissues by MPM did not reveal a preferred localization of the loci upon transcription induction. The genes could be found in the nuclear interior as well as at the nuclear periphery (94). Consistently, a GFP-tagged locus tethered to the nuclear periphery by a lamin B1 fusion maintained its transcriptional competence, indicating that sole peripheral or internal/central nuclear positioning does not influence transcription (51). However, another study suggested that expression of a subset of genes can reversibly be suppressed when tethered to the periphery, whereas many genes are not affected (30). Using DNA FISH, Reddy et al. (70) showed that genes can be silenced when targeted to the inner nuclear membrane. Together these results suggest that the nuclear periphery is not incompatible with active transcription but that it is not a primary determinant of whether genes are active. Different *cis*- and *trans*-acting factors are likely to determine whether peripherally localized genes in higher eukaryotes can be transcribed. However, chromatin movements in higher eukaryotes seem to actively play a role in regulating gene expression. Chromatin can frequently exhibit long-range movements of >2 μm during the cell cycle (90). Migration of an interphase chromosome site from the nuclear periphery to the interior has been observed 1–2 h after targeting a transcriptional activator to this site, showing a contrary localization to that in yeast (19). More surprisingly this movement was perturbed in specific actin and myosin I mutants, suggesting some kind of motor-driven movement. Similarly, actin-dependent intranuclear repositioning occurs with the U2 snRNA gene locus (25). If and how motor proteins mediate such long-distance chromatin movements still remain to be determined.

IMAGING ENDOGENOUS GENES

Steroid Receptors

Perhaps the most well-studied transcription factors of endogenous genes in living cells are nuclear receptor (NR) regulated. These ligand-activated transcription factors constitute the nuclear hormone receptor superfamily and are involved in regulating a vast array of eukaryotic genes. NR transcription is initiated by agonist binding to the receptor, forming either a homodimer or heterodimer complex. The corepressors (histone deacetylases, NR-specific corepressors) associated with the dimer are then replaced by coactivators such as histone acetylases (SRC/p160 family or CBP/p300) and histone methylases (CARM-1, PRMT-1). In addition, ATP-coupled chromatin remodeling complexes (SWI/SNF) are recruited. Eventually, the basal transcription machinery is assembled, followed by the initiation of Pol II. After initiation, transcription can be influenced by NR factors such as vitamin D receptor interacting protein and thyroid associated protein (38). Thus, NR transcription is an excellent model system for observing the cooperative interactions among enhancers, repressors, transcription factors, and basal transcription components (63). The view that has emerged from live cell studies utilizing fluorescence techniques such as FRAP, FRET, and FCS is that these NR complexes are highly dynamic: Individual species have dwelling times on the order of seconds to minutes. However, these same complexes can result in cycles of transcriptional progression that can last hours or days (56). NR-regulated transcription is therefore dynamically responsive to changes in agonist concentration and also capable of long-term changes of gene expression.

Live cell studies of NR-regulated transcription can be divided into those that study nuclear dynamics in general and those that focus on a particular locus. The first approach provides information about multiple possible transcription sites within the nucleus in addition to nonspecific interactions. The second approach has the benefit of providing specific information about interactions and dynamics at an active transcription site but usually requires modification of the locus—either multimerization of an endogenous gene (54) or creation of an artificial locus (85). The first example of this approach, which has been used by a number of investigators since its inception, was a large tandem array of a mouse mammary tumor virus/Harvey viral ras (MMTV/v-Ha-ras) reporter, which contains about 200 copies of the LTR and thus includes 800 to 1200 binding sites for the GR (54). This same array has been used for FRAP studies of the GR (6, 45, 55, 83), the androgen receptor (AR) (50), and the progesterone receptor (PR) (69). For each of those receptors, an agonist-dependent decrease in receptor mobility was observed [GR, $t_{1/2}$: 1–1.6 s (55); AR, $t_{1/2}$: 0.2–3.6 s (50); PR, $t_{1/2}$: 0.6–3.7 s; (69)]. A similar agonist-dependent decrease in mobility (increase in $t_{1/2}$) was also observed for general nuclear bleaching of the estrogen receptor [ER, $t_{1/2}$: 0.8–5.9 s (85)]. These observations demonstrate that the recovery time reflects the interaction of the NR with the locus in a specific fashion. In fact, Schaaf & Cidlowski (73) demonstrated that higher-affinity ligands result in slower recovery times, and Kino et al. (49) directly showed a positive correlation between FRAP $t_{1/2}$ times and transcriptional activity, with higher transcriptional activity corresponding to longer effective recovery times.

In contrast, other receptors do not show an agonist-dependent increase in $t_{1/2}$ for general nuclear recovery. The retinoic acid receptor (RAR), the thyroid hormone receptor (TR), the peroxisome proliferator-activated receptor (PPAR), and the retinoid X receptor (RXR) all have the same recovery time with or without ligand [RAR, $t_{1/2}$: 1.9–2.3 s; TR, $t_{1/2}$: 1.8–1.8 s (53); PPAR, $t_{1/2}$: 0.13–0.15 s; RXR, $t_{1/2}$: 0.2–0.25 s (29)]. In the case of PPAR, this lack of measurable difference may reflect some constitutive activity of the receptor (29).

In all FRAP experiments, the recovery dynamics will reflect both specific and nonspecific interactions. In the case of transient

NR: nuclear receptor

PR: progesterone receptor

transfections, in which an excess of receptor may be present, nonspecific interactions are likely to be a significant contribution to the dynamics for both locus-specific recovery and general nuclear recovery. The recovery curve is likely a convolution of more than one kinetic process. In computational models of AR dynamics, the recovery was separated into two distinct kinetic components: a fast component (due to diffusion or transient binding) of 1–5 s and a slow component of ~60 s (28, 50). This slow component presumably represents a longer-lived interaction in the vicinity of the gene such as with chromatin or nuclear matrix (28, 49, 55, 69, 73, 83), although the nature of this interaction is not clear and may vary between receptors.

In addition to receptor dynamics, several studies have addressed the kinetic behavior of coactivators involved in NR-regulated transcription. Becker et al. (6) observed the receptor coactivator GRIP1 (glucocorticoid receptor interacting protein 1) at the active MMTV array and measured a recovery time that was indiscernible from the GR $t_{1/2}$ (5 s), suggesting that the binding and release of these proteins may be coupled. CBP and SRC-1 (ER coactivators) have $t_{1/2}$ times of 4 s and 8 s, respectively (85); BRM and BRG1, subunits of the SWI/SNF chromatin remodeling complex, have $t_{1/2}$ times of 2 s and 4 s, respectively (45).

Taken together, the remarkable aspect of these data is that these recovery times are all less than or equal to 11 s (**Table 1**). Consider, for example, a typical NR transcription complex: NR $t_{1/2} = 5$ s, SRC1 $t_{1/2} = 8$ s, CBP $t_{1/2} = 4$ s (85), BRM $t_{1/2} = 2$ s, BRG1 $t_{1/2} = 4$ s (45), and GRIP1 $t_{1/2} = 5$ s (6). The only molecular species that has a dwell time on the order of minutes is the elongating polymerase ($t_{1/2} \sim 5$ min) (6). How might these transient interactions lead to transcriptional cycles that are observed in the timescale of hours? One idea that has been proposed is that of a transcriptional ratchet, in which permanent changes—methylation, acetylation, phosphorylation—accumulate at a transcription site as a result of the transient interactions described above (56). There are several suggestive directions about how such long-lived interactions might occur. SRC1 recovery becomes progressively slower at longer times after stimulation of ER with estradiol ($t_{1/2} = 8.0$–30.2 s) (85); chromatin decondensation seems to depend on polymerase elongation (59). Live cell experiments that follow the change in dynamics over an induction period are likely to be informative as well.

Imaging a Single Gene

Imaging the transcription of a single gene is potentially a powerful approach because it obviates the averaging inherent in gene array studies. This way, the behavior of individual transcription units can be quantified and their variability assessed. However, this has been difficult to achieve because of technical challenges: specifically detecting the desired locus and then observing the small numbers of factors involved in transcribing a single gene.

When a major challenge must be overcome, the tool of choice in vivo is fluorescence microscopy. Although a single fluorescent protein molecule can be detected when immobilized on a surface, it is difficult to resolve in the context of a living cell, where it undergoes fast diffusion or transport and where the fluorescent background can be high. So far, only a few experiments have managed to provide direct observation of gene expression at the single gene level.

A series of recent experiments demonstrated that it is possible to detect single protein products resulting from the expression of a single gene in live bacteria (15, 18, 97). From the distribution of proteins synthesized over time, it is then possible to test different models of transcription. In the first experiment (15), the reporter was a β-galactosidase protein, which produces a fluorescent product upon hydrolysis of a synthetic substrate. Hydrolysis of a large number of substrate molecules by a single enzyme provides the signal amplification necessary to observe a single protein. By observing discrete values in the rate of hydrolysis, the

Table 1 FRAP experiments summarized

Receptor	Cofactor	Nuclear/array	Ligand	Agonist = 1; partial antagonist = 2; antagonist = 3	$t_{1/2}$ (s)	Immobile fraction	Reference
AR		MMTV array	None		0.20	0	(50)
AR		MMTV array	R1881	1	3.60	0	
AR		MMTV array	DHT	1	5.30	0	
AR		MMTV array	TST	1	5.00	0	
AR		MMTV array	Bicalutamide	3	0.50	0	
AR		MMTV array	OHF	3	0.50	0	
AR		MMTV array	CPA	2	1.10	0	
AR		MMTV array	RU486	2	4.30	0	
AR		Nuclear	R1881	1	Two-component fit		(28)
ER		Nuclear	None		0.80		(86)
ER		Nuclear	Estradiol	1	5.90		
ER		Nuclear	4-HT	2	5.30		
ER		Nuclear	ICI	3	–		
ER	SRC	Nuclear	Estradiol	1	9.80		
ER		Artificial array—LacO	Estradiol	1	–		(Binding to lacO is tight)
ER	SRC	Artificial array—LacO	None		2.10		(85)
ER	SRC	Artificial array—LacO	Estradiol	1	8.00		
ER	CBP	Artificial array—LacO	Estradiol	1	4.20		
ER		Nuclear	None		1.60	0.14	
ER		Nuclear	Estradiol	1	5.80	0.44	(53)
GR		MMTV array	Dexamethasone	1	5.00		(6)
GR	GRIP (p160 family)	MMTV array	Dexamethasone	1	5.00		
GR	RNAPII	MMTV array	Dexamethasone	1	300.00		
GR		MMTV array	RU486	3	<5		(84)
GR		MMTV array	Corticosterone	3	<5		
GR		Nuclear	None		1.00		(73)
GR		Nuclear	Dexamethasone	1	1.90		
GR		Nuclear	Triamcinolone	1	1.87		
GR		Nuclear	Corticosterone	1	1.13		
GR		Nuclear	RU486	3	1.43		
GR		Nuclear	ZK98299	3	1.07		
GR		Nuclear	Dexamethasone	1	0.43–>1.11		(49)
GR		MMTV array	Dexamethasone	1	1.60		(55)

(*Continued*)

Table 1 (*Continued*)

Receptor	Cofactor	Nuclear/array	Ligand	Agonist = 1; partial antagonist = 2; antagonist = 3	$t_{1/2}$ (s)	Immobile fraction	Reference
GR	407C-GR mutant	MMTV array	Dexamethasone	1	1.30		
GR	ΔLBD	MMTV array	Dexamethasone	1	0.69		
GR		MMTV array	RU486	3	0.82		
GR		MMTV array	Dex-Mex		1.40		
GR	BRG1	MMTV array	Dexamethasone	1	3.90		(45)
GR	BRM	MMTV array	Dexamethasone	1	1.95		
RAR		Nuclear	None		1.90	0.18	(53)
RAR		Nuclear	Retinoic acid	1	2.30	0.18	
TR		Nuclear	None		1.80	0.12	(53)
TR		Nuclear	Triiodithyronine	1	1.80	0.14	
PR		MMTV array	None		0.60		(69)
PR		MMTV array	R5020	1	3.70		
PR		MMTV array	RU486	3	11.00		
PR		MMTV array	ZK98299	3	1.80		
PPAR α		Nuclear	None		0.13	0.017	(29)
PPAR α		Nuclear	wy14643	1	0.15	2.4	
PPAR α		Nuclear	None		0.10	0.014	
PPAR α		Nuclear	L-165041	1	0.16	0.03	
PPAR α		Nuclear	None		0.10	0.05	
PPAR α		Nuclear	Rosiglitazone	1	0.12	0.033	
RXR α		Nuclear	None		0.20	0.019	
RXR α		Nuclear	9-*cis* retinoic acid	1	0.25	0.064	

Abbreviations: AR, androgen receptor; ER, estrogen receptor; GR, glucocorticoid receptor; GRIP, glucocorticoid receptor interacting protein; MMTV, mouse mammary tumor virus; PPAR α, peroxisome proliferator-activated receptor α; PR, progesterone receptor; RAR, retinoic acid receptor; RNAPII, RNA polymerase II; RXR α, retinoid X receptor α; TR, thyroid hormone receptor.

authors could indeed resolve single protein numbers. In subsequent experiments, the reporter was a fluorescent protein fused to a membrane protein (18, 97). When bound to the membrane, the protein is slowly diffusing and it is therefore possible to accumulate enough fluorescence to resolve a single protein.

These experiments studied reporter genes under the control of the Lac repressor. In this classic system, two operator sequences on the DNA can be bound by a tetramer repressor. Upon full induction, the repressor unbinds the DNA and the cell fully expresses the *lac* genes downstream. In the absence of inducer, protein is produced in infrequent bursts (0.1–1 per cell cycle) in which a few (5–20) proteins are produced. The distribution of the number of proteins produced per burst is consistent with a model in which a burst results from the transcription of a single mRNA molecule, finally yielding a few proteins. In the regime of moderate inducer concentration, both low-expressing cells (1–10 proteins) and high-expressing ones (hundreds of proteins) are observed. This bimodal distribution results from the presence of frequent, small bursts (similar to the

noninduced state) and infrequent, large bursts of protein production. The authors proposed a model in which the small bursts consist of a partial dissociation of the repressor from one of the operator sequences; one RNA molecule is transcribed typically before the repressor binds back the operator sequence. In contrast, the large bursts correspond to full dissociation of the repressor from both operator sequences. In this case, many mRNA molecules are transcribed before another repressor binds the DNA, leading to the production of a large number of proteins.

A similar detection approach was used to study transcription factor dynamics in *Escherichia coli* at the single molecule level (27). Lac repressor (Lac I) molecules fused to a fluorescent protein could be detected when bound to their promoter sequence by imaging for long periods of time (~1 s) to average out the background of freely diffusing molecules. This made it possible to measure the kinetics of association of Lac I to its promoter in vivo. The authors also used short light excitation pulses to characterize the diffusion of the repressor as well as its nonspecific binding to DNA. From these results emerged a picture of Lac I dynamics: Searching for its target sequence, the protein spends 87% of its time in short events (<5 ms), where it is nonspecifically bound to DNA and undergoes 1D diffusion while it scans the DNA. These short events are separated by periods where the repressor diffuses in three dimensions between different DNA segments.

It is also possible to directly visualize mRNA molecules using a technique that exploits the high affinity between RNA stem loops and the bacteriophage MS2 coat protein (7). By introducing repeats of the stem loop coding sequence in the desired gene, and expressing the MS2 coat protein fused to a fluorescent protein, one can detect single mRNA molecules. This technique has been used in numerous studies to characterize single mRNA motion and localization in *E. coli* (34), yeast (5, 7), mammalian cells (33, 72, 77), and *Drosophila* oocytes (31, 91, 98).

Golding et al. (35) used this technique to study transcription in *E. coli* by utilizing an inducible reporter gene under the control of the $P_{lac/ara}$ promoter. By measuring the distribution of mRNA molecules per cell, the authors tested two models for transcription. The simplest model, in which transcription events were randomly initiated according to a Poisson process (with a constant probability per unit time), could not fit the data; a more elaborate model, in which the gene can switch between an "off" state (no transcription takes place) and an "on" state (transcription is randomly initiated) successfully described the data. The gene stays "on" for typically 6 min, during which it produces approximately two transcripts. In contrast, the "off" state lasts much longer (~37 min), which results in a burst-like transcription behavior, even in full-induction conditions.

Bursts of transcription have also been observed on *dscA*, an endogenous developmental gene in the social amoeba *Dictyostelium discoideum* (20). Although the authors could not detect single mRNA particles, they could resolve the sites of transcription because of the high fluorescence accumulated by the multiple nascent mRNAs. As differentiation occurred, they could observe the *dscA* gene switch between the "on" and "off" states, which displayed similar lifetimes (5.2 and 5.8 min, respectively), in contrast with the *E. coli* result. Over the course of development, variation was only observed in the fraction of the population expressing the gene, but the lifetime of the "on" and "off" states remained constant. In addition, the authors observed transcriptional memory, as a gene was more likely to enter the "on" state if it had been transcribing before than when it was undergoing de novo transcription.

In spite of their quantitative differences, both studies could be modeled the same way, using a simple two-state system. The nature of the event(s) that dictate the transition between the "on" and "off" states has yet to be discovered, but it could consist of DNA conformational change and/or chromatin remodeling, binding (or release) of an activator (or repressor), or transcription pausing/reinitiation.

These studies demonstrate the potential of single-molecule techniques in studying transcription in vivo and open avenues for future research. Upcoming directions could involve expanding these observations to different systems and/or trying to combine imaging of different factors at a given locus.

ANALYSIS OF KINETICS

Mobility of Transcription Factors

Transcription factor mobility represents the process of a genome-wide search for specific target sites. Since the discovery of the GFP, more precise observations of nuclear protein mobility have been enabled. In vivo techniques identify populations of transcription factors on a real-time timescale, breaking down many assumptions previously held about this topic.

Dynamics

The use of fluorescent proteins made it possible to conduct experiments on the transcription factors of yeast (47), bacteria (26), and mammalian cells (54), enlightening many aspects of the dynamic behavior of transcription factors, from Brownian motion to anomalous diffusion to cyclic binding (dynamic equilibria) at binding sites to dynamic complex formation. Transcription factors are generally impaired in their diffusion throughout the nucleus by unspecific interactions with other nuclear components (64). Furthermore, a tagged nuclear protein might exhibit more than one apparent diffusion constant due to complex formation. Most models of FRAP have been applied to homogeneously and globally distributed binding sites, easily approximated by diffusion.

The pioneering work on the construction of localized cluster binding sites in the genome (54, 89) made it possible to address specific binding of transcription factors in the nucleus. Sprague et al. (80) used such a system to prove that recoveries resulting from bleaching of the tandem array area could not fit a model accounting for only diffusion. The new model involves "on" and "off" rates of the transcription factor's binding to the promoter array and provides information on the binding dynamics of the system. A similar construct has been used by the Singer laboratory (21) to analyze transcriptional mechanisms by bleaching the Pol II and nascent RNAs on a tandem array. This particular system was demonstrated to be kinetically independent of the availability of the freely diffusing components, therefore making it possible to disregard the diffusing component (80) and to use first-order differential equations to model the reactions.

The work of the Natoli's laboratory (9) on NF-κB promoter binding microdynamics showed that stable bindings were actually states of dynamic equilibrium between promoter-bound and nucleoplasmic dimers (**Figure 3*c***). In a subsequent study, Karpova et al. (47) showed that the yeast transcription factor Ace1p fit an accessibility model in which the slow cycle of binding reflects the number of accessible binding sites at promoters and each accessible site can be bound by fast-cycling molecules.

Complex formation of transcription factors on their promoters is likewise of a highly dynamic nature: The factor and its partners do not associate with and are not released from target promoters as a single and stable complex (9, 36, 71). Bosisio et al. (9) specified that NF-κB residence time on specific sites defined a stochastic window during which general transcription factors and possibly additional activators must collide with the same regulatory region for transcription to occur. Furthermore, Gorski et al. (36) successfully proved the role of complex formation in regulating transcription. Remodeling factors also play a critical role in the regulation of gene expression and governing the dynamics of transcription factors (46, 47).

Modeling

FRAP analysis quantifies and sets a kinetic model characterized by parameters of diffusion coefficients, chemical rates, and residence times

translatable into differential equations. Taking into account only diffusion, a convenient way of displaying fluorescence recovery curves is the form defining the mobile fraction:

$$f_K(t) = [F_K(t) - F_K(0)]/[(F_K(\infty) - F_K(0)].$$

Under assumptions of a Gaussian intensity profile laser, this gives the closed form solution of the normalized fluorescence recovery:

$$F_K(t) = (q\, P_0 C_0/A)\nu K^{-\upsilon}\Gamma(\upsilon)P(2K|\upsilon),$$

where P_0 designates the total laser power, K the bleaching parameter, C_0 the initial fluorophore concentration, and A the attenuation factor of beam during the observation of the FRAP recovery. q is the product of all quantum efficiencies of light absorption, emission, and detection. Furthermore, parameter ν and the characteristic diffusion time τ_D are given by $\upsilon = (1 + 2t/\tau_D)^{-1}$ and $\tau_D = w^2/4D$, respectively, where t is time and w is the radius of the laser beam at e^{-2} intensity/height. $\Gamma(\upsilon)$ is given by the gamma function (4). $P(2K|\upsilon)$ is the probability distribution tabulated in Reference 1.

However, this might be only an approximation when it comes to systems in which specific binding cannot be ignored. A complete solution to FRAP reaction-diffusion equations has been proposed (81) in which various special cases of FRAP with binding (diffusion/binding dominant) can be covered by a set of differential equations including the diffusion terms given above, as well as the chemical kinetics of binding.

Whether a dynamic system is diffusion limited depends on the magnitude of two parameters: diffusion time and association rate. The relative magnitude of these two parameters reflects potential interplay between diffusion and binding and thus determines whether a recovery is diffusion coupled or uncoupled (79). A simple method of testing whether a system is diffusion limiting is to vary the spot size of the bleach: If the recovery is dependent on the spot size, the system is diffusion limiting and must be included in the analysis (83).

Apart from diffusion modeling with differential equations, Rino et al. (71) successfully portrayed modeling of splicing protein kinetics in the nucleus with a method involving k_{on} and k_{off} rates in a Monte Carlo simulation. Other types of modeling might explain dynamic behaviors (for more details, see Reference 65).

CONCLUSION

Controlled manipulation of the biological system by using drugs has proved to be a useful tool to perturb biological mechanisms in order to obtain deeper understanding of the mechanisms involved (32, 21). Other biological manipulation stems from the construction of binding defective mutants of the transcription factors under analysis (46, 82). Photoactiveable and photoswitchable fluorophores have a particular advantage, however, in that they can be used with single-cell, single-molecule sensitivity, and they produce photons, which are easily measured, quantified, and converted to the dynamic behavior of transcription (92). Further, hyper-resolution techniques provide a tool to produce single-molecule dynamic measurements at the subdiffraction level. Using these approaches, it will be possible to answer crucial questions about how transcription factor dynamics regulate gene expression, how transcription factors sort the right genes, and how they search for their targets. Short residence times, stochastic formation of complexes, anomalous diffusion with continuous assembly, and disassembly of the transcription factors is only the beginning of a complex story about the dynamic behavior of transcription factors.

THE FUTURE

The current conclusions regarding transcription dynamics are based mainly on synthetic genes and cell lines that give us some insight into the processes involved in gene expression. However, the next important step is to apply the technology to minimally perturbed systems, endogenous genes, and primary cells or tissues. To achieve this we will need more

sensitive systems capable of processing weak signals. Additionally, high-speed imaging will be required to separate transient and rapid events from the diffusional rates occurring in the background. Brighter fluorochromes with lower photobleaching or novel labeling systems capable of multiplexing will also be required. Finally, the digitization of the data will allow for the type of mining that is common with microarray databases, but now required are algorithms capable of extracting data from large image sets, particularly those that contain 4D information (a time series in three dimensions). The field therefore will assemble expertise from engineers, computer scientists, chemists, physicists, and biophysicists. As these explorations evolve, they will lead to leaps in understanding the biological basis of gene expression.

SUMMARY POINTS

1. MPM imaging of *Drosophila* polytene loci provides an experimental system of analyzing transcription factor dynamics and function at specific native gene loci in vivo and in real time.
2. NR-induced/regulated transcription is an excellent model system for observing the cooperative interactions among enhancers, repressors, transcription factors, and basal transcription components.
3. The view that has emerged from live cell studies, utilizing fluorescence techniques such as FRAP, FRET, and FCS, is that these NR complexes are highly dynamic: Individual species have dwell times on the order of seconds to minutes.
4. Single-molecule studies of a single gene, if technically challenging, offer the most detailed framework on which to test the models of transcription.
5. The cell may organize transcription machinery to efficiently produce mature transcripts.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We would like to thank Gerry Rubin and Kevin Moses for the support of this Consortium and Jon Lis and Watt Webb for reading portions of this manuscript and giving permission for use of published figures. Support for personnel is from the HHMI, the NIH, and the CNRS. The authors thank Shailesh Shenoy for his help in preparing the manuscript.

LITERATURE CITED

1. Abramowitz M, Stegun IA. 1965. *Handbook of Mathematical Functions*. New York: Dover
2. Abruzzi KC, Belostotsky DA, Chekanova JA, Dower K, Rosbash M. 2006. 3′-end formation signals modulate the association of genes with the nuclear periphery as well as mRNP dot formation. *EMBO J.* 25:4253–62
3. Akhtar A, Gasser SM. 2007. The nuclear envelope and transcriptional control. *Nat. Rev. Genet.* 8:507–17
4. Axelrod D, Koppel DE, Schlessinger J, Elson E, Webb WW. 1976. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys. J.* 16:1055–69

5. Beach DL, Salmon ED, Bloom K. 1999. Localization and anchoring of mRNA in budding yeast. *Curr. Biol.* 9:569–78
6. Becker M, Baumann C, John S, Walker DA, Vigneron M, et al. 2002. Dynamic behavior of transcription factors on a natural promoter in living cells. *EMBO Rep.* 3:1188–94
7. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. 1998. Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* 2:437–45
8. Boireau S, Maiuri P, Basyuk E, de la Mata M, Knezevich A, et al. 2007. The transcriptional cycle of HIV-1 in real-time and live cells. *J. Cell Biol.* 179: 291–304
9. Bosisio D, Marazzi I, Agresti A, Shimizu N, Bianchi ME, Natoli G. 2006. A hyperdynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF-kappaB-dependent gene activity. *EMBO J.* 25:798–810
10. Bres V, Yoh SM, Jones KA. 2008. The multi-tasking P-TEFb complex. *Curr. Opin. Cell Biol.* 20:334–40
11. Brickner DG, Cajigas I, Fondufe-Mittendorf Y, Ahmed S, Lee P-C, et al. 2007. H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol.* 5:e81
12. Brickner JH, Walter P. 2004. Gene recruitment of the activated INO1 locus to the nuclear membrane. *PLoS Biol.* 2:e342
13. Bridges CB. 1935. Salivary chromosome maps. *J. Hered.* 26:60–64
14. Cabal GG, Genovesio A, Rodriguez-Navarro S, Zimmer C, Gadal O, et al. 2006. SAGA interacting factors confine subdiffusion of transcribed genes to the nuclear envelope. *Nature* 441:770–73
15. Cai L, Friedman N, Xie XS. 2006. Stochastic protein expression in individual cells at the single molecule level. *Nature* 440:358–62
16. Carter DR, Eskiw C, Cook PR. 2008. Transcription factories. *Biochem. Soc. Trans.* 36:585–89
17. Casolari JM, Brown CR, Komili S, West J, Hieronymus H, Silver PA. 2004. Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell* 117:427–39
18. **Choi PJ, Cai L, Frieda K, Xie XS. 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322:442–46**
19. Chuang C-H, Carpenter AE, Fuchsova B, Johnson T, de Lanerolle P, Belmont AS. 2006. Long-range directional movement of an interphase chromosome site. *Curr. Biol.* 16:825–31
20. **Chubb JR, Trcek T, Shenoy SM, Singer RH. 2006. Transcriptional pulsing of a developmental gene. *Curr. Biol.* 16:1018–25**
21. **Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, et al. 2007. In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 14:796–806**
22. Darzacq X, Singer RH. 2008. The dynamic range of transcription. *Mol. Cell* 30:545–46
23. Dieppois G, Iglesias N, Stutz F. 2006. Cotranscriptional recruitment to the mRNA export receptor Mex67p contributes to nuclear pore anchoring of activated genes. *Mol. Cell. Biol.* 26:7858–70
24. **Dundr M, Hoffmann-Rohrer U, Hu Q, Grummt I, Rothblum LI, et al. 2002. A kinetic framework for a mammalian RNA polymerase in vivo. *Science* 298:1623–26**
25. Dundr M, Ospina JK, Sung M-H, John S, Upender M, et al. 2007. Actin-dependent intranuclear repositioning of an active gene locus in vivo. *J. Cell Biol.* 179:1095–103
26. **Elf J, Li GW, Xie XS. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316:1191–94**
27. Eskiw CH, Rapp A, Carter DR, Cook PR. 2008. RNA polymerase II activity is located on the surface of protein-rich transcription factories. *J. Cell Sci.* 121:1999–2007
28. Farla P, Hersmus R, Geverts B, Mari PO, Nigg AL, et al. 2004. The androgen receptor ligand-binding domain stabilizes DNA binding in living cells. *J. Struct. Biol.* 147:50–61
29. Feige JN, Gelman L, Tudor C, Engelborghs Y, Wahli W, Desvergne B. 2005. Fluorescence imaging reveals the nuclear behavior of peroxisome proliferator-activated receptor/retinoid X receptor heterodimers in the absence and presence of ligand. *J. Biol. Chem.* 280:17880–90
30. Finlan LE, Sproul D, Thomson I, Boyle S, Kerr E, et al. 2008. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.* 4:e1000039
31. Forrest KM, Gavis ER. 2003. Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in *Drosophila*. *Curr. Biol.* 13:1159–68

18. The number of proteins expressed from a gene under the control of LacO is linked to dissociation of the repressor.

20. This work demonstrates the stochastic behavior of an endogenous, developmentally regulated gene.

21. This work demonstrates the power of combining live cell imaging of transcription with mathematical modeling to determine rate constants for each of the various components of the process.

24. This is the first work to measure polymerase kinetics.

26. Specific binding, diffusion tracking, and colocalization of the transcription factor to its DNA binding domains are measured.

32. Fromaget M, Cook PR. 2007. Photobleaching reveals complex effects of inhibitors on transcribing RNA polymerase II in living cells. *Exp. Cell Res.* 313:3026–33
33. Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard JM, et al. 2003. Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr. Biol.* 13:161–67
34. Golding I, Cox EC. 2004. RNA dynamics in live *Escherichia coli* cells. *Proc. Natl. Acad. Sci. USA* 101:11310–15
35. Golding I, Paulsson J, Zawilski SM, Cox EC. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell* 123:1025–36
36. Gorski SA, Snyder SK, John S, Grummt I, Misteli T. 2008. Modulation of RNA polymerase assembly dynamics in transcriptional regulation. *Mol. Cell* 30:486–97
37. Hieda M, Winstanley H, Maini P, Iborra FJ, Cook PR. 2005. Different populations of RNA polymerase II in living mammalian cells. *Chromosome Res.* 13:135–44
38. Hinojos CA, Sharp ZD, Mancini MA. 2005. Molecular dynamics and nuclear receptor function. *Trends Endocrinol. Metab.* 16:12–18
39. Iborra FJ, Pombo A, Jackson DA, Cook PR. 1996. Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *J. Cell Sci.* 109(Pt. 6):1427–36
40. Iborra FJ, Pombo A, McManus J, Jackson DA, Cook PR. 1996. The topology of transcription by immobilized polymerases. *Exp. Cell Res.* 229:167–73
41. Jackson DA, Hassan AB, Errington RJ, Cook PR. 1993. Visualization of focal sites of transcription within human nuclei. *EMBO J.* 12:1059–65
42. Jackson DA, Iborra FJ, Manders EM, Cook PR. 1998. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell* 9:1523–36
43. Janicki SM, Tsukamoto T, Salghetti SE, Tansey WP, Sachidanandam R, et al. 2004. From silencing to gene expression: real-time analysis in single cells. *Cell* 116:683–98
44. Jeang KT, Xiao H, Rich EA. 1999. Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *J. Biol. Chem.* 274:28837–40
45. Johnson TA, Elbi C, Parekh BS, Hager GL, John S. 2008. Chromatin remodeling complexes interact dynamically with a glucocorticoid receptor-regulated promoter. *Mol. Biol. Cell* 19:3308–22
46. Karpova TS, Chen TY, Sprague BL, McNally JG. 2004. Dynamic interactions of a transcription factor with DNA are accelerated by a chromatin remodeller. *EMBO Rep.* 5:1064–70
47. Karpova TS, Kim MJ, Spriet C, Nalley K, Stasevich TJ, et al. 2008. Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science* 319:466–69
48. Kimura H, Sugaya K, Cook PR. 2002. The transcription cycle of RNA polymerase II in living cells. *J. Cell Biol.* 159:777–82
49. Kino T, Liou SH, Charmandari E, Chrousos GP. 2004. Glucocorticoid receptor mutants demonstrate increased motility inside the nucleus of living cells: time of fluorescence recovery after photobleaching (FRAP) is an integrated measure of receptor function. *Mol. Med.* 10:80–88
50. Klokk TI, Kurys P, Elbi C, Nagaich AK, Hendarwanto A, et al. 2007. Ligand-specific dynamics of the androgen receptor at its response element in living cells. *Mol. Cell Biol.* 27:1823–43
51. Kumaran RI, Spector DL. 2008. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J. Cell Biol.* 180:51–65
52. Lanctot C, Cheutin T, Cremer M, Cavalli G, Cremer T. 2007. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8:104–15
53. Maruvada P, Baumann CT, Hager GL, Yen PM. 2003. Dynamic shuttling and intranuclear mobility of nuclear hormone receptors. *J. Biol. Chem.* 278:12425–32
54. McNally JG, Muller WG, Walker D, Wolford R, Hager GL. 2000. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* 287:1262–65
55. Meijsing SH, Elbi C, Luecke HF, Hager GL, Yamamoto KR. 2007. The ligand binding domain controls glucocorticoid receptor dynamics independent of ligand release. *Mol. Cell Biol.* 27:2442–51
56. Metivier R, Reid G, Gannon F. 2006. Transcription in four dimensions: nuclear receptor-directed initiation of gene expression. *EMBO Rep.* 7:161–67
57. Molle D, Maiuri P, Boireau S, Bertrand E, Knezevich A, et al. 2007. A real-time view of the TAR:Tat:P-TEFb complex at HIV-1 transcription sites. *Retrovirology* 4:36

34. A real-time quantitative study of the expression of a single gene in bacteria demonstrating burst-like transcription.

36. This work demonstrates the power of combining live cell imaging of transcription with mathematical modeling to determine rate constants for each of the various components of the process.

47. This work demonstrates the rapid turnover of transcription factors at the promoter.

54. This seminal paper was the first to show the dynamic behavior of transcription factors at an active gene locus in a living cell.

58. Muller WG, Rieder D, Karpova TS, John S, Trajanoski Z, McNally JG. 2007. Organization of chromatin and histone modifications at a transcription site. *J. Cell Biol.* 177:957–67
59. Muller WG, Walker D, Hager GL, McNally JG. 2001. Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. *J. Cell Biol.* 154:33–48
60. Nalley K, Johnston SA, Kodadek T. 2006. Proteolytic turnover of the Gal4 transcription factor is not required for function in vivo. *Nature* 442:1054–57
61. Ni Z, Saunders A, Fuda NJ, Yao J, Suarez JR, et al. 2008. P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. *Mol. Cell Biol.* 28:1161–70
62. Pan D, Zhang L. 2008. Tandemly arrayed genes in vertebrate genomes. *Comp. Funct. Genomics* 2008:545269
63. Perissi V, Rosenfeld MG. 2005. Controlling nuclear receptors: the circular logic of cofactor cycles. *Nat. Rev. Mol. Cell Biol.* 6:542–54
64. Phair RD, Misteli T. 2000. High mobility of proteins in the mammalian cell nucleus. *Nature* 404:604–9
65. Phair RD, Misteli T. 2001. Kinetic modelling approaches to in vivo imaging. *Nat. Rev. Mol. Cell Biol.* 2:898–907
66. Pombo A, Cook PR. 1996. The localization of sites containing nascent RNA and splicing factors. *Exp. Cell Res.* 229:201–3
67. Pombo A, Jackson DA, Hollinshead M, Wang Z, Roeder RG, Cook PR. 1999. Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J.* 18:2241–53
68. Rafalska-Metcalf IU, Janicki SM. 2007. Show and tell: visualizing gene expression in living cells. *J. Cell Sci.* 120:2301–7
69. Rayasam GV, Elbi C, Walker DA, Wolford R, Fletcher TM, et al. 2005. Ligand-specific dynamics of the progesterone receptor in living cells and during chromatin remodeling in vitro. *Mol. Cell Biol.* 25:2406–18
70. Reddy KL, Zullo JM, Bertolino E, Singh H. 2008. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452:243–47
71. Rino J, Carvalho T, Braga J, Desterro JM, Luhrmann R, Carmo-Fonseca M. 2007. A stochastic view of spliceosome assembly and recycling in the nucleus. *PLoS Comput. Biol.* 3:2019–31
72. Rook MS, Lu M, Kosik KS. 2000. CaMKIIalpha 3′ untranslated region-directed mRNA translocation in living neurons: visualization by GFP linkage. *J. Neurosci.* 20:6385–93
73. Schaaf MJ, Cidlowski JA. 2003. Molecular determinants of glucocorticoid receptor mobility in living cells: the importance of ligand affinity. *Mol. Cell Biol.* 23:1922–34
74. Schmid M, Arib G, Laemmli C, Nishikawa J, Durussel T, Laemmli UK. 2006. Nup-PI: the nucleopore-promoter interaction of genes in yeast. *Mol. Cell* 21:379–91
75. Schneider DA, Nomura M. 2004. RNA polymerase I remains intact without subunit exchange through multiple rounds of transcription in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 101:15112–17
76. Sehgal PB, Derman E, Molloy GR, Tamm I, Darnell JE. 1976. 5,6-Dichloro-1-Beta-D-ribofuranosylbenzimidazole inhibits initiation of nuclear heterogeneous RNA chains in HeLa cells. *Science* 194:431–33
77. Shav-Tal Y, Darzacq X, Shenoy SM, Fusco D, Janicki SM, et al. 2004. Dynamics of single mRNPs in nuclei of living cells. *Science* 304:1797–800
78. Shav-Tal Y, Singer RH, Darzacq X. 2004. Imaging gene expression in single living cells. *Nat. Rev. Mol. Cell Biol.* 5:855–61
79. Sprague BL, McNally JG. 2005. FRAP analysis of binding: proper and fitting. *Trends Cell Biol.* 15:84–91
80. Sprague BL, Muller F, Pego RL, Bungay PM, Stavreva DA, McNally JG. 2006. Analysis of binding at a single spatially localized cluster of binding sites by fluorescence recovery after photobleaching. *Biophys. J.* 91:1169–91
81. Sprague BL, Pego RL, Stavreva DA, McNally JG. 2004. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys. J.* 86:3473–95
82. Sprouse RO, Shcherbakova I, Cheng H, Jamison E, Brenowitz M, Auble DT. 2008. Function and structural organization of Mot1 bound to a natural target promoter. *J. Biol. Chem.* 283:24935–48
83. Stavreva DA, McNally JG. 2004. Fluorescence recovery after photobleaching (FRAP) methods for visualizing protein dynamics in living mammalian cell nuclei. *Methods Enzymol.* 375:443–55

84. Stavreva DA, Muller WG, Hager GL, Smith CL, McNally JG. 2004. Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Mol. Cell Biol.* 24:2682–97
85. Stenoien DL, Nye AC, Mancini MG, Patel K, Dutertre M, et al. 2001. Ligand-mediated assembly and real-time cellular dynamics of estrogen receptor alpha-coactivator complexes in living cells. *Mol. Cell Biol.* 21:4404–12
86. Stenoien DL, Patel K, Mancini MG, Dutertre M, Smith CL, et al. 2001. FRAP reveals that mobility of oestrogen receptor-alpha is ligand- and proteasome-dependent. *Nat. Cell Biol.* 3:15–23
87. Sugaya K, Vigneron M, Cook PR. 2000. Mammalian cell lines expressing functional RNA polymerase II tagged with the green fluorescent protein. *J. Cell Sci.* 113(Pt. 15):2679–83
88. Taddei A, Van Houwe G, Hediger F, Kalck V, Cubizolles F, et al. 2006. Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature* 441:774–78
89. Tsukamoto T, Hashiguchi N, Janicki SM, Tumbar T, Belmont AS, Spector DL. 2000. Visualization of gene activity in living cells. *Nat. Cell Biol.* 2:871–78
90. Walter J, Schermelleh L, Cremer M, Tashiro S, Cremer T. 2003. Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J. Cell Biol.* 160:685–97
91. Weil TT, Forrest KM, Gavis ER. 2006. Localization of bicoid mRNA in late oocytes is maintained by continual active transport. *Dev. Cell* 11:251–62
92. Xie XS, Choi PJ, Li GW, Lee NK, Lia G. 2008. Single-molecule approach to molecular biology in living bacterial cells. *Annu. Rev. Biophys.* 37:417–44
93. Xu M, Cook PR. 2008. Similar active genes cluster in specialized transcription factories. *J. Cell Biol.* 181:615–23
94. Yao J, Ardehali MB, Fecko CJ, Webb WW, Lis JT. 2007. Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation. *Mol. Cell* 28:978–90
95. Yao J, Munson KM, Webb WW, Lis JT. 2006. Dynamics of heat shock factor association with native gene loci in living cells. *Nature* 442:1050–53
96. Yao J, Zobeck KL, Lis JT, Webb WW. 2008. Imaging transcription dynamics at endogenous genes in living *Drosophila* tissues. *Methods* 45:233–41
97. Yu J, Xiao J, Ren X, Lao K, Xie XS. 2006. Probing gene expression in live cells, one protein molecule at a time. *Science* 311:1600–3
98. Zimyanin VL, Belaya K, Pecreaux J, Gilchrist MJ, Clark A, et al. 2008. In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization. *Cell* 134:843–53
99. Zipfel WR, Williams RM, Webb WW. 2003. Nonlinear magic: multiphoton microscopy in the biosciences. *Nat. Biotechnol.* 21:1369–77

95. This work first describes the method of MPM imaging *Drosophila* polytene chromosomes to study transcription factor kinetics at endogenous gene loci.

Intragenic mapping of dense chromatin in THO mutants.

The product of transcription of class II genes is an mRNA. This mRNA is however not released on its own. It is incorporated into a messenger ribonucleoprotein (mRNP). The biogenesis of mRNP requires not only transcription, but the proper completion of many co-transcriptional processes, from the methylguanosine capping of the 5’ end to the polyadenylation of the 3’ end of the mRNA. THO is a multi subunit complex with an essential function in the formation of mRNPs (Libri et al. 2002; Rougemaille et al. 2007; Strässer et al. 2002). Although the mechanisms of action of THO are not at this time fully understood, it has been demonstrated that when THO is deficient, mRNA are retained at the site of their transcription and therefore very inefficiently exported.

This accumulation of mRNA is concurrent with an accumulation of polyadenylation factors, suggesting that THO defects do not influence termination of mRNA, but their release and transport from the transcription site. The absence of lethality of THO mutants also suggest that mRNA are terminated correctly. It is possible to identify genes targeted by THO by a cross-linking process: the accumulation of polyadenylation factors and nuclear pore complex (NPC) weights down the chromatin of these genes in their 3' region. A simple differential fractionation (DCF) is enough to isolate this heavy chromatin tethered to these

factors in the absence of functional THO. In this study, this methodology was used to offer a subgene resolution of heavy chromatin formation. The roles of the promoter and terminator regions of the genes was investigated. The specific case of HSP104, a heat shock responsive gene, was also studied using FISH and DCF. We have found during this study that although heavy chromatin formation is placed according to the termination site, its presence or absence is actually dictated by the promoter. More precisely, it seems that the promoter “strength” is involved. The formation of dense chromatin in THO mutants is in fact linked to the transcriptional activity on the said gene; probably due to the fact that because mRNA release and export is deficient, there are more terminating mRNA accumulating at the 3' end of the gene, each of these bringing an extra load of terminating and polyadenylating factors, thus the denser chromatin.

The use of FISH in this study enabled us to establish the transcriptional output of the HSP104 gene, thus demonstrating the link between transcriptional output and heavy chromatin formation. FISH was also used in our case to show the absence of mRNA retention when HSP104 was truncated. We could establish using our FISH technique that initiation of transcription on the HSP104 gene occurred every 4 seconds on average, using an elongation speed of 4kb/minute described in the litterature (Darzacq et al. 2007; Zenklusen et al. 2008). At the moment, this article is in an early stage of writing.

Chapter 4 Methodology Improvements

Live Cell Imaging.

The discovery of the green fluorescent protein (GFP) produced by *Aequoria victoria* is perhaps the discovery which has had the biggest impact on the development of fluorescence microscopy. The sequencing of the gene encoding GFP allowed the use of this protein to tag almost any other protein in a large number of species, since GFP itself does not require much post translational modification apart from a self mediated maturation and has no apparent effect on the function of many proteins. Random mutations of the GFP gene and active searching of other naturally occurring fluorescent proteins have yielded a large palette of fluorescent proteins with various absorption and emission wavelengths, high brightness and stability.

This gives microscopist the opportunity to perform more robust studies in which multiple proteins of interest can be followed in time. However, imaging living cells regularly for long periods of time requires more than that. In laboratories, cells are grown in very specific conditions of temperature, pH, osmotic pressure and oxygen levels. A variation of one of these parameters in the

environment of a cell results in stress with which the cell has to cope. This in turn may impact the phenomenon the experimentator is studying. It is therefore important to maintain cells in a very stable environment as close to the environment they are used to, all the while being imaged. We have customized our high magnification and high numerical aperture microscopes so that the temperature of the sample is maintained at 37°C. We have also added a sealed chamber system in which CO2 levels are kept at 5%. The microscope room temperature was kept at 20°C, in order to keep the cameras and illumination systems at their optimum. This however required the combined use of a perfect focus system (PFS) from Nikon and a focus retrieval program, which was necessary when imaging different stage positions with long intervals between image acquisitions.

The program we have created consists of a 3 step algorithm. The first step is verifying if the PFS can recover the focal plane on its own. If the PFS fails to recover the focal plane, the second steps' role is to assist the PFS by scanning a 50 µm range in the Z axis until the PFS finds the focal plane. The third step consists in memorizing this new Z axis coordinate, which will be used as reference during the next round of imaging at this stage position.

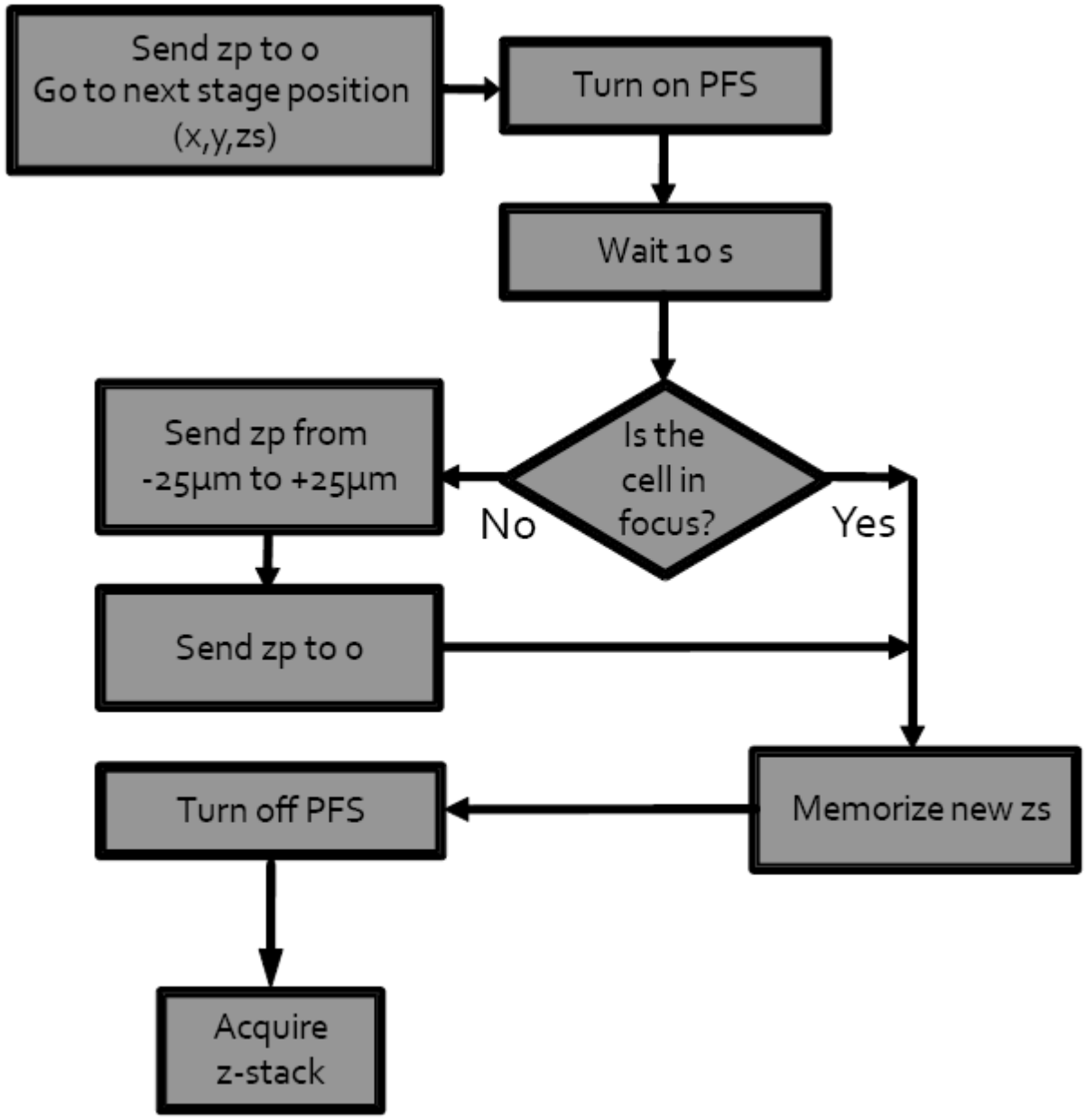


Figure S1: Algorithm of the focus retrieving macro. There are two z axis coordinates: zs is the absolute coordinate of the stage in the z direction, controlled by the stage servos. zp is a z direction coordinate relative to the zs coordinate, controlled by a piezzo-electric device. The total z coordinate is therefore z=zs+zp. The PFS controls the zs coordinate only. If the PFS detects the proper focal plane, as zp changes, zs will change to keep zs+zp constant. The z-stack is acquired using the faster and more precise piezzo-electric device.

Imaging in itself is also a huge source of stress for the cells, as the cell will be exposed to intense illumination. The bleaching of fluorophores obviously reduces the signal to noise ratio, but also produces free radicals which may damage the cell through oxydation of nearby cell components. We have adapted our imaging protocols accordingly to limit photobleaching, by using high numerical aperture objectives, with the lowest possible magnification to observe our structures of interest. Using high efficiency charge coupled device (CCD) cameras also permitted us to use less illumination and exposure time. When possible, we favored using longer exposure times with less illumination than low exposure times with high illumination, as I have seen that it has a strong effect on photobleaching and phototoxicity. We also established the “loss of signal” we could expect as we lose focus of a PSF. By producing large z-stacks with a spacing of 100nm between each image, we produced an “optimal” z-stack, from which we constructed different z-stacks, in which the spacing and the phase was modified. We projected the maximum intensities in z and measured the signal of the PSF in question. Using this, we could establish the loss of signal when using suboptimal z-stacks. Thus we could make a knowledgeable decision when it comes to Z-stack spacing according to the PSF size and wavelength. This gave us the possibility to have quantitative data with fast time-resolution and as little light exposure as possible.

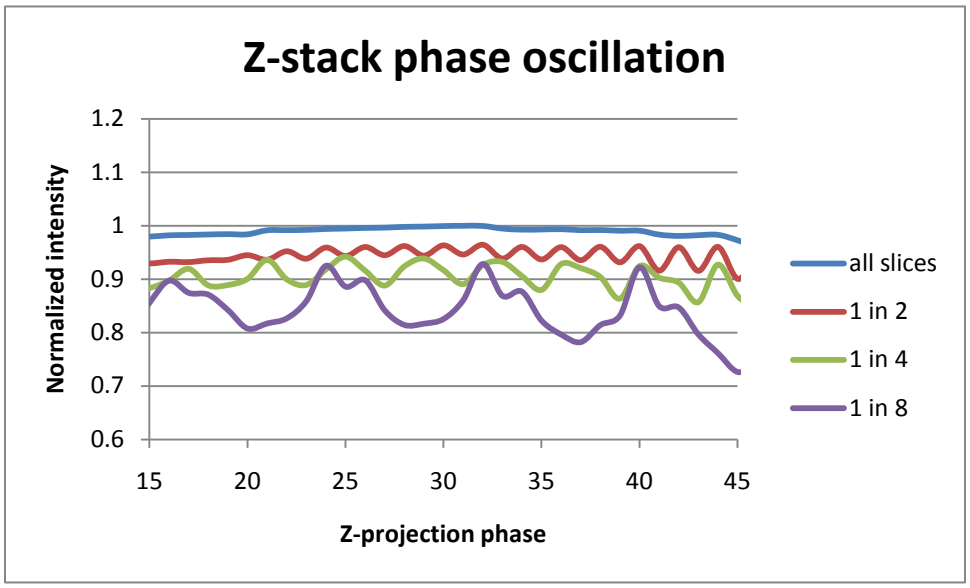


Figure S2: Signal loss from out of phase imaging in different z-stack intervals. The graph indicates the general loss of signal intensity and variance of signal that can be expected from using different stack intervals. A stack of 100 images of a genetic locus tagged with mRFP-LacI in U2OS cells was taken using an interval of 100 nm between two successive images in the z-direction. 51 Substacks of 50 consecutive images were extracted from the original image, each of these are shifted by one image compared to the previous stack to simulate a drift in the z-axis. Maximum intensity projections were then done on the whole substacks, or by using only one in two, one in four or one in eight slices of these substacks (simulating stacks of 200 nm, 400 nm or 800 nm interslice distances respectively). Values were normalized to the maximum value obtained. Signal is lost as the stack drift out of focus. Signal is also lost as interslice distances increase. Furthermore, signal oscillations appear as interslice distance increases, due to dephasing of the stack with the focal plane. The signal obtained is not symmetrical. Drifting the stack the same distance above or below the focal plane will produce different results, due to the specific shape of the PSF and spherical aberrations.

FISH.

Fluorescent In Situ Hybridization is a powerful technique that allows the visualization of specific nucleic acid sequences. It's main advantage is the fact that it is a non-invasive technique, however, its use is limited to fixed cells, and therefore no direct dynamic information can be extracted from the samples. In recent years, FISH has evolved from a nucleic acid localization technique to become a powerful semi-quantitative tool for the counting of single messenger RNAs in single yeast cells (Zenklusen et al. 2008). By comparison of signal from nascent transcripts and full length mRNAs, the authors were also able to estimate the number of mRNAs being transcribed in each cell at the time the sample was fixed.

Transposition of these feats to metazoans has proven to be more difficult. During hybridization, fluorescent probes may randomly bind to non-specific targets, and not be washed away during post-hybridization washes. This results in an increase of a speckled background, which increases the number of false positives. Because mammalian cells have a much larger volume than yeast cells, washes have a reduced effect, and the background is therefore very high. This speckled background is therefore much more present in the larger mammalian cells than in the smaller yeast cells. Increasing the number of probes targeting the same

RNA is a way to overcome this issue, as the hybridization of multiple probes, usually 5, on a same RNA will provide sufficient signal to noise to selectively pick out the higher intensity spots corresponding to RNA (Zenklusen et al. 2008; Gandhi et al. 2011).

However, this requires the generation of multiple 50mer probes of high specificity, which is not always possible, especially when studying short RNA molecules. I have used a high stringency FISH protocol using single multiply labeled probes to increase our signal to noise and reduce the number of false positives. I have calculated the melting temperature of our probes using the Casey and Davidson formula (Casey & N. Davidson 1977). I have chosen to adapt our stringency and hybridization temperature so that it is equivalent to the melting temperature of a DNA/RNA hybrid with 20 homologous bases in the same conditions. By doing so, we create conditions in which our probes may bind to RNA molecules with more than 30 homologous bases to our probe have little chances of being washed out. RNA molecules with 50 bases of homology would bind sufficiently strongly to our probe to avoid being washed out at all. Since 25 base stretches of our probes already exhibit sufficient specificity, this level of stringency should be enough to avoid any binding of our probes to non specific RNA targets. There remains the possibility of our probes simply being stuck in our sample, which we reduced by using very low quantities of probes and large

volumes of washing buffer. This reduction of non-specific binding allows us to target RNA molecules with a single probe.

However, we still need to detect that probe. The use of high numerical aperture objectives with highly efficient CCD cameras in this case is usually sufficient to detect the signal from 5 cyanine fluorophores bound to each of our probes. However, cyanines are first generation fluorophores with relatively poor photostability and brightness, and are sensitive to their environment, especially for cyanine 5 which is sensitive to ozone (Fare et al. 2003). Furthermore, cyanines in close proximity to each other tend to aggregate into hydrophobic structures, causing increased FRET and self quenching (Berlier et al. 2003; Randolph & Waggoner 1997). I therefore preferred the labeling of our probes with the more resilient Alexa fluorophores, which expresses less self-quenching at high degrees of labeling, and are much more photostable and brighter (Berlier et al. 2003). All in all, the combination of these strategies has enabled us to do efficient quantitative analysis on single molecule RNA FISH.

Hyperresolution work in progress.

In fluorescence microscopy, the maximal resolution of an image depends on

the numerical aperture of the objectives and the emission wavelength of the fluorescent molecules. The Rayleigh limit of lateral resolution formula states the distance at which the Rayleigh criterion is reached, *i.e.* the maximal resolution of a microscope: ——, where d is the lateral resolution in nanometers, NA the numerical aperture and λ the wavelength in nanometers. Thus, when two adjacent fluorescent molecules are in close proximity (inferior to d), as is usually the case in densely labeled biological samples, there is no way to make out these two fluorescent molecules from one another.

In recent years however, microscopy techniques capable of imaging single fluorescent molecules have made their appearance. Although PALM, STORM and even TIRF are different techniques, they share the same fundamental concept. The idea is in fact to ensure that only a very small proportion of the fluorescent molecules emit light, in order to reduce the density of point spread functions. If that density is low enough, it becomes possible to distinguish the individual point spread functions from independent fluorescent molecules. By successive imaging of different subpopulations of fluorescent molecules, and pinpointing of the point spread functions obtained, we obtain images of higher resolutions than that dictated by the Rayleigh formula. By imaging living cells in this way, we can even achieve single particle tracking without having to label the molecules of interest with

cumbersome nanoparticles. The differences between these three hyperresolution techniques are however quite important, and each offer advantages and disadvantages that a researcher must take into account when planning experiments.

Total Internal Reflection Fluorescence (TIRF) was the first single molecule technique to have been described, by Axelrod and colleagues in the 1980s. It relies on the creation of an evanescent field of excitation light near the interface between the cover slip and the sample. This evanescent field is only 200nm deep, and therefore only a very few fluorophores will be excited at a given time. The main drawback of TIRF is that studies are limited the extremely thin section closest to the cover slip.

Photoactivation Light Microscopy (PALM), on the other hand, gives the possibility to study any part of a sample that a regular wide field microscope may image. It relies on the use of photoconvertible fluorescent proteins, *i.e.* fluorescent proteins whose excitation and emission wavelengths can be irreversibly modified when hit by a photon of the proper wavelength (usually 405nm). By using very low photoconverting light intensities, only a few proteins will turn to the photoconverted state where they can be imaged in hyperresolution. When these proteins have bleached, a new burst of photoconverting light will allow the imaging of a new set of proteins (Betzig et al. 2006). At this time, only very few

photoconvertible fluorescent proteins are known and most have overlapping wavelengths. This means that only one type of protein may be studied per experiment. Furthermore, these photoconvertible proteins have very low brightness compared to classical fluorescent proteins, requiring highly sensitive CCD cameras to achieve sufficient signal to noise. The fact that PALM relies on fluorescent proteins also implies that it requires genetic engineering, that only proteins may be marked directly and that the influence of the tag on the protein function must be verified.

Finally, Stochastic Optical Reconstruction Microscopy (STORM) relies on an intrinsic property of organic fluorophores. Organic fluorophores are in fact prone to blinking when they are in an oxygen-free environment and in the presence of reducing agents. In classical microscopy, this is an unwanted side-effect. STORM however takes advantage of this property. In effect, this property can be somewhat controlled using 405nm wavelength illumination. In these conditions, a fluorophore can recover from a bleached state when it receives a photon with a wavelength of 405nm. Like PALM, low pulses of 405nm illuminations will allow the imaging of low densities of fluorophores, and thus permit hyperresolution (Rust et al. 2006). The use of organic fluorophores by STORM is both an advantage and a drawback: organic fluorophores are usually much brighter than fluorescent proteins; they also come in a wide array of wavelengths, allowing STORM imaging

of many different components in a same sample. However, tagging molecules of interest in the sample can be troublesome. Furthermore, the conditions for STORM (an oxygen free, reducing environment) are normally not compatible with living processes. This means that STORM is usually limited to fixed cell techniques such as FISH or immunofluorescence (IF).

PALM and STORM offer the possibility to study transcription in a whole new way. Molecules of interest involved in transcription may be studied independently, and their nuclear localization may be determined with an extremely high resolution. I have created various PALM capable cell lines based on either my 4A cell line or the 2-6-3 cell line that has been extensively used for various live cell studies of gene expression (Darzacq et al. 2007; Janicki et al. 2004; Rafalska-Metcalf et al. 2010). In these cells, the largest endogenous subunit of RNA Polymerase 2, Rpb1, was replaced by an alpha-amanitin resistant mutant, fused to Dendra2, a photoconvertible fluorescent protein. These modified cells served us in the study of the dynamics of RNA Polymerase 2 in the nucleus, whether they are transcribing or not. They also give us a better view of clustered RNA polymerase 2 in nuclear foci. We are in the process of combining information extracted by PALM with STORM immunofluorescence in which we specifically target phosphorylation patterns of the RNA Polymerase 2 CTD in the nucleus.

Chapter 5 New types of data require novel analysis methodologies.

The fundamental advantage of microscopy techniques such as those described above is the fact that each cell or even each molecule may be analyzed separately from the next. Instead of having only access to a “mean value” of a sample (for example the mean number of mRNA or the number of nascent transcripts), we can now have access to these same values for every individual cell of our sample. This allows us to conduct precise studies of the variability between cells. A typical example of this is the study of single molecule sensitivity RNA-FISH data: for each cell, we have access to the number of mRNAs and the fluorescence intensity of the transcription site, which is a direct function of the number of nascent transcripts that were being produced at the time the cell was fixed.

We have developed a resampling method for the study of distributions of a value between two different samples. This method is particularly adapted for the study of population profiles in different factors. For a given factor and the various levels of the said factor, the two samples will have differences for each factor. Our H0 hypothesis is that these differences can be reasonably obtained by chance, *i.e.* our two samples are not different in respect to this factor.

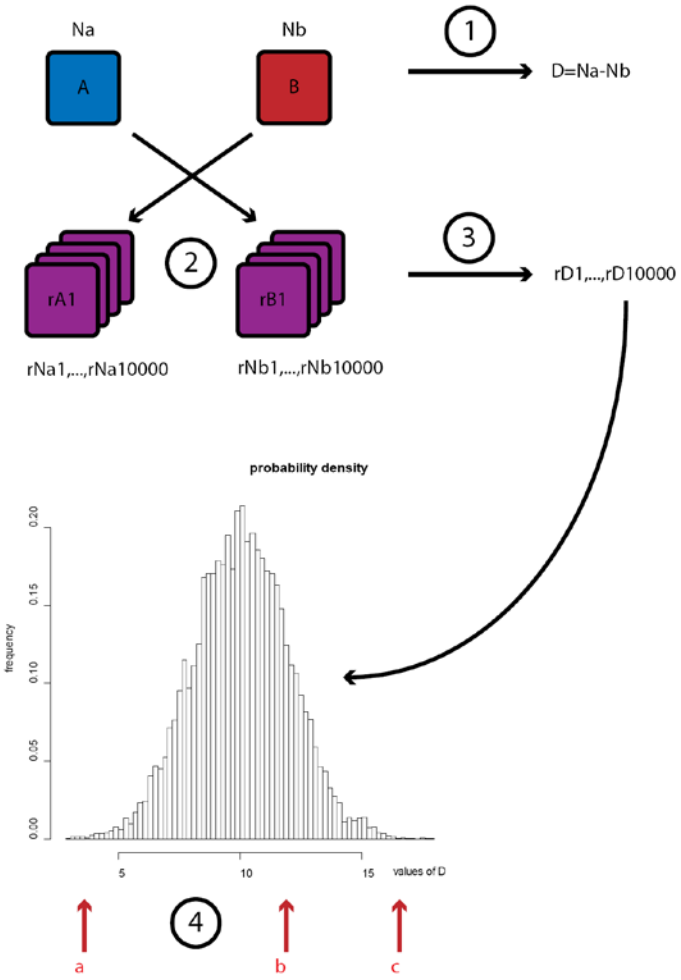


Figure S3: Statistical resampling method developed for transcription site analysis. This method allows the comparison of a quantifiable value "N" (Na, Nb) between two populations. In our studies we chose to use population counts and population profiles as our "N" values. This method will of course also work on calculated values such as mean or median values, and could be of use if parametric tests are inadapted. The method is not sensitive to differences in population sizes. A and B may therefore be of different sizes. 1) The difference in "N" between the two populations is called "D" (1). The individuals of both populations are pooled, and redistributed randomly into two populations of the same size as the original populations (2). This provides us with random "N" values (rNa, rNb), and therefore a random "D" value (rD). This resampling procedure is repeated a large number of times (10 000 in this demonstration) (3). The frequency of the random "D" is plotted. Because the resampling is random, this is equivalent to a probability density. 4) According to the value we originally observed for D, we may decide if D is significative or not. If D is smaller than 95% of the random D obtained by resampling (a), D is significantly smaller than random. If D is neither above nor below 95% of the random D (b), D is not significative. Finally if D is above the 95th percentile of the random D (c), then D is significantly higher than expected by random.

My method consists in pooling the values of the individuals of both samples, and then redistributing these values randomly into two populations of the same sizes than the original samples. This gives us random differences for each factor. We generate in this way a large number of random differences, and plot the probability density of differences obtained by chance, but inferred by our original samples. If our observed difference is more extreme than 95% of the differences obtained by chance, then we may reject the H0 hypothesis.

This method allows us not only to see if two populations are different on the whole, but also to point out precisely which parts of the populations are different from one another. This is a more sensitive and precise method than the usual method for analyzing profiles, the Chi2 test. This is in part due to the fact that Chi2 methods are affected by very low or very unbalanced population counts. With a resampling method, the contingency table cases do not require to have a minimum population. Only an assay based on Fisher's exact test would be more precise, but is impossible to perform with large population of individuals. The drawbacks of our method compared to Chi2 tests are that they need to be adapted for each statistical question, and they also require much more calculations and therefore time.

This method has allowed us to establish differences in the transcription of the TEF4 gene in yeast, between a wild type strain and strains with the same genetic

background, but with the protein XRN1 either deleted or mutated into an exonuclease-deficient protein. These three strains have a constant level of TEF4 mRNA, confirmed both by Northern blot and by RNA-FISH. However, the XRN1 deleted and mutated strains have a strongly impaired mRNA decay rate. We therefore checked if the mutation or deletion impacted the transcriptional levels of the TEF4 mRNA. Using our resampling method, we could indeed verify not only that strains in which XRN1 was mutated or deleted had a reduced transcription of the TEF4 gene. But we could also see differences in the transcription patterns in these two strains. We could conclude that strains where XRN1 was mutated had a significantly higher proportion of cells with no active transcription site, and those cells which did have active transcription sites had very few nascent transcript levels compared to the wild type strain. On the other hand, the XRN1 deleted strain had the same proportion of cells with an active transcription site than the wild type strain, but these transcribing cells and had levels of nascent transcripts equivalent to the XRN1 mutant strains.

The combination of advanced microscopy techniques with newer methods of analysis could therefore detect differences between strains where a sample-wise technique would have only seen that the mutant and deleted XRN1 strains had a reduced transcriptional activity compared to the wild type strain.

A variety of descriptive multifactorial analysis exist, which produces very intuitive geometrical representations of populations and individuals described by a large number of factors. These statistical methods are very well adapted to large data sets such as those produced by microscopy where individuals (single cells, or single molecules) can be described by a large number of descriptors. Correlations can be easily established using such methods, and these may serve as guides for the researcher to narrow his investigation, make decisional statistic tests and hopefully help him turn the statistical correlation into a verifiable cause and effect hypothesis.

One such method, the Principal Component Analysis (PCA) was conducted on images of single molecule RNA FISH experiments. We extracted from these images a series of descriptors of signals we referred to as “spots”. These spots are potential candidates for single molecules of RNA tagged by 5 probes, each of these labeled by 5 organic fluorophores. However, these spots may also correspond to single probes or accumulations of a few probes that are not bound to their target and that have not been washed away in the FISH process. The aim of our analysis is to characterize these spots. By using a second set of probes aimed at the same RNA, labeled with a different dye, we would identify which of these spots were actual mRNAs. From there we would extract using multifactorial analysis the characteristics of “true positives”, *i.e.* spots which correspond to mRNA. This would greatly improve the precision of our analysis, by reducing the number of

false positives and false negatives. Characterizing true positive spots will have another important benefit when it comes to the study of RNA accumulations. Using a reconstruction algorithm, and the “characteristic spot” as a basic building block, we could have a precise count and positioning of each of the mRNAs in an accumulation, such as transcription sites, perhaps even translation sites or p-bodies.

Chapter 6 Materials and methods.

Cell Culture.

U2OS cells and their derived cell-lines were cultured in Dulbecco’s Modified Eagle Medium (Gibco 11880) without phenol red, supplemented with 10% v/v tetracycline free foetal calf serum (Clontech 631106), 100 U/mL penicillin, 100µg/mL streptomycin (Gibco 15140) and 2 mMGlutamax (Gibco 35050-038). Cells were incubated at 37°C in a humidified chamber, with CO2 regulated at 5%.

Transfection.

Transient transfection were made using Fugene 6 (Roche 11 814 443 001),

according to supplier’s recommendations.

Alternatively, for stable cell line generation, the DNA molecule to insert in solution in 2X HeBSCl pH7 was precipitated by addition of calcium chloride up to 133mM under strong agitation. The mix was left to incubate at room temperature for 20 minutes. Cells were washed in 2X HeBSCl, and then incubated for 20 minutes at room temperature in the transfection mix. Regular culture media was then added to the cells, which were returned to their incubator.

Generation of derived cell-lines.

Cells were grown to 80% confluence in 6cm dishes, and transfected using calcium precipitate transfection. The following day, cells were passed in three 10cm dishes in different dilutions, and the selection antibiotic was added. Isolated colonies were picked using cloning cylinders.

Cell line	**Original cell line**	**Transfected DNA**	**Selection antibiotic**
4A	U2OS	45% 256-LacO 45% Reporter gene 10% Puromycin resistance plasmid	Puromycin
4AYP	4A	YFP-Rpb1 Amr	Alpha amanitin
4ADP	4A	Dendra2-Rpb1 Amr	Alpha amanitin
4ATOD	4A	50% Dendra2-Rpb1 Amr 50% pTet-on	Alpha amanitin Geneticin
4ATOMY	4A	50% pTet-on 50% MS2-YFP	Geneticin
4ATY	4A	50% pTet-on 50% YFP-Rpb1 Amr	Geneticin Alpha amanitin

Table 1 : Generated cell lines.

Luciferase.

Cells were lysed in a Tris-H3PO4 pH7.8 buffer containing 10mM MgCl2, 1%v/v Triton, 15%v/v glycerol and 1mM EDTA. DTT was added just prior to lysing, to a final concentration of 1mM. Samples were stored at -20°C. Luciferin reaction buffer consisted of Tris-H3PO4 pH7.8 buffer with 1.25mM ATP and 0.3mM D-Luciferin (Sigma L9504). Measures were taken using a TriStar LB941 luminometer (Berthold Technologies GmbH and Co. KG, Bad Wilbad, Germany).

FISH.

Cells growing on N°1.5 cover slips were fixed using 4% paraformaldehyde in 1x PBS with 2mM VRC (New England Biolabs S 1402S) at room temperature for 10 minutes. Cells were washed in PBS before being stored in 70% ethanol at 4°C.

Cells were then rehydrated in PBS once for 10 minutes, at room temperature. Samples were then equilibrated in a 60% deionized formamide (Carl Roth P040.1), 2X SSPE buffer. Each cover slip of cells was then hybridized overnight at 37°C against 0.1ng of probes in 40 µL of hybridization solution.

Component	Concentration
SSPE	2X
Deionized formamide	60 % v/v
Bovine Serum Albumin	10 µg/mL
Probe	0.1ng/cover slip/probe
tRNA	25µg/mL
ssDNA	25µg/mL
VRC	2mM

Table 2: Hybridization solution composition.

After hybridization, cover slips were washed 2 times in 60% formamide, 2X SSPE for 15 minutes each time, at 37°C. They were then equilibrated in PBS for one hour, stained with DAPI and mounted in PBS with 90% glycerol and 1mg/mL paraphenylenediamine (Sigma Aldrich 78429).

Chemical coupling of FISH probes.

Oligonucleotide probes were from Eurogentec. They were designed to have 50 bases of homology to the target RNA sequence, with approximately 50% GC content. During synthesis, we planned that 5 approximately equidistant dT were modified so that they carried an amino-allyl moiety. The oligonucleotide probes

were coupled to succinimidyl esters of fluorophores. Cyanine fluorophores were from GE Healthcare (PA23001 (Cy3), PA25001 (Cy5)) and Alexa fluorophores were from Invitrogen (A20009 (Alexa 555), A20006 (Alexa 647), A20000 (Alexa 488)). The coupling was carried out in a 100mM sodium carbonate buffer at pH 9.3, overnight at room temperature. Coupled probes were separated from uncoupled fluorophores by size exclusion chromatography using a Sephadex G50 column (GE Healthcare 17-0573-02), in a 5mM triethylamine buffer (Fluka 90335). Probes were finally vacuum dried and resuspended in water.

Name	**target**	**Sequence**
5'CMV	7-56	**t**CCGGTGTCT**T**CTATGGAGG**T**CAAAACAGCG**T**GGATGGCGTC**T**CCAGGCGA
β-g 5'	120-170	**t**ACCATGGTG**T**CTGTTTGAGG**T**TGCTAGTGAACACAG**T**TGTGTCAGAAGCA**t**a
Luc 1	348-398	GCGG**T**TCCATCCTC**T**AGAGGATAGAA**T**GGCGCCGGGCC**T**TTCTTTATG**T**T
Luc 2	400-450	TGT**T**CCAGGAACCAGGGCG**T**ATCTCTTCA**T**AGCCTTATGCAG**T**TGCTCTC**t**a
Luc 3	473-523	**t**CCAACCGAACGGACA**T**TTCGAAGTA**T**TCCGCGTACG**T**GATGTTCACC**T**CG
Luc 4	809-859	**t**AACCGGGAGG**T**AGATGAGATG**T**GACGAACGTGTACA**T**CGACTGAAA**T**CCC
Luc 5	963-1013	**t**AAAATAGGA**T**CTCTGGCA**T**GCGAGAATC**T**GACGCAGGCAGTTCTA**T**GCGG
MS2-NBX	2809-3958	C**T**AGGCAAT**T**AGGTACCT**T**AGGATCTAA**T**GAACCCGGGAATAC**T**GCAGAC

Table 3: FISH probes. Underscored and bold T indicate amino-allyl modified dT, chemically coupled to fluorophores. Capital letters indicate nucleotides homologous to the target sequence. The region targeted is relative to the transcription start site. The MS2-NBX probe has 11 binding sites in the target area.

Immunofluorescence.

Cells were fixed using 4% Paraformaldehyde in 1x PBS, for 10 minutes at room temperature. They were then washed 3 times in 1x PBS for 5 minutes per wash. Cells were permeabilized by a 10 minute treatment in 1x PBS containing 0.5% (v/v) Triton X100. Cells were then washed 3 times in 1x PBS for 5 minutes per wash. Cells were then blocked with 1x PBS containing 3% BSA (w/v) and 100mM glycine. Incubation with the primary antibody was carried out in 0.1% Tween (v/v), 0.3% BSA (w/v) in 1x PBS at 4°C overnight.

Samples were then washed with 0.1% Tween in 1x PBS for 10 minutes, then twice in 1x PBS for 10 minutes per wash. Incubation with the secondary antibody was then carried out in 0.1% Tween, 0.3% BSA in PBS, for 2 hours at room temperature.

Samples were then washed once in 1x PBS, 0.1 % Tween, then twice in 1x PBS, each wash lasting 10 minutes. When combined with FISH, samples followed the FISH procedure first. Once the final 1 hour PBS wash was done, they went through the immunofluorescence procedure, starting at the 30 minute 3%BSA in

PSB blocking step.

Fixed cell imaging.

Cells labeled through IF and/or FISH were imaged using a Nikon Ti inverted microscope (Nikon Instruments, Melville NY), using either a 100x, 1.4 numerical aperture objective, or a 60x, 1.4 numerical aperture objective. Illumination was provided by a Nikon Intensilight illuminator (Nikon Instruments, Melville NY). Images were acquired using a HQ2 Coolsnap camera (Photometrics, Tucson AZ).

Live cell imaging.

Live cell imaging was carried out on a Nikon TE2000 microscope (Nikon Instruments, Melville NY), with an 100x, 1.4 numerical aperture objective. Illumination was provided by a DG4 (Sutter instruments, Novato CA). Images were acquired with a 512SC QuantEM camera (Princeton Instruments, Trenton NJ). Cells were kept in a closed humidified environment at 37°C, with the atmospheric CO2 maintained at 5%, using a CO2 controller, a stage and objective heating system from Bioscience Tools (BioScience Tools, San Diego CA).

Chapter 7 Bibliography

Aoyagi, S. et al., 2002. Nucleosome remodeling by the human SWI/SNF complex requires transient global disruption of histone-DNA interactions. *Molecular and Cellular Biology*, 22(11), pp.3653-3662.

Audibert, A., Weil, D. & Dautry, F., 2002. In vivo kinetics of mRNA splicing and transport in mammalian cells. *Molecular and Cellular Biology*, 22(19), pp.6706-6718.

Baron, U., Gossen, M. & Bujard, H., 1997. Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential. *Nucleic Acids Research*, 25(14), pp.2723-2729.

Becker, M. et al., 2002. Dynamic behavior of transcription factors on a natural promoter in living cells. *EMBO Reports*, 3(12), pp.1188-1194.

Bentley, David L, 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Current Opinion in Cell Biology*, 17(3), pp.251-256.

Berlier, J.E. et al., 2003. Quantitative comparison of long-wavelength Alexa Fluor dyes to Cy dyes: fluorescence of the dyes and their bioconjugates. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society*, 51(12), pp.1699-1712.

Betzig, E. et al., 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science (New York, N.Y.)*, 313(5793), pp.1642-1645.

Blau, J. et al., 1996. Three functional classes of transcriptional activation domain. *Molecular and Cellular Biology*, 16(5), pp.2044-2055.

Blazek, D. et al., 2005. Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. *Nucleic Acids Research*, 33(22), pp.7000-7010.

Boeger, H., Griesenbeck, J. & Kornberg, R.D., 2008. Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell*, 133(4), pp.716-726.

Boireau, S. et al., 2007. The transcriptional cycle of HIV-1 in real-time and live cells. *The Journal of Cell Biology*, 179(2), pp.291-304.

Bosisio, D. et al., 2006. A hyper-dynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF-kappaB-dependent gene activity. *The EMBO Journal*, 25(4), pp.798-810.

Brady, J. & Kashanchi, F., 2005. Tat gets the “green” light on transcription initiation. *Retrovirology*, 2, p.69.

Brzozowski, A.M. et al., 1997. Molecular basis of agonism and antagonism in the oestrogen

receptor. *Nature*, 389(6652), pp.753-758.

Casey, J. & Davidson, N., 1977. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucleic Acids Research*, 4(5), pp.1539-1552.

Chen, R., Yang, Zhiyuan & Zhou, Qiang, 2004. Phosphorylated positive transcription elongation factor b (P-TEFb) is tagged for inhibition through association with 7SK snRNA. *The Journal of Biological Chemistry*, 279(6), pp.4153-4160.

Chrivia, J.C. et al., 1993. Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature*, 365(6449), pp.855-859.

Core, L.J. & Lis, John T, 2008. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science (New York, N.Y.)*, 319(5871), pp.1791-1792.

Czermin, B. et al., 2002. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, 111(2), pp.185-196.

Darzacq, X. et al., 2007. In vivo dynamics of RNA polymerase II transcription. *Nature Structural & Molecular Biology*, 14(9), pp.796-806.

Doi, M., Hirayama, J. & Sassone-Corsi, P., 2006. Circadian regulator CLOCK is a histone acetyltransferase. *Cell*, 125(3), pp.497-508.

Dulac, C. et al., 2005. Transcription-dependent association of multiple positive transcription elongation factor units to a HEXIM multimer. *The Journal of Biological Chemistry*, 280(34), pp.30619-30629.

Dynan, W.S. & Tjian, R., 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell*, 35(1), pp.79-87.

Egloff, S. & Murphy, S., 2008. Cracking the RNA polymerase II CTD code. *Trends in Genetics: TIG*, 24(6), pp.280-288.

Egloff, S., Van Herreweghe, E. & Kiss, Tamás, 2006. Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. *Molecular and Cellular Biology*, 26(2), pp.630-642.

Elbi, C. et al., 2004. Molecular chaperones function as steroid receptor nuclear mobility factors. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9), pp.2876-2881.

Elf, J., Li, G.-W. & Xie, X.S., 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science (New York, N.Y.)*, 316(5828), pp.1191-1194.

Fare, T.L. et al., 2003. Effects of atmospheric ozone on microarray data quality. *Analytical Chemistry*, 75(17), pp.4672-4675.

Farla, P. et al., 2004. The androgen receptor ligand-binding domain stabilizes DNA binding in living cells. *Journal of Structural Biology*, 147(1), pp.50-61.

Ferreira, M.E. et al., 2005. Mechanism of transcription factor recruitment by acidic activators. *The Journal of Biological Chemistry*, 280(23), pp.21779-21784.

Fletcher, T.M. et al., 2002. ATP-dependent mobilization of the glucocorticoid receptor during chromatin remodeling. *Molecular and Cellular Biology*, 22(10), pp.3255-3263.

Freeman, B.C. & Yamamoto, K.R., 2002. Disassembly of transcriptional regulatory complexes by molecular chaperones. *Science (New York, N.Y.)*, 296(5576), pp.2232-2235.

Fu, T.J. et al., 1999. Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. *The Journal of Biological Chemistry*, 274(49), pp.34527-34530.

Fujinaga, K. et al., 2004. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Molecular and Cellular Biology*, 24(2), pp.787-795.

Gallo, G.J., Schuetz, T.J. & Kingston, R E, 1991. Regulation of heat shock factor in Schizosaccharomyces pombe more closely resembles regulation in mammals than in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 11(1), pp.281-288.

Gandhi, S.J. et al., 2011. Transcription of functionally related constitutive genes is not coordinated. *Nature Structural & Molecular Biology*, 18(1), pp.27-34.

Gerber, H.P. et al., 1994. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science (New York, N.Y.)*, 263(5148), pp.808-811.

Glover, J.N. & Harrison, S.C., 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature*, 373(6511), pp.257-261.

Goler-Baron, V. et al., 2008. Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes & Development*, 22(15), pp.2022-2027.

Gorman, J. & Greene, E.C., 2008. Visualizing one-dimensional diffusion of proteins along DNA. *Nature Structural & Molecular Biology*, 15(8), pp.768-774.

Gorski, S.A., Dundr, M. & Misteli, Tom, 2006. The road much traveled: trafficking in the cell nucleus. *Current Opinion in Cell Biology*, 18(3), pp.284-290.

Gossen, M. & Bujard, H., 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12), pp.5547-5551.

Gowers, D.M., Wilson, G.G. & Halford, S.E., 2005. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44), pp.15883-15888.

Greenblatt, J., 1997. RNA polymerase II holoenzyme and transcriptional regulation. *Current Opinion in Cell Biology*, 9(3), pp.310-319.

Halford, S.E. & Marko, J.F., 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Research*, 32(10), pp.3040-3052.

Hayden, M.S. & Ghosh, S., 2004. Signaling to NF-kappaB. *Genes & Development*, 18(18), pp.2195-2224.

He, F. et al., 2003. Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Molecular Cell*, 12(6), pp.1439-1452.

Hell, S.W. & Wichmann, J., 1994. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19(11), pp.780-782.

Hinow, P. et al., 2006. The DNA binding activity of p53 displays reaction-diffusion kinetics. *Biophysical Journal*, 91(1), pp.330-342.

Hoffmann, A, Natoli, G & Ghosh, G., 2006. Transcriptional regulation via the NF-kappaB signaling module. *Oncogene*, 25(51), pp.6706-6716.

Hoffmann, Alexander & Baltimore, D., 2006. Circuitry of nuclear factor kappaB signaling. *Immunological Reviews*, 210, pp.171-186.

Holmberg, C.I. et al., 2001. Phosphorylation of serine 230 promotes inducible transcriptional activity of heat shock factor 1. *The EMBO Journal*, 20(14), pp.3800-3810.

Hoogstraten, D. et al., 2002. Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. *Molecular Cell*, 10(5), pp.1163-1174.

Isel, C. & Karn, J., 1999. Direct evidence that HIV-1 Tat stimulates RNA polymerase II carboxyl-terminal domain hyperphosphorylation during transcriptional elongation. *Journal of Molecular Biology*, 290(5), pp.929-941.

Ivanov, D. et al., 2000. Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Molecular and Cellular Biology*, 20(9), pp.2970-2983.

Janicki, S.M. et al., 2004. From silencing to gene expression: real-time analysis in single cells. *Cell*, 116(5), pp.683-698.

John, S. et al., 2009. Kinetic complexity of the global response to glucocorticoid receptor action. *Endocrinology*, 150(4), pp.1766-1774.

John, S. et al., 2008. Interaction of the glucocorticoid receptor with the chromatin landscape. *Molecular Cell*, 29(5), pp.611-624.

Johnson, T.A. et al., 2008. Chromatin remodeling complexes interact dynamically with a glucocorticoid receptor-regulated promoter. *Molecular Biology of the Cell*, 19(8), pp.3308-3322.

Jones, N., 1990. Transcriptional regulation by dimerization: two sides to an incestuous relationship. *Cell*, 61(1), pp.9-11.

Juven-Gershon, T. et al., 2008. The RNA polymerase II core promoter - the gateway to transcription. *Current Opinion in Cell Biology*, 20(3), pp.253-259.

Kang, Z. et al., 2002. Involvement of proteasome in the dynamic assembly of the androgen receptor transcription complex. *The Journal of Biological Chemistry*, 277(50), pp.48366-48371.

Kaplan, C.D. et al., 2000. Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. *Genes & Development*, 14(20), pp.2623-2634.

Karpova, T.S. et al., 2008. Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science (New York, N.Y.)*, 319(5862), pp.466-469.

Kim, D.-K. et al., 2003. Structure-function analysis of human Spt4: evidence that hSpt4 and hSpt5 exert their roles in transcriptional elongation as parts of the DSIF complex. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms*, 8(4), pp.371-378.

Kim, E. et al., 1997. Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. *The Journal of Cell Biology*, 136(1), pp.19-28.

Kim, J.B. & Sharp, P.A., 2001. Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *The Journal of Biological Chemistry*, 276(15), pp.12317-12323.

Kimura, H., Sugaya, K. & Cook, P.R., 2002. The transcription cycle of RNA polymerase II in living cells. *The Journal of Cell Biology*, 159(5), pp.777-782.

King, I.F.G. & Kingston, Robert E., 2001. Molecular biology: Specifying transcription. *Nature*, 414(6866), pp.858-861.

Klemm, J.D. et al., 1994. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell*, 77(1), pp.21-32.

Kornberg, R.D., 2007. The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 104(32), pp.12955-12961.

Kornblihtt, A.R. et al., 2004. Multiple links between transcription and splicing. *RNA (New York, N.Y.)*, 10(10), pp.1489-1498.

Kugel, J.F. & Goodrich, J.A., 2006. Beating the heat: A translation factor and an RNA mobilize the heat shock transcription factor HSF1. *Molecular Cell*, 22(2), pp.153-154.

Kuznetsova, A.V. et al., 2003. von Hippel-Lindau protein binds hyperphosphorylated large subunit of RNA polymerase II through a proline hydroxylation motif and targets it for ubiquitination. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5), pp.2706-2711.

Lamb, P. & McKnight, S.L., 1991. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends in Biochemical Sciences*, 16(11), pp.417-422.

Lania, L, Majello, B. & De Luca, P., 1997. Transcriptional regulation by the Sp family proteins. *The International Journal of Biochemistry & Cell Biology*, 29(12), pp.1313-1323.

Lee, M.S. et al., 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science (New York, N.Y.)*, 245(4918), pp.635-637.

Li, Q. et al., 2005. Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *The Journal of Biological Chemistry*, 280(31), pp.28819-28826.

Libri, D. et al., 2002. Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Molecular and Cellular Biology*, 22(23), pp.8254-8266.

Lipford, J.R. et al., 2005. A putative stimulatory role for activator turnover in gene expression. *Nature*, 438(7064), pp.113-116.

Liu, Y. et al., 2006. Thyroid hormone-regulated target genes have distinct patterns of coactivator recruitment and histone acetylation. *Molecular Endocrinology (Baltimore, Md.)*, 20(3), pp.483-490.

Lotan, R. et al., 2007. The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *The Journal of Cell Biology*, 178(7), pp.1133-1143.

de la Mata, M. et al., 2003. A slow RNA polymerase II affects alternative splicing in vivo. *Molecular Cell*, 12(2), pp.525-532.

Mayo, L.D. & Donner, D.B., 2002. The PTEN, Mdm2, p53 tumor suppressor-oncoprotein network. *Trends in Biochemical Sciences*, 27(9), pp.462-467.

McCracken, S., Fong, N., et al., 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, 385(6614), pp.357-361.

McNally, J G et al., 2000. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science (New York, N.Y.)*, 287(5456), pp.1262-1265.

Mellor, J., 2005. The dynamics of chromatin remodeling at promoters. *Molecular Cell*, 19(2), pp.147-157.

Michels, Annemieke A et al., 2004. Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *The EMBO Journal*, 23(13), pp.2608-2619.

Misteli, T, 2001. Protein dynamics: implications for nuclear architecture and gene expression. *Science (New York, N.Y.)*, 291(5505), pp.843-847.

Misteli, T & Spector, D L, 1999. RNA polymerase II targets pre-mRNA splicing factors to

transcription sites in vivo. *Molecular Cell*, 3(6), pp.697-705.

Mizzen, C.A. et al., 1996. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell*, 87(7), pp.1261-1270.

Morillon, A. et al., 2003. Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. *Cell*, 115(4), pp.425-435.

Mortillaro, M.J. et al., 1996. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proceedings of the National Academy of Sciences of the United States of America*, 93(16), pp.8253-8257.

Muchardt, C. & Yaniv, M., 1999. ATP-dependent chromatin remodelling: SWI/SNF and Co. are on the job. *Journal of Molecular Biology*, 293(2), pp.187-198.

Mueller, F., Wach, P. & McNally, James G, 2008. Evidence for a common mode of transcription factor interaction with chromatin as revealed by improved quantitative fluorescence recovery after photobleaching. *Biophysical Journal*, 94(8), pp.3323-3339.

Muse, G.W. et al., 2007. RNA polymerase is poised for activation across the genome. *Nature Genetics*, 39(12), pp.1507-1511.

Muñoz, M.J., de la Mata, M. & Kornblihtt, A.R., 2010. The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends in Biochemical Sciences*, 35(9), pp.497-504.

Myer, V.E. & Young, R.A., 1998. RNA polymerase II holoenzymes and subcomplexes. *The Journal of Biological Chemistry*, 273(43), pp.27757-27760.

Müller, J. et al., 2002. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*, 111(2), pp.197-208.

Nagaich, A.K. et al., 2004. Rapid periodic binding and displacement of the glucocorticoid receptor during chromatin remodeling. *Molecular Cell*, 14(2), pp.163-174.

Nardelli, J. et al., 1991. Base sequence discrimination by zinc-finger DNA-binding domains. *Nature*, 349(6305), pp.175-178.

Neil, H. et al., 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, 457(7232), pp.1038-1042.

Nguyen, V T et al., 2001. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature*, 414(6861), pp.322-325.

Peng, J, Marshall, N.F. & Price, D H, 1998. Identification of a cyclin subunit required for the function of Drosophila P-TEFb. *The Journal of Biological Chemistry*, 273(22), pp.13855-13860.

Peng, J, Zhu, Y., et al., 1998. Identification of multiple cyclin subunits of human P-TEFb. *Genes*

& Development, 12(5), pp.755-762.

Perlmann, T., Eriksson, P. & Wrange, O., 1990. Quantitative analysis of the glucocorticoid receptor-DNA interaction at the mouse mammary tumor virus glucocorticoid response element. *The Journal of Biological Chemistry*, 265(28), pp.17222-17229.

Phair, R.D. et al., 2004. Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Molecular and Cellular Biology*, 24(14), pp.6393-6402.

Pratt, W.B., 1997. The role of the hsp90-based chaperone system in signal transduction by nuclear receptors and receptors signaling via MAP kinase. *Annual Review of Pharmacology and Toxicology*, 37, pp.297-326.

Prives, C., 1998. Signaling to p53: breaking the MDM2-p53 circuit. *Cell*, 95(1), pp.5-8.

Proudfoot, Nick & Gullerova, M., 2007. Gene silencing CUTs both ways. *Cell*, 131(4), pp.649-651.

Pufall, M.A. et al., 2005. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science (New York, N.Y.)*, 309(5731), pp.142-145.

Qiu, Y. et al., 2006. HDAC1 acetylation is linked to progressive modulation of steroid receptor-induced gene transcription. *Molecular Cell*, 22(5), pp.669-679.

Rafalska-Metcalf, I.U. et al., 2010. Single cell analysis of transcriptional activation dynamics. *PloS One*, 5(4), p.e10272.

Randolph, J.B. & Waggoner, A.S., 1997. Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes. *Nucleic Acids Research*, 25(14), pp.2923-2929.

Rayasam, G.V. et al., 2005. Ligand-specific dynamics of the progesterone receptor in living cells and during chromatin remodeling in vitro. *Molecular and Cellular Biology*, 25(6), pp.2406-2418.

Ringrose, L., Ehret, H. & Paro, R., 2004. Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. *Molecular Cell*, 16(4), pp.641-653.

Rondón, A.G. et al., 2003. Molecular evidence for a positive role of Spt4 in transcription elongation. *The EMBO Journal*, 22(3), pp.612-620.

Rougemaille, M. et al., 2007. Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. *The EMBO Journal*, 26(9), pp.2317-2326.

Rust, M.J., Bates, M. & Zhuang, X., 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10), pp.793-795.

Salghetti, S E et al., 2001. Regulation of transcriptional activation domain function by ubiquitin. *Science (New York, N.Y.)*, 293(5535), pp.1651-1653.

Selitrennik, M. et al., 2006. Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways. *Eukaryotic Cell*, 5(12), pp.2092-2103.

Shamovsky, I. et al., 2006. RNA-mediated response to heat shock in mammalian cells. *Nature*, 440(7083), pp.556-560.

Shang, Y. et al., 2000. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell*, 103(6), pp.843-852.

Sharma, D. & Fondell, J.D., 2002. Ordered recruitment of histone acetyltransferases and the TRAP/Mediator complex to thyroid hormone-responsive promoters in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), pp.7934-7939.

Sharp, Z.D. et al., 2006. Estrogen-receptor-alpha exchange and chromatin dynamics are ligand- and domain-dependent. *Journal of Cell Science*, 119(Pt 19), pp.4101-4116.

Shiama, N., 1997. The p300/CBP family: integrating signals with transcription factors and chromatin. *Trends in Cell Biology*, 7(6), pp.230-236.

Singh, J. & Padgett, R.A., 2009. Rates of in situ transcription and splicing in large human genes. *Nature Structural & Molecular Biology*, 16(11), pp.1128-1133.

Sorger, P.K., Lewis, M.J. & Pelham, H.R., 1987. Heat shock factor is regulated differently in yeast and HeLa cells. *Nature*, 329(6134), pp.81-84.

Spiluttini, B. et al., 2010. Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *Journal of Cell Science*, 123(Pt 12), pp.2085-2093.

Sprague, B.L. et al., 2004. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophysical Journal*, 86(6), pp.3473-3495.

Sprouse, R.O. et al., 2008. Regulation of TATA-binding protein dynamics in living yeast cells. *Proceedings of the National Academy of Sciences of the United States of America*, 105(36), pp.13304-13308.

Stavreva, D.A. et al., 2004. Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Molecular and Cellular Biology*, 24(7), pp.2682-2697.

Stenoien, D.L. et al., 2001. FRAP reveals that mobility of oestrogen receptor-alpha is ligand- and proteasome-dependent. *Nature Cell Biology*, 3(1), pp.15-23.

Strässer, K. et al., 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature*, 417(6886), pp.304-308.

Tago, K. et al., 2004. Regulation of nuclear retention of glucocorticoid receptor by nuclear Hsp90. *Molecular and Cellular Endocrinology*, 213(2), pp.131-138.

Treisman, J. et al., 1992. The homeodomain: a new face for the helix-turn-helix? *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 14(3), pp.145-150.

Triezenberg, S.J., 1995. Structure and function of transcriptional activation domains. *Current Opinion in Genetics & Development*, 5(2), pp.190-196.

Tupler, R., Perini, G. & Green, M.R., 2001. Expressing the human genome. *Nature*, 409(6822), pp.832-833.

Uesugi, M. et al., 1997. Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science (New York, N.Y.)*, 277(5330), pp.1310-1313.

Vincent, M. et al., 1996. The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. *Nucleic Acids Research*, 24(23), pp.4649-4652.

Voss, T.C., John, S. & Hager, Gordon L., 2006. Single-Cell Analysis of Glucocorticoid Receptor Action Reveals that Stochastic Post-Chromatin Association Mechanisms Regulate Ligand-Specific Transcription. *Mol Endocrinol*, 20(11), pp.2641-2655.

Väisänen, S. et al., 2005. Spatio-temporal activation of chromatin on the human CYP24 gene promoter in the presence of 1alpha,25-Dihydroxyvitamin D3. *Journal of Molecular Biology*, 350(1), pp.65-77.

Whetstine, J.R. et al., 2006. Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell*, 125(3), pp.467-481.

Wilkins, R.C. & Lis, J T, 1997. Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. *Nucleic Acids Research*, 25(20), pp.3963-3968.

Willmann, T. & Beato, M., 1986. Steroid-free glucocorticoid receptor binds specifically to mouse mammary tumour virus DNA. *Nature*, 324(6098), pp.688-691.

Winston, F., 2001. Control of eukaryotic transcription elongation. *Genome Biology*, 2(2), p.REVIEWS1006.

Wong, J, Shi, Y.B. & Wolffe, A.P., 1997. Determinants of chromatin disruption and transcriptional regulation instigated by the thyroid hormone receptor: hormone-regulated chromatin disruption is not sufficient for transcriptional activation. *The EMBO Journal*, 16(11), pp.3158-3171.

Wood, M.J., Andrade, E.C. & Storz, G., 2003. The redox domain of the Yap1p transcription factor contains two disulfide bonds. *Biochemistry*, 42(41), pp.11982-11991.

Wood, M.J., Storz, G. & Tjandra, N., 2004. Structural basis for redox regulation of Yap1 transcription factor localization. *Nature*, 430(7002), pp.917-921.

Wu, C. et al., 1979. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell*, 16(4), pp.797-806.

Wurtz, J.M. et al., 1996. A canonical structure for the ligand-binding domain of nuclear receptors. *Nature Structural Biology*, 3(2), p.206.

Wysocka, J. & Herr, Winship, 2003. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends in Biochemical Sciences*, 28(6), pp.294-304.

Xu, Z. et al., 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457(7232), pp.1033-1037.

Yamada, T. et al., 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular Cell*, 21(2), pp.227-237.

Yamaguchi, Y et al., 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, 97(1), pp.41-51.

Yamaguchi, Yuki et al., 2002. Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. *Molecular and Cellular Biology*, 22(9), pp.2918-2927.

Yamamoto, Y. & Gaynor, Richard B, 2004. IkappaB kinases: key regulators of the NF-kappaB pathway. *Trends in Biochemical Sciences*, 29(2), pp.72-79.

Yamane, K. et al., 2006. JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell*, 125(3), pp.483-495.

Yang, Z et al., 2001. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature*, 414(6861), pp.317-322.

Yao, J. et al., 2007. Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation. *Molecular Cell*, 28(6), pp.978-990.

Yao, J. et al., 2006. Dynamics of heat shock factor association with native gene loci in living cells. *Nature*, 442(7106), pp.1050-1053.

Yik, J.H.N. et al., 2003. Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Molecular Cell*, 12(4), pp.971-982.

Zenklusen, D., Larson, D.R. & Singer, R.H., 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature Structural & Molecular Biology*, 15(12), pp.1263-1271.

Zobeck, K.L. et al., 2010. Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Molecular Cell*, 40(6), pp.965-975.

Zou, J. et al., 1998. Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell*, 94(4), pp.471-480.

My name is Ozymandias, king of kings:
Look on my works, ye mighty, and despair!