UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé

IGBMC - CNRS UMR 7104 - Inserm U 964

THÈSE présentée par:
Akos Gyenis

Soutenue le: **19 décembre 2012**

Pour obtenir le grade de : **Docteur de l'Université de Strasbourg**
Discipline/ Spécialité: Aspects moléculaires et cellulaires de la biologie

Etudes génomiques de la dynamique de l'ARN polymérase II pendant l'étape de terminaison de la transcription et après un stress causé par les UV-B

(Genome-wide characterization of RNA polymerase II behavior during transcription termination and upon UV-B stress)

THÈSE dirigée par:
Dr. Laszlo Tora — *Institut de génétique et de biologie moléculaire et cellulaire, Strasbourg*

RAPPORTEURS:
Dr. Olivier Bensaude — *Institut de biologie de l'école normale supérieure, Paris*
Dr. Matthieu Gerard — *Institut de biologie et de technologies de Saclay*
Dr. Evi Soutoglou — *Institut de génétique et de biologie moléculaire et cellulaire, Strasbourg*

Contents

Acknowledgement

First of all I would like to thank Dr Laszlo Tora for accepting me in his lab, and for all the support, advices and guidance for my projects in the past years. The time I spent in his lab was very instructive for me.

I thank the jury members Dr Olivier Bensaude, Dr Matthieu Gerard and Dr Evi Soutoglou for spending time to read my thesis and to evaluate my work.

I am also grateful to Didier Devys for the constant scientific, non-scientific and climbing discussions, critics and help.
I specially thank every present and former lab members: Elisabeth Scheer, Jacques Bonnet, Marjorie Fournier, Krishanpal Karmodya, Eda Suer, David Umlauf, Sarina Ravens, Anne Riss, Nikolaos Vosnakis, Matthieu Stierle, Chen-Yi Wang, Guillaume Lang, Arnaud Krebs, Monika Ballarino, Anne Helmrich, Anamika Krishanpal, and Meritxell Orpinell for their help in my experiments and life in France. I thank Dr Frédéric Coin for providing me equipments for my experiments. I thank Dr Evi Soutoglou and Dr Dirk Eick for antibodies. I specially thank Tao Ye, Stéphanie Le Gras and the sequencing platform for helping me with the ChIP-seq analyses. I thank Betty Heller and the cell culture facility for providing me cells and medium for my experiments. I also thank Gabor Papai, Judit Osz Papai, Orban Komonyi, Zita Nagy, Attila Oravecz, Emese Gazdag, Tibor Pankotai, Gabriella Pankotai-Bodo and Maria Takacs for the help to arrange my life in France. I thank for my family and friends for the constant support.

The work presented herein was carried out in Dr. Laszlo Tora's laboratory in IGBMC Strasbourg.

My studies were supported by the Fondation Pour La Recherche Médicale (FRM).

List of abbreviations

6

6-4PPs pyrimidine 6-4 pyrimidone photoproducts

A

ATAC

ATM Ataxia-telangiectasia-mutated

B

BRE TFIIB-recognition element

C

CAK Cyclin-Activating Kinase complex

CARM1 coactivator-associated arginine methyl-transferase 1

CDC Cell Division Cycle

CDK Cyclin-Dependent Kinase

ChIP Chromatin Immunoprecipitation

ChIP-seq ChIP coupled High Throughput Sequencing

CPDs Cyclobutane pyrimidine dimers

CPSF cleavage and polyadenylation specificity factor

CS Cockayne syndrome

CstF Cleavage stimulation factors

CTD Carboxyl-Terminal Domain

D

DCE Downstream core element

DDB DNA Damage binding protein

DP DNA polymerase

DPE Downstream promoter element

DSB Double strand break

DSIF DRB Sensitivity-Inducing Factor

E

EAG End of annotated gene

G

GGR Global genome repair

GRO-seq Global Run-On assay coupled sequencing

GTF General Transcription Factor

GW Genome-wide

H

H Histone

HR Homologous recombination

I

Inr Initiator

K

kb Kilobase

L

LEC Little elongation complex

M

m Meter

MAZ MYC-associated zinc-finger protein

mb Megabase

MD Mega Dalton

MED Mediator

MLL mixed lineage leukemia

mRNA messenger RNA

MTE Motif ten element

N

NC2 Negative cofactor

NELF Negative elongation factor

NER Nucleotide excision repair

NHEJ Non-homologous end-joining

nm Nanometer

NMR Nuclear Magnetic Resonance

nt Nucleotide

O

O3 Ozone

O-GlcNAc O-linked N-acetylglycosamine

ORF Open reading frame

P

P Phosphorylation

PA Polyadenylated

PAR Promoter associated RNA

PARP1 Poly ADP ribose polymerase 1

PIC Pre Initiation Complex

PIN1 peptidyl–prolyl cis/trans-isomerase

Pol I, II, III RNA polymerase I, II, III

Pol IIa, hypophosphorylated Pol II

Pol IIo hyperphosphorylated Pol II

P-TEFb Positive Transcription Elongation factor complex

R

RNAi RNA interference

ROS Reactive oxygen Species

RPA Replication protein A

RPAP2 RNA Polymerase II-associated protein

rRNA Ribosomal RNA

S

SAGA Spt-Ada-Gcn5 acetyltransferase

SEC Super elongation complex

snoRNAs small nucleolar RNA

snRNP Small nuclear ribonucleoprotein

T

TCR Transcription coupled repair

TF Transcription factor

tiRNA Transcription initiation related RNA

TRF TBP-related factors

tRNA Transfer RNA

TSS Transcription start site

TTS Transcription termination site

U

UAS	Upstream activating sequence	**X**	
UB	Ubiquitin	XCPE1	X core promoter element1
USA	Upstream stimulatory activity	XP	Xeroderma pigmentosum
UV	Ultraviolet		

Résumé de thèse

Analyse sur l'ensemble du génome de la dynamique de l'ARN polymérase II lors de la terminaison de la transcription et en réponse à un stress aux UV-B.

Introduction:

Les cellules vivantes sont en permanence soumises à des signaux internes (processus métaboliques de base) et externes (stress environnementaux). En réponse à ces stimuli les cellules disposent de mécanismes très précisément contrôlés afin d'accéder et de transcrire les informations génétiques contenues dans l'ADN permettant ainsi de maintenir l'état physiologique des cellules. Ceci est réalisé chez les eucaryotes par la machinerie de transcription associée à l'ARN polymérase II qui permet l'expression des gènes codant pour les protéines et pour certains ARNs régulateurs non-codants.

La transcription par l'ARN polymérase II (Pol II) est un mécanisme régulé de facon extrêmement fine nécessitant l'action séquentielle d'un grand nombre de complexes protéiques. Chaque cycle de transcription peut être divisé en trois phases principales : initiation, élongation et terminaison. Chacune de ces phases peut également être divisée en plusieurs étapes correspondant chacune à des possibilités de régulation pour l'expression des gènes et la synthèse d'ARNm.

La transition entre les phases d'initiation et d'élongation de la transcription est une étape limitante au cours de laquelle la Pol II s'arrête (pause) en aval du site d'initiation après la synthèse des 25-50 premiers nucléotides. Cette pause de la polymérase a d'abord été décrite comme une étape limitante entre initiation et élongation précoce lors de la transcription de certains gènes de choc thermique chez la drosophile. Plus récemment des études sur l'ensemble du génome ont montré que ce phénomène de pause de la Pol II à proximité du promoteur constitue un mécanisme de régulation qui est conservé et observé sur un très grand nombre de promoteurs chez les eucaryotes supérieurs. Une pause ou un ralentissement de l'ARN Pol II a également été découvert en aval de la fin des gènes annotés (end of annotated genes : EAGs) mais cette observation n'a jamais été analysée de façon

globale par des études sur l'ensemble du génome qui se sont concentrées sur la caractérisation des pauses de la Pol II à proximité du promoteur. Ainsi, une caractérisation détaillée des gènes potentiellement régulés par une pause de la polymérase en aval des EAGs n'a pas été réalisée sur l'ensemble du génome. Il reste donc à démontrer si ce phénomène est mécanisme général observé sur l'ensemble des gènes chez les mammifères.

La terminaison de la transcription correspond à l'arrêt de la synthèse de l'ARN par la Pol II et au relargage de la molécule d'ARNm et de la Pol II de la matrice d'ADN. La terminaison de la transcription est cruciale pour la physiologie de la cellule. En effet, elle empêche une interférence entre la Pol II et des éléments fonctionnels de séquence d'ADN comme les promoteurs évitant ainsi que la Pol II n'initie la transcription d'un gène voisin. La terminaison de la transcription est également liée au clivage et à la polyadénylation du transcrit natif d'ARN. Il a été montré que le ralentissement de la Pol II au site de terminaison permet au domaine C-terminal (C-terminal domain : CTD) de la Pol II de recruter et d'augmenter la concentration locale des facteurs de maturation de l'extrémité 3' et de servir de plateforme pour l'assemblage des complexes de clivage de l'ARNm.

La terminaison de la transcription n'a pas lieu sur des sites conservés ou à une distance constante de l'extrémité 3' des ARNm matures. Chez les mammifères, la terminaison peut se situer à une distance très variable de l'extrémité 3' des ARNm, de quelques paires de bases à plusieurs kb. En effet des analyses par ChIP ont montré sur plusieurs gènes modèles, des densités de Pol II plus élevées en aval des extrémités 3' des gènes (EAG) que sur l'ensemble de la séquence transcrite. Cependant il reste à déterminer ci cet enrichissement correspond à une pause, un arrêt ou un ralentissement de la polymérase. En effet, il a été montré que le signal de polyadénylation mais également des éléments de séquence en aval peuvent ralentir la progression de la Pol II en aval.

Durant ma thèse j'ai réalisé deux projets utilisant des techniques d'immunoprécipitation de la chromatine associées à des méthodes de séquençage à haut débit (ChIP-seq) afin d'analyser la distribution de l'ARN polymérase II dans deux conditions :

D'une part j'ai caractérisé sur l'ensemble du génome les pauses de la Pol II en aval du site EAG des unités de transcription dans les cellules humaines.

D'autre part, j'ai étudié les effets de stress génotoxiques sur la machinerie de transcription associée à la Pol II en analysant les modifications de profils de distribution de la Pol II.

Résultats 1.: Caractérisation des pauses de la Pol II en aval de la fin des unités de transcription

Afin de caractériser les profils de distribution de l'ARN Pol II en aval des EAGs, j'ai réalisé des expériences de ChIP-seq en utilisant un anticorps reconnaissant toutes les formes d'ARN Pol II humaine. J'ai analysé les profils de Pol II en aval de 13787 gènes qui n'ont pas de gène flanquant à +/- 4kb en amont ou en aval. Nos résultats ont été analysés en comparaison avec des données disponibles de séquençage à haut débit d'ARN naissants (Global Run On assay coupled sequencing : GRO-seq). Nos résultats montrent qu'un enrichissement de la Pol II en aval de l'extrémité des unités de transcription est une caractéristique partagée par tous les gènes exprimés et reflète la présence d'ARN Pol II active. Des analyses bioinformatiques (K-means clustering) m'ont permis de distinguer quatre groupes de gènes : le premier groupe (H) est caractérisé par un profil de pause étroit alors que les trois autres groupes (PA1-PA3) montrent un profil large ou très large, pouvant aller jusqu'à 6kb en aval des EAGs. Des analyses d'annotations (Gene Ontology) révèlent que le groupe H contient pratiquement exclusivement des gènes d'histones qui ne contiennent pas d'intron et dont les transcrits ne sont pas polyadénylés. A l'inverse, les groupes PA1-PA3 contiennent des gènes codant pour des transcrits polyadénylés. J'ai confirmé par des expériences de ChIP couplées à une analyse par qPCR les différents types de profils de distribution de Pol II décrits par analyse bioinformatique. Nos résultats sont en accord avec d'autres publications et suggèrent un lien entre le profil de distribution de la Pol II à l'extrémité 3' des gènes histones et les mécanismes particuliers de maturation de l'extrémité 3' de ces transcrits. Cette idée est renforcée par nos analyses fonctionnelles montrant que l'inhibition des mécanismes de polyadénylation augment la présence de l'ARN Pol II en 3' des EAGs pour les gènes codant pour des transcrits polyadénylés. A l'inverse, cette inhibition ne change pas la distribution de la Pol II sur les gènes d'histones dont les transcrits ne sont pas polyadénylés. L'ensemble de ces résultats suggère un mécanisme qui augmenterait le

temps de résidence de la Pol II en aval de l'extrémité 3' des gènes en cas de défaut de polyadénylation.

L'analyse des profils de distribution de l'ARN Pol II en aval des unités de transcription sur l'ensemble du génome, nous ont permis de montrer que le phénomène de pause de la Pol II est un mécanisme général retrouvé pour tous les gènes exprimés et qui parait conservé chez les vertébrés. Sur les gènes histones qui ne contiennent pas d'introns et dont les transcrits ne sont pas polyadénylés, le profil d'enrichissement de la Pol II en aval de l'extrémité 3' est étroit et pourrait correspondre à des mécanismes très particuliers de maturation de l'extrémité 3' de ces ARNm.

Résultats 2.: Analyse de la dynamique de l'ARN polymérase II au cours du stress par les UV-B.

Afin d'étudier les effets de stress génotoxiques sur la dynamique de l'ARN Pol II, j'ai utilisé une approche ChIP-seq en utilisant un anticorps reconnaissant toutes les formes de Pol II sur des chromatines préparées à partir de lignées de cellules humaines à différents temps après traitement par les UV-B.

L'analyse bioinformatique des résultats de distribution de la Pol II sur l'ensemble du génome m'a permis de montrer que des doses sub-létales d'UV-B induisent temporairement un arrêt global de la transcription. J'ai pu monter qu'un traitement par les UV-B entraine une perte progressive et majeure du signal Pol II des promoteurs des gènes exprimés. Cette perte de signal s'étend ensuite sur l'ensemble de l'unité de transcription jusqu'à quatre heures après irradiation. Cependant, la densité du signal Pol II en aval des EAGs n'est que très faiblement diminuée après irradiation. Des analyses par western blot utilisant différents anticorps reconnaissant spécifiquement les différentes formes de phosphorylation du CTD de l'ARN Pol II m'ont permis de montrer que les niveaux de ces différentes formes de Pol II ne sont pas affectées excluant un mécanisme actif de dégradation. Nous avons également identifié un groupe de gènes caractérisés par une augmentation du signal Pol II après irradiation par les UV-B. Ces gènes correspondent à des gènes de réponses aux dommages de l'ADN.

L'ensemble de nos résultats sont en accord avec des observations montrant qu'après irradiation par les UV-B l'arrêt temporaire de la transcription est nécessaire pour la mise en place de la réparation couplée à la transcription (transcription-coupled repair: TCR). Nous avons également observé que, six heures après traitement par les UV-B, les gènes initialement réprimés montrent ensuite une augmentation du recrutement de l'ARN Pol II par rapport au contrôle. Cette observation suggère que la machinerie transcriptionnelle de l'ARN Pol II pourrait compenser la réduction initiale de l'expression de ces gènes lors des étapes de réparation TCR.

Conclusions:

La transcription par l'ARN polymérase II a été étudiée en détail au cours des dernières décennies mais de nombreux mécanismes sont encore mal compris, empêchant ainsi de décrire un modèle décrivant tous les mécanismes de régulation de l'expression des gènes. Les projets que j'ai abordés pendant ma thèse nous ont permis d'élargir nos connaissances à propos de la transcription par la Pol II. Nous avons tout d'abord caractérisé sur l'ensemble du génome l'accumulation de la Pol II atour de l'extrémité 3' des gènes. Nous avons observé qu'une augmentation de la présence de la Pol II en aval des gènes exprimés est une caractéristique générale conservée chez les mammifères. Nous avons également montré que les profils de distribution de la Pol II sur les gènes des histones canoniques sont différents ce qui pourrait correspondre à des mécanismes spécifiques de maturation des extrémités 3' des ARNm correspondants. Nos études fonctionnelles renforcent l'hypothèse d'un lien entre augmentation de la densité de Pol II en aval des unités de transcription et terminaison de la transcription ainsi qu'avec la maturation de l'extrémité 3' des ARNm.

L'étude des effets de stress génotoxiques sur la distribution de l'ARN Pol II sur l'ensemble du génome, nous a permis de montrer que des doses sub-létales d'UV-B altèrent de façon globale la machinerie de transcription. Nos résultats confirment d'autres observations suggérant que les complexes Pol II sont enlevés à la fois des sites de lésion de l'ADN mais également des promoteurs des gènes afin d'inhiber la transcription pendant les étapes de TCR. De plus, des modèles récents suggèrent qu'en présence de lésions persistantes de l'ADN qui seront lentement réparées, des sous-unités de l'ARN Pol II seront poly-ubiquitinées et dégradées pour permettre l'accès des facteurs de réparation sur ces

lésions et afin d'inhiber de nouveaux cycles de transcription. A l'inverse, nos résultats montrent une absence de dégradation de la Pol II après traitement avec des doses sub-létales d'UV-B ce qui pourrait s'expliquer par un mécanisme permettant une dissociation temporaire de la Pol II des promoteurs sans dégradation. Cette observation suggère que différents mécanismes pourraient être mis en jeu pour enlever la Pol II des unités de transcription pendant la réparation des lésions. L'activation de l'un ou l'autre de ces mécanismes d'éviction de la Pol II dépendrait du type de lésion impliquée et de leur persistance dans le temps.

ENGLISH VERSION

Genome-wide characterization of RNA polymerase II behavior during transcription termination and upon UVB stress

Summary of Thesis

Introduction:

Living cells are continuously exposed to stimuli from internal (basic metabolic processes) and external (environmental or chemical stress) sources. Therefore a very accurate and tightly regulated cellular process is needed to access and express the DNA-encoded information in order to maintain the normal physiological state of a cell. In eukaryotes this is carried out by the RNA polymerase II (Pol II) machinery, which is responsible for the expression of thousands of genes coding for proteins and non-coding regulatory RNAs.

RNA Polymerase II based transcription is one of the most highly regulated cellular process, which requires a well orchestrated, sequential action of multiple different protein complexes. Transcription cycle can be divided into three main phases: initiation, elongation and termination, however these phases are also built up from multiple sub-steps, which all represent a checkpoint and a regulatory possibility for proper mRNA synthesis and gene expression.

The transition from transcription initiation into elongation is an inefficient process, where Pol II shows tendency to stop (pause) after transcribing the first 25-50 nucleotides. This polymerase pause was first described as a rate limiting regulatory step between initiation and early elongation at the promoters of certain Drosophila heat shock genes. Recently, genome-wide studies demonstrated that promoter proximal Pol II pausing is a conserved regulatory step that is present at almost every promoters in higher eukaryotes. Surprisingly, paused or slowed down polymerases were also discovered downstream from end of genes (EAGs) however, in mammalian systems, the main focus of genome-wide study with respect to Pol II pausing was only promoter proximal pausing. Therefore the full spectrum of genes regulated by pausing downstream of EAGs has not yet been investigated

at a genome-wide level. Thus, it was not clear whether this phenomenon is commonly occurring among mammalian genes.

Transcription termination occurs when Pol II ceases RNA synthesis and both Pol II and the RNA molecule are released from the DNA template. Termination serves many vital functions in the cell. It prevents Pol II from interfering with downstream DNA elements, such as promoters and thus, Pol II will not enter into an unnecessary transcription of a neighbour gene. In addition, termination is functionally linked to the cleavage and polyadenylation of the nascent RNA transcript. It has been shown that the terminating or slowed down Pol II C-terminal domain (CTD) raises the local concentration of 3' end processing factors near the nascent transcript and also serves as a platform for the assembly of the cleavage complexes.

Termination does not occur at a conserved site or constant distance from the 3' end of the mature mRNAs. In mammals, termination can occur anywhere from a few base pairs to several kb downstream from the 3' end of mRNA. Indeed, ChIP assays showed higher Pol II densities downstream from 3' end of genes (EAG) than throughout the gene body on several model genes, however, it is uncertain, whether it indicates pausing, arrested or slowed down polymerases. In turn, it is known that the polyadenylation signal and certain downstream sequence elements can further negatively influence Pol II progression downstream of genes.

During my Ph.D. I carried out projects using chromatin immunoprecipitation assay coupled to high-throughput sequencing techniques (ChIP-seq) to analyse genome-wide Pol II behaviour in two aspects:

First, we characterized Pol II pausing downstream of the EAG of the transcription units in human cells genome-wide, by using high-resolution occupancy profiling.

Second, we investigated the effect of genotoxic stress on the Pol II transcription machinery by following the global alteration of Pol II occupancy profiles.

Results 1.: Characterization Pol II pausing downstream of the end of transcription units

In order to characterize Pol II occupancy profiles downstream of EAGs, we carried out ChIP-seq technique with an antibody, which recognizes all forms of human Pol II. We analyzed Pol II occupancy profiles downstream of 13787 genes which have no neighboring genes +/- 4kb up and downstream. Our results together with a Global Run on assay coupled sequencing (GRO-seq) data show that Pol II occupancy downstream of 3' end of transcription units is a common feature of expressed genes and reflects transcriptionally active Pol IIs. We subdivided the expressed genes by K-means clustering and distinguished four main clusters, where cluster one (H) showed narrow pause profile, the rest (PA1-PA3) showed broad and very broad sometimes up to 6 kb long Pol II profiles downstream of their EAG. Gene ontology analyses showed that genes in cluster H are almost exclusively core histone genes, which are intronless and code for non-polyadenylated transcripts. The PA1-PA3 clusters represent genes coding for polyadenylated transcripts. We validated the bioinformatically isolated Pol II pause patterns with ChIP followed by qPCR detection. Our results are in good agreement with previous studies and may suggest a link between the Pol II occupancy profiles 3' from histone genes and the different 3' end processing mechanisms of the corresponding transcripts. This idea was corroborated by the functional studies, which showed that inhibition of polyadenylation increased Pol II occupancy 3' of the EAGs of genes with polyadenylated transcripts, while it had no significant effect on Pol II drop at the 3' end of the core histone genes. In addition these results suggest a mechanism, which increases the residency time of Pol II downstream of genes upon defective polyadenylation.

By characterizing Pol II occupancy profiles downstream from transcription units, we found that Pol II pausing is a genome-wide feature of expressed genes and it is conserved in mammals. On histone genes, which are coding for intronless and non polyadenylated transcripts, the Pol II occupancy profile is narrow, which may reflect the different mRNA 3' end processing mechanism.

Results 2.: Investigation of RNA polymerase II behavior upon UVB stress.

To study the effect of genotoxic stress on the Pol II transcription machinery, we carried out ChIP-seq technique with an antibody which recognizes all form of Pol II on human cells in different time points after UVB treatment.

By generating heat maps from the global Pol II occupancy datasets, we found that a sub-lethal dose of UVB can temporary interrupt transcription genome-wide. Following UVB treatment we observed a progressive and massive Pol II signal loss from the promoters of expressed genes, which will then extend through the entire transcription unit, up to four hours after irradiation. Surprisingly, Pol II tag densities downstream from EAGs showed only a slightly decrease upon irradiation. To determine whether Pol II is degraded upon UVB treatment, we carried out western blot analyses with antibodies which recognize different phosphorylated forms of Pol II. Surprisingly we did not detect any decrease in the global level of Pol II. In addition, we found a subset of genes with increased Pol II signal upon irradiation, which were identified as DNA damage response genes. Our results are in good agreement with the observation that after UV irradiation transcription is arrested during the period of transcription-coupled repair (TCR). Interestingly, six hours after UVB treatment, genes that were negatively influenced, showed increased Pol II signal compared to the control state, suggesting that the Pol II transcription apparatus may compensate for the reduced gene expression associated with TCR.

Conclusions:

RNA polymerase II transcription has been extensively studied in the past decades, but lots of details are still missing to establish a model, which covers every regulatory mechanism during gene expression. With my Ph.D. projects, we managed to expand the current knowledge about Pol II transcription machinery. First we characterized the Pol II accumulation around the 3' end of genes that was not yet investigated at a genome wide level. We found that increased Pol II occupancy downstream from expressed genes is a general feature that is conserved in mammalians. We also reported that Pol II profiles at core histone genes are different, which can be due to their unique mRNA 3' end processing mechanism. With our functional study, we further support the link between the increased Pol II presence downstream from transcription units, transcription termination and 3' end processing of mRNA. We also uncovered a possible regulatory mechanism which negatively influences transcription initiation upon defective polyadenylation.

By studying the effect of genotoxic stress on Pol II mediated transcription genome-wide, we show that a sublethal dose of UVB can disturb the global transcription machinery.

Our results are in good agreement with studies suggesting that Pol II complexes are removed not only from the DNA lesion site, but also from promoters to prevent further transcription initiation during TCR. Moreover, recent models suggest that in the presence of persistent DNA lesions, which are repaired slowly, Pol IIs can be polyubiquitinated and degraded to provide access for the repair factors to the lesion and to prevent new transcription cycles. In contrast, we did not detect Pol II degradation upon sublethal dose of UVB, which might reflect the activity of a mechanism that may facilitate the temporary dissociation of Pol II complexes from the promoters of expressed genes without degradation. This idea further suggests that different mechanisms may exist to remove Pol II from the transcriptional units during the DNA repair processes and their activation depends on the persistency and types of DNA lesions.

INTRODUCTION

Chapter one: RNA polymerase II regulation and pause during transcription

I. The expression of genetic information

The genetic information carried by DNA macromolecules, resides in the linear order (sequence) of nucleotides along a strand and divided into functional units (genes). Every cellular function and events require the access and activation of only a small subset of genes out of the whole genome. Eukaryotic genomes are complex (up to 25000 genetic loci in human) and organized within compact nucleoprotein (chromatin) structures. Therefore, transcription has a tightly regulated spatiotemporal control, which requires a well orchestrated recruitment of multiple protein complexes at a precise location in the nucleus in order to express the encoded information in the genome.

Living cells are constantly exposed to different intrinsic and extrinsic stimuli coming from the basic metabolic and cellular processes or from external environmental sources. In order to maintain the normal physiological state, cells have to access their DNA-encoded instructions and transcribe RNA molecules from particular genes. Most of these RNAs are processed and are translated during protein synthesis or they can act as regulators during different biochemical processes, which are required for normal cell functions such as development, differentiation and homeostasis in eukaryotic cells.

1. DNA dependent RNA polymerases

In prokaryotes, one DNA-dependent RNA polymerase is sufficient to transcribe genes into RNA molecules required by the cell. In eukaryotes, expression of genes is shared by three distinct multi-subunit enzymes, each of which is dedicated to transcribe specific gene types. In eukaryotes RNA polymerase (Pol) I transcribes genes encoding 18S, 28S and 5.8S ribosomal RNAs (rRNAs). RNA Polymerase III is responsible for the transcription of genes coding for tRNAs 5S RNA and other small RNAs found in the nucleus and cytosol. RNA

Polymerase II is responsible for the transcription of all the protein-coding genes which constitute the largest group of distinct individual genes in the eukaryotic genome. In addition, it also transcribes genes coding for small non-coding RNAs, like micro RNAs (miRNAs), spliceosomal RNAs, small nucleolar RNA (snoRNAs) and a yet still unidentified fraction of intergenic or non-coding transcripts. These polymerases are the products of 31 genes: Pol I contains 14, Pol III 17 and Pol II has 12 subunits.

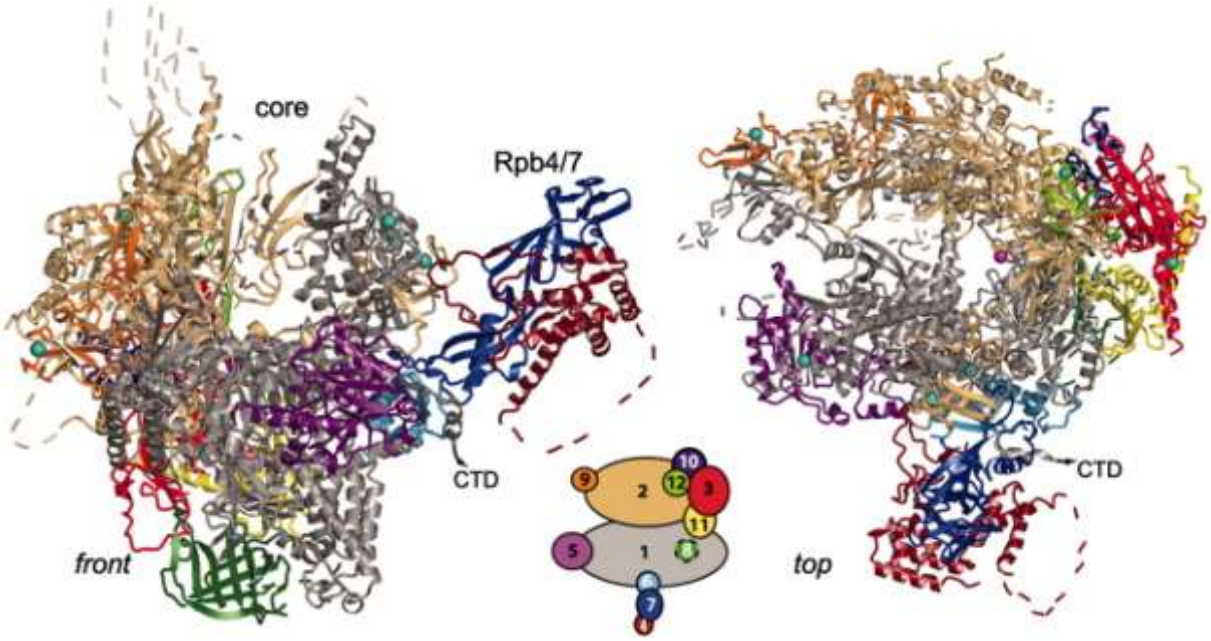


Figure 1: RNA polymerase II structure. Ribbon diagram shows two standard views front *(left)* and top *(right*). The 12 subunits Rpb1-Rpb12 are colored according to the key *(middle)* between the views. *Dashed lines* represent disordered loops. Eight zinc ions and the active site magnesium ion are depicted as *cyan spheres* and a *pink sphere*, respectively. The full CTD tail is not presented here. Figure adopted from (Armache et al, 2005).

Yeast and human Pol II both contain 12 subunits, (RPB1 to RPB12, with a total mass of > 0.5 MD) by decreasing order of their molecular mass (**Figure 1**) (Young, 1991). The 12 subunits of Pol II are highly conserved in sequence, architecture, and function thus, seven subunits of human Pol II can either partially (RPB4, RPB7, and RPB9) or completely (RPB6, RPB8, RPB10, and RPB12) substitute for the their yeast homologues. Five subunits (RPB5, RPB6, RPB8, RPB10, and RPB12) are shared between the three polymerases. The largest catalytic subunits of these polymerases (RPB1, RPB2, RPB3, and RPB6) share homology with each other and with the largest subunit of bacterial RNA polymerase subunits β-, β, α and ω, respectively. The similar sequences between RPB1 and β- as well as between RPB2 and β also refer to functional similarity: RPB1 and β-are involved in DNA binding, while RPB2 and β bind nucleotide substrates (Hampsey, 1998; Lee & Young, 2000). RPB1 and RPB2 are responsible for most of the catalytic activity of polymerase and are essential for

phosphodiester bond formation. Only RPB4, RPB7, RPB9 and the Carboxyl-Terminal Domain (CTD) of RPB1 are unique to Pol II (Minakhin et al, 2001; Mitsuzawa & Ishihama, 2004; Wild & Cramer, 2012).

A. Structure of Pol II

In the past twenty years, the structure of eukaryotic Pol II has been studied intensively and a wealth of structural information is provided by photocrosslinking, X-ray crystallography, NMR, and cryo-electron microscopy. These structure studies showed that the yeast Pol II easily dissociates into a 10-subunit catalytic core and a heterodimer of RPB4 and RPB7 subunits. The RPB4/RPB7 heterodimer is not essential for RNA chain elongation, although it is required for Pre-Initiation Complex (PIC) formation and mRNA maturation (Pankotai et al, 2010). [Also reviewed by (Cramer, 2004a; Cramer, 2004b; Hampsey, 1998; Lee & Young, 2000)] The two large polymerase subunits in Pol II (Rpb1 and Rpb2) form the opposite sides of the active centre. The small core subunits RPB5, RPB6 and RPB8 can bind to RPB1, RPB9 binds to RPB2 and the remaining subunits (RPB3, RPB10, RPB11 and RPB12) form a distinct subassembly that bridges between RPB1 and RPB2. Therefore, the polymerase core may be divided into three interacting subassemblies, which we refer to as RPB1 (composed of RPB1, RPB5, RPB6 and RPB8), RPB2 (consisting of RPB2 and RPB9) and RPB3 (RPB3, RPB10, RPB11 and RPB12) subassemblies, respectively (**Figure 1**). The central mass formed by RPB1 and RPB2 can be subdivided into four mobile elements, termed “core”, “clamp”, “shelf” and “jaw” lobe which bunk up around a positively charged “cleft” (**Figure 2**). The active center is buried at the base of the “cleft”. A “pore” beneath the active centre widens towards the outside, creating an inverted “funnel” through RPB1. DNA is suggested to enter at the “cleft” down the middle of the enzyme, passing between the mobile “jaws”. Beyond the active site, the DNA path is blocked by a protein “wall”. DNA-RNA hybrid formed in the active site would have to pass up the wall, at nearly right angles to the incoming DNA in the cleft. Both DNA and RNA are held in place by a massive “clamp” swinging over the cleft and active-center region. A hole in the floor of the cleft below the active site (“pore”) would allow entry of substrate nucleoside triphosphates and would also allow exit of RNA during retrograde movement of the polymerase on the DNA.

The location of RPB4/RPB7 was found close to the RNA exit channel and an additional role in transcriptional initiation was suggested (Armache et al, 2003; Boeger et al, 2005; Bushnell & Kornberg, 2003).

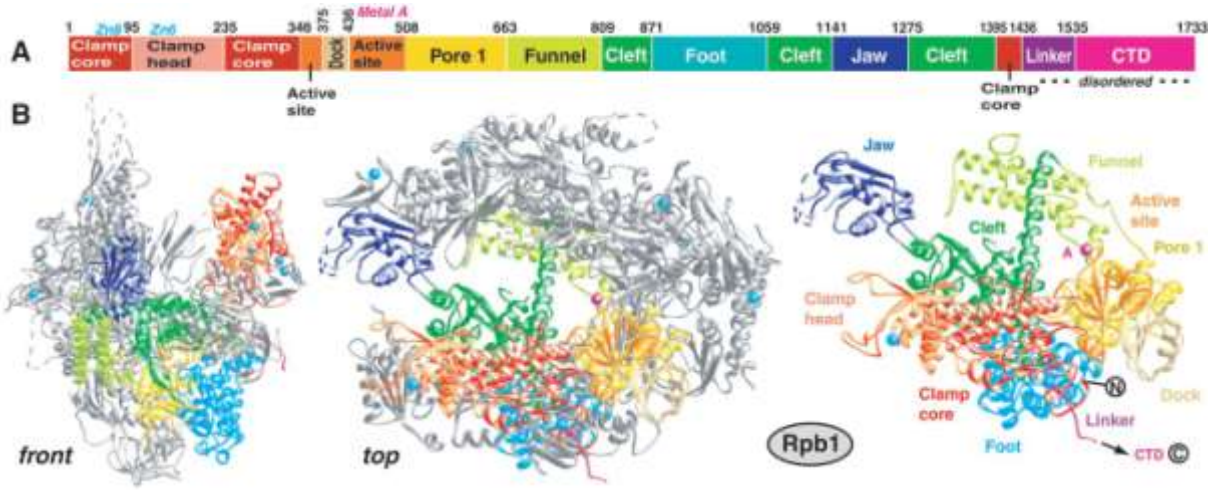


Figure 2: Structure of Rpb1. (**A**) Domains and domain like regions of Rpb1. The amino acid residue numbers at the domain boundaries are indicated. (**B**) Ribbon diagrams, showing the location of Rpb1 within Pol II ("front" and "top" views of the enzyme) and Rpb1 alone. Locations of NH2- and COOH-termini are indicated. Color-coding as in (**A**). Figure adopted from (Cramer et al, 2001).

The CTD is an unstructured extension from the catalytic core of Pol II. It is flexibly linked via an 80-residue linker to the close proximity of the RNA exit channel and provides an exposed surface to interact with proteins involved in 5' capping, mRNA splicing, termination and 3'-end processing (Meinhart et al, 2005).

B. Carboxyl-terminal domain of Pol II

Compared to the other DNA-dependent RNA polymerases, the eukaryotic Pol II has an unusual structure what is uniquely found at the C terminus of the largest subunit (RPB1). This carboxyl-terminal domain is evolutionarily conserved in eukaryotes even as distant as yeast and humans. CTD comprises multiple tandem repeat of heptapeptides with the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser ($Y_1S_2P_3T_4S_5P_6S_7$), with up to 52 repeats in the mammalian proteins (**Figure 3**). Deletion of the CTD in mouse, Drosophila or yeast is lethal, demonstrating that this structure is essential for living cells.

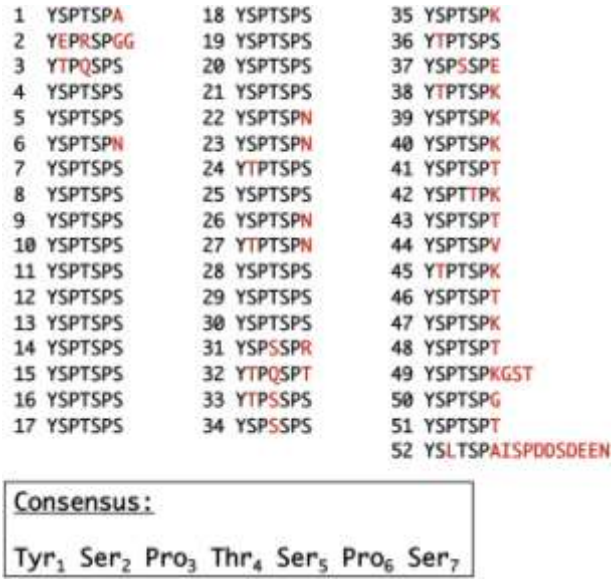


Figure 3: The carboxyl-terminal domain (CTD) of Pol II. The sequence of human CTD (amino acids 1593–1970) varies mainly at position 7 in the heptad motif in the distal (C-terminal) section of CTD. Amino acid residues that vary from the consensus motif are depicted in red. The number of repeats as well as the amino acid sequence is 100% conserved in mammals. Figure adopted from (Heidemann et al, 2012).

Interestingly, transcription, in some conditions can proceed without the CTD (Corden, 1990; McCracken et al, 1997), pointing to an ancillary, rather than fundamental role for this structure. It is also important that the heptads have to be in tandem: insertion of an alanine residue between heptads is lethal in yeast, whereas insertion of an alanine between "heptad pairs" can be tolerated (Stiller & Cook, 2004). While the CTD is essential for life, it is frequently not required for General Transcription Factor (GTF)-mediated initiation and RNA synthesis *in vitro* (Buratowski & Sharp, 1990; Zehring et al, 1988). This suggested that CTD tail of the RPB1 does not form part of the catalytic essence of Pol II, thus it may have important regulatory roles. Later, it became clear, that CTD serves as a scaffold for the interaction of a wide range of nuclear factors and plays a major role in transcription and co-transcriptional RNA processing of protein-coding genes, mammalian snRNA genes and yeast snoRNA genes (Gudipati et al, 2008). The CTD extends from the Pol II core enzyme close to the RNA exit channel, and this localization makes it capable for direct or indirect interaction with components of the RNA processing machinery.

C. Posttranslational modifications of CTD

Phosphorylation of the heptapeptide residues is the best-studied CTD modification. RPB1 can exist in two forms that can be distinguished on SDS PAGE: Pol IIa, when the CTD is hypophosphorylated and Pol IIo with hyperphosphorylated CTD. Pol IIa was shown to associate with the PIC at the promoter, and any phosphorylation of the CTD before this point

will prevent incorporation in the PIC and initiation. Transcription elongation and 3' processing require sequential phosphorylation of Pol II. The CTD is the site of phosphorylations, with up to five potential phosphorylation sites on the tandem heptapeptide (Y_1, S_2, T_4, S_5, S_7) (Baskaran et al, 1993) (**Figure 3**). *In vivo* phosphorylation occurs mainly on serine residues (Zhang & Corden, 1991).

Proline-directed serine kinases are creating the characteristic pattern of Ser2 and Ser5 phosphorylation, which correlates with the phase of transcription on protein-coding genes in yeast and mammals (**Figure 4 and 5**) (Phatnani & Greenleaf, 2006). Phosphorylation of Ser5 is carried out by Cyclin-Dependent Kinase (CDK7) subunit of the TFIIH complex (see section E). Ser5 phosphorylation is needed for the recruitment of enzymes (*Ceg1* and *Abd1* in yeast) that will add the methyl guanosine cap to the nascent RNAs (Komarnitsky et al, 2000; Schroeder et al, 2000). Phosphorylation level of Ser5 is enriched at the promoter and decrease successively towards the 3' end of genes. Ser5-P is dephosphorylated by Ssu72 phosphatase around 3' end of genes which is essential for 3' end processing of mRNAs (Steinmetz & Brow, 2003).

Phosphorylation of Ser2 is mediated by CDK9, which is the catalytic subunit of the Positive Transcription Elongation factor b (P-TEFb) complex (see section K/d). Ser2 phosphorylation increases on the elongating Pol II toward the 3' end of genes (**Figure 4 and 5**) (Peterlin & Price, 2006). Ser2 phosphorylation plays an important role in the transition from early elongation block into a productive elongating form. In addition, *in vitro* it was shown that splicing and polyadenylation is activated by Ser2 and Ser5 phosphorylated CTD (Hirose & Manley, 1998; Hirose et al, 1999). In higher eukaryotes, inhibition of Ser2 phosphorylation will end up in biased elongation, splicing and polyadenylation (Bird et al, 2004). However, Ser2 phosphorylation seems to play little part in the elongation of transcription of mammalian snRNA and replication-activated histone genes, (encoding relatively short RNAs) which are neither spliced nor polyadenylated (Medlin et al, 2005). In addition, the mammalian p21 protein-coding gene does not require P-TEFb activity for either elongation or RNA processing (Gomes et al, 2006), indicating that the requirement for Ser2 phosphorylation can be bypassed. Histone mRNA 3' end formation is directed by the specialized, replication-activated histone mRNA processing signal and seems to be phospho-Ser2 (Ser2-P) independent (Discussed in section 7). In contrast, 3' processing of snRNA genes require Ser2-P. Ser2 and Ser5 can also be phosphorylated by CDK8 and Cell Division Cycle 2

protein (CDC2). These kinases create the highly phosphorylated Pol IIo and Pol IIm forms that are thought to be transcriptionally inactive (Palancade & Bensaude, 2003). Dynamic dephosphorylation of Ser2 and Ser5 by CTD-specific phosphatases during the transcription cycle is essential for recycling Pol II. Two evolutionary conserved proteins (FCP1 and SSU72) dephosphorylate phospho-Ser2 and Ser5 respectively (Meinhart et al, 2005). RPAP2 (RNA Polymerase II-associated protein) also can dephosphorylate CTD-Ser5 and creates heptapeptides with phospho-Ser7, which facilitate the recruitment of P-TEFb and the subsequent phosphorylation of Ser2 (Czudnochowski et al, 2012; St Amour et al, 2012).

Ser7 is also a target of dynamic phosphorylation during the Pol II transcription cycle in eukaryotic cells (Akhtar et al, 2009; Chapman et al, 2007). Phosphorylated Ser7 is detected at the Transcription Start Site (TSS) and its levels remain high close to the 3' end of Pol II genes (Brookes et al, 2012; Kim et al, 2010; Mayer et al, 2010) (**Figure 5**). Substitution of Ser7 to alanine is not lethal in yeast, but impairs viability of human cells by affecting the processing of snRNAs (Egloff et al, 2007; Schwer & Shuman, 2011). For the recruitment of the functional Integrator complex [a multi-subunit complex responsible for snRNA 3' processing (Baillat et al, 2005)] two adjacent Ser7-P and Ser2-P marks are needed (Egloff et al, 2010). Ser7-P is further known to be a gene-specific "CTD-code" for its role in the expression and processing of snRNAs, although this mark is also present in protein-coding genes, where the Ser7-P is placed early in transcription, similar to Ser5-P, but its levels remain stable until the transcription termination site. Recent studies demonstrated that CDK7 is also the primary kinase for Ser7 phosphorylation in human and yeast cells (Boeing et al, 2010; Glover-Cutter et al, 2009; Kim et al, 2009). In the later phases of the transcription, the role of CDK7 is taken over by BUR1 a kinase which travels with the elongating Pol II placing phospho-Ser7 marks on its CTD. Although high levels of CTD Ser7-P correlate with high transcription rates (Bataille et al, 2012; Tietjen et al, 2010), the exact function of this CTD mark in transcription of protein coding genes is still unclear. Nevertheless, there might be a strong connection between Ser5 and Ser7 phosphorylation, which is suggested by data showing that CDK7 and Ssu72 act as a common kinase/phosphatase for both residues. Ssu72 removes Ser7-P immediately after cleavage and polyadenylation and thereby contributes to the reconstitution of the hypophosphorylated state of Pol II (Bataille et al, 2012; Zhang et al, 2012a).

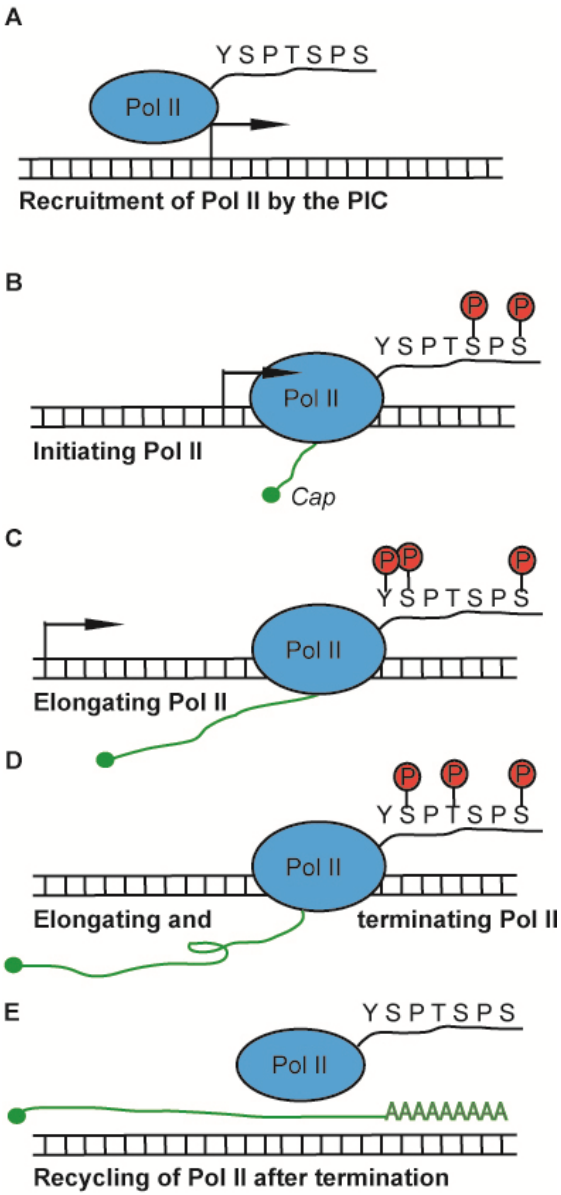


Figure 4: Pol II CTD phosphorylations during transcription. (**A**)The CTD of Pol II which is recruited by the PIC at the promoter is unphosphorylated. (**B**) Phosphorylation of Ser5 and Ser7 by the CDK7 subunit of TFIIH just after initiation helps to recruit and activate enzymes that add a methylguanosine cap (green filled circle) to the 5' end of the nascent RNA. (**C**) Subsequent phosphorylation of Ser2 by CDK9 subunit of P-TEFb activates elongation and RNA processing. Phosphorylation of Tyr1 prevents the binding of termination factors to the CTD. Ser5 is dephosphorylated toward the 3' end of the transcription unit by SSU72. (**D**) As Pol II reaches the termination site, the Tyr1-P signal weakens, but Thr4 is phosphorylated by PLK3 in humans and facilitates recruitment of termination factors. (**E**) After cleavage and polyadenylation of the 3' end of the pre-mRNA, dephosphorylation of the CTD may help Pol II to disengage and to be recruited for another round of transcription.

Recent studies revealed that CTD threonine 4 is phosphorylated in yeast and higher eukaryotes. In humans Thr4 is phosphorylated by PLK3, but so far in yeast, no specific kinase was identified (Hsin et al, 2011) (Hintermair et al, 2012; Mayer et al, 2012). The replacement of Thr4 to alanine is not lethal in yeast (Stiller et al, 2000) however, the presence of Thr4 is essential for viability in human cells (Hintermair et al, 2012). Chromatin immunoprecipitation assay coupled sequencing (Chip-seq) datasets in yeast revealed that Thr4-P can be found mainly on the body region of genes and shows a low profile at the polyadenylation site (Mayer et al, 2012), while in human cells, phospho-Thr4 is enriched in the 3' region of genes (**Figure 5**). Experiments demonstrated that Thr4 phosphorylation may have a role in the recruitment of termination factors like PCF11 (Meinhart & Cramer, 2004). When Thr4 was

changed into valine in chicken cells, biased processing of core histone mRNA was observed, while the basal transcription was not affected (Hsin et al, 2011).

Phosphorylation of CTD-Tyr1 was reported on human Pol II almost two decades ago (Baskaran et al, 1993), but its functional role was unknown. Genome-wide occupancy profiling revealed that this modification is associated with active genes (Mayer et al, 2012). The distribution profile correlates with the genome-wide phospho-Ser2 Pol II profile: Tyr-1 occupancy is low at promoters, then increases on the gene bodies and drops before reaching the polyadenylation site (**Figure 5**). This is in good agreement with experiments which demonstrated that Tyr1-P prevents CTD binding to the conserved CTD-interacting domain of termination factors. In contrast, Tyr1-P does not impair CTD binding to the distinct CTD-binding domain of elongation factor SPT6, consistent with SPT6 occupancy within the Tyr1-phosphorylated region of genes (Mayer et al, 2012). Therefore studies about phospho-Tyr1 mark explain how termination factor activities are restricted to the late phases of the Pol II transcription.

Each CTD repeat contains two prolines, embedded between phosphorylation sites. The two prolines at positions 3 and 6 of the CTD heptads can undergo conformational changes mediated by peptidyl–prolyl cis/trans-isomerases (PIN1) (Egloff & Murphy, 2008). Proline isomerization is also part of the "CTD code" which creates various binding scaffolds for other CTD associating factors. For example the cleavage and polyadenylation factor PCF11 is recruited to Ser2-P modified CTD repeats (Licatalosi et al, 2002) in combination with the relevant prolines in the energetically favored trans conformation (Meinhart & Cramer, 2004; Noble et al, 2005). Thus, the isomerization status of prolines can greatly influence the CTD phosphorylation pattern.

Serine and threonine residues can carry O-linked N-acetylglycosamine (O-GlcNAc), which is mutually exclusive with phosphorylation, meaning that this mark may be important to prevent CTD phosphorylation (Comer & Hart, 2001). Current studies propose that glycosylation of CTD Ser5 and Ser7 is carried out by O-GlcNAc transferase (OGT) and O-GlcNAc aminidase (OGA) during assembly of PIC (Ranuncolo et al, 2012).

The distal part of the Pol II tail comprises largely non-canonical CTD heptads with aberrations mainly occurring at position 7 (Chapman et al, 2008). Experiments showed that lysines at position 7 of the heptads and arginine 1810 of CTD repeat 31 can be specifically methylated by coactivator-associated arginine methyl-transferase 1 (CARM1) (Sims et al,

2011). This CTD mark can be found in hyperphosphorylated CTD *in vivo*. The enzymatic activity of CARM1 toward Arg1810 is repressed by phospho-Ser5 and Ser2 residues *in vitro*, meaning that the methylation is placed before early initiation but is still present during transcription (Sims et al, 2011).

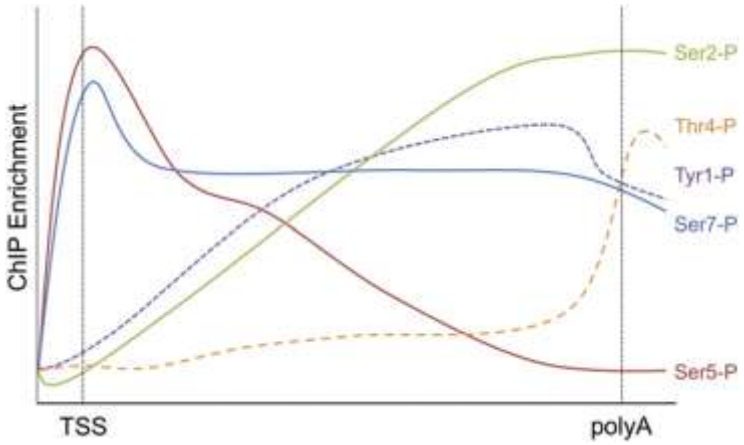


Figure 5: Average profile of CTD phosphorylation marks in genes. Schematic representation of genome-wide occupancy profiles for all CTD phosphorylation marks. Tyr1-P, Ser2-P, Thr4-P, Ser5-P and Ser7-P show similar distribution of the modified Pol II along certain gene regions. While Ser5-P and Ser7-P peak at the transcription start site (TSS) of genes, Tyr1-P, Ser2-P and Thr4-P signals increase toward the 3' end and polyadenylation site. Adopted from (Heidemann et al, 2012)

CTD residue	Type of modification	Modification added by	Removed by	Role	Localization
Serine / Threonine	O-linked N-acetylglycosamine	O-GlcNAc transferase	O-GlcNAc aminidase	Prevent CTD phosphorylation during PIC formation	Promoters
Serine 2	Phosphorylation	CDK9 /CDK8 / CDC2	FCP1	Promotes elongation / 3' end processing	Gene body / End of genes
Serine 5	Phosphorylation	CDK7 / CDK8 / CDC2	SCP1 / PRAP2 / SSU72	Recruits RNA-capping enzymes / Promoter clearance	Promoters
Serine 7	Phosphorylation	CDK7 / BUR1	SCP1	snRNA processing	Promoter / Gene body
Tyrosine 1	Phosphorylation	c-Abl proto-oncogene		Prevents termination factor binding to CTD	Gene body / End of genes
Threonine 4	Phosphorylation	PLK3 / CDK9?		Recruits 3'-end processing / termination factors	End of genes
Proline 3/6	Isomerization	PIN1		Recruits 3'-end processing / termination factors	End of genes

Table 1.: Posttranslational modifications of RNA Polymerase II C-terminal domain

II. RNA polymerase II transcription cycle

Transcription of genes is the most highly regulated process in eukaryotic gene expression. Transcription cycle of Pol II can be divided into three phases: initiation, when the polymerase is recruited at a promoter with the help of a set of DNA binding proteins and begins to synthesize RNA, elongation, when Pol II extends the RNA transcript and termination when the polymerase and the RNA product disengage from the template.

2. Transcription initiation

Transcription initiation starts with the recognition of a promoter region, which will serve as a platform for the assembly of the transcription factors and indicates the starting site of the transcription.

D. Core promoter elements

The Pol II core promoters contain special, evolutionally conserved sequence motifs, which are extended approximately 70 nucleotides around the Transcription Start Site (TSS) and direct the initiation of the transcription. A core promoter could be as simple as a single motif that serves as a universal transcription start site, or as complex as having a unique combination of a set of sequence instructions for each promoter (**Figure 6**).

We can distinguish focused promoters, where there is either a single transcription start site or a distinct cluster of start sites in a relatively short region and dispersed promoters, in which there are a number of transcription start sites distributed over a broad region with a range from up to 100 nucleotides. Some other genes might have alternate promoter sites, which are distinct and sometimes involve different type of regulation. These promoters are typically located hundreds or thousands of nucleotides apart.

In vertebrates, only about one-third of core promoters are focused core promoters, whereas the vast majority of genes contain dispersed core promoters (Juven-Gershon et al, 2008).

The presence of core promoters is necessary for accurate transcription initiation. These regions and their combination of short sequence elements represent an ancient and the most basic level of transcription regulation.

A core promoter can be built up from the following elements:

- **TATA box** was the first eukaryotic promoter element to be identified. It is the most ancient and the most widely-used core promoter motif in nature. It has an A/T-rich consensus sequence [TATA(A/T)A(A/T)(A/G)], which is located approximately 20-30 nucleotides (nt) upstream of TSS in humans. The upstream T nucleotide is commonly located at -31 or -30 relative the Inr (Initiator). TATA box is present in less than 10% of human promoters. Surprisingly a wide range of A/T-rich sequences are capable of functioning as TATA boxes, consistent with the fact that TATA-Binding Protein (TBP) recognizes DNA through sequence-independent minor groove contacts. Many promoters do not contain consensus TATA boxes, or even non-consensus TATA boxes. Some TATA-less promoters retain the ability to direct transcription initiation from a specific nucleotide, whereas others can direct transcription initiation from multiple start sites (Smale, 1997; Smale & Kadonaga, 2003).

- The initiator (**Inr**) is generally a pyrimidine-rich sequence around the TSS, PyPyA+1N(T/A)PyPy, (from nucleotide -6 to +11). In mammalians, computational analysis of thousands of transcription start sites suggests that the mammalian Inr consensus is YR, where R corresponds to the +1 start site (Carninci et al, 2006). The Inr is necessary *in vitro* and *in vivo* for a low level of accurate transcription by RNA polymerase II. Inr is capable to direct accurate transcription initiation either alone or with additional core promoter elements. These characteristics make Inr and TATA box functionally similar in two respects: both direct accurate transcription initiation by Pol II in the absence of other control elements, and both direct a high level of accurately-initiated transcription when stimulated by an upstream activator (Concino et al, 1984). The Inr is probably the most commonly occurring sequence motif in focused core promoters and it is a recognition site for a number of proteins.

- The downstream promoter element (**DPE**) can be mainly found at TATA-less promoters, precisely +28 to +32 nt from TSS. It has a consensus sequence of (A/G)G(A/T)CGTG what was shown to be bound by Transcription Factor II D (TFIID). DPE is functionally dependent on Inr, therefore either promoter features a DPE sequence, it also has an Inr. TFIID binds cooperatively to DPE and Inr, therefore any mutation in either motif or alteration of the spacer region will lead to diminished level of basal transcription (Kutach & Kadonaga, 2000). Photocrosslinking studies revealed that the DPE is in close proximity to the TFIID subunits TAF6 and TAF9 (TAF stands for: TBP Associated Factor). Both contain histone fold motifs, thus it is possible that these subunits of TFIID interact with the DPE in a manner that is similar to binding of histones H3-H4 to DNA in nucleosomes (Shao et al, 2005).

- Motif Ten Element (**MTE**) with a consensus sequence C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C) can be found between +18 to +29 from TSS. It can also functionally substitute the loss of the TATA box and in addition, with the TATA box and DPE in proximity, it enhances the promoter activity in an Inr-dependent manner.

- Downstream Core Element (**DCE**) consists of three sub-elements: SI, SII, and SIII, with a core sequence of CTTC, CTGT, and AGC, respectively. DCE occurs frequently with the TATA box, and appears to be distinct from the DPE. In a photo-crosslinking experiment TAF1 was shown to bind specifically the DCE.

- **XCPE1** (X core promoter element 1) motif is located from −8 to +2 relative to the +1 start site. It is present in about 1% of human core promoters, most of which are TATA-less. XCPE1 exhibits little activity by itself. [Reviewed in (Juven-Gershon et al, 2008)]

- TFIIB-recognition element (**BRE**) with its (G/C)(G/C)(G/A)CGCC consensus sequence, can be found immediately upstream of the TATA box of 12% eukaryotic TATA-promoters. As it is in the name, TFIIB recognizes and binds BRE and its presence can

both negatively and positively influence basal promoter strength (Deng & Roberts, 2007).

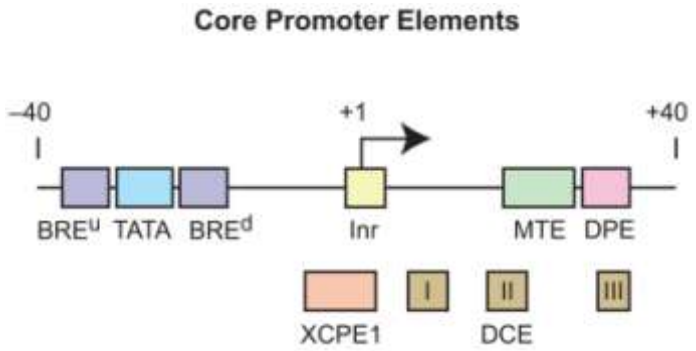


Figure 6: Location of core promoter motifs. Each of these elements is found in only a fraction of all core promoters. The Inr is probably the most commonly occurring core promoter motif. TATA box, MTE, DPE, DCE are recognition sites for TFIID. BREu and BREd interact with TFIIB. Figure adopted from (Juven-Gershon et al, 2008).

Large majority of the mammalian promoters, 60% of the human protein coding genes have multiple alternative start sites spread across a region. The advantage of this alternative promoter usage is that in one cell, different proteins can be generated from the same locus for different requirements.

The presence and combination of these sequence elements 5' of transcription units determines not only the start site of transcription, but influences the promoter strength, direction of transcription, and may indicate functional differences between genes with different promoter architecture.

E. General transcription factors

The RNA polymerase enzymes do share a common property in transcribing DNA sequences, although they lack sequence-specific recognition ability to correctly specify the transcription start site unique to each gene. For site-specific initiation, additional proteins are necessary to form an initiation competent RNA polymerase complex. These essential accessory factors were collectively defined as general transcription factors (GTFs). Namely TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (TF means transcription factor, II indicates Pol II driven transcription).

- TFIID was first identified as a multiprotein complex which contains TBP and another 13 TBP-associated factors (TAFs) which support the promoter recognitions by binding to different promoter elements around the TSS (Fire et al, 1984; Young, 1991). Later, biochemical experiments demonstrated that the TFIID subunit composition can be different in cells at certain stages of development to provide different gene expression patterns (Muller et al, 2010). It is known that TAFs bind to the Inr, DPE, and DCE element, which allow TFIID to recognize TATA-less promoters. TBP exhibit nonspecific DNA-binding activity, which can lead to the formation of nonproductive transcription complexes. Thus, to prevent false initialization, TBP is affected by multiple negative regulations such as: TBP homodimer formation, inhibition by TAF1, BTAF1, Negative Cofactor2 (NC2) and nucleosome barrier (Goppelt et al, 1996; Kim & Burley, 1994; Kokubo et al, 1993; Nikolov & Burley, 1994; Timmers et al, 1992). The promoter recognition and TBP-TATA complex formation is positively regulated by TFIIA and TFIIB complex.

- TFIIA is a complex of three proteins (α, β, γ subunit), which are encoded by two genes in higher eukaryotes. TFIIA increases the precise promoter recognition of TBP via promotion of the dissociation of TBP dimers. Furthermore, TFIIA stabilizes the forming PIC by competing with the TAF1/BTAF1 inhibition (Auble & Hahn, 1993; Coleman et al, 1999).

- TFIIB is a single, 316 amino acid long polypeptide, which stabilizes the promoter bound TBP by competing with NC2 and it has also an important role in the recruitment of Pol II/TFIIF (Goppelt et al, 1996). TFIIB was further shown to recognize BREu element in a TBP independent manner via a non-conserved helix-turn-helix motif (Evans et al, 2001).

- TFIIF is a heterodimer comprising two subunits, each of RAP30 and RAP74 (RNA Pol II Associated Proteins). It has multiple essential roles during the PIC formation process. Once it is bound to Pol II, it will facilitate its recruitment to the promoters and will further increase the TBP-TATA-TFIIB complex stability. As for the next step in the PIC formation it will facilitate the recruitment of TFIIE/TFIIH complexes. Later, it is

required for TSS selection, Pol II promoter escape and in transcription elongation (Flores et al, 1990; Yan et al, 1999).

- TFIIE contains two subunits, α and β, which form a heterotetramer. TFIIE interacts with TFIIF, TFIIB, Pol II and promoter DNA. TFIIE's α subunit is needed for the interaction with Pol II and binds close to the active center. TFIIE helps to recruit TFIIH, while it stimulates the helicase activities and the CDK7-mediated phosphorylation of the Pol II CTD (Ohkuma & Roeder, 1994; Watanabe et al, 2003).

- TFIIH complex is composed of ten subunits: p89/XPB, p80/XPD, p62, p52, p44, p40/CDK7, p38/Cyclin H, p34, p8 and p32/MAT1 (Compe & Egly, 2012; Conaway & Conaway, 1989). The whole complex can be divided into two functional subcomplexes: a Cyclin-Activating Kinase complex (CAK) and a core complex. TFIIH has multiple roles during transcription: First, the core complex (XPB which is a helicase, p62, p52, p44, p34 and p8) is responsible for a DNA dependent ATPase activity. This activity is required for stable promoter opening and for the first phosphodiester bond formation which will further catalyze promoter clearance of Pol II(Kumar et al, 1998). Second, CAK subcomplex (CDK7, Cyclin H, MAT1) can phosphorylate Pol II CTD through CDK7 (Lu et al, 1992). This phosphorylation step mediates the transition from initiation to elongation phase. Third, the XPB and XPD subunits have ATP-dependent helicase activity, which is essential for promoter clearance and has an important role in DNA-damage repair too (Schaeffer et al, 1994). Fourth, E3 ubiquitin ligase activity of p44 subunit was also detected during DNA damage response (Takagi et al, 2005).
-

F. Sequential assembly of the PIC complex

The stepwise assembly of these GTFs at the recognized promoter will lead to a formation of a transcriptionally productive PIC. According to the sequential assembly model (Buratowski et al, 1989), first, the TATA-element is recognized and bound by TBP and the TFIID complex. Then the entry of TFIIA and TFIIB further stabilize the TBP–TFIID complex at

the promoter and provide a platform for the integration of the TFIIF bound Pol II complex. This promoter bound TFIID-TFIIA-TFIIB-TFIIF/Pol II pre-PIC complex shows resistance to any negative regulatory events, which target the promoter recognition and PIC assembly, and also supports the recruitment of TFIIE. TFIIE binds close to the Pol II active center and once it is present, it recruits TFIIH. TFIIH facilitates the melting of the double stranded DNA into a single stranded bubble, which establishes an open promoter complex. Transcription initiation occurs upon incorporation of the two initiating nucleotides dictated by the template and formation of the first phosphodiester bond.

G. Activation dependent transcription and coactivators

Some eukaryotic genes expression is regulated by promoter and gene-specific activators and transcription factors. The activators contain DNA binding and activator type domains that are required for stimulating transcription. In addition, a set of proteins called "coactivators" are needed for activator-dependent initiation. The coactivators facilitate the communication between gene-specific transcription factors and the components of the GTFs (**Figure 7**).

a) Mediator complex

Mediator (MED) was first identified in yeast and found to consist of approximately 21 polypeptides which can form different sub-complexes. Eleven of the MED subunits are essential for yeast viability. In mammals, the first Mediator-related complex was the human TRAP (Thyroid hormone Receptor-Associated Protein) complex as it has been initially identified as a group of protein associated to thyroid hormone receptor. Later, approximately 30 MED subunits were identified which can form multiple metazoan Mediator-related complexes. The term ''Mediator'' was first proposed for its role as an adaptor between transcription factors bound at regulatory elements, and Pol II to activate transcription in a reconstituted system. Detailed mechanisms of how MED participates in transcription have not been fully determined. However, Mediator is believed to act mainly during PIC assembly, either by increasing the efficiency and rate of PIC assembly, or by

affecting the recruitment of Pol II and other general initiation factors (Kornberg, 2005; Takagi & Kornberg, 2006) (**Figure 7**).

b) Upstream stimulatory activity (USA)

The USA (upstream stimulatory activity) was initially defined as a set of proteins which are able to stimulate activator-dependent transcription. Later USA was further purified and a series of positive cofactors (PC1, PC2, PC3 and PC4) and a negative cofactor (NC1) were identified [PC2 is an active form of MED (Chen & Roeder, 2011)]. Interestingly, these USA-derived cofactors have dual roles: to function as stimulators during activator-dependent transcription and also as a repressor to inhibit basal transcription, when specific activators are absent (Kaiser & Meisterernst, 1996).

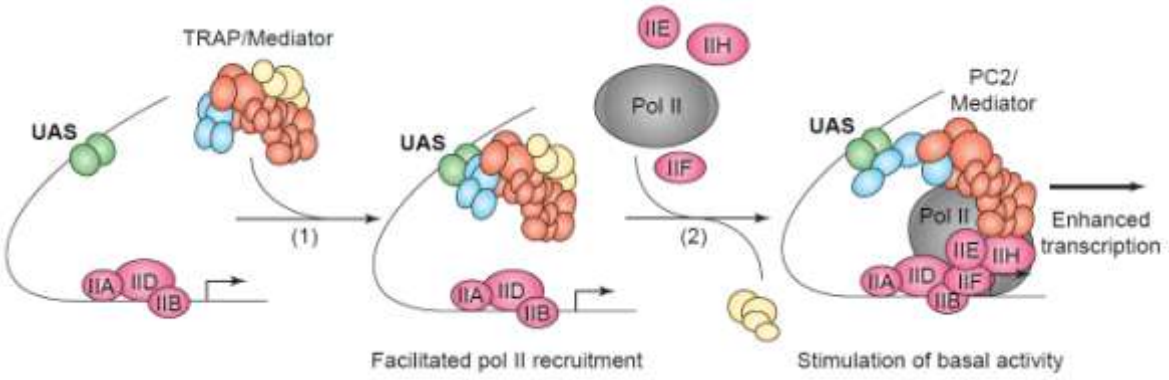


Figure 7: Model for activated transcription. Transcriptional activators bind to their target sites [upstream activating sequence (UAS)] and recruit the intact form of Mediator. (**1**). The Mediator facilitates the recruitment of Pol II into the nascent PIC (**2**). This might cause the loss of certain MED subunits (yellow bubbles). The remaining MED complex is the activated, PC2-like form. Figure adopted from: (Malik & Roeder, 2005).

c) Chromatin remodeling and modifications

Once activators bind to the promoter, they trigger a cascade of recruitment of factors such as chromatin-remodeling complexes and histone-modification enzymes. These factors not only facilitate stronger binding of activators to DNA, but also make nucleosomal DNA elements more accessible to GTFs.

Nucleosomes appear to affect all stages of transcription. A nucleosome is composed of 147 bp of DNA wrapped 1.65 turns around the histone octamer (a dimer of Histone H2A,

H2B H3 and H4). Beside the packaging function, nucleosomes possess dynamic properties that are tightly regulated by various protein complexes.

Both histone tails and globular domains are subject to various posttranslational modifications like: methylation of arginine residues; methylation, acetylation, ubiquitination, ADP-ribosylation, sumoylation of lysines (K); phosphorylation of serines and threonines, etc.

Increased histone acetylation at the promoter region has been linked to active transcription (Workman & Kingston, 1998). Studies demonstrated that acetylation of H3 and H4 peaks sharply at active promoters. In addition, it has been shown that HATs such as p300, SAGA and ATAC (depends on the target genes and the state of differentiation etc.) are recruited to the promoters and histone acetylation occurs prior to PIC formation (Bhaumik, 2011; Bhaumik & Green, 2001; Nagy et al, 2010).

Typically, histone acetylation occurs at multiple lysine residues and is usually carried out by a variety of HATs (Brown et al, 2000). Some modifications can directly influence higher-order chromatin structure. For instance, acetylation of H4 K16 inhibits the formation of compact 30 nm fibers (Shogren-Knaak et al, 2006). With the exception of methylation, histone modifications result in a change in the net charge of nucleosomes, which could loosen histone-DNA interactions.

To make DNA more accessible, promoter-bound activators also target a set of protein complexes that utilize ATP hydrolysis to transiently unwrap the DNA from histone octamers, to form DNA loop, or moving nucleosomes to different positions. These complexes are called ATP-dependent chromatin remodelers. Not only individual histone subunits can be moved or evicted, but also histone dimers, and even entire octamers can be temporary removed (Workman, 2006).

The mechanisms of histone eviction involve numerous activities. Studies suggest that cooperative binding and action of transcription factors, chromatin-remodeling complexes such as Swi/Snf and RSC, and actively transcribing Pol II can all mediate histone displacement. To prevent removed histones to rebind to the same DNA molecule, the presence of histone chaperones like ASF1, NAP1, and nucleophosmin are necessary. On active genes different histone variants H2A.Z (Htz1) or H3.3 can be enriched (Guillemette et al, 2005; Schwartz & Ahmad, 2005). High-resolution mapping revealed that Htz1-containing nucleosomes flank a 200 bp nucleosome-free region at promoters (Raisner et al, 2005) and these Htz1-containing nucleosomes are resistant to remodeling (Li et al, 2005). In yeast,

upon transcription activation, Htz1 is rapidly evicted from the promoter, and its loss is required for full transcription (Zanton & Pugh, 2006; Zhang et al, 2005). However, it should be noted that although there is solid evidence for histone loss, the promoter is not completely nucleosome free. Acetylated histones H3 and H4 continue to accumulate during gene activation (Pokholok et al, 2005).

3. Promoter escape

Once PIC formed, the polymerase begins the synthesis of the nascent mRNA chain and it enters a critical stage referred to as promoter escape, which is characterized by physical and functional instability of the transcription complex. Before entering into a productive elongation, initiating Pol II must undergo structural and functional maturation.

Promoter escape begins with the onset of transcription initiation at which point the transcription complex is called: initially transcribing complex (ITC). Promoter escape is considered complete when the nascent RNA associates stably with the transcription complex. The transcription complex is then called an early elongation complex (EEC), however, it is still different than the fully functional elongation complex.

After transcribing the first 4 nt, the Pol II complex takes a ‘‘committed’’ conformation with the interaction of TFIIB, however transcription complex still remains physically and functionally unstable until the nascent transcripts are 10 nt or longer (Kugel & Goodrich, 1998; Kugel & Goodrich, 2002). At this phase, ITC can undergo upstream transcript slippage, which means that the initial RNA transcript unpairs from the DNA template and rehybridizes upstream (Pal & Luse, 2003).

The polymerase complex becomes further stabilized through RNA–protein contacts in the exit channel. 9–10 nt long RNA transcript can begin to form contacts with several protein loops that both stabilize the transcription complex and separate the RNA transcript from the DNA template, causing conformational changes within the polymerase that converts Pol II to the EEC.(Dvir et al, 1997).

At the same time, with the 3’-5’ helicase activity, XPB subunit of TFIIH continues unwinding the promoter region in an ATP-dependent manner (Moreland et al, 1999). Due to this helicase activity, the downstream edge of the transcription bubble moves together with

transcription. Collapse of the upstream portion of the transcription bubble produces a transcription bubble that is more characteristic of a productively elongating polymerase. TFIIE was shown to follow the early elongating complex, until the 10th nt, where it dissociates from the complex. TFIIH might escort the Pol II complex longer approximately until the 25th or 30th nt. The TFIID and the other PIC components remain stably bound to the promoter (Dvir, 2002). [Initiation and promoter escape are very similar, and perhaps even indistinguishable processes]

4. Early elongation

The transition from initiation to early elongation requires a tightly controlled exchange of factors bound to the promoter region. These changes around the PIC are orchestrated in part through phosphorylation of the Pol II CTD. As it was discussed above, CTD is mostly unphosphorylated during the recruitment into the PIC, which favors interactions between PIC stabilizing factors, like coactivators. Pol II is then phosphorylated at Serine-5 by CDK7, which is thought to destabilize the interactions between Pol II and promoter-bound factors, helping the promoter escape. Phospo-Ser5 is targeted by a number of factors involved in early elongation and modification of promoter proximal histones. When the nascent transcript reaches ~20 nucleotides, its 5'-end is modified with a 7-methyl-guanosine cap (Rasmussen & Lis, 1993). This modification is critical for RNA stability, further RNA processing, export from the nucleus and protein translation (Proudfoot et al, 2002). The co-transcriptional capping appears to be stimulated by Serine-5-phosphorylation. In addition, RNA capping may also be facilitated by inefficient early elongation, which could provide enough time for this step to take place (Pei et al, 2003).

Early elongation is a slow, inefficient process *in vitro*. The early elongation complex shows tendency to pause, arrest and terminate transcription (abortive transcription), with the presence of initiated Pol II complexes failing to generate full-length RNA (Marshall & Price, 1992). The slow characteristic of early elongation can be due to the appearance of negative elongation factors that inhibit synthesis through the promoter-proximal region. In addition, the early elongation complex needs time for the final conformational changes to become fully stable and enable for transcription.

5. Pol II pause

In addition, detailed *in vitro* and in vivo analyses proved that Pol II can pause during early elongation. This phenomenon was first documented at the *Drosophila* heat shock genes, where Pol II synthesizes a 25–50 nt RNA before pausing in elongation (Rougvie & Lis, 1988). This paused elongation complex is phosphorylated at Serine-5 and the nascent RNA is capped (Boehm et al, 2003). Pol II pausing at hsp70 gene is rate-limiting for gene expression, with release of polymerase into the gene approximately once every 10 minutes prior to heat shock (Lis, 1998).

Later, by the spreading of the genome-wide techniques, evidences proved that pausing is a widespread post-recruitment regulation of transcription in metazoans (Guenther et al, 2007) (**Figure 8**).

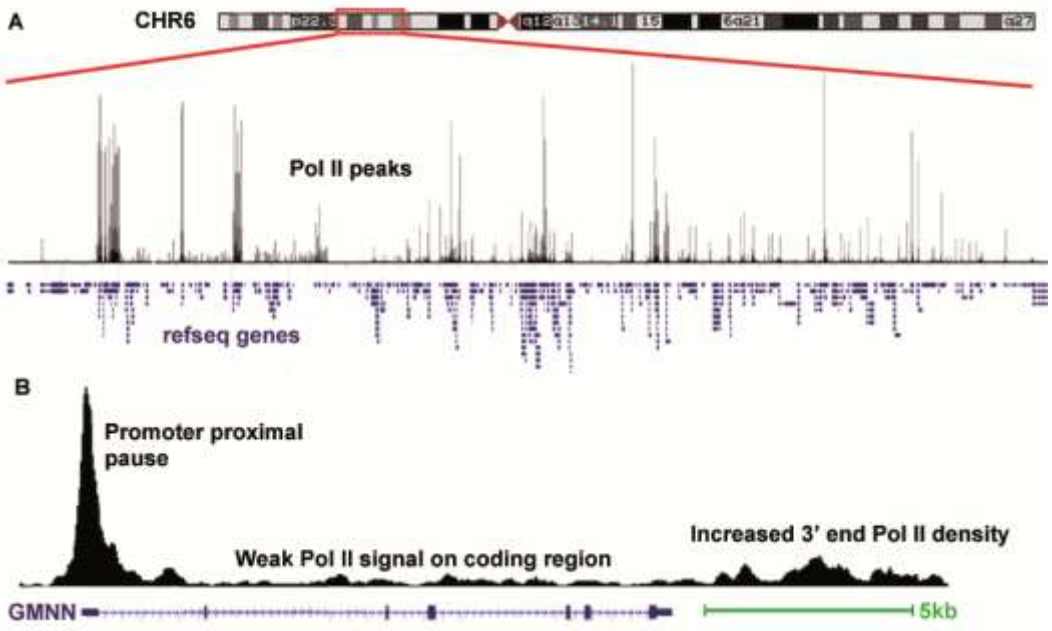


Figure 8: RNA polymerase II ChIP seq data visualization. Human Pol II ChIP-seq data was uploaded and visualized in UCSC genome browser. The antibody recognizes all forms of Pol II (N20). (**A**) Pol II peaks (black bars) are visible at almost all genes (blue blocks). The randomly selected genomic interval on chromosome 6 is approx. 14mb long. (**B**) Pol II occupancy profile on an average polyadenylated gene from the selected genomic interval. The black and sharp peak (left) represent the promoter proximal pause. The Pol II signal is weak on the gene body. Downstream of the End of Annotated Gene (EAG) a wide-spreading medium strength Pol II density can be observed.

The Pol II pause has become a target of intensive investigation in the recent years, although, there is still no common agreement about its exact role in transcription regulation. In addition, different status of promoter-proximal Pol II pause can be described. For

example, recruited polymerase can fail to initiate RNA synthesis and can be trapped as a PIC. This would involve substantially different mechanisms for release than in a case when Pol II is blocked after synthesizing a short transcript. Moreover, early elongation complexes that accumulate downstream of promoters can be present in several conformations that are not all competent to resume RNA synthesis. These observations led to the introduction of different terms, which all refer to the same Pol II, which does not efficiently transcribe into the gene but they are used to describe the different mechanistic background.

We can distinguish poised, paused, backtracked, arrested and stalled form of Pol II pause. All of these terms have been used to describe Pol II enrichment at promoters or at different regions of a gene (**Figure 9**).

Poised Pol II can be detected at promoters without distinguishable PIC or elongation complexes. This term is used when the data comes from ChIP assays, which cannot determine whether Pol II is engaged in transcription.

Paused Pol II can be found in elongation complexes that have halted transcription temporarily, but remain competent to resume transcription.

Backtracked Pol II elongation complexes halted synthesis and moved backward several nucleotides to displace the growing 3'-end of the nascent RNA from the active center.

Arrested Pol IIs are basically referred to backtracked Pol II complexes which cannot spontaneously restart transcription in the absence of factors that induce cleavage of the RNA. Formation of a new RNA 3'-end in the active center allows resumption of transcription, but these elongation complexes can further pause before synthesis.

The term **stalled** Pol II refers to elongation complexes which stopped RNA synthesis, but the exact mechanistic background is unknown, thus the term stalled may also refer to Pol IIs which are paused, backtracked, arrested or undergoing termination. Stalled Pol IIs can be detected also at DNA lesions caused by genotoxic stress.

In most cases, paused Pol II can be induced to restart transcription. Arrested or terminating elongation complexes cannot be restarted and require either rescue or reinitiation to generate a productive transcript.

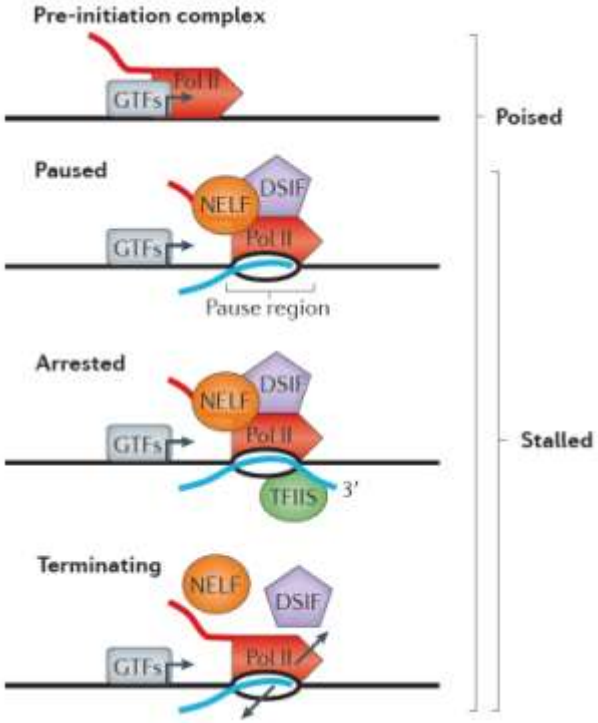


Figure 9: Terms used to describe promoter proximal Pol II complexes. The Pol II (red) is recruited by the GTFs (grey) to the promoter. The pause-inducing factors are: NELF (orange), DSIF (purple) and TFIIS (green). The nascent RNA is shown in blue, and a bracket indicates the pausing region, usually 20–60 nucleotides downstream from the TSS. **Pre-initiation complex:** Pol II is bound to the TSS, but has not yet initiated RNA synthesis. **Paused Pol II:** an early elongation complex which transiently halted RNA synthesis. Paused polymerase is fully competent to resume elongation. Two protein complexes, DSIF and NELF, reduce the rate of elongation and facilitate the establishment of the stably paused state. **Arrested Pol II:** a stably engaged but backtracked elongation complex. In this case the RNA 3′ end might be displaced from the active site. Restart of an arrested complex usually requires TFIIS. **Terminating Pol II:** an unstable elongation complex that is in the process of dissociating from the DNA template and releasing the nascent RNA. **Poised Pol II:** describes a Pol II which is located near the TSS, but does not specify anything about its transcriptional status. **Stalled Pol II:** indicates a transcriptionally engaged Pol II but makes no assumptions about its ability to resume synthesis. Figure, adopted from (Adelman & Lis, 2012).

H. Factors contributing in Pol II pause establishment

Factors were later identified at promoters, which can facilitate Pol II pausing. The first factor is a heterodimeric complex called DRB Sensitivity-Inducing Factor (DSIF), which comprises the homologues of the yeast Spt4 and Spt5 proteins (Wada et al, 1998). Second is the Negative ELongation Factor (NELF) complex (Yamaguchi et al, 1999). It is known that DSIF and NELF together are required to inhibit and/or regulate transcription elongation (Renner et al, 2001) (**Figure 9; 10**).

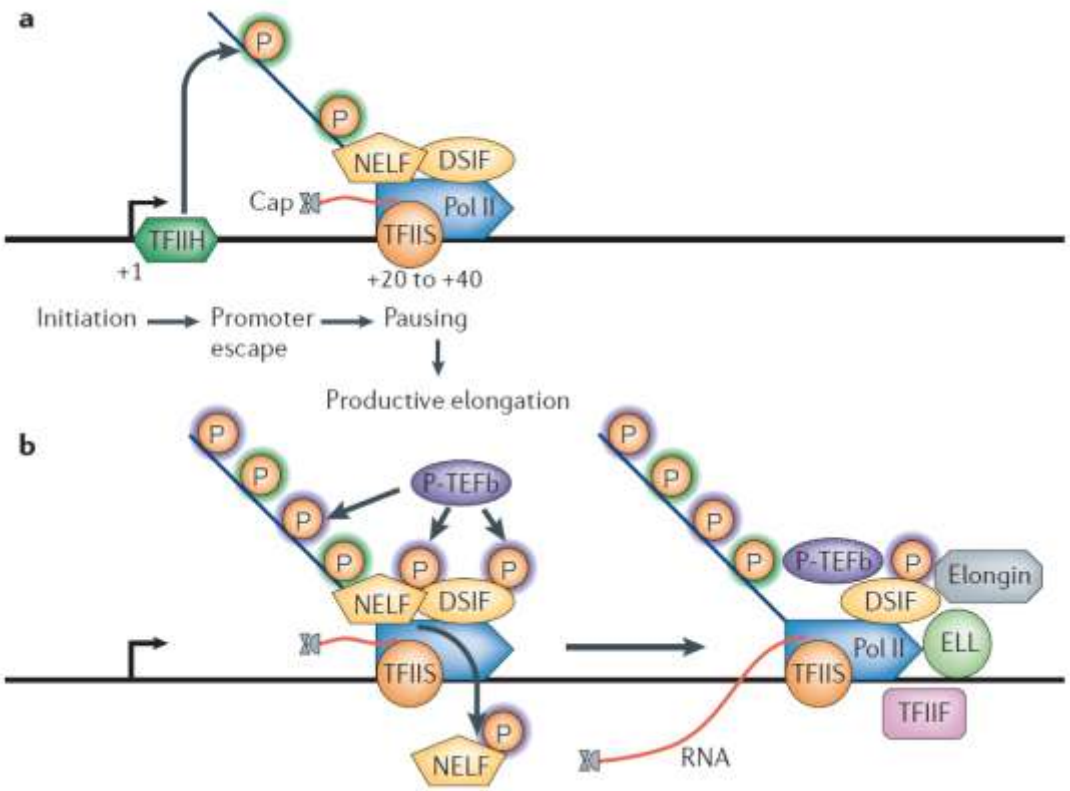


Figure 10: Promoter-proximal pausing and escape to productive elongation. (**A**) TFIIH-mediated phosphorylation of Ser5 of Pol II CTD occurs before promoter-proximal pausing. DSIF and NELF facilitate Pol II pausing. TFIIS stimulates the intrinsic RNA-cleavage activity of Pol II to create a new RNA 3' end in the Pol II active site after backtracking. Capping enzyme associates with the Ser5-P CTD and the nascent RNA becomes capped. (**b**) P-TEFb mediates phosphorylation of DSIF, NELF and Ser2 of the Pol II CTD, which stimulates productive elongation. TFIIS facilitates efficient release of Pol II from the pause site by aiding the escape of backtracked transcription complexes NELF dissociates from the transcription complex and DSIF, TFIIS and P-TEFb moves together with Pol II along the gene. In addition, TFIIF, ELL, and elongin, will further stimulate Pol II elongation activity. Figure adopted from: (Saunders et al, 2006).

NELF is a complex of four subunits, NELF-A, B, C or D and E, that is conserved in higher eukaryotes. (Narita et al, 2003). The mammalian NELF-A protein, also called WHSC2, helps to anchor the NELF complex to the polymerase (Yamaguchi et al, 2001). The NELF-B subunit, also known as COfactor of BRCA1 (COBRA-1), can bind to several other transcription factors at promoters (Aiyar et al, 2004; Zhong et al, 2004). NELF C and D are translation variants of the same mRNA, and their function is unknown. NELF-E subunit contains RNA recognition motif, which through NELF may inhibit elongation by interacting with certain structures in the nascent RNA (Narita et al, 2003). It is in good agreement with the observation that NELF is able to inhibit transcription only after the RNA is sufficiently long to emerge from Pol II. In addition, nascent RNA is also necessary for the association of DSIF with Pol II, which make uncertain if DSIF, or NELF, recognizes the transcript (Cheng & Price,

2008; Missra & Gilmour, 2010). Roles of DSIF and NELF have been shown also in RNA processing of short, non-polyadenylated histone transcripts (Narita et al, 2007). ChIP-chip technique was carried out in both *Drosophila* and mammalian cells, which indicated that NELF generally occupies Pol II-bound genes. Indeed, ChIP-chip analysis of Pol II after RNA interference (RNAi) treatment against NELF revealed that loss of NELF reduces Pol II promoter occupancy at most promoters (Muse et al, 2007), supporting the idea that NELF mediates inhibition of early elongation. ChIP assays show that NELF is present only around promoters, but does not advance into the gene with the elongating Pol II, thus dissociation of NELF accompanies pause release (Core et al, 2008; Rahl et al, 2010; Wu et al, 2003). In contrast, DSIF remains with polymerase during elongation, eliciting both negative and positive effects on transcription (Andrulis et al, 2000; Kaplan et al, 2000). DSIF complex is present in all eukaryotes, thus, DSIF is likely to be fundamental for Pol II transcription and to serve as a general platform for regulatory factors. Taken together DSIF and NELF may play a role in establishing pausing during early elongation specifically in metazoans (**Figure 10**).

Histones and nucleosome occupancy pattern may also play a supplementary role in Pol II pausing and backtracking. Nucleosomes can be highly phased upstream and downstream of promoters. This organization leaves the promoter open and facilitates transcription initiation (Mavrich et al, 2008; Schones et al, 2008). Pol II must go through the +1 nucleosome for successful transcription. The +1 nucleosome can retard Pol II progression and is linked to polymerase pausing or backtracking (Margaritis & Holstege, 2008). In the review of (Taft et al, 2009)), a model was proposed that Pol II generates a short nascent RNA and proceeds to the 5' border of the +1 nucleosome. Then elongation is retarded by the +1 nucleosome, or by contact with other polymerases paused there, and backtracks ~18 nt, which displaces the 3' end of the nascent RNA from the Pol II active site. TFIIS (an important elongation factor) cleaves the nascent transcript to generate a new 3' end, positioned properly in the Pol II active site for further synthesis. The 3' ~18 nt of the nascent transcript fragment is released (tiRNA biogenesis), and the elongation eventually proceeding through the +1 nucleosome from the point of cleavage. This sequence may be repeated until Pol II is able to bypass the nucleosomal barrier.

Pausing might involve the cooperation between DNA sequences and the pause-enhancing factors. Sequences are known to influence elongation in two aspects: first, the thermodynamic stability of the 9 nucleotide long RNA-DNA hybrid, formed by nascent RNA

and the DNA template. Weaker hybrids (more A+T) increase the likelihood of pausing. Second, structures formed by the nascent RNA facilitate pausing by displacing the 3'-end of RNA from the polymerase catalytic site (Artsimovitch & Landick, 2000; Landick, 2006).

So far, no general sequence motif is identified that specifically influences Pol II pausing. However, DPE and Pause button motifs [Pause button motif is a CG rich sequence element that is over-represented at stalled gene promoters in Drosophila (Hendrix et al, 2008)] are clearly enriched at genes where Pol II pausing is prominent. These G+C-rich sequences are preferentially located between positions +26 and +33 with respect to the TSS, precisely aligned with the peak of Pol II pausing. Downstream of this G+C-rich region, the sequence of paused genes is characterized by stretches of A+T, leading to the suggestion that the initially transcribed sequence contains a signal that specifies pausing and backtracking to Pol II. According to this model, promoter-proximal pausing comprises two stages: first, an initial halt in transcription is induced by an A+T-rich, weak RNA-DNA hybrid; and second, the polymerase backtracks to reach the more thermodynamically stable, G+C-rich sequence upstream, which coincides with the DPE or pause button motif. Therefore, the intrinsic inefficiency of early elongation and the tendency of Pol II to undergo backtracking at these genes provide a kinetic window of opportunity for the binding of NELF and DSIF, and the establishment of a stably paused state (Hendrix et al, 2008; Nechaev et al, 2010).

I. Possible roles of Pol II pause

As every step during transcription, pausing also represents an additional regulatory opportunity during transcription cycle. Accordingly, this could allow activators that influence pause release to work together with factors that stimulate recruitment in a combinatorial manner. Therefore Pol II pause is a checkpoint for elongation and RNA processing with the possibility of post-initiation regulation. (Mandal et al, 2004; Moore & Proudfoot, 2009; Nechaev & Adelman, 2008).

In metazoans, non-expressed genes that exhibit pausing are enriched in signal-responsive pathways, including development, cell proliferation and stress or damage responses. However, it is important to note that the prevalence of pausing Pol II on promoters can greatly vary by using different methods and different statistical criteria.

Investigating Pol II enrichment in pluripotent cells, such as ESCs, has suggested that pausing might have a role in cell differentiation (Guenther et al, 2007). Genomic analysis of Pol II distribution also indicates that pausing occurs at genes across the range of expression levels (Gilchrist et al, 2010; Nechaev et al, 2010). In fact, recent global analyses of Pol II distribution by Global Run-On assay coupled sequencing (GRO-seq) indicate that few paused genes are transcriptionally inactive (<1%) (Core et al, 2008). This argues against the idea, that Pol II pausing is predominantly a mechanism to silence gene expression (Hargreaves et al, 2009). Data show that all of the traditionally defined paused genes (such as *Drosophila Hsp* genes and mammalian *Myc* and *Fos*) exhibit considerable basal expression. In addition, the *Drosophila Hsp* genes continue to undergo pausing during activation. On the basis of these data, Pol II pausing is rather a mechanism for tuning expression of active genes and perhaps for poising them for future changes in expression (Lis, 1998).

Pol II can also reduce its elongation speed and can pause during productive elongation. In this case, the factors involved and the mechanisms that govern pausing within the gene are different from those that regulate promoter-proximal Pol II pausing (Peterlin & Price, 2006). Pausing within the gene body has been reported, to be coupled to splicing and elongation rates can have an impact on alternative splicing (de la Mata et al, 2003).

Pausing Pol II was also shown to play a role in establishing permissive chromatin. Wrapping promoter DNA around histone proteins to form nucleosomes can present a barrier to transcription by rendering crucial recognition sequences inaccessible. Therefore Pol II pause at promoters could help to maintain an open and accessible promoter structure to facilitate binding of transcription factors. The link between paused Pol II and maintenance of a nucleosome-deprived promoter has been demonstrated at a genome-wide level in *Drosophila* (Gilchrist et al, 2010). Genes with paused polymerase were globally shown to possess low levels of promoter nucleosome occupancy, which was dependent on the presence of promoter-associated Pol II. This result may indicate that highly regulated promoters, with paused Pol II, have evolved DNA sequences that enable a dynamic competition between Pol II pause and nucleosomes (Gilchrist & Adelman, 2012). This idea raised a model, describing that the presence of paused Pol II poises genes in ESCs for expression during development, in part by altering promoter chromatin. As such, the loss of paused Pol II later in development could permit nucleosome occlusion and permanent gene repression (Guenther et al, 2007). With the chromatin-opening function, paused Pol II can

allow genes that are transcribed at lower basal levels to be continually accessible and primed for bursts of transcription activation.

J. Promoter associated RNAs

Promoter-associated RNAs (PARs) are defined as short transcripts localizing around the TSSs. PARs have been described in all major eukaryotic organisms (Taft et al, 2009). PAR density and position may be related to the widespread, and bidirectional, activity of Pol II. A recent genome-wide study has shown that transcript initiation occurs at almost all genes, regardless of their expression levels, yielding small transcripts at least 70 nt long coincident with the 5' ends of protein-coding genes (Davis & Ares, 2006).

Three classes of PARs have been described to date in animals:

Promoter-associated small RNAs, are generally 20–200 nt long, capped, with 5' ends that coincide with the TSSs of protein- and non-coding genes (Fejes-Toth et al, (2009; Kapranov et al, 2007).

Second, a similar and perhaps overlapping class of small RNAs, transcription start site associated RNAs, are 20–90 nt long and localized within -250 to +50 around of TSSs (Seila et al, 2008).

A third class, transcription initiation RNA (tiRNA), has been identified in human, mouse, chicken and *Drosophila*. These species are predominantly 18 nt in length, have their highest density just downstream of TSSs, and show patterns of positional conservation. All three classes are strongly associated with highly expressed genes (Seila et al, 2008).

It is possible that tiRNAs and other PARs have evolved to serve an active role in the regulation of transcription, or in the communication between the different steps of transcription to regulate other processes associated with it. However, the overall function and evolution of PARs is still unclear.

6. Transcription elongation

Once Pol II is released from the promoter proximal pause, it travels into the coding region. As Pol II advances on the gene, it creates a transcription bubble, where the DNA template forms a hybrid duplex with the nascent pre-mRNA. To maintain the movement of the transcription bubble, the RNA polymerization and the mRNA processing, the assembly of a productive Elongation complex is essential. In addition, efficient elongation must overcome several blocks that are present naturally, coming from Pol II catalytic activity and from the chromatinized DNA template. During elongation Pol II has to face transcriptional pause, arrest, and termination (Shilatifard et al, 2003; Uptain et al, 1997). Transcriptional pausing occurs when the RNA polymerase stops the addition of NTPs to the nascent transcript for a time, then it resumes productive elongation on its own. Transcriptional arrest can be defined as an irreversible halt to RNA synthesis, when RNA polymerase cannot resume productive elongation without the help of other factors. During termination, the RNA polymerase and RNA transcript are released from the DNA, ending the elongation stage of transcription. Transcriptional pause and arrest in vivo are most likely caused by a combination of DNA sequences, protein factors, and the nascent transcript. The cells exploit a very sophisticated array of factors, which follow the traversing Pol II, control the chromatin architecture and diminish the pausing of the elongation. In the next paragraphs I will describe these important factors:

K. Factors to diminish Pol II pause during elongation

d) P-TEFb and DSIF

The transition from pause to early elongation requires the change of transcription factors binding to the promoter. One of the most important elongation factors is P-TEFb, which helps to reverse the effects of the negative elongation factors and allow Pol II to enter into productive elongation (Price, 2000). P-TEFb is a heterodimer of the CDK9 kinase and CyclinT, and localized in the nucleus (Peterlin & Price, 2006). P-TEFb is an essential kinase,

which was found in eukaryotes from yeast to humans. In metazoans it is regulated by an unusual mechanism involving reversible association with a small nuclear ribonucleoprotein (snRNP) complex (Diribarne & Bensaude, 2009).

The amount and availability of active P-TEFb in the cell is controlled through sequestering of P-TEFb into an inactive complex with 7SK RNA and HEXIM protein (Diribarne & Bensaude, 2009). Activation events lead to the release of P-TEFb from this inactive complex, resulting in a rapid increase in P-TEFb activity. Beside this regulation, general P-TEFb inhibitors can be present in certain cell types (like germ lines in *Drosophila)*, suggesting that blocking P-TEFb function may broadly suppress germ line transcription prior to the appropriate developmental stage (Hanyu-Nakamura et al, 2008).

P-TEFb phosphorylates the Pol II CTD at Ser2 residues, which cancels inhibition of early elongation (Cheng & Price, 2007; Marshall & Price, 1995). In addition, Ser2-P will then provide a platform for assembly of complexes that travel with the polymerase through the gene. This includes factors that regulate transcription elongation, RNA processing and termination (Peterlin & Price, 2006), as well as the modification and remodeling of histones (Li et al, 2007) (**Figure 10**).

As mentioned above, in addition to its role in phosphorylation of Ser2, P-TEFb has also been shown to phosphorylate NELF and DSIF, which facilitates the release of NELF. Accordingly, *in vitro* transcription systems indicate that the kinase activity of P-TEFb cancels the repressive effect of NELF (Peterlin & Price, 2006). Without NELF, DSIF converts from a negative to a positive elongation factor (Ivanov et al, 2000; Kim & Sharp, 2001). DSIF will then interact with factors implicated in mRNA maturation and surveillance through its SPT5 subunit (Lindstrom et al, 2003; Pei & Shuman, 2002).

P-TEFb can be directly recruited to genes by transcription activators like NF-κB and c-myc (Barboric et al, 2001; Eberhardy & Farnham, 2002). Moreover, artificial recruitment of P-TEFb enhanced transcription from the *Drosophila hsp70* gene, suggesting that delivery of P-TEFb to promoters may be a key role of transcription activators (Lis et al, 2000). Recently, a Super Elongation Complex has been described, which contains P-TEFb together with elongation factors like ELL and a number of chromatin remodeling factors, like PAF1 (Lin et al, 2010; Sobhian et al, 2010). These studies suggest that P-TEFb may be often associated with the elongating polymerase as part of a large, multi-protein complex.

Once phosphorylation of DSIF by P-TEFb removes NELF from the elongation complex, other factors will be recruited to Pol II which will provide the ability to elongate with approximately ~3.8 kb/min rate (Singh & Padgett, 2009). After transition to productive elongation, the Pol II complex is very stable, and can transcribe tens, or even hundreds of kb without dissociating from the DNA template (Singh & Padgett, 2009).

e) TFIIS

In most cases, a transcriptionally arrested Pol II cannot resume elongation without help. Transcription arrests result from the Pol II "backtracking". During backtracking, Pol II moves backward relative to the DNA template, which results in misalignment of the catalytic active site and 3' end of the nascent RNA transcript. This pause turns gradually into arrest in a time-dependent fashion (Gu & Reines, 1995). Arrested Pol II complexes resume productive elongation via the elongation factor TFIIS. This eukaryotic factor promotes Pol II readthrough at transcriptional arrest sites (Conaway et al, 2003; Fish & Kane, 2002). TFIIS was initially identified by its ability to stimulate transcription *in vitro*, and was shown to function after transcription initiation, as well as to stimulate elongation by reducing Pol II pausing (Bengal et al, 1991; Sluder et al, 1989) (**Figure 10**).

Studies further demonstrated that TFIIS stimulates RNA polymerase-mediated transcript cleavage. The cleavage reaction is intrinsic to the RNA polymerase itself, but is enhanced in the presence of accessory factors, such as TFIIS (Orlova et al, 1995; Rudd et al, 1994). The generation of a new 3' end for the nascent RNA will position the transcript properly in the catalytic site (Fish & Kane, 2002). The active site of Pol II undergoes extensive structural changes during TFIIS binding, and these structural changes are consistent with a realignment of the RNA in the active center. TFIIS was observed to extend along the surface of Pol II, insert into a pore, and reach the active site from the bottom face of Pol II (Kettenberger et al, 2003).

In general, TFIIS localizes to the coding regions of genes as shown by ChIP and physically interacts with the Spt5 subunit of DSIF (Lindstrom et al, 2003; Pokholok et al, 2002). In addition, TFIIS is expected to play a critical role also in promoting the activity of early elongation complexes at promoters. Indeed, global analysis of paused genes in *Drosophila*

reveals that Pol II backtracking and TFIIS induced transcript cleavage are common amongst early elongation complexes (Nechaev et al, 2010). RNAi mediated depletion of TFIIS delays the induction of heat-shock genes; but surprisingly, global gene expression profiles are not significantly affected (Adelman et al, 2005).

f) TFIIF

Additionally to its important role in PIC formation, TFIIF reduces the time Pol II is paused and stimulates the rate of elongation. Both subunits of TFIIF contribute to transcription initiation and elongation (Bengal et al, 1991; Izban & Luse, 1992; Tan et al, 1994). Following its release from the PIC complex, TFIIF associates with the elongation complex, in particular when the Pol II complex has stalled. Consistent with this finding, TFIIF has the ability to associate with both the Pol IIa present in initiation complexes, and the Pol IIo found in the elongation complex (Zawel et al, 1995) (**Figure 10)**. TFIIF localizes predominantly near the promoter region, not uniformly throughout the coding region, and it appears also at 3'untranslated region (Krogan et al, 2002; Pokholok et al, 2002). These findings are in agreement with mechanistic studies demonstrating that TFIIF does not travel with Pol II, but re-associates with those which encounter a block to elongation.

g) Elongin complex

The complex has three subunits: the transcriptionally active elongin A, the regulatory B and C proteins. Similar to TFIIF, elongin A stimulates the rate of Pol II transcription but it is not required for PIC formation or initiation (Takagi et al, 1996)(**Figure 10**).

Elongin B is a member of the ubiquitin homology gene family, and elongin C contains homology to a subunit of the yeast SCF complex that targets proteins for polyubiquitination (Ivan & Kaelin, 2001). Therefore, it was previously speculated that elongin may facilitate the ubiquitination of Pol II or other transcriptional cofactors (Shilatifard et al, 2003). The large subunit of Pol II becomes ubiquitinated upon Ultraviolet (UV) radiation, or other DNA damaging agents in a manner dependent on CTD phosphorylation (see section 12). It is contradictory but possible, that the function of elongin *in vivo* is related to the disassembly

of the Pol II elongation complex that encounters blocks to elongation, such as lesions on the DNA. Elongin may associate with stalled Pol II and promote its degradation, allowing the repair machinery to gain access and subsequently repair the DNA. However, a link between the elongin complex and specific ubiquitination events remains to be elucidated (Mitsui & Sharp, 1999; Ratner et al, 1998).

h) The ELL family

ELL is functionally analogous to TFIIF and Elongin. It suppresses Pol II pause, and stimulates the transcriptional rate. It was shown that ELL associates with active sites of transcription following heat shock and co-localizes with the elongation-competent form of Pol II in *Drosophila* (Gerber et al, 2001; Shilatifard et al, 1996).

The ELL family contains ELL1, ELL2 and ELL3. ELL1 directly increases the catalytic rate of Pol II by suppressing transient pausing of Pol II along the DNA template (Shilatifard et al, 1996). ELL1 had been also identified as a fusion partner of the mixed lineage leukemia (MLL) protein (Thirman et al, 1994). MLL plays an important role in maintaining the appropriate expression of the *Hox* gene loci in hematopoietic cells and is a key regulator of hematopoietic stem cell development (Smith et al, 2011a). ELL2 and ELL3 were identified and shown to have similar elongation stimulatory activity (Miller et al, 2000).

In *Drosophila*, there is only one ELL that is colocalizing with Pol II at heat-shock gene loci upon heat shock (Gerber et al, 2001).

The ELL family members ELL1-2-3, together with:

- the MLL translocation partners AF4/ FMR2 family member 1 and4 (AFF1, AFF4),
- eleven-nineteen leukemia (ENL)
- ALL1-fused gene from chromosome 9 (AF9)
- and the Pol II elongation factor P-TEFb

can form a Super elongation complex (SEC) following translocation of *MLL* to any of the genes encoding SEC components (Smith et al, 2011a).

SEC was first found to be required for the aberrant activation of *MLL* chimera target genes through misregulation of transcription elongation checkpoint control (TECC), which supports a role for abnormal transcription elongation control during leukemogenesis. SEC also has a

major role in the regulated release of paused Pol II and gene activation in normal physiology (Luo et al, 2012).

ELL can be a part of a second complex named little elongation complex (LEC), which different subunit composition: LEC contains ICE1 (interacts with the C terminus ELL subunit 1) and ICE2 (also known as NARG2) LEC might have a role in the expression of Pol II transcribed snRNA genes (Smith et al, 2011b).

i) CSB

CSB (Cockayne Syndrome) is a DNA-dependent ATPase that has been linked to transcriptional elongation. CSB gene was shown to stimulate the rate of elongation *in vitro*, to directly bind Pol II, and to affect the activity of TFIIS. Aside from transcriptional elongation, CSB has a role in transcription-coupled nucleotide repair and base excision repair. Patients with CS exhibit defects in nucleotide excision repair (discussed at section 10/U.) (Licht et al, 2003).

L. Elongation factors with chromatin remodeling ability

In addition to the transcriptional pause and arrest, the requirement of Pol II to pass through a nucleosome represents a major block to transcriptional elongation. It is known that the chromatin structure within active genes is altered to increase DNA accessibility. Later, it became clear that there are histone modifiers which support Pol II transcription on chromatin templates in a reversible manner.

j) SWI/SNF

SWI/SNF (homothallic switching deficient/sucrose non-fermenting) is an ATP-dependent chromatin-remodeling complex. SWI/SNF can be recruited to a gene via interaction with a transcriptional activator (Narlikar et al, 2002).

It has been shown that SWI/SNF has a role in releasing promoter proximal paused Pol II in vitro, moreover SWI/SNF components were found in the body of transcribed genes, not only in the promoter regions (Armstrong et al, 2002; Subtil-Rodriguez & Reyes, 2010).

Based on these observations, it has been proposed that SWI/SNF is required to help RNAPII overcome a nucleosomal barrier in the body of the genes, in vivo. According to a recently established model, SWI/SNF is recruited to the arrested Pol II, and then the translocase activity of the complex might draw DNA into the nucleosome from one side, creating a loop that traverses or evicts the nucleosome. The transient destabilization of the interactions (pass of the loop) around the dyad would facilitate progression of RNAPII across this region and across the nucleosome (Subtil-Rodriguez & Reyes, 2011; Zhang et al, 2006).

k) Chd1

The ATPase CHD1 (chromo-ATPase/helicase-DNA binding domain) remodels nucleosomes *in vitro* and appears to function in both elongation and termination. Studies performed on polytene chromosomes in flies revealed that Chd1 associates with highly active sites of transcription (Stokes et al, 1996; Tran et al, 2000). Moreover, Chd1 physically associates with PAF, DSIF and FACT complex. A recent model proposes that Chd1 to regulate chromatin structure by preventing trans-histone exchange from taking place over coding regions (Smolle et al, 2012).

l) FACT and Spt6

There are two well known factors, FACT and Spt6 which are central to regulating chromatin structure during elongation. FACT and Spt6 have been shown experimentally to be important in maintaining chromatin structure disrupted by active transcription (Mason & Struhl, 2003).

The FACT (facilitates chromatin transcription) complex is a heterodimer comprised of hSpt16 and the HMG-box protein SSRP1 (Orphanides et al, 1999). During transcription FACT destabilizes the nucleosome by selectively removing one H2A/H2B dimer, thereby allowing Pol II to traverse a nucleosome.

Spt6 promotes nucleosome assembly *in vitro* and interacts with histones, preferentially histone H3. Spt6 specifically co-localizes to actively transcribed regions together with the elongating form of Pol II (Bortvin & Winston, 1996; Kaplan et al, 2000).

M. Elongation factors which maintain histone modifications

PAF complex was identified by its ability to associate with Pol II in yeast. It has five subunits: Paf1 (RNA polymerase-associated factor 1), Ctr9, Cdc73, Rtf1, and Leo1 (Krogan et al, 2002; Squazzo et al, 2002). The PAF complex has been demonstrated to cross-link throughout the entire length of genes, consistent with its functioning as an elongation factor (Krogan et al, 2002; Pokholok et al, 2002). PAF is central to maintain histone posttranslational modifications during transcription.

PAF is required for the recruitment of Set2 (Histone methyltransferase) to coding regions, which results in H3K36 methylation. It associates with the Set1 histone methyltransferase. Additionally, histone methylation at H3K4 and H3K79 depends on the PAF complex.

Components of PAF are also required for histone H2B monoubiquitination, a histone modification shown to be a prerequisite for H3K4 and H3K79 methylation (Briggs et al, 2002; Krogan et al, 2003; Wood et al, 2003).

H2B is deubiquitinated by Ubp8 component of SAGA complex therefore it attracts SAGA together with Elongator acetyltransferase complex, they may responsible for the permissive acetylation mark through the gene body (Rodriguez-Navarro, 2009; Wittschieben et al, 2000).

7. Transcription termination and 3' end processing

Transcription termination occurs when the transcribing Pol II releases the DNA template and the nascent RNA. Termination by Pol II does not occur at a certain conserved sequence or constant distance from the 3'-end of mature RNAs. In mammals, termination can occur within, and up to several kilobases downstream from the ORF (Proudfoot, 1989). It is long known that Pol II transcription termination is coupled to 3'-end processing of the

nascent RNA, and that an intact polyadenylation (poly(A)) signal is necessary for transcription termination of protein-coding genes in human and yeast cells (Buratowski, 2005; Connelly & Manley, 1988). Transcription termination serves many vital functions in gene expression. For example, as Pol II continues transcription after completing the pre-mRNA, termination prevents Pol II from interfering with downstream DNA elements, such as promoters, and promotes Pol II recycling. Biased termination leads to decreased splicing, increased rate of RNA degradation and reduced initiation at the promoter of the given gene.

N. 3' end processing of pre-mRNA

In mammals, cleavage and polyadenylation occur around 10–30 nt downstream from a conserved sequence: AAUAAA, and 30 nt upstream of a less conserved U- or GU-rich region. Once Pol II transcribes the AAUAAA signal, Pol II pauses and the poly(A) signal is recognized by the cleavage and polyadenylation specificity factors (CPSF), together with Cleavage stimulation factors (CstF).

CPSF contains five subunits: CPSF30, CPSF73, CPSF100, CPSF160, and Fip1. CstF contains three subunits: CstF50, CstF77, and CstF64 which directly recognizes the U/GU-rich element.

The endonucleolytic cleavage that precedes polyadenylation is carried out by CPSF-73 (Mandel et al, 2006). In addition, CPSF-73 is also likely responsible for cleavage of non-polyadenylated histone pre-mRNAs (Yang et al, 2009). CPSF160 directly recognizes AAUAAA, binds to and delivers CPSF-100, CstF-77, poly(A) polymerase (PAP), and Fip1 to the poly(a) signal (Kaufmann et al, 2004)(**Figure 11**).

The cleavage complex contains additional factors, like CFI (cleavage factor), which responsible for the regulation of poly(A) site selection and 3' end processing, CFII which mediates RNA kinase activity and symplekin which is a scaffold-like protein(de Vries et al, 2000; Takagaki & Manley, 2000). Once the poly(A) is enough long, it is bound by PABP (poly(A)-binding protein) which is required for correct poly(A) tail synthesis and protects the pre-mRNA from exonuclease degradation (Minvielle-Sebastia et al, 1997).

O. Different 3’ end processing and termination pathway on short, none-polyadenylated transcripts

Mammalian replication-dependent histone proteins are encoded by mRNAs that do not possess a poly(A) tail at their 3’-end, but rather a stem loop structure. The 3’-ends of histone mRNAs are generated by endonucleolytic cleavage 3’ to a dinucleotide CA, positioned downstream from a conserved stem–loop and upstream of a purine-rich histone downstream element (HDE). The conserved stem-loop is recognized and bound by SLBP protein (stem-loop binding protein), while HDE is recognized by the 5’-end of the U7 snRNA (Marzluff et al, 2008). The 3’ cleavage is most likely performed by CPSF-73. In addition to these, another factor is required for histone pre-mRNA processing (**Figure 11**). It is called Heat-Labile Factor (HLF), which is required for histone pre-mRNA cleavage and was found to contain polyadenylation factors, including CPSF subunits. Histone transcription continues only a few hundred nucleotides past the 3’-end of the mRNA and the termination of transcription is, like other protein-coding genes, dependent on 3’-end processing (Chodchoy et al, 1991).

Mammalian spliceosomal snRNAs (U1, U2, U4, and U5) are also intronless and non-polyadenylated RNAs. Formation of their 3’-ends depends on specific snRNA promoters but does not require a poly(A) site or HDE, but rather a 3’-box element located 9–19 nt downstream from the mature 3’-end of snRNAs. After endonucleolytic cleavage downstream from the 3’-box, snRNAs carry an extended 3’-end that is trimmed after their export to the cytoplasm (Egloff et al, 2008).

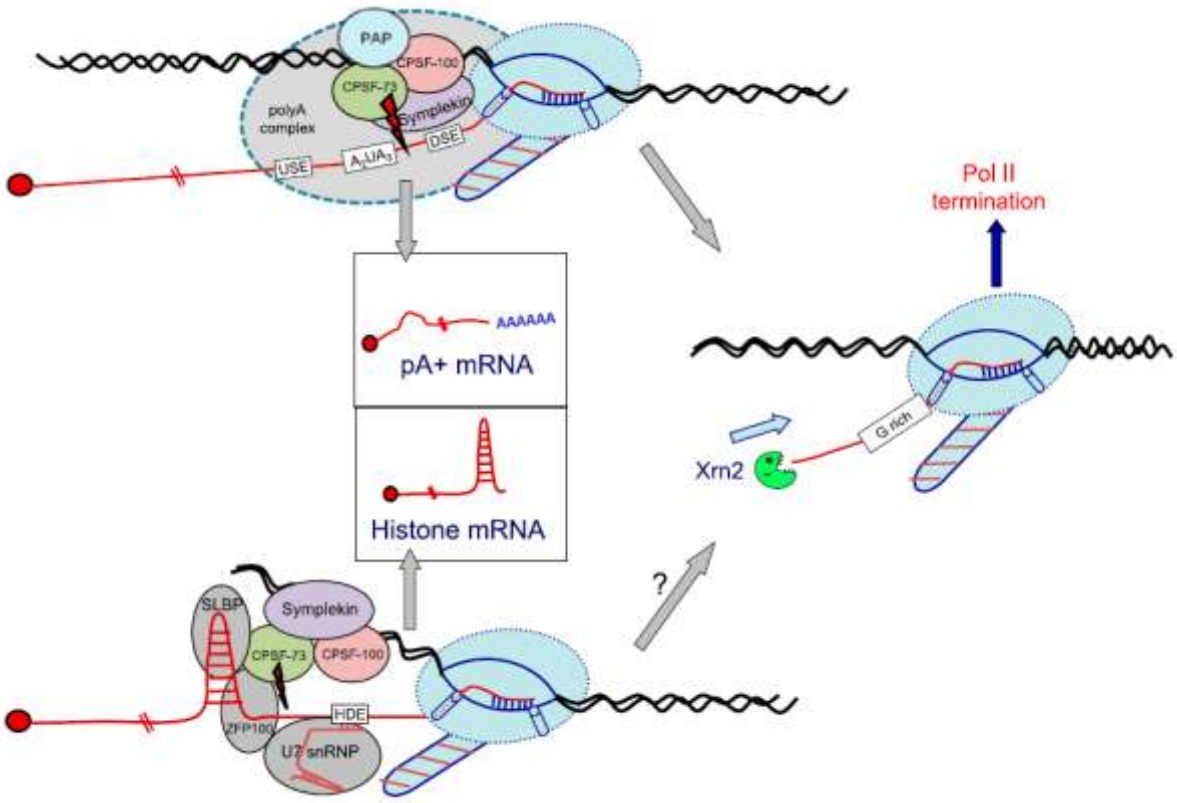


Figure 11: Comparing the mechanism of 3'-end formation polyadenylated mRNA versus none-polyadenylated histone mRNA. (RNA is in red, and DNA is in black). For poly(A)$^+$ mRNA, only factors conserved with histone pre-mRNA processing are shown. The rest of the poly(A) complex is depicted by a large grey shadow. Pol II (in light blue) is depicted with striped CTD, and the position of the ssDNA bubble is depicted with associated nascent RNA. Histone 3' processing factors are indicated. The cap at the 5' end of mRNA is depicted as a red ball. The histone mRNA 3' hairpin is shown, as is the U7 RNA hairpin and the interaction with the histone downstream element (HDE). Positions of cis RNA 3'-end processing sequence elements are shown in white boxes. mRNA products of poly(A)$^+$ and histone pre-mRNA 3'-end processing are shown (in boxes), as is the subsequent coupled termination of Pol II by Xrn2-mediated torpedo effects [established for poly(A)$^+$ genes, but only inferred for histone genes]. (I adopted the figure together with the complete figure legend from (Proudfoot, 2011) without any modification, as the legend is essential to fully understand the illustrated models).

P. Transcription termination I: Torpedo model

The role of 3' end processing in transcription termination was explained by several different models. One of the oldest, but strongly supported model is the "torpedo model". According to this theory, the cleavage after the poly(A) site creates an uncapped 5' monophosphate, which is an entry site for a 5'–3' exonuclease. This exonuclease (Rat1 in yeast/Xrn2 in mammals) follows Pol II and degrades the unnecessary RNA product (**Figure 12**). According to this model, it promotes Pol II release when it catches up and collides Pol II destabilizing the elongation complex (Connelly & Manley, 1988).

The torpedo model is supported by studies that demonstrated the role for the 5'–3' exonuclease in Pol II termination. These studies showed that inactivation of the exonuclease resulted in defects in termination, as measured by ChIP and/or nuclear run-on experiments (Kim et al, 2004; West et al, 2004).

XRN2 activity is likely not sufficient to force RNA and Pol II release from the template. Depletion of XRN2/Rat1 inhibits but does not abolish termination, and digestion of polymerase-associated RNA is not sufficient to disassemble the elongation complex *in vitro* (Kim et al, 2004; Kireeva et al, 2000). Therefore transcription termination may occur due to the combination of the mentioned model and additional events mentioned in a second proposed model, called allosteric/anti-terminator model.

Q. Transcription termination II: Allosteric/anti-terminator model

The allosteric model proposes that transcription through the poly(A) site triggers conformational changes and/or factor changes in the productive elongation complex, which facilitates its release from the DNA template. For example, by recruitment of the 3' processing factors, and/or negative elongation factors, or release of an anti-termination factor. The following changes would destabilize the complex, resulting in termination (Logan et al, 1987)(**Figure 12)**.

Termination induced by conformational change in the elongation complex after the poly(A) signal might be mediated by the CTD of Pol II (Rosonina et al, 2006).

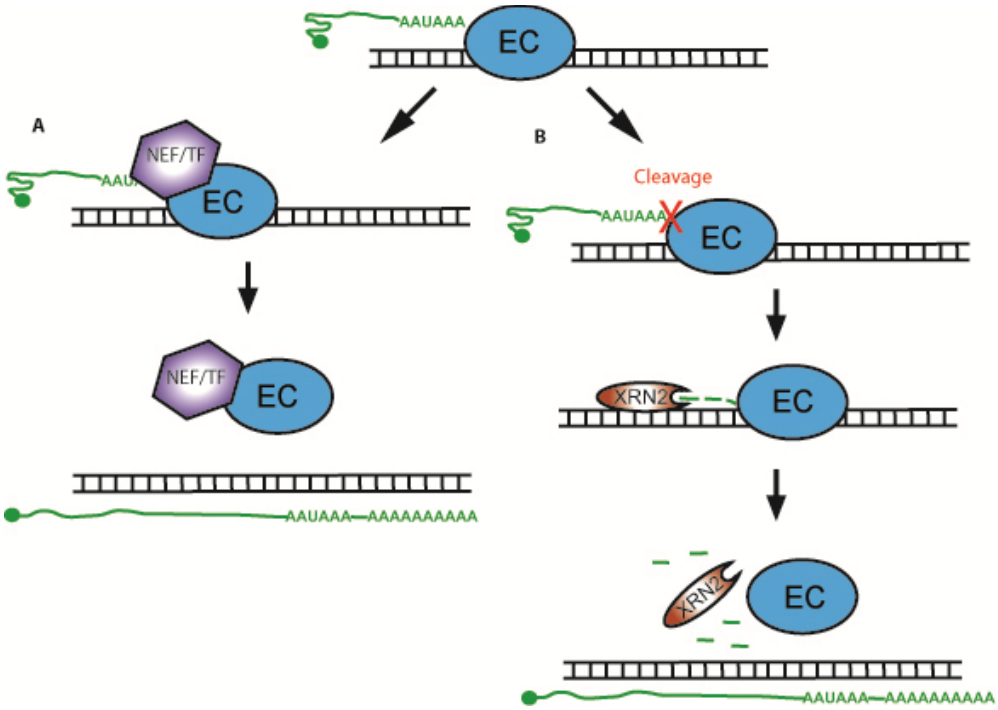


Figure 12: Comparing the (A) Allosteric/Anti-terminator model with the (B) Torpedo model. (**A**) The transcription through the poly(A) signal causes conformational changes and/or factor changes in the elongation complex (EC). Positive elongation factors may dissociate and termination factors/negative elongation factors may be recruited, which will facilitate the release of EC/Pol II from the DNA template. (**B**) Cleavage of the nascent transcript generates a new, uncapped 5' end that is targeted and degraded by Rat1/XRN2 nuclease. XRN2 follows the EC and upon "collision" EC releases the template DNA.

R. Chromatin remodeling factors at the 3' end of genes

Chromatin architecture plays an important role in gene expression. Chromatin remodeling coordinates elongation with termination and 3'-end processing. Most commonly, chromatin remodeling occurs via histone modifications or via the action of ATP-dependent remodeling complexes. ATP-dependent remodelers like SWI/SNF play crucial roles in transcription initiation and recently, evidence has accumulated which showing that these factors may also function in transcription termination. Yeast Chd1 was identified as a termination as deletion caused alterations in chromatin structure at the 3'-end of genes.

ChIP analysis confirmed the presence of the CHD/Mi2 and ISWI complexes at the terminator regions of yeast protein-coding genes. Additionally, an ATPase-deficient version of Isw1 resulted in a termination defect at the 3'-end of protein-coding genes (Alen et al, 2002; Morillon et al, 2003).

These data show that chromatin remodeling is important for termination and suggest that Pol II pause and termination requires presence of remodelers.

8. Pol II pause downstream from 3'-end of genes

ChIP and genome-wide analyzes of human genes frequently shows a high density of Pol II at promoter regions that decreases throughout the gene body and increases downstream from the poly(A) site (Glover-Cutter et al. 2008). This increased Pol II density at the 3'-end may reflect pausing or slowed down Pol II that undergoes termination. These long, sometimes up to 6 kb long Pol II pauses were first suggested to be due to transcription pause sites located downstream from poly(A) sites, which appear to play a role in Pol II 3'-end processing and termination, at least in some mammalian genes (Enriquez-Harriset al. 1991; Plant et al. 2005). These pause sites might be G-rich sequences like the MAZ element (G5AG5), which was identified downstream from the poly(A) site of the liver-specific C2 complement gene (Ashfield et al. 1994; Yonaha and Proudfoot 1999, 2000). This sequence is a potential binding site for the transcription factor MAZ (MYC-associated zinc-finger protein) that was suggested to promote poly(A)-dependent termination. Additionally, it has been suggested that the poly(A) signal itself can direct Pol II pausing that is mediated by CPSF, and the strength of the poly(A) site correlates with the efficiency of pausing-dependent termination (Plant et al. 2005; Gromak et al. 2006).

These observations, combined with the two termination model (torpedo/allosteric) raised a new model (**Figure 13**) which suggest that the poly(A) signal promotes Pol II pause by turning CPSF ,into a negative elongation factor. CPSF might travel with Pol II as a part of the elongation complex (Calvo & Manley, 2003; Dantonel et al, 1997), or can be recruited upon the AAUAAA sequence. Then, on the slowed down/paused Pol II, CPSF would recruit CstF in the presence of Ser-2 phosphorylated CTD. Binding of CstF is followed by the appearance of CFII. Once the cleavage apparatus is mature enough, it will cut the 3' end of the nascent mRNA. The upstream RNA product will be polyadelyated and the downstream RNA with uncapped 5' terminus will be an entry point for the XRN2 exonuclease. XRN2 degrades the unnecessary RNA product and follows Pol II that is slowed down by the cleavage factors and the G-rich sequences until they collide. Upon collision (which is rather

stochastic, and explains the long Pol II densities), the unstable Pol II will release the template and the nascent mRNA (Nag et al, 2006; Nag et al, 2007).

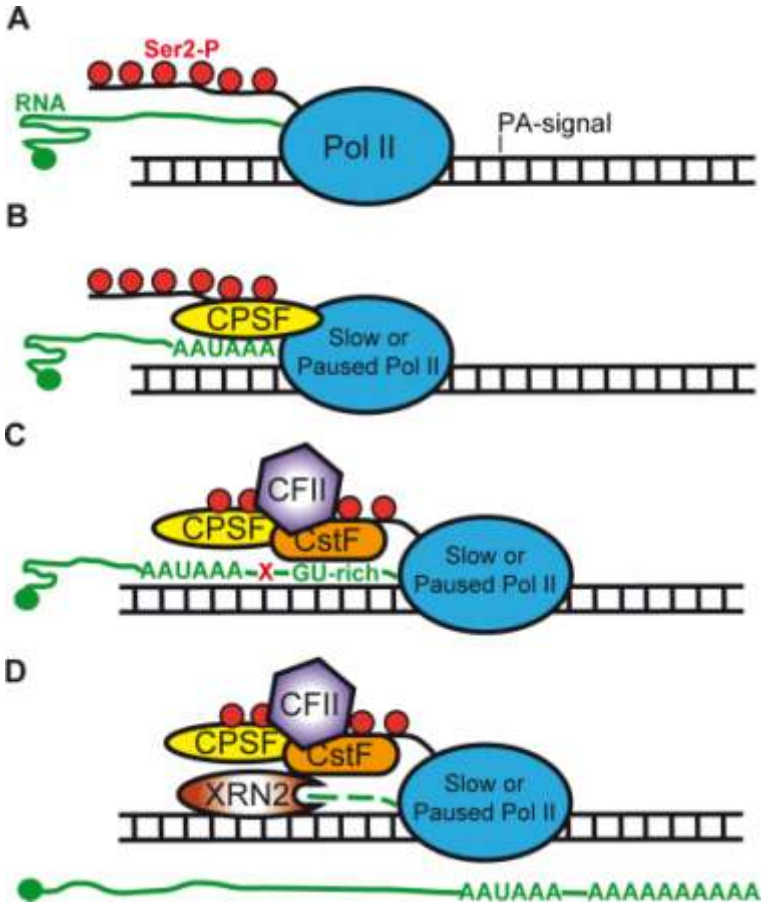


Figure 13: Suggested model for Pol II transcription termination on polyadenylated genes. When a(**A**) Pol II elongation complex with decreased Ser5-P and increased Ser2-P signal on its CTD transcribes the poly(A) signal (PA). (**B**) The exposed AAUAAA sequence and the Pol II body is bound by CSPF and slows down Pol II as a negative elongation factor. (**C**) CPSF recruits CstF and CFII with the help of Ser2-P CTD of Pol II. Upon exposure of the GU-rich binding site, CstF dislodges CPSF which cuts the nascent RNA between AAUAAA and the GU-rich sequence (red X). Following cleavage, (**D**) XRN2 degrades the downstream RNA products, and upon "collision" with Pol II it may facilitate transcription termination.-

Pol II terminates differently on genes coding for small non-coding transcripts and non polyadenylated transcripts which can be due to the different 3' end processes. In yeast the Pol II termination on these short transcripts is suggested to be mediated on the R-loops formed during transcription by the cooperative function of Nab3 (nuclear polyadenylated RNA-binding 3) Sen1 (putative RNA and DNA helicase) and Nrd1(RNA-binding protein). In this model instead of Rat1 collision, Sen1 mediates termination by unwinding the RNA-DNA hybrid in the active center of Pol II (Kuehner et al, 2011).

RESULTS I.

III. Genome-wide characterization of Pol II pause downstream from 3' end of genes

As illustrated above, many efforts have been made to fully understand Pol II transcription termination however, there are still many missing details to build a final model, which explains how Pol II finish transcription. It seems that transcription termination cannot be handled as an individual and independent process as more and more evidence links it to other nuclear processes. Furthermore, the diversity of transcripts that Pol II synthesizes and the continuous need for re-initiation have probably placed a significant pressure on the cell to evolve alternative and redundant termination mechanisms as it is the case for short non-coding transcripts. This idea, together with the powerful evidences showing links between Pol II pause, pre-mRNA 3' end processing and termination urge the genome wide investigations of transcriptional events downstream from 3' end of genes.

Genome wide studies in mammalian systems were mainly focusing on early transcriptional activities, such as initiation and promoter proximal pausing. Therefore to better understand the Pol II behavior during termination, I carried out a genome wide project to characterize Pol II pause genome wide downstream of 3'end of genes.

The results of these studies are presented and discussed in the two following publications:

IV. Publication I.

OPEN ACCESS Freely available online

PLoS one

RNA Polymerase II Pausing Downstream of Core Histone Genes Is Different from Genes Producing Polyadenylated Transcripts

Krishanpal Anamika[1,2☯], Àkos Gyenis[1☯], Laetitia Poidevin[2], Olivier Poch[2], Làszlò Tora[1]*

1 Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 964, Université de Strasbourg, Illkirch, France, **2** Department of Structural Biology and Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 964, Université de Strasbourg, Illkirch, France

Abstract

Recent genome-wide chromatin immunoprecipitation coupled high throughput sequencing (ChIP-seq) analyses performed in various eukaryotic organisms, analysed RNA Polymerase II (Pol II) pausing around the transcription start sites of genes. In this study we have further investigated genome-wide binding of Pol II downstream of the 3′ end of the annotated genes (EAGs) by ChIP-seq in human cells. At almost all expressed genes we observed Pol II occupancy downstream of the EAGs suggesting that Pol II pausing 3′ from the transcription units is a rather common phenomenon. Downstream of EAGs Pol II transcripts can also be detected by global run-on and sequencing, suggesting the presence of functionally active Pol II. Based on Pol II occupancy downstream of EAGs we could distinguish distinct clusters of Pol II pause patterns. On core histone genes, coding for non-polyadenylated transcripts, Pol II occupancy is quickly dropping after the EAG. In contrast, on genes, whose transcripts undergo polyA tail addition [poly(A)$^+$], Pol II occupancy downstream of the EAGs can be detected up to 4–6 kb. Inhibition of polyadenylation significantly increased Pol II occupancy downstream of EAGs at poly(A)$^+$ genes, but not at the EAGs of core histone genes. The differential genome-wide Pol II occupancy profiles 3′ of the EAGs have also been confirmed in mouse embryonic stem (mES) cells, indicating that Pol II pauses genome-wide downstream of the EAGs in mammalian cells. Moreover, in mES cells the sharp drop of Pol II signal at the EAG of core histone genes seems to be independent of the phosphorylation status of the C-terminal domain of the large subunit of Pol II. Thus, our study uncovers a potential link between different mRNA 3′ end processing mechanisms and consequent Pol II transcription termination processes.

Citation: Anamika K, Gyenis À, Poidevin L, Poch O, Tora L (2012) RNA Polymerase II Pausing Downstream of Core Histone Genes Is Different from Genes Producing Polyadenylated Transcripts. PLoS ONE 7(6): e38769. doi:10.1371/journal.pone.0038769

Editor: Nicholas S. Foulkes, Karlsruhe Institute of Technology, Germany

Received December 13, 2011; **Accepted** May 13, 2012; **Published** June 11, 2012

Funding: AG is supported by a fellowship from Fondation pour la Recherche Médicale (FRM) France. This work was supported by grants from the EU (European Transcriptome Regulome & Cellular Commitment Consortium EUTRACC, LSHG-CT-2007-037445 and EPIDIACAN), SkinChroma-CNRS and the Institut National du Cancer (INCA; 2008 Ubican) to LT and Agence nationale de la recherche (ANR-09-BLAN-0266) to LT and OP. 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.

* E-mail: laszlo@igbmc.fr

☯ These authors contributed equally to this work.

Introduction

RNA polymerase II (Pol II) transcription is a highly regulated process that requires coordinated action of numerous transcription factors. It can be divided into initiation, promoter escape, elongation and termination phases. Transcription initiation is a complex series of ordered multistep process that involves the recruitment of Pol II to a promoter, local melting of the DNA around the transcription start site (TSS), and formation of the first few phosphodiester bonds of mRNA. Recognition of promoters begins with the assembly of a large protein complex, called the preinitiation complex (PIC), containing Pol II and multiple general transcription factors (GTFs) on the promoter [1]. Recruitment of Pol II to eukaryotic gene promoters by specific transcription factors is a key regulatory step in transcription initiation [2,3].

Several studies analysed Pol II binding at the promoter and its entry in transcription genome-wide by using chromatin immunoprecipitation (ChIP) assay coupled either to genomic DNA microarrays (ChIP-chip) or to high throughput sequencing (ChIP-seq). These reports uncovered a significant fraction of genes that have high levels of Pol II density at or around their TSSs when compared to the transcribed region of the genes [4,5,6,7,8,9,10]. Many of these genome-wide experiments demonstrated that these high peaks, reflecting Pol II binding accumulation at promoters, were centered around a region 20–50 nucleotides downstream of the TSSs of the genes. Thus, it is now accepted that these high Pol II densities or occupancy signals reflect Pol II pausing at promoter-proximal sites of many transcribed, but also non-transcribed, genes in different organisms. Promoter proximal pausing (PPP) of Pol II was first described at the *Drosophila* heat shock gene (*HSP70*) promoters, and shown that Pol II stalls 20–50 nucleotides downstream from the TSS [11,12]. PPP was once considered a rare phenomenon, but the recent genome-wide reports have demonstrated that it is a common and widespread regulatory step in eukaryotic Pol II transcription [6,7,8,9,13]. In some of these earlier analyses, Pol II pausing has

been observed at the 3′ end of the genes however none of these genome-wide Pol II binding studies analysed and characterized in details Pol II binding and pausing around the 3′ of the transcription units.

It is long established that 3′-end processing is absolutely required for transcription termination [14]. Transcription termination is defined as the cessation of RNA synthesis and release of Pol II from its DNA template [15]. The analysis of transcription termination event carried out on few model genes showed that termination is dependent on polyadenylation (polyA) signals and downstream terminator sequences [16]. Pol II has been suggested to terminate transcription at sites positioned between 100 bp and several kbps downstream of the 3′ end of the annotated genes (EAGs) [17,18,19,20,21]. Downstream of the EAGs two classes of terminator sequences have been identified in human genes: G-rich transcription pause sites and co-transcriptionally cleaved (CoTC) RNA sequences [16,21,22,23]. Earlier reports suggest a link between transcription termination and 3′ end processing of Pol II transcripts in which two general models were put forward: the ''allosteric'' or ''anti-terminator'' model and the ''torpedo'' model [24,25]. The allosteric model proposes that the presence of a polyadenylation sequence on the RNA triggers a change in the factors associated with the polymerase [26]. In this model, binding of the cleavage and polyadenylation specificity factors (CPSFs) and the cleavage stimulation factors (CstFs) to the AAUAAA polyadenylation signal on the nascent pre-mRNA favors transcription termination by displacing elongation factors and consequently rendering Pol II less processive [24,27,28,29]. In the torpedo model, the cleavage event at the polyadenylation site generates a new 5′ end [30]. Unlike the capped 5′ end of the pre-mRNA, this extremity could act as an entry point for an activity (such as Xrn2 exonuclease or a helicase) that would track along the RNA and dissociate the polymerase from the DNA template. A recent work has shown that the 5′–3′ exonuclease Xrn2 in human and Rat1 in yeast are required for efficient termination [31,32]. The 3′-end processing coupled termination of Pol II by Xrn2-mediated torpedo effects are established for poly(A)$^+$ genes, but only inferred for histone genes [14]. However, Luo et al. (2006) reported that neither the ''anti-terminator'' nor the ''torpedo'' model is sufficient to cause Pol II termination, instead a termination mechanism, which is a hybrid of the two models, may exist [33].

Besides the existence of a link between pre-mRNA 3′ end processing and termination, Pol II pausing has also been implicated in promoting transcription termination. Studies in mammalian cells indicate that termination can be separated into two steps: pausing and polymerase release [34]. Recently, a study performed on the β-globin/β-actin reporter gene suggested that R-loops formed by DNA/RNA hybrids near G-rich pause sites, downstream of polyA signals may be involved in transcription termination [35]. In contrast, other studies argue that Pol II pausing is exclusively a function of polyadenylation signals and does not require any additional elements or pausing sites in the DNA sequence [36,37,38,39]. In mammalian systems, the main focus of genome-wide study with respect to pausing was PPP and pause/stalling at or 3′ of EAGs has been studied just for few individual genes [40,41]. However, the full spectrum of genes regulated by pausing and different pausing profiles at or downstream of EAG has not yet been investigated at a genome-wide level. Thus, it is not clear whether this phenomenon is commonly occurring among mammalian genes.

To characterize Pol II pausing downstream of the EAG of the transcription units in human MCF7 cells genome-wide, we used high-resolution occupancy profiling by ChIP-seq of Pol II in human cells. Comparison of our ChIP-seq results with (i) recently published Gro-seq data [42], where the authors mapped genome-wide the position, amount, and orientation of transcriptionally engaged Pol II in MCF7 cells, and (ii) published ChIP-seq data from mouse embryonic stem cell (ESCs) with several different forms of Pol II [5,43] indicates that active Pol II pauses genome-wide downstream of the EAGs in mammalian cells. Interestingly, Pol II occupancy downstream of the EAGs can be detected up to 6 kb from the EAGs, but core histone genes, which encode non-polyadenylated transcripts, show a very narrow Pol II pause downstream of their EAGs.

Results

The accumulation of both Pol II and the corresponding transcripts downstream of the EAGs is a genome-wide feature of Pol II transcription

To better understand Pol II behavior at the 3′ end of the transcription units and in transcription termination, we have analysed the patterns and profiles of Pol II binding downstream of EAGs in differentiated human cells. To this end, we have generated ChIP-seq data for Pol II from the MCF7 human breast cancer cell line (GSE34001) by using an antibody that binds to the N-terminus of the largest subunit of Pol II (N-20; Santa Cruz, H-224X). This antibody allows the detection of Pol II independently of the phosphorylation status of the C-terminal domain (CTD) of its largest subunit. To test for non-specific binding, we carried out a control ChIP-seq using an antibody raised against a yeast factor that does not recognize any human proteins (Mock; GSE34001). Sequencing reads were mapped to human genome, and uniquely mapped reads were considered for further analyses. To avoid the overlapping of Pol II occupancy signals downstream from the EAGs with signals coming from neighboring transcription units, only those human refseq genes (13787 genes) were analysed, which were at least 4 kb away from the neighboring transcription units. Pol II tag density −/+4 kb around the gene body and in the transcribed regions was then calculated. The average tag density calculated on 13787 genes, which do not have genes in −/+4 kb neighboring region, shows that Pol II binding profiles are relatively low in the transcribed regions, but higher at both, around the TSSs and 3′ from the EAGs (Figure 1A). These data suggest that genome-wide Pol II pauses not only at the TSSs, as previously reported [40], but also downstream of the EAGs (Figure 1A). Furthermore, a comparison of the Pol II ChIP-seq data with publicly available Gro-seq data demonstrates that transcriptionally engaged Pol II is present not only in the transcription units [42], but also downstream of the EAGs (Figure 1B). The genome-wide comparison of both, the Pol II occupancy and the corresponding transcript production mapped by Gro-seq around the EAGs (−500 bp to +4 kb from EAG; Figure 1C) of 500 highly expressed genes (Table S1) shows that indeed transcriptionally active Pol II is bound in the regions downstream from the EAGs. Interestingly, in these regions all the transcripts identified by Gro-seq are mapping in the sense orientation 3′ from the EAGs suggesting that they are produced by Pol II molecules that have been transcribing pre-mRNAs. As observed previously, by analyzing the transcriptionally engaged Pol II obtained by Gro-seq [44], we did not detect any significant antisense transcription 3′ from the EAGs of the genes. Taken together our analysis, in good agreement with previous studies [5,9,44], suggests that the accumulation of Pol II downstream of the EAGs is a genome-wide feature of Pol II transcription.

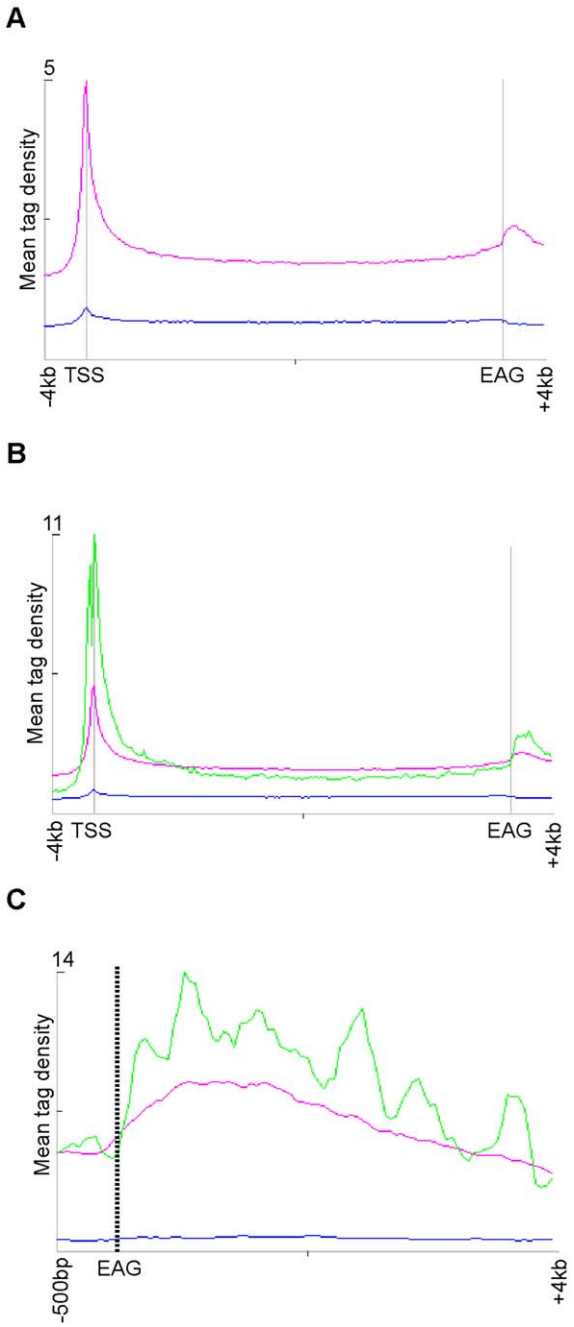


Figure 1. Pol II pausing and the corresponding transcripts downstream of the EAGs are genome-wide features of Pol II transcription termination. Chip-seq data using an anti-Pol II antibody (N-20, our study) and Gro-seq [42] was carried out using human MCF7 cells. **A**) Mean tag densities of ChIP-seq data of Mock (Blue) and Pol II (Pink) in average genes and −/+4 kb around them are represented. Pol II enrichment density of 13787 "non-overlapping and isolated" refseq genes were calculated. TSS: transcription start site; EAG: end of annotated gene. **B**) Mean tag densities of Mock (Blue) and Pol II ChIP-seq (Pink) and Gro-seq (Green) data on average 13787 "non-overlapping and isolated" refseq genes, and −/+4 kb upstream and downstream are represented. **C**) Mean Pol II tag density from the ChIP-seq of Mock and Pol II (Blue and Pink, respectively) and Gro-seq (Green) data in the region −500 bp to +4 kb around the EAG of 500 highly expressed genes (Table S1). Note that all the Gro-seq RNA reads map in the sense orientation when compared to the pre-mRNA.
doi:10.1371/journal.pone.0038769.g001

Different patterns of genome-wide Pol II pausing downstream of the EAGs

As in our initial analysis (Figure 1A and B), both expressed and non-expressed genes were analysed to characterize Pol II pausing, we next subdivided the non-neighboring 13787 isolated genes relative to their i) 3′ Pol II pause densities and ii) expression. Pol II enrichment signals −/+4 kb around the EAGs were used for K-means clustering (Figure S1A) [45]. Average Pol II tag density −/+4 kb of EAGs was calculated (Figure S1B) and average gene density profiles, considering regions −/+4 kb upstream and downstream of the transcription unit were also generated (Figure S1C). The identified Pol II enrichment patterns in combination with gene expression data allowed us to subdivide the non-neighboring 13787 refseq genes in two clusters (Figure S1A). These data together with the analysis of the expression of the genes in the two clusters show that genes in Cluster 2 are not or very weakly expressed as compared to genes in the Cluster 1 (Figure S1D). Thus, in order to study Pol II occupancy only at transcribed genes we analysed the 3495 expressed genes (Cluster 1 in Figure S1) by re-clustering them (Figure 2). Using K-means clustering and seqMINER, a ChIP-seq data interpretation platform [45], four new clusters based upon Pol II tag densities and patterns were generated. Distinct patterns of Pol II binding profiles downstream of the EAGs can be divided as follows: narrow (Cluster H), very broad (Cluster PA1) and broad pause (Cluster PA2) (Figure 2A and B). Interestingly, gene ontology (GO) analyses of members in these categories indicated that genes in the narrow cluster are almost exclusively core histone genes, which are intronless, code for non-polyadenylated transcripts and involved in replication-dependent nucleosomal assembly (Table 1), hereafter called Cluster H. Note that the other three categories contained genes coding principally for polyadenylated (PA) transcripts, thus these clusters are called PA1-3. In the narrow peak-containing Cluster H, Pol II occupancy signals decrease very rapidly after the EAG. In contrast, in Cluster PA1 and PA2, Pol II occupancy is quite widespread and can be detected until 4–6 kb downstream from the EAGs (Figure 2B). The difference between PA1 and PA2 seems to be related to gene expression as genes belonging to Cluster PA1 are higher expressed than those belonging to Cluster PA2 (Figure 2C) which is in good agreement with their higher Pol II occupancy downstream of the EAGs (Figure 2B). Genes in the Cluster PA3, which have less (or very less) Pol II enrichment, also show detectable Pol II occupancy downstream of the EAGs (Figure 2B).

To validate the bioinformatically isolated Pol II pause patterns downstream of the EAG (shown in Figure 2A and B), we have carried out ChIP followed by quantitative PCR detection (ChIP-qPCR). To this end we have randomly chosen two genes from each cluster and designed primer pairs to detect Pol II occupancy downstream from their EAGs (Figure 2D and E). For Cluster H, which contains mainly histone genes, primers were designed about 100–300 bps and 1.5–2 kb downstream from EAGs (I and II in Figure 2D, upper most panel; Figure 2E left panel). For Clusters PA1-3 primer pairs were chosen on the selected genes to detect peaks in regions approximately 0.5–1 kb, 2–3 kb and 4–5 kb downstream of the EAGs (see I, II and III respectively in Figure 2D; Figure 2E right panel). These validation experiments confirmed the bioinformatically defined Pol II occupancy patterns in each of the four clusters. Further in good agreement with the bioinformatics analysis of ChIP-seq, on the narrow peak containing histone cluster (Cluster H), ChIP-qPCR using primer pair II (situated 1.5–2 kb from EAG) did not amplify any significant product, indicating that Pol II occupancy is rapidly dropping after the 3′end of histone genes. In contrast, on the other three clusters, ChIP-qPCR confirmed that Pol II occupancy downstream of the

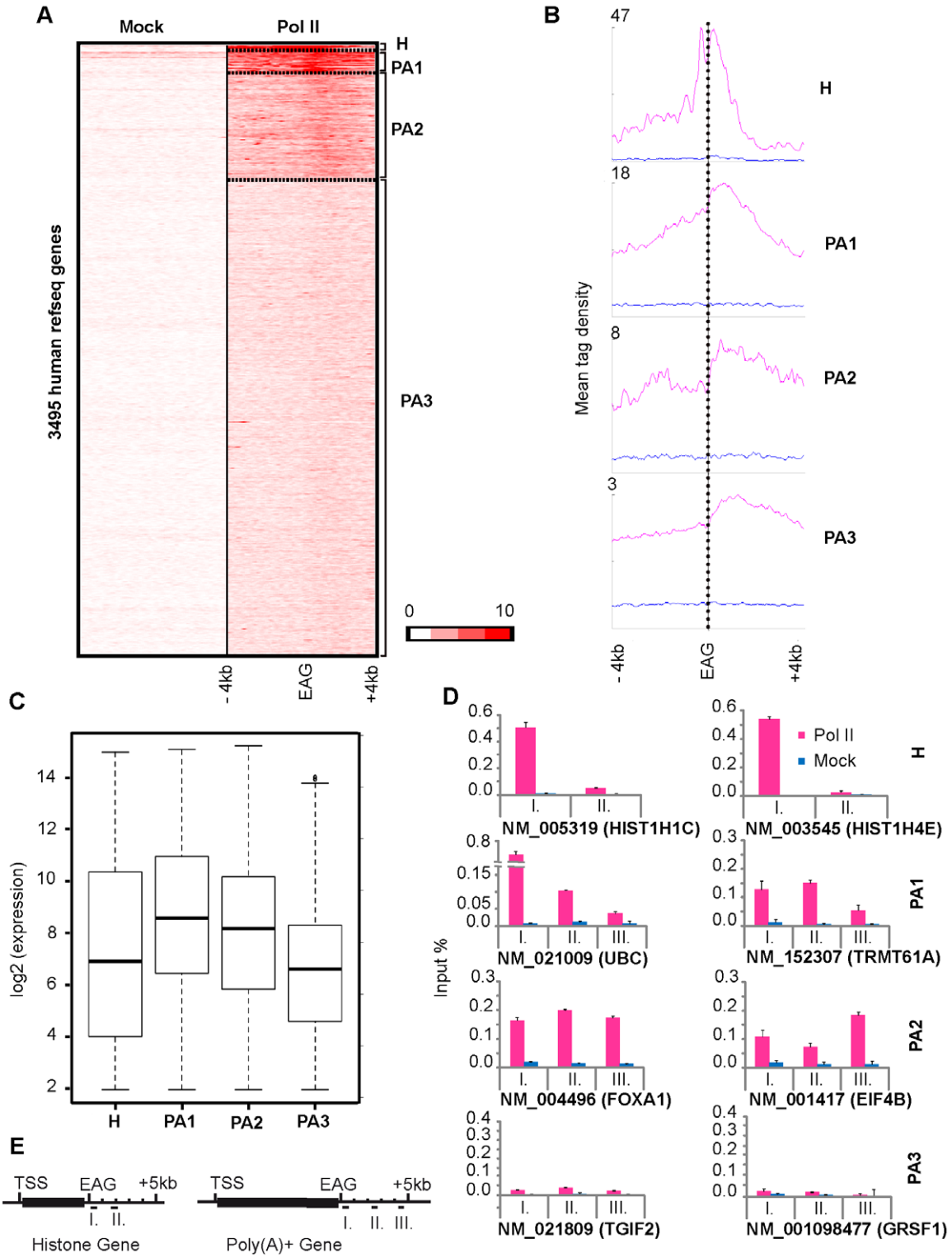


Figure 2. Genome-wide Pol II pauses with different patterns downstream of the EAGs. Clustering of genes, which have relatively high Pol II enrichment 3′ of their EAGs and high microarray expression value (considering genes from Cluster 1 of Figure S1) generates four clusters: H, PA1, PA2 and PA3. Total number of non-redundant refseq genes is 3495. Number of genes (or n) in each cluster is: Cluster H, n = 39; Cluster PA1, n = 74; Cluster PA2, n = 492 and Cluster PA3, n = 2890. **A**) Heatmap generated after K-means clustering of Mock and Pol II reads in the regions −/+4 kb upstream and downstream of the EAG. Color scale indicates the level of enrichment. **B**) Mean tag densities of Mock (Blue) and Pol II (Pink) on genes −/+4 kb upstream and downstream of the EAG in each cluster. **C**) The distribution of the expression levels of genes belonging to the H, PA1, PA2 and PA3 clusters (see A and B) is displayed by Whisker plot. The plots represent relative mRNA expression level of each cluster. The median is indicated with a horizontal line in each box showing that genes in Cluster PA1 have higher relative mRNA expression level than Cluster PA2 and PA3. **D**) ChIP-qPCR validation of the ChIP-seq data on two randomly selected genes (refseq ids and gene names are given) from each cluster. Pol II occupancy (Pink bars) compared to the mock (Blue bars) at different distances downstream from the EAGs are represented in input %. Distances from EAGs on the

indicated genes: Cluster H: I: +0.1–0.3 kb, II: +1.5–2 kb; Cluster PA 1–3: I: +0.5–1 kb, II: +2–3 kb, III: +4–5 kb. Error bars represent +/− standard deviations. **E**) The locations of oligonucleotides, which were used to validate Pol II pause profiles, are represented schematically.
doi:10.1371/journal.pone.0038769.g002

EAG in general could be detected on a large region, often covering 4–6 kb downstream of the EAGs. Thus, similarly to PPP (see Introduction), we interpret these relatively high Pol II binding signals downstream of genes as Pol II pausing. Moreover, these experiments suggest that Pol II pausing 3′ from genes is different on genes from which the transcribed pre-mRNA is polyadenylated (Clusters PA1 to 3), or not (Cluster H). This genome-wide observation is in good agreement with previous suggestions based on single gene analyses [40] and may suggest a possible link between Pol II pausing and the 3′ end processing of the corresponding transcripts.

Pol II pausing downstream from the EAGs on highly expressed genes

Next we analysed Pol II occupancy on 100 highly expressed poly(A)$^+$ genes (HEPA), independently from the distance of the next neighboring gene(s), and found that on the majority of highly expressed genes Pol II pausing downstream of the EAG sites is similar to that described above in Cluster PA1 (see Cluster HEPA1, Figure 3A and B). However, a small subset of highly expressed genes (about 10%) have high Pol II enrichment at their TSSs, but no or very little Pol II occupancy downstream of their EAGs (Cluster HEPA2 of Figure 3A and B).

To validate Pol II occupancy on genes in Cluster HEPA1 and HEPA2 we have carried out ChIP-qPCR. To this end we have compared Pol II occupancy downstream of the EAGs on two randomly chosen highly expressed genes with high 3′ peak with those which have no or only very low 3′ peaks (Figure 3C and D). For genes belonging to each cluster, primer pairs were chosen to detect peaks both at their TSSs and in regions approximately 0.5–1 kb, 2–3 kb and 4–5 kb downstream of the EAGs (I, II and III, respectively in Figure 3E). These ChIP-qPCR experiments confirmed the absence of significant Pol II occupancy downstream of the EAGs in a small fraction of the highly expressed genes suggesting that on a smaller subset of genes Pol II pausing at the 3′ end of the genes might be differentially regulated.

Pol II pausing downstream from the EAGs of core histone genes is different from those transcription units producing polyadenylated transcripts

Replication dependent core histone genes are intronless and coding for mRNAs that have a different 3′ processing mechanism than mRNAs transcribed from poly(A)$^+$ genes. Instead of polyA tail addition, the cleavage-only mRNA 3′-end formation of core histone genes involves stem-loop formation, the U7 snRNP, hairpin-binding protein and specific components of the cleavage/poly(A) complex [46,47,48,49,50]. As described in our above analysis, looking for genes that have no neighboring genes in the 4 kb vicinity, the narrow Pol II peak-containing Cluster H contained mainly core histone genes, prompted us to map Pol II profile in and around all the known histone genes from the human genome. As core histone genes are often found in clusters we have analysed Pol II occupancy only −/+1 kb upstream and downstream of histone genes. These new K-means clustering and profiling analyses of Pol II 1 kb around all histone genes show a high Pol II enrichment throughout the gene body and a sharp drop in the Pol II occupancy 3′ of the EAGs of core histone genes (Cluster H1 of Figure 4A, B). In contrast, Pol II occupancy on genes encoding non-replication dependent histone variants (which generally have introns, are weaker expressed than core histone genes and their transcripts undergo polyadenylation) is different from the core histone genes (Cluster H2 of Figure 4A, B). On the variant histone genes we did not observe the sharp drop of Pol II occupancy downstream of the EAGs (Figure 4B) as in the core histone genes, suggesting that 3′ end processing on core histone genes involving stem-loop formation may be in favor of a rapid Pol II release. Moreover, our results further suggest that the polyadenylation of variant histone transcripts may participate in Pol II pausing on the corresponding variant histone genes before transcription termination may occur.

As core histone genes are intronless, we have analysed Pol II occupancy on all intronless genes from the human genome to test whether the narrow Pol II pause profile downstream of the EAG was characteristic of core histone genes or, rather, a common feature of genes producing transcripts that do not undergo splicing. Interestingly, the sharp drop of Pol II occupancy occurred only downstream of EAGs of core histone genes suggesting that the narrow Pol II pausing pattern 3′ of EAGs is the characteristic of core histone genes (Figure 5). Taken together these results with the above qPCR validation results, suggest that Pol II pausing downstream from the EAGs of core histone genes is different from those transcription units producing polyadenylated transcripts.

Table 1. Gene Ontology (GO) terms (at significant P-values), associated with the clusters represented in Figure 2.

Cluster	Number of genes	Top GO biological process	P-value
Cluster H	39	Nucleosome assembly	1.48E-012
Cluster PA1	74	Structural constituent of cytoskeleton	6.90E-003
		Translational elongation	4.68E-002
		Cell morphogenesis involved in neuron differentiation	4.68E-002
Cluster PA2	492	Translational elongation	6.15E-032
Cluster PA3	2890	Modification dependent protein catabolic process	5.10E-010
		Chromatin modification	2.01E-009

doi:10.1371/journal.pone.0038769.t001

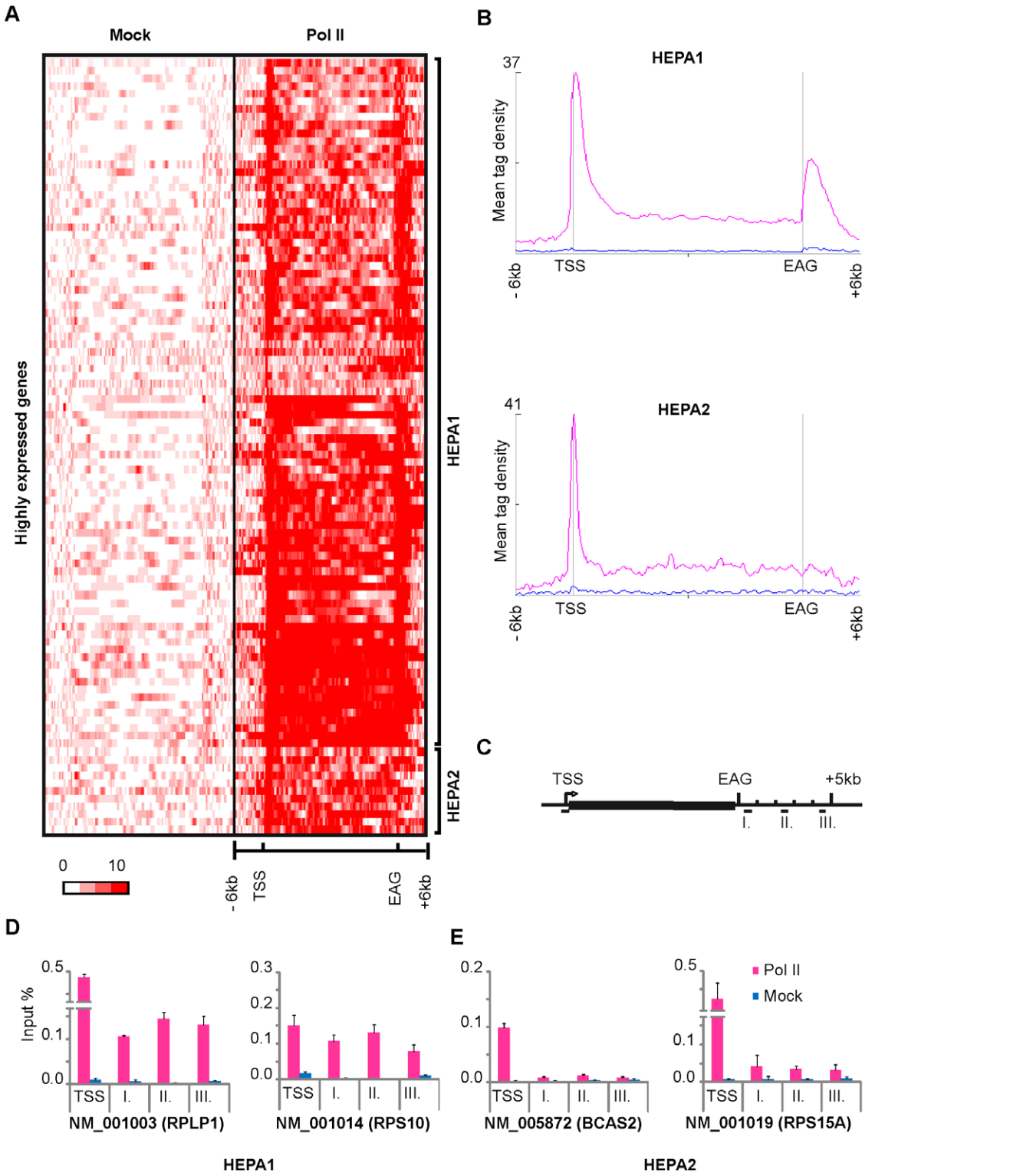


Figure 3. Pol II pause on highly expressed genes. K-means clustering of Mock (blue) and Pol II (pink) reads on 100 highly expressed genes from MCF7 cells (for exact gene names see Table S2) mainly generated two distinct clusters in terms of Pol II occupancy at the corresponding EAGs. **A**) Heatmap generated after the K-means clustering of Mock (Blue) and Pol II (Pink) reads in average gene body and −/+6 kb upstream and downstream of the genes. Color scale indicates the level of enrichment. **B**) Mean tag densities of Mock (Blue) and Pol II (Pink) signals on the two clusters of genes and −/+6 kb upstream and downstream of the gene body. **C**) The locations of oligonucleotides, which were designed to validate Pol II pause profile, are represented schematically. **D, E**) ChIP-qPCR validation of the ChIP-seq data for two randomly selected genes from each cluster (as indicated). Pol II occupancy (Pink bars) compared to mock (Blue bars) on the TSS and at different distances downstream from the EAGs are represented in input %. Distances from EAGs of the indicated genes: I: +0.5–1 kb, II: +2–3 kb, III: 4–5 kb. Error bars represent +/− standard deviations.
doi:10.1371/journal.pone.0038769.g003

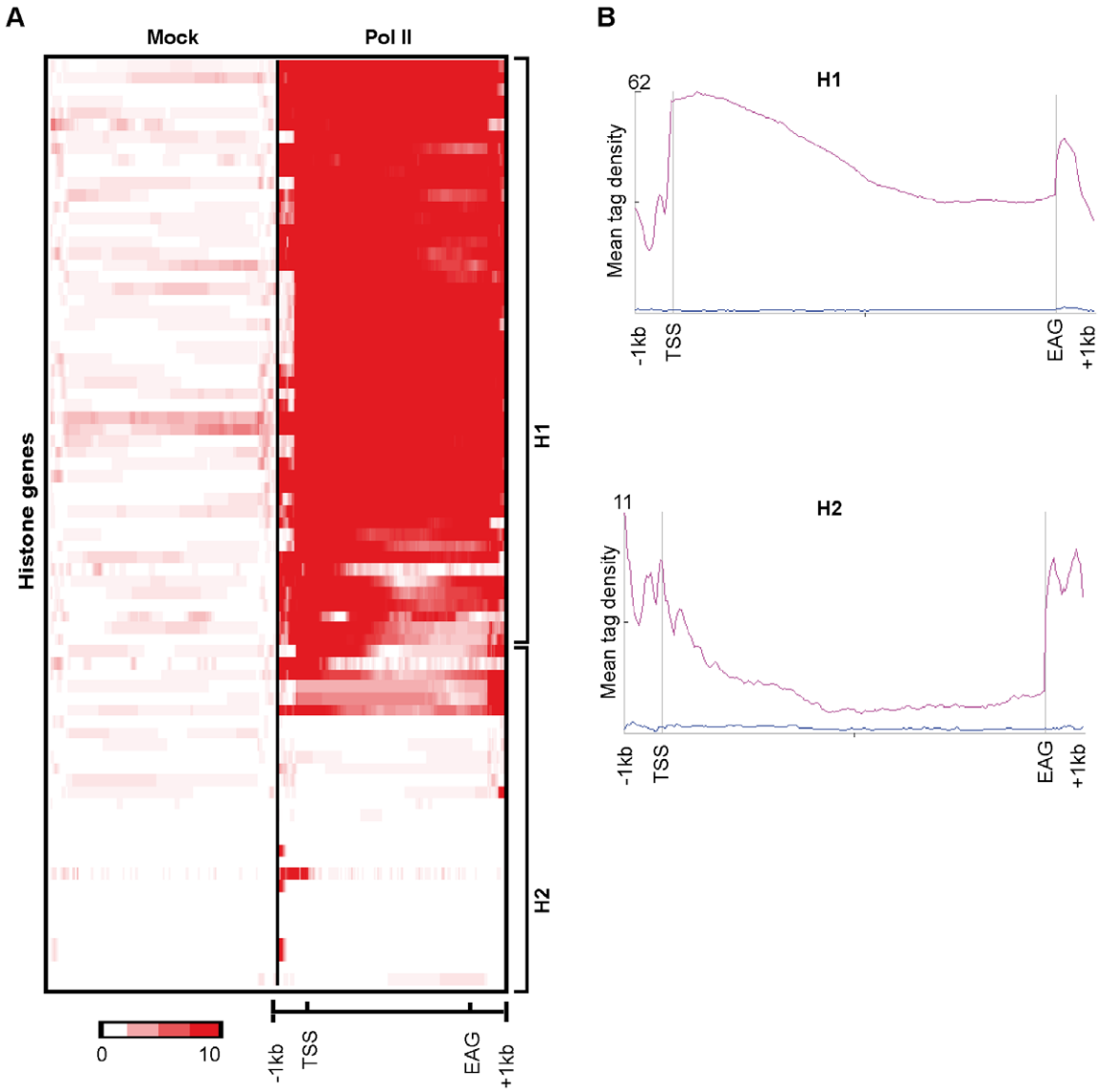


Figure 4. On core histone genes Pol II occupancy downstream of the EAGs is quickly dropping. Clustering of reads obtained following anti-Mock ChIP-seq and anti-Pol II ChIP-seq on all human histone genes generates two clusters. **A**) Heatmap generated after the K-means clustering of Mock and Pol II reads in average gene body and −/+1 kb upstream and downstream of the genes. Color scale indicates the level of enrichment. **B**) Mean tag densities of Mock (Blue) and Pol II (Pink) signals in two clusters of genes (H1 = core histone; H2 = variant histone) in the average gene body and −/+1 kb upstream and downstream of the genes.
doi:10.1371/journal.pone.0038769.g004

Differential Pol II pausing downstream from the EAGs of core histone genes and poly(A)$^+$ genes is conserved between mouse and human cells and seems to be independent from the developmental stage of the cells

In order to further investigate the differential pattern observed for Pol II occupancy downstream of the EAGs on core histone and poly(A)$^+$ genes in differentiated human cells and to analyse the state of the phosphorylation of the C-terminal domain (CTD) of the largest subunit of Pol II, we analysed four different published Pol II ChIP-seq data sets from pluripotent mouse embryonic stem cells (mESs) [5,43]. In one of these studies, genome-wide Pol II occupancy was investigated by using the antibody, which can recognize the N-terminus of the largest subunit of Pol II (as above for human MCF7 cells), allowing monitoring Pol II independent of the phosphorylation status of its CTD (see also "Total Pol II" Figure 6A). In addition, ChIP-seq data were also available for the Ser2, Ser5 and Ser7 phosphorylated form of CTD of Pol II from mES cells [5,43]. Using these data sets, we created two categories of genes for each form of Pol II: actively transcribed poly(A)$^+$ genes and core histone genes (Figure 6). In mES cells, similarly to human cells, we observed (i) a differential Pol II pausing downstream from the EAGs (Figure 6A, "Total Pol II) and (ii) that "total" Pol II pausing at the 3′ end of histone genes is narrow as compared to the broad pause observed downstream of the expressed poly(A)$^+$ genes (Figure 6A). We also investigated which form of Pol II is recruited to the 3′ end of core histone genes (Figure 6B, C and D). Our analyses show that Pol II occupancy profiles 3′ from the EAGs are very comparable between differentiated human and pluripotent mouse cells (compare Figure 1 and 2 to Figure 6A). Moreover, we find that Ser2, Ser5 and Ser7 phosphorylated forms

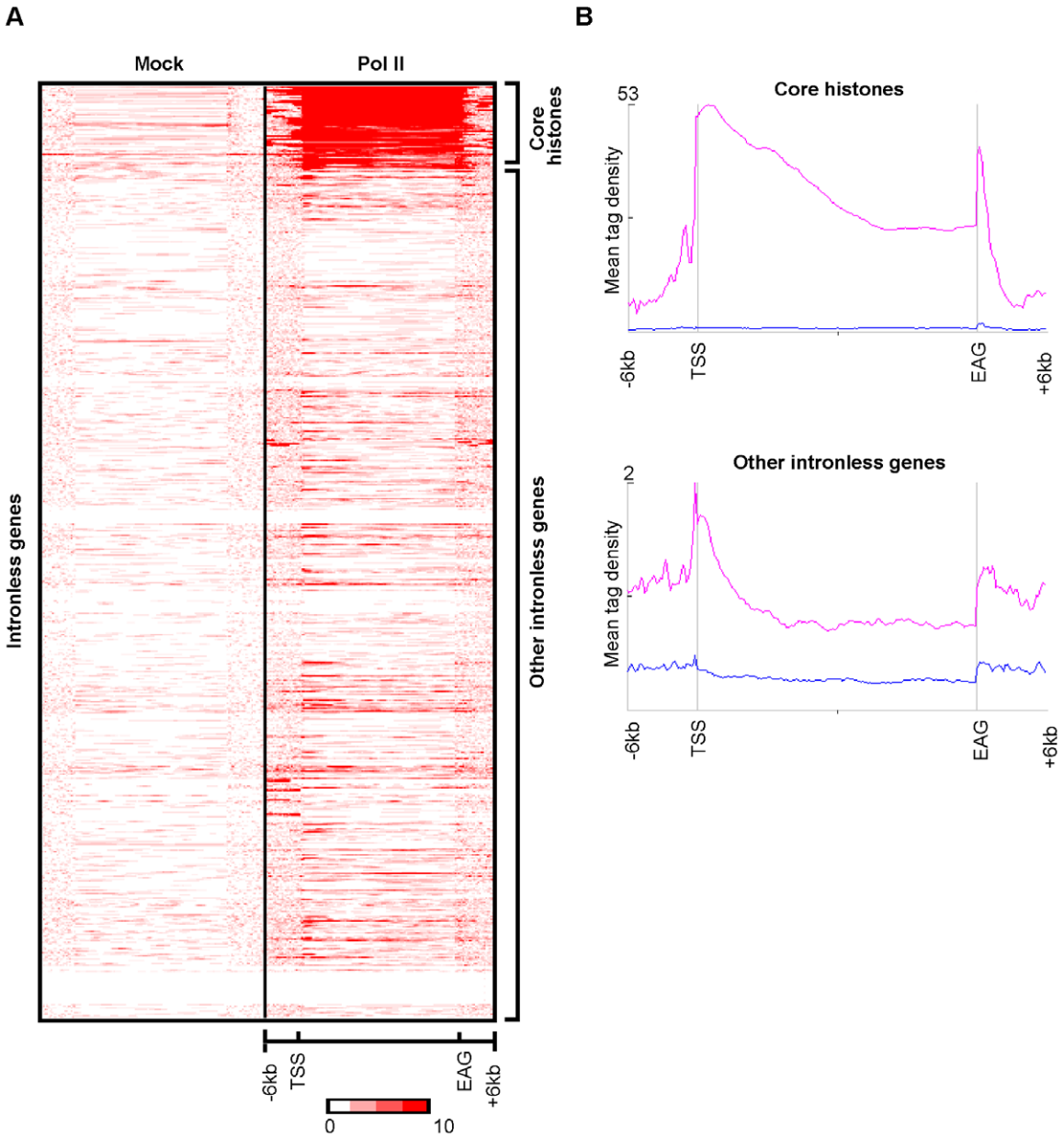


Figure 5. Amongst intronless genes only histone genes have narrow Pol II pause peaks downstream of the EAGs. Clustering of Mock and Pol II reads on all human intronless genes generates two clusters. **A**) Heatmap generated after the K-means clustering of Mock and Pol II reads in average gene body and −/+6 kb upstream and downstream of the genes. Color scale indicates the level of enrichment. **B**) Mean tag densities of Mock (Blue) and Pol II (Pink) in two clusters (Core histones and Other intronless genes) −/+6 kb upstream and downstream of the gene body.
doi:10.1371/journal.pone.0038769.g005

of Pol II are present, but drop rapidly at the 3′ end of the core histone genes (Figure 6C and D). In contrast, but in good agreement with previous studies [5,43], Ser2 phopshorylated form of Pol II is mainly present in the gene body and peaks downstream of the EAGs of poly(A)$^+$ genes (Figure 6B). Taken together these results suggest that differential Pol II pausing downstream of the EAGs of either core histone or poly(A)$^+$ genes is conserved in vertebrate cells and seems to be independent from the developmental stage of the cells. The observed conservation of the distinct Pol II pausing downstream from the EAGs between core histone and poly(A)$^+$ genes further suggest a possible differential link between Pol II pausing and the 3′ end processing of the corresponding transcripts.

Inhibition of polyadenylation increases Pol II occupancy downstream of the EAGs on poly(A)$^+$ genes, but not on core histone genes

To test whether there is a functional link between Pol II occupancy downstream of the EAGs and the 3′ end processing of the corresponding transcripts we have inhibited poladenylation using cordycepin ([51] and refs therein), and tested by ChIP-qPCR whether we can observe a change in Pol II occupancy downstream of the EAGs on poly(A)$^+$ genes when compared to core histone genes. To test the effect of polyadenylation inhibition, MCF7 cells were either not treated, or treated for 3 hours with cordycepin, as described earlier [51]. Following the treatment, cells were subjected to ChIP-qPCR analysis as described above. We

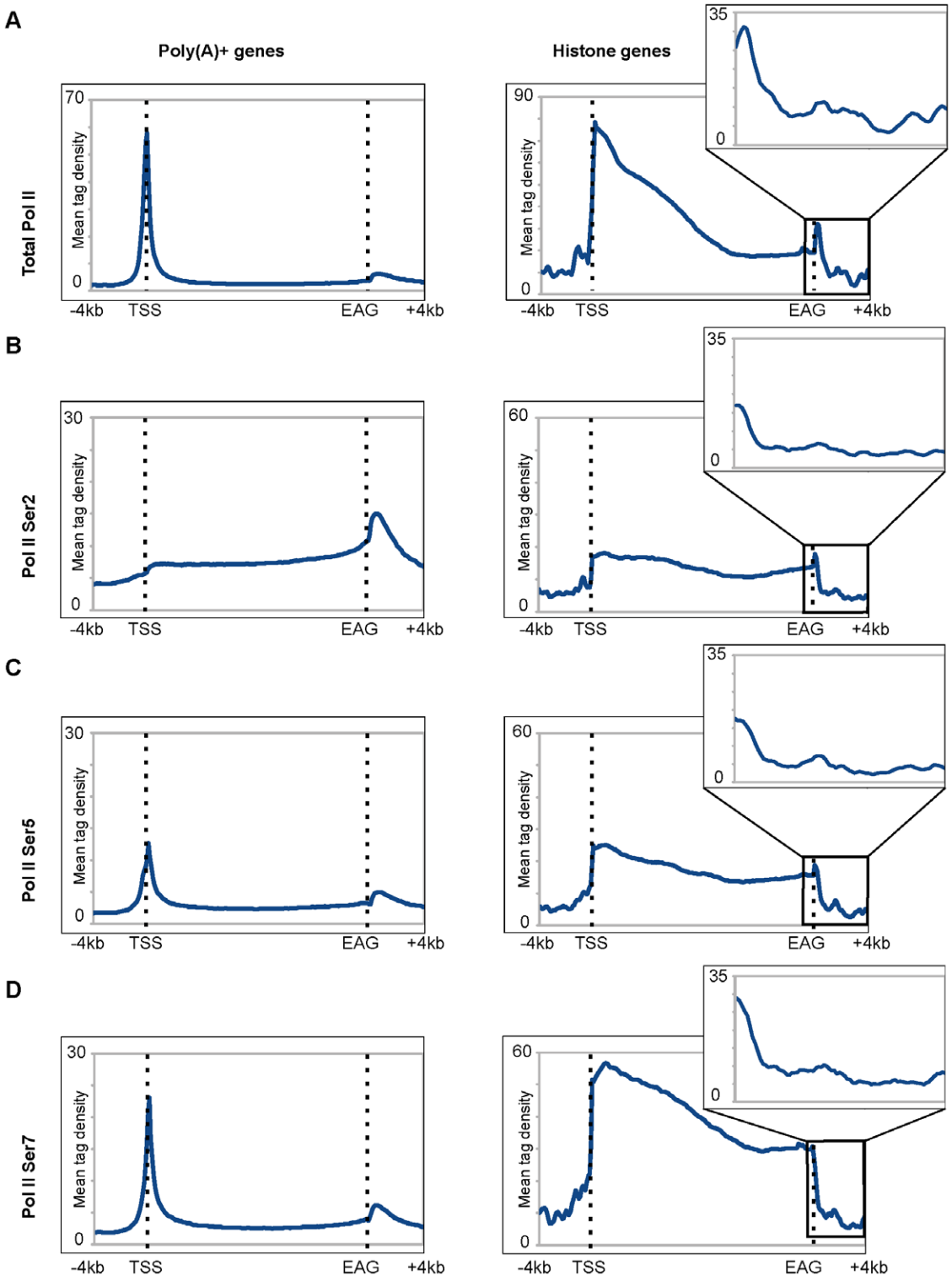
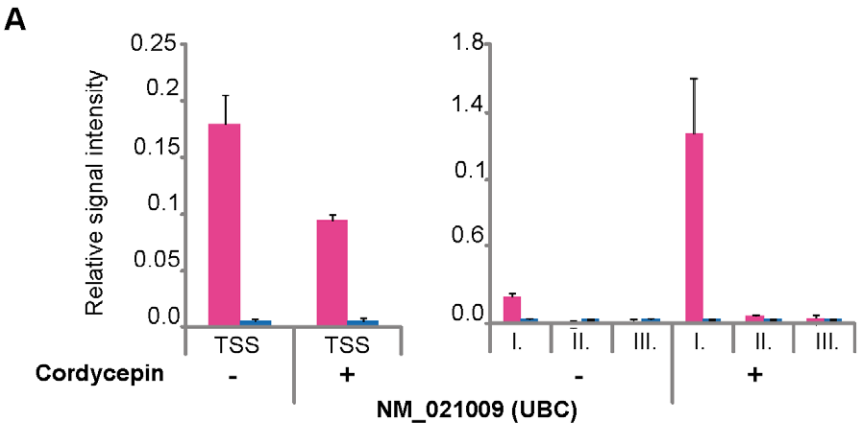
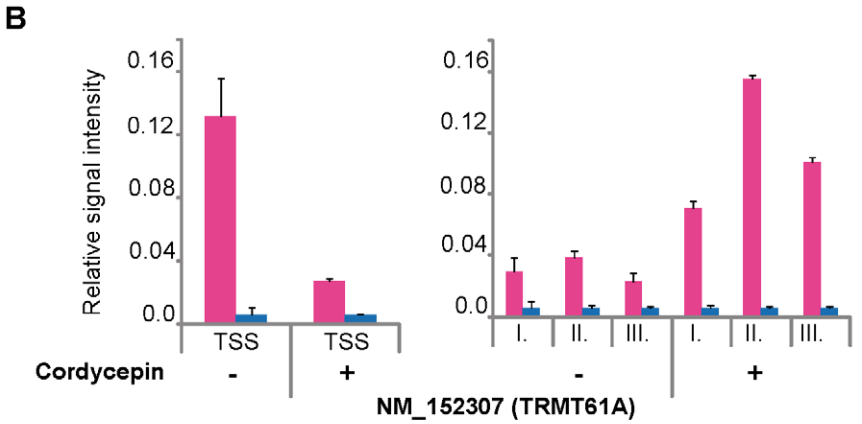
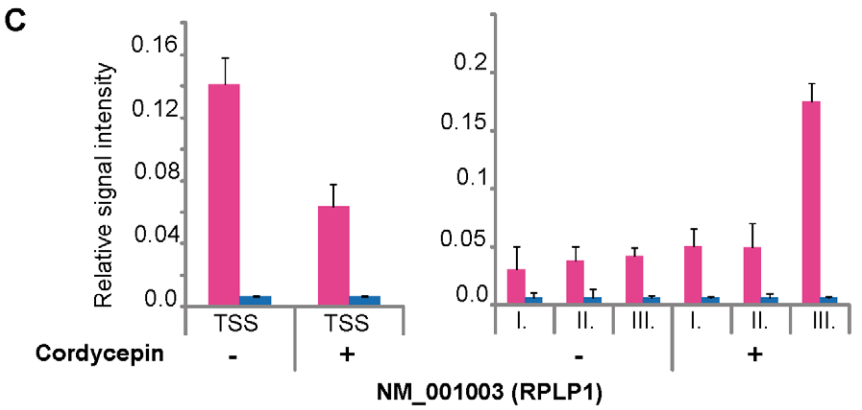


Figure 6. Differential Pol II pausing downstream from the EAGs is conserved between mouse and human cells and seems to be independent from the developmental stage of the cells. A–D) Four different published Pol II ChIP-seq data sets from mES cells [5,43] were used to generate average gene profiles for different forms of Pol II (Total, Ser2, Ser5 and Ser7 phosphorylated form of the CTD of the largest subunit of Pol II) for Poly(A)$^+$ and core Histone genes, as indicated on the top of the figure and on the left of the panels. Y-axis represents mean tag densities.
doi:10.1371/journal.pone.0038769.g006

observed that on the tested poly(A)$^+$ genes inhibition of polyadenylation increased Pol II occupancy downstream of the EAGs (Figure 7 A, B and C; right panels). In contrast, on the tested core histone gene the cordycepin treatment had no significant effect on the Pol II signal at the 3′end of the EAG (Figure 7 D). These observations show that on poly(A)$^+$ genes there is a link between polyadenylation and the broad Pol II pausing downstream of the EAGs and further suggest that Pol II pausing 3′ of the EAGs, and


A
Relative signal intensity
0.25
0.2
0.15
0.1
0.05
0.0
1.8
1.4
0.1
0.6
0.0
TSS
TSS
I.
II.
III.
I.
II.
III.
Cordycepin
-
+
-
+
NM_021009 (UBC)


B
Relative signal intensity
0.16
0.12
0.08
0.04
0.0
0.16
0.12
0.08
0.04
0.0
TSS
TSS
I.
II.
III.
I.
II.
III.
Cordycepin
-
+
-
+
NM_152307 (TRMT61A)


C
Relative signal intensity
0.16
0.12
0.08
0.04
0.0
0.2
0.15
0.1
0.05
0.0
TSS
TSS
I.
II.
III.
I.
II.
III.
Cordycepin
-
+
-
+
NM_001003 (RPLP1)

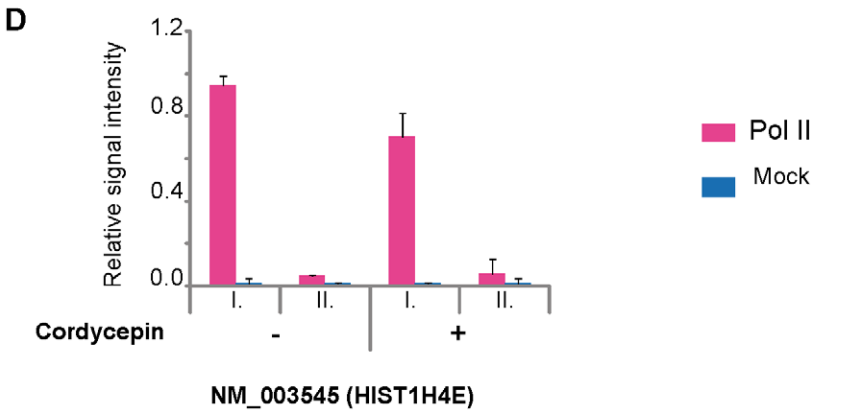

D
Relative signal intensity
1.2
0.8
0.4
0.0
I.
II.
I.
II.
Cordycepin
-
+
Pol II
Mock
NM_003545 (HIST1H4E)

Figure 7. Inhibition of polyadenylation increases Pol II occupancy downstream of the EAGs on poly(A)$^+$ genes, but not on core histone genes. ChIP-qPCR validation of Pol II occupancy following cordycepin treatment on poly(A)$^+$ genes (**A–C**) and histone gene (**D**). Pol II occupancy (Pink bars) is compared at promoters and at different distances downstream from the EAGs before and after cordycepin treatment. Values are normalized to mock controls (Blue bars) and represented as relative signal intensity values. Distances: A, B, C: +0.5–1 kb, II: +2–3 kb, III: +4–5 kb from EAGs of the indicated poly(A)$^+$ genes; D: I: +0.1–0.3 kb, II: +1.5–2 kb from EAG the indicated core histone gene. Error bars represent +/− standard deviations.
doi:10.1371/journal.pone.0038769.g007

consequent termination, may undergo different types of regulation depending on the 3′ end processing mechanisms of the transcripts.

Interestingly, the inhibition of polyadenylation by cordycepin not only increased Pol II occupancy downstream of the EAGs of poly(A)$^+$ genes, but reduced Pol II occupancy at the TSSs of the genes tested (Figure 7A, B and C; left panels). This observation is in good agreement with previous findings suggesting a functional link between transcription initiation and termination by Pol II [52]. Note that at the TSSs of core histone genes the q-PCR reactions could not be carried out because the sonicated and ChIP-ed DNA fragments are longer (500–1500 bps) than the distance would be between the TSS of core histone genes and their EAGs (300–500 bp), where the first q-PCR primer pair has been designed (see Figure 7D).

Discussion

Pol II pausing on long regions downstream of the EAGs is a common feature of genes producing polyadenylated transcripts

Various RNA-processing events have been shown to occur cotranscriptionally (reviewed in [14,24,28,53,54]). Despite its apparent simplicity, the mechanism of Pol II termination is not yet well understood. Moreover, different terminator sequences in the downstream regions of EAGs and different transcription termination mechanisms coupling 3′ end processing and termination exists (see Introduction). However, in the case of core histone genes the link between the lack of Pol II pausing and pre-mRNA 3′ end processing is less well characterized.

By analyzing genome-wide Pol II occupancy together with Gro-seq data [42] in human MCF7 cells and Pol II occupancy in mouse ES cells we report that on most of the expressed human and mouse poly(A)$^+$ genes transcriptionally active Pol II pausing downstream of EAGs is a common event. On these genes, Pol II occupancy often covers a very long region, as ChIP signals can be mapped up to 4–6 kb downstream of the EAGs. The fact that majority of the expressed poly(A)$^+$ genes have significant Pol II occupancy throughout a long region after the 3′ of the genes and that core histone genes have a very sharp drop of Pol II occupancy at 3′ of their EAGs, suggests that the different 3′ processing regulatory mechanisms influence the residency time of Pol II downstream of the EAGs. These differential Pol II occupancy mechanisms 3′ of the distinct gene categories seem to be conserved in vertebrates and not influenced by the differentiation state of the cells.

According to a recent report, expressed genes are linked to different specialized transcriptional factories and the size of the "factory" depend on the strength of the expression of the genes connected to these sites of transcription [55]. Thus, it is conceivable that different transcription factories might have a link with the different pause patterns observed in this analysis. In agreement, poly (A)$^+$ genes belonging to the broad category are higher expressed than those belonging to the very broad category (Figure 2). Note, however, that in terms of DNA sequence, amongst the downstream regions of poly(A)$^+$ genes belonging to the very broad and broad categories we could not identify any significant differences and we do not know whether "allosteric-/anti-terminator" or "torpedo" model type termination mechanisms, or a hybrid of the two mechanisms, would play a role.

Our genome-wide results also suggest that the pausing step downstream of the EAGs reflects a slowing down of the Pol II elongation complex. It seems that on genes that produce polyadenylated transcripts Pol II binding downstream of EAGs is in general different than during transcription in the gene body in terms of detectable Pol II occupancy. Moreover, in most of the cases after the cleavage event at the polyadenylation site, transcriptionally active Pol II can be detected for a long time and distance on the DNA template. In good agreement, all the transcripts identified by Gro-seq are mapping in the sense orientation 3′ from the EAGs, suggesting that they are produced by Pol II molecules that continue to transcribe after having finished the pre-mRNAs. In this respect it is interesting that inhibition of polyadenylation increased Pol II occupancy downstream of EAGs on poly(A)$^+$ genes suggesting that defective polyadenylation can signal to the terminating Pol II, to slow down. Such a signaling may be necessary to reduce transcription on poly(A)$^+$ genes, where polyadenylation would be defective. The increase of Pol II residency time 3′ of poly(A)$^+$ genes in turn may reduce the amount of Pol II that can be released from each gene to enter in a new reinitiation cycle. This hypothesis is in good agreement with our findings showing that inhibition of polyadenylation reduces Pol II occupancy at the TSSs of the tested genes (Figure 7). Thus, Pol II pausing downstream of the EAGs may be implicated in a feedback regulation of 3′ processing of the nascent transcripts, where Pol II 3′ pausing would be controlled by the completeness of the polyadenylation of the pre-mRNA and thus, may prevent immediate Pol II release from the gene [16,25,56]. Consequently, longer Pol II pausing in regions downstream of the EAGs may play a role in the establishment of nuclear pools of Pol II that can be engaged in new rounds of transcription. These observations further point towards a link between transcription initiation and termination as previously suggested [57,58]. In agreement with this hypothesis, several general transcription factors, such as TFIIB, TFIID, TFIIH and the Mediator complex, have been described in linking transcription initiation and Pol II termination [59,60,61,62,63]. Moreover, these studies also hypothesized that on actively transcribed genes looping may occur [57,58,64]. The finding that a small subset of highly expressed poly(A)$^+$ genes exist with no, or weak, Pol II pause downstream of the EAGs may suggest that on these genes transcription initiation and termination are not linked or that on these subset of highly expressed (often ribosomal protein coding) genes a very efficient Pol II recycling is required.

On core histone genes Pol II occupancy is dropping sharply downstream of the EAGs

Core histone genes are generally small in size, intronless and encoding for transcripts that do not undergo polyadenylation. The 3′ end of core histone mRNAs is formed by a cleavage reaction between the stem–loop and the purine-rich sequence [65], with transcription continuing for at least a few hundred nucleotide past the 3′ end of the mRNA [66]. In contrast to poly(A)$^+$ genes, at core

histone genes Pol II occupancy downstream of the EAGs quickly declines (Figure 2 and 4). Such short Pol II pause may be of high importance since often core histone genes can be found in clusters, separated by only short distances (i.e. 0.8–1 kb) and thus, Pol II occupancy has to decline rapidly to prevent Pol II from running into the neighboring gene in the cluster. Our observations suggest that in spite the fact that also common core cleavage factors are required for processing of core histone and polyadenylated pre-mRNAs [67], there are important differences remaining in the way how Pol II is released from the different templates coding for the two types of transcripts during transcription termination. Thus, as the 3′ processing of core histone transcripts are carried out by a distinct machinery than those transcripts generated from poly(A)$^+$ genes, it is conceivable that the differential Pol II pause profiles observed on core histone genes versus poly(A)$^+$ genes represents a differential interaction between Pol II and the two different 3′ processing machineries. This idea is also corroborated by the functional studies, which showed that inhibition of polyadenylation increased Pol II occupancy 3′ of the EAGs of poly(A)$^+$ genes, while it had no significant effect on Pol II drop at the 3′ end of the core histone genes.

Our findings are in good agreement with previous reports describing a connection between polyadenylation and Pol II pause followed by transcription termination (reviewed in [24,39]) and in addition suggests that this connection is differently regulated on core histone genes, where 3′ processing of the transcripts is distinct from other protein coding genes. It seems that Pol II release on histone genes is an actively regulated and quick event. It may be in relation with the high transcription rate and the possible recycling of Poll II on these genes for quick transcription re-initiation. Interestingly, histone variant genes which are different from core histone genes, because they contain introns and their encoded transcripts are polyadenylated, show extended Pol II pause downstream of their EAGs in contrast to core histone genes, further suggesting that polyadenylation might have a role in the Pol II slow down and hence a longer pause.

Taken together, our study suggests that in mammalian cells Pol II pausing downstream of the EAGs and mRNA 3′ processing are not independent, but linked. Thus, it seems that pausing of Pol II 3′ from the EAG is part of transcription regulatory mechanisms at different type of genes.

Materials and Methods

Chromatin immunoprecipitation

Human MCF7 cells (obtained from American Type Culture Collection; ATCC; reference number HTB-22) were grown up to 85–90% confluence, washed with PBS, and cross-linked with 1% formaldehyde for 20 min at room temperature. The reaction was stopped with 0.5 M glycine, and cells were washed three times with ice-cold PBS supplemented with 0.5 mM phenylmethylsulfonyl fluoride (PMSF), scraped, and resuspended in swelling buffer (25 mM HEPES, pH 7.8, 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% NP-40, 1 mM dithiothreitol [DTT], 0.5 mM PMSF, protease inhibitor cocktail [PIC], Amersham). Cells were broken with a Dounce homogenizer, and the nuclear fraction was resuspended in sonication buffer (50 mM HEPES, pH 7.8, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 1% sodium dodecyl sulfate [SDS], 0.5 mM PMSF, PIC). The chromatin was sonicated with a Bioruptor (Diagenode) sonicator into 100–500-bp fragments and centrifuged to avoid any remaining cell debris.

From the supernatant, 30 μg chromatin diluted with sonication buffer (without SDS) up to 1 ml (0.05% SDS concentration) was used for one IP. For ChIP-seq 5 samples were added together. Protein G Sepharose beads were washed and blocked with sonication buffer containing cold-water fish skin gelatin (SIGMA) and yeast tRNA. Chromatin samples were pre-cleared with unblocked beads at 4°C, for 2 hours. The precleared chromatin samples were rotated overnight at 4°C with the Pol II antibody (Santa Cruz, H-224X), and then blocked beads were added for 2 hours to the samples to pull down specific protein-DNA complexes. After immunoprecipitation, samples were washed two times at 4°C with the following buffers: twice with Sonication buffer (without SDS), twice with buffer A (50 mM HEPES, pH 7.8, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS, 0.5 mM PMSF, PIC), twice with buffer B (20 mM Tris, pH 8, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 0.5 mM PMSF, PIC), and finally twice with Tris-EDTA buffer (10 mM Tris, pH 8, 1 mM EDTA). Bound fraction of the chromatin was eluted with 2×100 μl of elution buffer (50 mM Tris, pH 8, 1 mM EDTA, 1% SDS) at 65°C for 2×10 min and elutions were pooled. RNase A treatment (5 μg/ml), and reverse cross-linking (125 mM NaCl) was carried out at 65°C overnight. Elutions were finally incubated with proteinase K. DNA was phenol-chloroform extracted and precipitated by ethanol. Validation of the ChIP was performed by quantitative PCR (qPCR) analysis using a Roche LightCycler 480 with Sybr green (Roche) master mix.

As a negative control, we immunoprecipitated the cross-linked material with a yeast antibody. The ChIP experiments were repeated at least twice, and all the qPCR reactions were done in triplicates.

To inhibit polyadenylation human MCF7 cells were grown up to 85–90% confluence and the medium was changed with fresh one containing cordycepin (Sigma, 50 μg/ml final concentration) [51]. After three hours of treatment, ChIP was carried out as described above.

Solexa high throughput sequencing

Sample preparation was performed as described by the manufacturer. The 32 base pair tags generated from Illumina/Solexa were mapped to the human genome Build 36.1 (UCSC hg18) using the eland program allowing two mismatches. Only sequences that mapped uniquely to the genome with maximum of two mismatches were used for further analysis. We obtained 10.9 and 4.5 millions uniquely mapped reads for Pol II and Mock, respectively. Using the liftover tool from UCSC (www.ucsc.org), tags were mapped onto the human genome hg19. The obtained Pol II ChIP-seq data from human MCF7 cells was deposited in GEO under GSE34001number.

Validation of ChIP-Seq by comparison to real-time PCR

qPCR was performed by Roche LightCycler 480 with Sybr green (Roche) master mix. The sequences of the oligonucleotides are available upon request.

Genome Annotations

Genome annotations were downloaded from the UCSC Genome Browser (www.ucsc.org), human genome Build 37 (hg19 assembly). Gene definitions were given by the refseq genes [68] track. For the analysis mentioned in the paper, we have considered only those refseq genes which are reviewed and validated.

Heatmap and Clustering of Pol II patterns

For all the refseq genes, which were analysed, we extracted the tag density in a 4 kb windows surrounding the EAG and

surrounding the gene body using the program seqMINER which generates heatmap as well as the profiles [45]. The sequenced ChIP-seq reads represent only the end of each immunoprecipitated fragments instead of the precise protein-DNA binding sites. To illustrate the entire DNA fragment, basically before analysis, 3′ end of each ChIP-seq read was extended to 200 bp in the direction of the reads. In order to get an average gene profile of ChIP-seq tags, genes from its start (5′ end) to end (3′ end) [according to refseq annotation] were averaged in a 50 bp window. While doing this analysis, the strand orientation is taken in account in order to orientate all analysed features in the same direction. Pol II tag densities were subjected to K-means clustering in order to organize or cluster genes in a same group based upon similar tag enrichment within a defined region. In K-means clustering, number of clusters is fixed a priori and hence we define the number of clusters based upon the tag enrichment and patterns of Pol II.

Microarray Expression Data

Cel files of MCF7 cells were downloaded from Gene Expression Omnibus (GSE18912, http://www.ncbi.nlm.nih.gov/geo/). In this experiment, MCF7 cells were tested as five replicates using Affymetrix U133 Plus 2.0 arrays. The cel files were normalized by gcrma method [69], and calculations were performed using R (http://www.r-project.org/). The expression level for each probeset was calculated as the mean of the five replicates. While processing the data, we applied the following filters. Firstly, we excluded probesets assigned to several genes localized on different places on the genome to avoid annotation artifacts. Secondly, we eliminated all low intensity probesets where signal cannot be distinguished from noise. These filters allowed us to obtain curated list of expressed probesets. A single relative expression value was computed for each gene and based upon relative expression values we selected lists of highly expressed genes.

Gene Ontology

Enrichment of GO categories was determined using WebGestalt software [70] and functional categories with stringent p-value (p-value$<$0.01) have been considered for the analysis.

Supporting Information

Figure S1 Showing genome-wide Pol II occupancy on ''non-neighboring'' human genes.
(DOC)

Table S1 List containing 500 highly expressed genes from MCF7 cells used in the analysis in **Figure 1C**.
(DOC)

Table S2 List containing 100 highly expressed genes from MCF7 cells used in **Figure 3**.
(DOC)

Acknowledgments

We are grateful to M. Ballarino, A.R. Krebs, D. Devys and J. Corden for comments and carefully reading the manuscript. We thank M. Carmo-Fonseca and A.R. Grosso for discussions on the analyses. We also thank the IGBMC High throughput-sequencing platform for data analysis.

Author Contributions

Conceived and designed the experiments: KA ÁG LT LP OP. Performed the experiments: KA ÁG. Analyzed the data: KA ÁG LP LT. Contributed reagents/materials/analysis tools: OP. Wrote the paper: KA ÁG LT.

References

1. Orphanides G, Lagrange T, Reinberg D (1996) The general transcription factors of RNA polymerase II. Genes Dev 10: 2657–2683.
2. Ptashne M, Gann A (1997) Transcriptional activation by recruitment. Nature 386: 569–577.
3. Hochheimer A, Tjian R (2003) Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev 17: 1309–1320.
4. Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, et al. (2010) Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science 327: 335–338.
5. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, et al. (2010) c-Myc regulates transcriptional pause release. Cell 141: 432–445.
6. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130: 77–88.
7. Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, et al. (2007) RNA polymerase is poised for activation across the genome. Nat Genet 39: 1507–1511.
8. Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, et al. (2007) RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet 39: 1512–1516.
9. Core LJ, Lis JT (2008) Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 319: 1791–1792.
10. Spies N, Nielsen CB, Padgett RA, Burge CB (2009) Biased chromatin signatures around polyadenylation sites and exons. Mol Cell 36: 245–254.
11. Rougvie AE, Lis JT (1988) The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell 54: 795–804.
12. Lis J (1998) Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. Cold Spring Harb Symp Quant Biol 63: 347–356.
13. Lee C, Li X, Hechmer A, Eisen M, Biggin MD, et al. (2008) NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Mol Cell Biol 28: 3290–3300.
14. Proudfoot NJ (2011) Ending the message: poly(A) signals then and now. Genes Dev 25: 1770–1782.
15. Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. Cell 108: 501–512.
16. Gromak N, West S, Proudfoot NJ (2006) Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol Cell Biol 26: 3986–3996.
17. Hagenbuchle O, Wellauer PK, Cribbs DL, Schibler U (1984) Termination of transcription in the mouse alpha-amylase gene Amy-2a occurs at multiple sites downstream of the polyadenylation site. Cell 38: 737–744.
18. Ashfield R, Enriquez-Harris P, Proudfoot NJ (1991) Transcriptional termination between the closely linked human complement genes C2 and factor B: common termination factor for C2 and c-myc? EMBO J 10: 4197–4207.
19. Enriquez-Harris P, Levitt N, Briggs D, Proudfoot NJ (1991) A pause site for RNA polymerase II is associated with termination of transcription. EMBO J 10: 1833–1842.
20. Tantravahi J, Alvira M, Falck-Pedersen E (1993) Characterization of the mouse beta maj globin transcription termination region: a spacing sequence is required between the poly(A) signal sequence and multiple downstream termination elements. Mol Cell Biol 13: 578–587.
21. Dye MJ, Proudfoot NJ (2001) Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. Cell 105: 669–681.
22. Ashfield R, Patel AJ, Bossone SA, Brown H, Campbell RD, et al. (1994) MAZ-dependent termination between closely spaced human complement genes. EMBO J 13: 5656–5667.
23. Dye MJ, Proudfoot NJ (1999) Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. Mol Cell 3: 371–378.
24. Buratowski S (2005) Connections between mRNA 3′ end processing and transcription termination. Curr Opin Cell Biol 17: 257–261.
25. Kaneko S, Rozenblatt-Rosen O, Meyerson M, Manley JL (2007) The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3′ processing and transcription termination. Genes Dev 21: 1779–1789.
26. Logan J, Falck-Pedersen E, Darnell JE , Jr.., Shenk T (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proc Natl Acad Sci U S A 84: 8306–8310.
27. Calvo O, Manley JL (2001) Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. Mol Cell 7: 1013–1023.
28. Proudfoot N (2004) New perspectives on connecting messenger RNA 3′ end formation to transcription. Curr Opin Cell Biol 16: 272–278.

29. Richard P, Manley JL (2009) Transcription termination by nuclear RNA polymerases. Genes Dev 23: 1247–1269.
30. Connelly S, Manley JL (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev 2: 440–452.
31. Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, et al. (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature 432: 517–522.
32. West S, Gromak N, Proudfoot NJ (2004) Human 5′–>3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature 432: 522–525.
33. Luo W, Johnson AW, Bentley DL (2006) The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes Dev 20: 954–965.
34. Park NJ, Tsao DC, Martinson HG (2004) The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II carboxyl-terminal repeat domain. Mol Cell Biol 24: 4092–4103.
35. Skourti-Stathaki K, Proudfoot NJ, Gromak N (2011) Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Mol Cell 42: 794–805.
36. Tran DP, Kim SJ, Park NJ, Jew TM, Martinson HG (2001) Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. Mol Cell Biol 21: 7495–7508.
37. Orozco IJ, Kim SJ, Martinson HG (2002) The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. J Biol Chem 277: 42899–42911.
38. Nag A, Narsinh K, Kazerouninia A, Martinson HG (2006) The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity. RNA 12: 1534–1544.
39. Kazerouninia A, Ngo B, Martinson HG (2010) Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro. RNA 16: 197–210.
40. 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.
41. Lian Z, Karpikov A, Lian J, Mahajan MC, Hartman S, et al. (2008) A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3′ end RNA polyadenylation. Genome Res 18: 1224–1237.
42. Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, et al. (2011) A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell 145: 622–634.
43. Brookes E, de Santiago I, Hebenstreit D, Morris KJ, Carroll T, et al. (2012) Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell 10: 157–170.
44. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, et al. (2008) Divergent transcription from active promoters. Science 322: 1849–1851.
45. Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, et al. (2011) seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res 39: e35.
46. Schaufele F, Gilmartin GM, Bannwarth W, Birnstiel ML (1986) Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3′ end of H3 messenger RNA. Nature 323: 777–781.
47. Dominski Z, Marzluff WF (2007) Formation of the 3′ end of histone mRNA: getting closer to the end. Gene 396: 373–390.
48. Marzluff WF, Wagner EJ, Duronio RJ (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet 9: 843–854.
49. Schumperli D (1988) Multilevel regulation of replication-dependent histone genes. Trends Genet 4: 187–191.
50. Dominski Z, Marzluff WF (1999) Formation of the 3′ end of histone mRNA. Gene 239: 1–14.
51. Ioannidis P, Courtis N, Havredaki M, Michailakis E, Tsiapalis CM, et al. (1999) The polyadenylation inhibitor cordycepin (3′dA) causes a decline in c-MYC mRNA levels without affecting c-MYC protein levels. Oncogene 18: 117–125.
52. Lykke-Andersen S, Mapendano CK, Jensen TH (2011) An ending is a new beginning: transcription termination supports re-initiation. Cell Cycle 10: 863–865.
53. Bentley DL (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol 17: 251–256.
54. Neugebauer KM (2002) On the importance of being co-transcriptional. J Cell Sci 115: 3865–3871.
55. Eskiw CH, Fraser P (2011) Ultrastructural study of transcription factories in mouse erythroblasts. J Cell Sci 124: 3676–3683.
56. Yonaha M, Proudfoot NJ (1999) Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. Mol Cell 3: 593–600.
57. El Kaderi B, Medler S, Raghunayakula S, Ansari A (2009) Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. J Biol Chem 284: 25015–25025.
58. Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH (2010) Crosstalk between mRNA 3′ end processing and transcription initiation. Mol Cell 40: 410–422.
59. Dantonel JC, Murthy KG, Manley JL, Tora L (1997) Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. Nature 389: 399–402.
60. Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, et al. (2009) TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Mol Cell Biol 29: 5455–5464.
61. Wang Y, Fairley JA, Roberts SG (2010) Phosphorylation of TFIIB links transcription initiation and termination. Curr Biol 20: 548–553.
62. Hampsey M, Singh BN, Ansari A, Laine JP, Krishnamurthy S (2011) Control of eukaryotic gene expression: gene loops and transcriptional memory. Adv Enzyme Regul 51: 118–125.
63. Mukundan B, Ansari A (2011) Novel role for mediator complex subunit Srb5/Med18 in termination of transcription. J Biol Chem 286: 37053–37057.
64. Krishnamurthy S, Ghazy MA, Moore C, Hampsey M (2009) Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. Mol Cell Biol 29: 2925–2934.
65. Gick O, Kramer A, Keller W, Birnstiel ML (1986) Generation of histone mRNA 3′ ends by endonucleolytic cleavage of the pre-mRNA in a snRNP-dependent in vitro reaction. EMBO J 5: 1319–1326.
66. Chodchoy N, Pandey NB, Marzluff WF (1991) An intact histone 3′-processing site is required for transcription termination in a mouse histone H2a gene. Mol Cell Biol 11: 497–509.
67. Sullivan KD, Steiniger M, Marzluff WF (2009) A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. Mol Cell 34: 322–332.
68. Pruitt KD, Tatusova T, Maglott DR (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 33: D501–504.
69. Wu ZJ, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F (2004) A model-based background adjustment for oligonucleotide expression arrays. Journal of the American Statistical Association 99: 909–917.
70. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33: W741–748.

V. Publication II.

Transcription 4:1, 1–6; January/February 2013; © 2013 Landes Bioscience

How to stop: the mysterious links among RNA polymerase II occupancy 3' of genes, mRNA 3' processing and termination

Krishanpal Anamika,† Àkos Gyenis† and Làszlò Tora*

Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC); CNRS UMR 7104; INSERM U 964; Université de Strasbourg; Illkirch Cedex, France

†These authors contributed equally to this work.

Eukaryotic genes are transcribed by RNA polymerase II (RNAP II) through cycles of initiation, elongation and termination. Termination remains the least understood stage of transcription. Here we discuss the role of RNAP II occupancy downstream of the 3' ends of genes and its links with termination and mRNA 3' processing.

Keywords: genome-wide mapping, RNA Polymerase II pause, core histone genes, End of Annotated Gene, polyadenylation, transcription termination, R-loops

Submitted: 08/21/12

Revised: 09/17/12

Accepted: 09/19/12

http://dx.doi.org/10.4161/trns.22300

*Correspondence to: Làszlò Tora;
Email: laszlo@igbmc.fr

Introduction

RNA polymerase II (RNAP II) transcription is a highly regulated process that requires coordinated action of numerous transcription factors. It can be divided into initiation, promoter escape, elongation and termination phases.

Several studies analyzed RNAP II binding at promoters genome-wide (GW) by using chromatin immunoprecipitation (ChIP) assays coupled to high-throughput sequencing (ChIP-seq). These reports revealed that thousands of transcribed genes in many different metazoans display accumulation of RNAP II just downstream of their promoters.[1-8] Many of these GW experiments demonstrated that these high and narrow RNAP II occupancy peaks, probably reflecting a static RNAP II binding at promoters, were centered on regions 50 nucleotides downstream of the transcription start sites (TSSs) of the genes. It is now accepted that these high RNAP II occupancy signals reflect RNAP II pausing at promoter-proximal sites of virtually all transcribed genes in various organisms and, thus, may be considered a general feature of RNAP II behavior in preinitiation complexes formed at the 5' end of transcribed genes. Moreover, it seems that maintaining a static RNAP II in the scaffold of general transcription factors associated with the promoter is important to establish an open chromatin conformation that facilitates RNA synthesis.[8]

In most of the GW analyses, a higher RNAP II occupancy profile has also been observed at the 3' end of the genes; however, none of the studies analyzed and characterized RNAP II binding downstream of the 3' end of annotated genes (EAGs) in detail. Here, we will discuss advances in our understanding of RNAP II occupancy downstream of EAGs and its possible link with 3' processing mechanisms and with transcription termination.

RNAP II occupancy increases downstream of transcribed genes genome-wide

To better understand RNAP II behavior at the 3' ends of transcription units and in transcription termination, in our recent study we analyzed different GW data sets representing RNAP II binding downstream of EAGs in pluripotent mouse embryonic stem cells and differentiated mammalian cells.[9] Our analyses in both cell types demonstrated that, in general, RNAP II detectability by ChIP-seq increases downstream of the EAGs. According to the definition, the 3' end of a gene (EAG) terminates at the 3' end of its 3' untranslated region. Interestingly, our analyses showed that downstream of genes producing polyadenylated

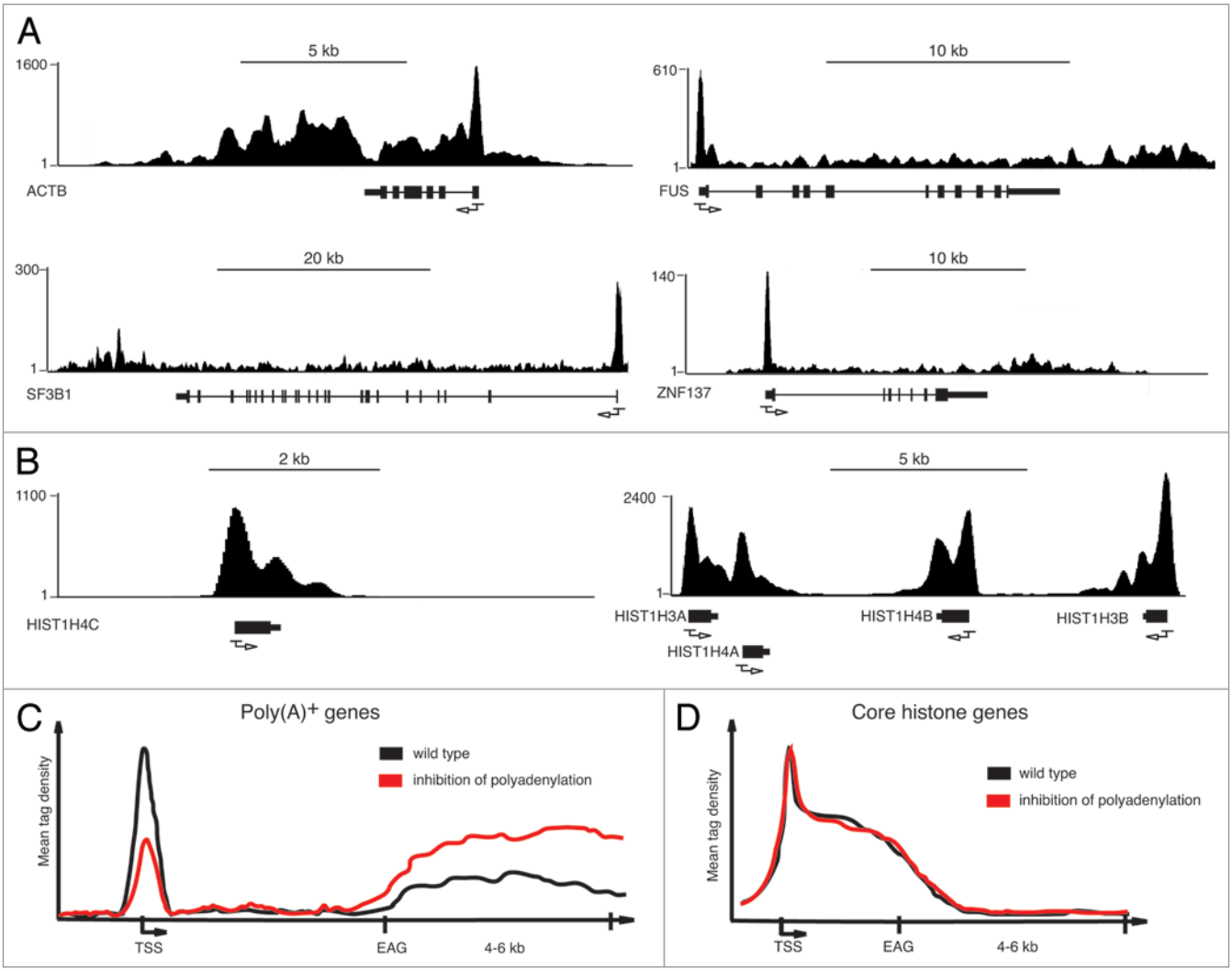


Figure 1. RNAP II occupancy at the TSS, gene body and downstream of EAG under normal and polyadenylation inhibition conditions. (A) and (B) RNAP II profile on poly(A)$^+$ genes and core histone genes. Examples of RNAP II occupancy profiles are shown as custom tracks from the UCSC genome browser for (A) poly(A)$^+$ genes and (B) core histone genes. (**A**) For poly(A)$^+$ genes one highly expressed (*ACTB*), two medium expressed (*FUS* and *SF3B1*) and one weakly expressed (*ZNF137*) genes are shown as examples. (**B**) For core histone genes an individual histone gene and a histone gene cluster is represented. In each panel on the y-axis RNAP II ChIP tag density is represented; on the top of the panels the corresponding distances are labeled in kb-s as horizontal bars. Arrows show the direction of the transcription of each transcription unit. Gene names are represented in capital letters. Gene structures are shown. Thick boxes represent exons, medium thick boxes untranslated regions, thin lines introns. (**C**) and (**D**) Extrapolation of RNAP II occupancy at the TSS, gene body and downstream of EAG under normal and polyadenylation inhibition conditions based on our ChIP-PCR results.[9] (**C**) RNAP II occupancy profile on average poly(A)$^+$ genes: in the wild type conditions, RNAP II occupancy shows a high narrow peak at transcription start sites (TSSs), often low occupancy throughout the gene body, and a broad enrichment 3' of the end of annotated genes (EAGs), which can extend up to 4–6 kb from the EAG (black line). According to the used definition 3' end of a gene (EAG) finishes at the 3' end of its 3' untranslated region. Upon inhibition of polyadenylation by cordycepin there is an increase in the RNAP II occupancy downstream of the EAGs and a parallel decrease at the TSSs (red line). (**D**) RNAP II occupancy profile on core histone genes: high RNAP II occupancy can be seen at the TSS and throughout the gene body and RNAP II occupancy declines rapidly 3' from the EAG. In wild type and cordycepin treated conditions there is no change in the RNAP II occupancy. Wild type and cordycepin treated conditions are represented in black and red lines, respectively.

transcripts [poly(A)$^+$ genes], RNAP II occupancy is often quite extensive and can be detected up to 4–6 kb downstream from the EAGs (**Fig. 1A and 1C**).[9] A possibility is that RNAP II slows down once it reaches the 3' end of a poly(A)$^+$ gene and therefore may be more likely to cross-link to its given genomic position during the ChIP procedure (**Fig. 2A**). Consequently, several RNAP II molecules may accumulate in these regions (**Fig. 2A**). Nevertheless, it is not clear what the broad RNAP II enrichment represents in the regions between the EAG and the termination site of a transcribed poly(A)$^+$ gene. Is it the sum of arrested RNAP II molecules at various locations in different cells of the population used for the ChIP or the RNAP II population that is transcribing with a reduced speed, compared with RNAP IIs in the gene body, in every cell? However, neither scenario accounts exactly for the observed distribution of RNAP II downstream of the EAGs often

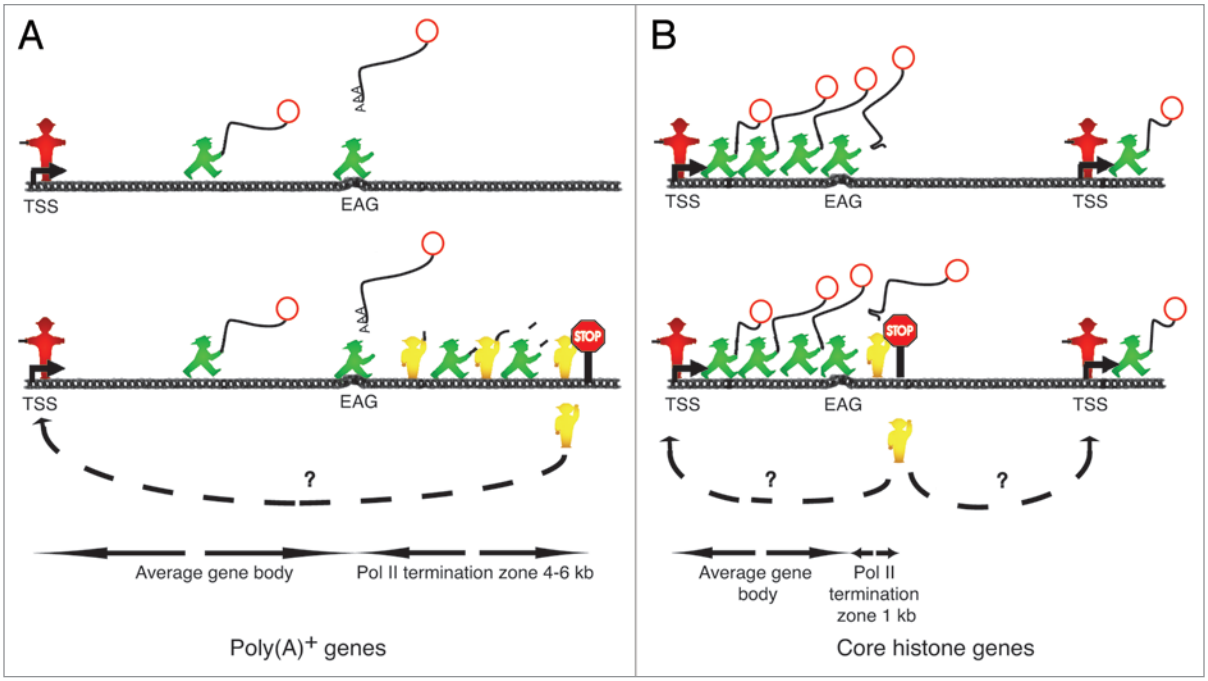


Figure 2. Model representing the dynamic behavior of RNAP II during different phases of transcription on poly(A)$^+$ genes (**A**) and replication-activated core histone genes (**B**). The traffic men represent different forms of the RNAP II during the transcription cycle. The red "traffic light men" represent static polymerases, yellow "traffic light men" represent slow motion polymerases waiting for a putative signal and green "traffic light men" represent polymerase molecules that move with full speed. Upper panels show RNAP II molecules (green traffic light men) before or when they reach the processing sites, more "green traffic light men" mean higher gene expression. Lower panels show the genes with RNAP II throughout the gene bodies and those that reach the termination sites or signals. Transcription start sites (TSSs, are represented by an arrow head), average gene bodies, and 3' of the end of annotated genes (EAGs) are shown. Strings with a red balloon represent 5' capped nascent mRNAs synthesized by the polymerases. mRNAs transcribed from poly(A)$^+$ genes get either polyadenylated (AAA) [shown in (**A**)] or when transcribed from core histone genes form a 3' hairpin [shown in (**B**)]. Transcription termination sites are represented with a stop sign. For core histone genes the TSSs of a second histone gene are also represented depicting that core histone genes are often found in clusters. The main factors involved in cotranscriptional 3'-end formation of mammalian polyadenylated mRNA vs. non-polyadenylated core histone mRNA are reviewed in.[17] Short black sticks in the hands of the traffic men represent short non-coding transcripts mapped by Gro-seq either to the TSSs of the genes or to the regions downstream of the EAGs. The dotted arrows represent the hypothetic (question mark) recycling of the terminating RNAP II to the TSSs of the genes (in good agreement with our experiments and with Mapendano et al., 2010).[9,32]

showing such broad occupancy peaks. Further experiments will be necessary to measure RNAP II transcription speed in these 4–6 kb regions downstream of EAGs to better understand RNAP II behavior in the termination phase. Thus, at present it is probably not accurate to call the increased RNAP II occupancy 3' of the genes "RNAP II pausing."

A comparison of the RNAP II ChIP-seq and Gro-seq data demonstrated that RNAP II is transcriptionally engaged downstream of the EAGs.[9,10] Interestingly, in these regions all the transcripts identified by Gro-seq[10] map in the sense orientation 3' from the EAGs, suggesting that they are produced by RNAP II molecules that have transcribed the pre-mRNAs previously. Moreover, no significant antisense transcription 3' from the EAGs was detected.

These observations further suggest that in most cases after synthesis and cleavage of the pre-mRNA the transcriptionally engaged RNAP II remains on the DNA template for a long distance (4–6 kb), as if it has difficulties in leaving or being released from the DNA template (**Fig. 1A, 1C and 2A**). The increased occupancy of RNAP II in these long non-coding regions downstream of the EAG of poly(A)$^+$ genes suggests that RNAP II might be waiting for a signal that would help to close the transcription bubble and result in RNAP II dissociation from the DNA template. In this respect, it is interesting to note that we were able to separate two clusters of RNAP II patterns (broad and very broad) probably representing two differential modes of RNAP II termination and release. Moreover, our analyses did not reveal any correlation between the number of alternative polyA sites present in a given gene and the broadness of RNAP II occupancy (our unpublished data). In conclusion, from our analyses, it seems that the accumulation of RNAP II downstream of the EAGs of transcribed poly(A)$^+$ genes is rather a genome-wide feature of RNAP II transcription, independent of the cell type.[9]

Another recent study using different bioinformatics tools employing a combination of three peak calling methods concluded that accumulation of RNAP II at the 3' ends of genes is not a global characteristic of RNAP II transcription, but rather occurs on a subset of actively transcribed genes.[11] However, using peak calling algorithms (such as MACS, QuEST and SISSR) may not be the most accurate method to detect very broad and diffuse enrichment peaks (4–6 kb, **Figure 1A and**

C). Nevertheless, Grosso et al.[11] suggest that the formation of 3' peaks is dependent on the transcription rate of the given gene, with highly expressed genes having 3' peaks more often, and that dynamic transitions exist in the relative enrichment of RNAP II at the 3' ends of the genes. Note, however, that this study did not analyze the highly expressed core histone genes, producing non-polyadenylated transcripts, separately (see below).

RNAP II occupancy downstream from the EAGs of core histone genes is different from those producing polyadenylated transcripts

Replication-dependent core histone genes differ from poly(A)$^+$ genes in terms of gene structure and the 3' processing mechanism. Core histone genes lack introns and their transcripts are not polyadenylated. Instead the 3' end formation of core histone gene mRNAs involves stem-loop formation, and includes the U7 snRNP, hairpin-binding protein and specific components of the cleavage and polyadenylation specificity factor complex.[12-16] During our GW data analyses, based upon RNAP II occupancy profiles downstream of the EAGs, we identified distinct patterns of RNAP II binding profiles. One of these clusters showed a rapid drop in RNAP II occupancy after the EAGs (**Fig. 1B and 1D**). Interestingly, gene ontology analyses indicated that this cluster contains almost exclusively core histone genes. As core histone genes are often found in clusters, we tested RNAP II occupancy -/+ 1 kb downstream of histone genes and observed that all core histone genes showed high RNAP II enrichment throughout the gene body and a relatively sharp drop in RNAP II occupancy 3' of the EAGs (**Fig. 1B**).[9] These rapidly diminishing RNAP II occupancy signals are in contrast to the broad RNAP II peaks seen in the majority of expressed poly(A)$^+$ genes (**Fig. 1**). Taken together, our results show that RNAP II occupancy downstream of the EAGs of core histone genes is different from those transcription units producing polyadenylated transcripts (**Figs. 1 and 2**).

At EAGs of core histone genes quickly dropping RNAP II occupancy signals may be of significance since core histone genes are often found in clusters, separated by only short distances (i.e., 0.8–1 kb) (**Fig. 1B and 2B**); thus, RNAP II occupancy may have to decline rapidly to prevent RNAP II from running into the neighboring gene in the cluster. In agreement with the fact that core histone mRNAs have a different 3' processing mechanism than polyadenylated mRNAs, our observations suggest that the two distinct mRNA 3' end-processing machineries differentially influence RNAP II termination on the two types of templates. Thus, it is conceivable that the differential RNAP II occupancy profiles observed on core histone genes vs. poly(A)$^+$ genes is a result of differential interactions between RNAP II and the two 3' processing machineries. In addition, it is possible that the efficiency of RNA processing of the core histone genes also plays an important role in regulating RNAP II occupancy once it has transcribed the processing site.

Inhibition of polyadenylation increases RNAP II occupancy downstream of the EAGs on poly(A)$^+$ genes, but not on core histone genes

It has long been established that intricate connections exist between pre-mRNA 3' end processing and transcription termination (reviewed in Richard and Manley).[17] According to the current views, at least two distinct mechanisms seem to contribute to transcription termination.[18-20] The first involves conformational changes in the RNAP II complex caused by dissociation of elongation factors and/or association of termination factors as the polymerase transcribes through the poly(A) site. The second, called the "torpedo" model, involves rapid exonuclease-mediated degradation of the 5' uncapped RNA produced by RNAP II after cleavage at the poly(A) site. A hybrid of the two models has also been suggested for efficient transcription termination.[21]

Based on these models, we tested whether inhibition of polyadenylation by cordycepin would affect RNAP II occupancy downstream of the EAGs.[9] Cordycepin (3' deoxyadenosine) is an analog of adenosine lacking the 3'-OH, thus preventing further elongation when incorporated into a growing RNA polyA tail, and producing a prematurely-terminated RNA molecule. Interestingly, inhibition of polyadenylation increased RNAP II occupancy downstream of the EAGs of poly(A)$^+$ genes (**Fig. 1C**). In contrast, on the tested core histone genes, inhibition of polyadenylation had no significant effect on the RNAP II signal at the 3' end of the EAG (**Fig. 1D**). These experiments show that for poly(A)$^+$ genes there is a link between polyadenylation and the broad RNAP II occupancy downstream of the EAGs. Thus, our observations further substantiate the link between polyadenylation and RNAP II transcription termination and suggest that transcription termination by RNAP II may undergo different types of regulation depending on the 3' end processing mechanisms of the transcripts.

These observations establish a new type of communication between the polyadenylation machinery and the terminating RNAP II that remains associated with the DNA template downstream of the EAGs. Following inhibition of polyadenylation, the increased RNAP II occupancy downstream of EAGs on poly(A)$^+$ genes suggests that defective polyadenylation can signal to the terminating RNAP II to slow it down or to make it pause more often. Such signaling may be necessary to delay termination on poly(A)$^+$ genes, where polyadenylation would be defective. Under normal conditions, RNAP II occupancy 3' of poly(A)$^+$ genes may be necessary to receive the green light from the polyadenylation machinery indicating that the 3' end processing of the nascent transcript is correct (**Figs. 1 and 2**). Thus, RNAP II deceleration downstream of the EAGs may be implicated in the feedback regulation of 3' processing of the nascent transcripts, where RNAP II 3' occupancy is controlled by the completeness of the polyadenylation reaction and thus, may act as a regulator of RNAP II termination.[9,11,20,22,23] In agreement with this hypothesis, poly(A) signal-dependent pausing as a checkpoint for coupling 3' processing and termination has previously been suggested.[35] Thus, sustained communication of the poly(A) signal to the polymerase over long distances downstream

may be required to trigger and/or control polymerase release.[36]

Additional features influencing RNAP II accumulation 3' of EAGs

Poly(A) signals, terminator sequences and R-loops. Efficient RNAP II termination is supposed to be dependent on both a functional poly(A) signal and a terminator sequence situated downstream of EAGs. Moreover, two terminator-type sequences, which aid in the termination process, have been described: co-transcriptional cleavage (CoTC) sequences, and transcription pause sites and G-rich MAZ termination elements.[17,23-25] While analyzing the significance of the poly(A) signal in the 3' end peak formation of RNAP II, Grosso et al.[11] found that for genes with similar expression levels, the presence of the canonical poly(A) signal appears to promote RNAP II accumulation, whereas the presence of variant hexamers is more frequently associated with the lack of accumulation. This suggests that DNA sequences can also contribute to the process of RNAP II enrichment 3' of the EAGs. Recently, it was also suggested that R-loops formed by DNA/RNA hybrids near G-rich pause sites, downstream of poly(A) signals, may be involved in transcription termination.[26] Nevertheless, further studies are required to investigate whether R-loop formation participates GW in RNAP II termination and/or whether R-loops are involved in the formation of RNAP II peaks 3' of the EAGs.

Gene loops and transcription "factories". Several general transcription factors have been described linking transcription initiation and RNAP II termination.[27-29] Gene looping and communication between the TSS and the 3' end of transcribed genes has also been described.[11,30-33] Thus, it is plausible that for poly(A)$^+$ genes the longer time RNAP II molecules spend downstream of the genes to go from the EAGs to the termination sites, the lesser amount of RNAP II is released from each gene to enter a new transcription reinitiation cycle on the same gene. This hypothesis is in good agreement with our findings that inhibition of polyadenylation not only results in increased RNAP II occupancy 3' of poly(A)$^+$ genes, but also in reduced RNAP II occupancy at the TSSs of the tested genes (**Fig. 2**). Consequently, longer RNAP II pausing in regions downstream of the EAGs may play a role in the establishment of nuclear pools of RNAP II that can be engaged in new rounds of transcription. In line with this hypothesis, the finding that for a subset of expressed poly(A)$^+$ genes, no increase of RNAP II occupancy downstream of EAGs can be detected,[9,11] suggest that rapid RNAP II recycling may be required for this subset of genes.

Recently, it has been described that active genes are linked to different specialized transcription "factories," or foci, in the cell nucleus and the size of a "factory" depends on the level of expression of the genes transcribed in a given "factory."[34] Thus, it is possible that different transcription "factories" might be linked with the different RNAP II pause patterns observed downstream of the EAGs.[9]

Chromatin. To determine whether nucleosomes interfere with RNAP II occupancy downstream of the 3' ends of genes, nucleosome organization in genes with and without 3' peaks were compared.[11] The results show that nucleosome occupancy in the region downstream from the poly(A) site is significantly higher in genes with 3' RNAP II peaks than in genes devoid of RNAP II occupancy. Furthermore, in genes with 3' RNAP II peaks, nucleosomes align at regular intervals from the poly(A) site, whereas nucleosome alignment is much less obvious in genes devoid of detectable RNAP II downstream of the EAGs. Thus, the presence of nucleosomes after the poly(A) site might contribute to RNAP II deceleration, leading to its higher occupancy in these regions. Conversely, it is conceivable that RNAP II moving slowly downstream from the EAGs does not disturb nucleosome occupancy, while quickly moving RNAP II on 3' peaks lacking poly(A) sites does disrupt nucleosome occupancy.

The complex network of interactions between mRNA 3' processing factors, elongation factors, chromatin remodeling complexes, histone chaperones, termination factors, DNA elements and possibly still uncharacterized factors provide numerous nodes for the regulation of RNAP II transcription termination. Many connections have already been established between RNAP II and the 3' processing and termination machineries, but there are still many open questions about transcription termination. As more systems are developed to study these fascinating processes, we expect and look forward to the discovery of additional interactions and/or activities participating in RNAP II termination.

Acknowledgments

We apologize to colleagues whose work could only be cited indirectly due to space limitations. We thank P. Bheda for helpful comments and critically reading the manuscript. A.Gy. was supported by a fellowship from Fondation pour la Recherche Médicale en France. This work was supported by funds from CNRS, INSERM, Strasbourg University, INCA, ANR (GenomATAC ANR-09-BLAN-0266; TFIID complexes ANR-09-BLAN-0052) and European Community (EUTRRAC) grants.

References

1. Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science 2010; 327:335-8; PMID:20007866; http://dx.doi.org/10.1126/science.1181421.
2. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, et al. c-Myc regulates transcriptional pause release. Cell 2010; 141:432-45; PMID:20434984; http://dx.doi.org/10.1016/j.cell.2010.03.030.
3. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 2007; 130:77-88; PMID:17632057; http://dx.doi.org/10.1016/j.cell.2007.05.042.
4. Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, et al. RNA polymerase is poised for activation across the genome. Nat Genet 2007; 39:1507-11; PMID:17994021; http://dx.doi.org/10.1038/ng.2007.21.
5. Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, et al. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet 2007; 39:1512-6; PMID:17994019; http://dx.doi.org/10.1038/ng.2007.26.
6. Core LJ, Lis JT. Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 2008; 319:1791-2; PMID:18369138; http://dx.doi.org/10.1126/science.1150843.
7. Spies N, Nielsen CB, Padgett RA, Burge CB. Biased chromatin signatures around polyadenylation sites and exons. Mol Cell 2009; 36:245-54; PMID:19854133; http://dx.doi.org/10.1016/j.molcel.2009.10.008.
8. Gilchrist DA, Fromm G, dos Santos G, Pham LN, McDaniel IE, Burkholder A, et al. Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. Genes Dev 2012; 26:933-44; PMID:22549956; http://dx.doi.org/10.1101/gad.187781.112.

9. Anamika K, Gyenis A, Poidevin L, Poch O, Tora L. RNA polymerase II pausing downstream of core histone genes is different from genes producing polyadenylated transcripts. PLoS One 2012; 7:e38769; PMID:22701709; http://dx.doi.org/10.1371/journal.pone.0038769.
10. Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, et al. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell 2011; 145:622-34; PMID:21549415; http://dx.doi.org/10.1016/j.cell.2011.03.042.
11. Grosso AR, de Almeida SF, Braga J, Carmo-Fonseca M. Dynamic transitions in RNA polymerase II density profiles during transcription termination. Genome Res 2012; 22:1447-56; PMID:22684278; http://dx.doi.org/10.1101/gr.138057.112.
12. Schaufele F, Gilmartin GM, Bannwarth W, Birnstiel ML. Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of H3 messenger RNA. Nature 1986; 323:777-81; PMID:3022153; http://dx.doi.org/10.1038/323777a0.
13. Schümperli D. Multilevel regulation of replication-dependent histone genes. Trends Genet 1988; 4:187-91; PMID:3070866; http://dx.doi.org/10.1016/0168-9525(88)90074-1.
14. Dominski Z, Marzluff WF. Formation of the 3' end of histone mRNA. Gene 1999; 239:1-14; PMID:10571029; http://dx.doi.org/10.1016/S0378-1119(99)00367-4.
15. Dominski Z, Marzluff WF. Formation of the 3' end of histone mRNA: getting closer to the end. Gene 2007; 396:373-90; PMID:17531405; http://dx.doi.org/10.1016/j.gene.2007.04.021.
16. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet 2008; 9:843-54; PMID:18927579; http://dx.doi.org/10.1038/nrg2438.
17. Proudfoot NJ. Ending the message: poly(A) signals then and now. Genes Dev 2011; 25:1770-82; PMID:21896654; http://dx.doi.org/10.1101/gad.17268411.
18. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. Genes Dev 2009; 23:1247-69; PMID:19487567; http://dx.doi.org/10.1101/gad.1792809.
19. Buratowski S. Connections between mRNA 3' end processing and transcription termination. Curr Opin Cell Biol 2005; 17:257-61; PMID:15901494; http://dx.doi.org/10.1016/j.ceb.2005.04.003.
20. Kaneko S, Rozenblatt-Rosen O, Meyerson M, Manley JL. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. Genes Dev 2007; 21:1779-89; PMID:17639083; http://dx.doi.org/10.1101/gad.1565207.
21. Luo W, Johnson AW, Bentley DL. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes Dev 2006; 20:954-65; PMID:16598041; http://dx.doi.org/10.1101/gad.1409106.
22. Yonaha M, Proudfoot NJ. Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. Mol Cell 1999; 3:593-600; PMID:10360175; http://dx.doi.org/10.1016/S1097-2765(00)80352-4.
23. Gromak N, West S, Proudfoot NJ. Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol Cell Biol 2006; 26:3986-96; PMID:16648491; http://dx.doi.org/10.1128/MCB.26.10.3986-3996.2006.
24. Dye MJ, Proudfoot NJ. Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. Mol Cell 1999; 3:371-8; PMID:10198639; http://dx.doi.org/10.1016/S1097-2765(00)80464-5.
25. Dye MJ, Proudfoot NJ. Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. Cell 2001; 105:669-81; PMID:11389836; http://dx.doi.org/10.1016/S0092-8674(01)00372-5.
26. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell 2011; 42:794-805; PMID:21700224; http://dx.doi.org/10.1016/j.molcel.2011.04.026.
27. Wang Y, Fairley JA, Roberts SG. Phosphorylation of TFIIB links transcription initiation and termination. Curr Biol 2010; 20:548-53; PMID:20226668; http://dx.doi.org/10.1016/j.cub.2010.01.052.
28. Hampsey M, Singh BN, Ansari A, Lainé JP, Krishnamurthy S. Control of eukaryotic gene expression: gene loops and transcriptional memory. Adv Enzyme Regul 2011; 51:118-25; PMID:21036187; http://dx.doi.org/10.1016/j.advenzreg.2010.10.001.
29. Mukundan B, Ansari A. Novel role for mediator complex subunit Srb5/Med18 in termination of transcription. J Biol Chem 2011; 286:37053-7; PMID:21921038; http://dx.doi.org/10.1074/jbc.C111.295915.
30. Krishnamurthy S, Ghazy MA, Moore C, Hampsey M. Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. Mol Cell Biol 2009; 29:2925-34; PMID:19332564; http://dx.doi.org/10.1128/MCB.01655-08.
31. El Kaderi B, Medler S, Raghunayakula S, Ansari A. Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. J Biol Chem 2009; 284:25015-25; PMID:19602510; http://dx.doi.org/10.1074/jbc.M109.007948.
32. Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. Crosstalk between mRNA 3' end processing and transcription initiation. Mol Cell 2010; 40:410-22; PMID:21070967; http://dx.doi.org/10.1016/j.molcel.2010.10.012.
33. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 2012; 148:84-98; PMID:22265404; http://dx.doi.org/10.1016/j.cell.2011.12.014.
34. Eskiw CH, Fraser P. Ultrastructural study of transcription factories in mouse erythroblasts. J Cell Sci 2011; 124:3676-83; PMID:22045738; http://dx.doi.org/10.1242/jcs.087981.
35. Orozco IJ, Kim SJ, Martinson HG. The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. J Biol Chem 2002; 277:42899-911; PMID:12196547; http://dx.doi.org/10.1074/jbc.M207415200.
36. Kim SJ, Martinson HG. Poly(A)-dependent transcription termination: continued communication of the poly(A) signal with the polymerase is required long after extrusion in vivo. J Biol Chem 2003; 278:41691-701; PMID:12933817; http://dx.doi.org/10.1074/jbc.M306304200.

VI. DISCUSSION I.

Additional discussion for Publication I.: RNA Polymerase II Pausing Downstream of Core Histone Genes Is Different from Genes Producing Polyadenylated Transcripts

RNA Polymerase II driven transcription is one of the most highly regulated cellular process, which requires a well orchestrated, sequential action of multiple different protein complexes. Transcription cycle can be divided into three main phases: initiation, elongation and termination, however these phases are also built up from multiple sub-steps, which all represent a checkpoint and a regulatory possibility for proper mRNA synthesis and gene expression.

The transition from transcription initiation into elongation is an inefficient process, where Pol II shows tendency to stop (or pause) after transcribing the first 25-50 nucleotides. This polymerase pause was first described as a rate limiting regulatory step between initiation and early elongation at the promoters of certain *Drosophila* heat shock genes. Recently, genome-wide studies demonstrated that promoter proximal Pol II pausing is a conserved regulatory step that is present at almost every transcribed gene promoters in higher eukaryotes. Surprisingly, paused or slowed down polymerases were also discovered downstream from EAGs. However, in mammalian systems, the genome-wide Pol II studies were mainly focusing on promoter proximal pausing and on investigations of different CTD-phosphorylation patterns on genes. Therefore the full spectrum of genes regulated by pausing downstream of EAGs has not yet been investigated at a genome-wide level. Thus, it was not clear whether this phenomenon is commonly occurring among mammalian genes.

Transcription termination occurs when Pol II ceases RNA synthesis and both Pol II and the RNA molecule are released from the DNA template. Termination serves many vital functions in the cell. It prevents Pol II from interfering with downstream DNA elements, such as promoters and other terminating Pol IIs. In addition, termination can influence the rate of Pol II recycling. Finally, termination is functionally linked to the cleavage and polyadenylation of the nascent RNA transcript. It has been shown that the terminating or slowed down Pol II CTD raises the local concentration of 3’ end processing factors near the nascent transcript and also serves as a platform for the assembly of the processing complexes.

Termination does not occur at a conserved site or constant distance from the 3' end of the mature mRNAs. In mammals, termination can occur anywhere from a few base pairs to several kb downstream from the 3' end of mRNA. Indeed, ChIP assays showed higher Pol II densities downstream from 3' end of genes than throughout the gene body on several model genes, however, it is uncertain, whether it indicates pausing, arrested or slowed down polymerases. In turn, it is known that the polyadenylation signal is essential for termination and certain downstream sequence elements (GU-rich region, MAZ element) further decrease the speed of Pol II downstream of genes.

By analyzing genome wide Pol II occupancy together with gene expression data and Gro-seq data in human cells, we reported that downstream from most of the expressed genes increased presence of transcriptionally active Pol II is a common event. Interestingly, while promoter proximal Pol II pauses show similar profile on single genes (mainly sharp peaks), Pol II densities downstream from genes show often 4-6kb long, moderately dense, non-uniform profiles. This characteristic of Pol II can make the analyses of Pol II behavior downstream of 3' end of genes difficult. The results coming from high-throughput sequencing techniques can be controversial depending on the chosen bioinformatical approach. Most of the publicly available ChIP-seq data analysis softwares are designed to detect and analyze static DNA binding proteins, which give narrow and sharp peaks or histone marks, which show strong, widespread signals. As Pol II "pauses" downstream from EAGs do not form distinct peaks, "peak-finder" algorithms might exclude a large proportion of positive signals and pause sites, therefore using these tools might lead to biased results (Grosso et al, 2012). Currently, there is no appropriate "peak-finder" algorithm, which identifies broad Pol II signals downstream from genes. Therefore it is not advised to involve "peak finding" step during the analysis of 3' end Pol II pauses. By using cluster generating tools together with manually selected genes and regions, one can receive more reliable results. Four our analyses, we selected every gene, which has no neighboring gene in a +/- 4kb distance together with K-mean clustering method to characterize Pol II occupancy profiles downstream from EAGs.

When we subdivided these genes into clusters based on Pol II tag densities and patterns, we found that, Pol II occupancy is a general feature downstream from 3' end of

expressed genes, and can cover a very long region, as ChIP signals can be mapped up to 6 kb downstream of the EAGs. However, by subdividing the expressed genes into additional clusters, we discovered one cluster, showing a narrow Pol II profile at 3' of a set of genes (few hundred bp long Pol II signal). Gene ontology analysis identified the member of this cluster (cluster H) as almost exclusively, replication dependent, intronless, core histone genes. The remaining clusters (PA1, PA2 and PA3) contained genes coding for polyadenylated transcripts [poly(A)$^+$ genes] with broad and very broad Pol II occupancy profiles at their EAG (**Figure 2A and B from publ. 1**).

The observation that the expressed poly(A)$^+$ genes have a widespread Pol II occupancy profile after the 3' of the genes and that core histone genes have a very sharp drop of Pol II occupancy at 3' of their EAGs, suggested that the different 3' processing regulatory mechanisms (described in section 7) may differently influence the residency time and the termination of Pol II downstream of the EAGs. When we extended our analyses to datasets coming from mouse ES cells, we found that the mechanism responsible for the differential Pol II occupancy profiles 3' of the distinct gene categories seems to be conserved in vertebrates.

The 3' end formation of the mRNA of poly(A)$^+$ genes is carried out by the cooperative action of cleavage factors at the poly A site followed by the polyadenylation step. In contrast, the 3' end of core histone mRNA is formed by a cleavage reaction between the stem–loop and the purine-rich sequence, and does not undergo polyadenylation. By taking this fact into consideration, the widespread Pol II profiles downstream from poly(A)$^+$ genes can be explained by the presence of the polyadenylation step, which might be the reason for the increased residency time of Pol II before termination. It is also corroborated by our functional studies, which showed that inhibition of polyadenylation increased Pol II occupancy 3' of the EAGs of poly(A)$^+$ genes, while it had no significant effect on Pol II drop at the 3' end of the core histone genes. This observation suggests that the polyadenylation step is also linked to transcription termination, and a defective polyadenylation step can further increase the residency time of Pol II downstream from EAGs. The postponed polyadenylation and termination can decrease the number of released and reinitiated Pol II. Such a feedback control might be necessary to reduce a transcription on a gene with defective polyadenylation. Our observation is in good agreement with this idea, since we detected

decreased Pol II occupancy on promoters of poly(A)$^+$ genes upon polyadenylation inhibition, but not on histone gene promoters.

The narrow Pol II profiles downstream from core histone genes might also reflect a more effective termination process. Since core histone genes can often be found in clusters, separated by only short distances (i.e. 0.8–1 kb) and thus, Pol II has to terminate rapidly to prevent Pol II from running into the neighboring gene in the cluster.

Rapid termination can be also characteristics of highly expressed genes with quick rounds of reinitiating Pol II. This idea is supported by our findings, when we characterized the 3' end Pol II occupancy profile on the most highly expressed genes, we found a subset of genes which have no or very little detectable Pol II signal downstream from their EAGs (**Figure 3 from publ. 1**). Therefore transcription termination and 3' end processing downstream from this subset of genes might be different and more effective which facilitates a quick round of Pol II recycle at promoters. Another explanation could be that these genes might have shorter poly(A) tail (Meijer et al, 2007), therefore long Pol II residency time is not needed downstream from these genes.

Active poly(A)$^+$ genes (from PA1, PA2 and PA3 clusters) have broad often 6kb long Pol II density profile downstream from their EAGs. The three cluster show difference with respect to Pol II occupancy profile downstream from EAG: PA1 has broad profile (~2-3 kb long), PA2 and PA3 have very broad profile (~4-6kb long) however, PA3 has less mean tag density (**Figure 2A and B from publ. 1**). Gro-seq data suggest that like on the gene body, the Pol II complexes are also transcriptionally active downstream from EAGs too, however, the accumulation of short transcripts suggest a slowed down and very inefficient elongation and transcription during 3' end formation and termination (**Figure 1 from publ.1**). Comparing our GW results with gene expression data, we found that genes from PA1-PA3 having shorter Pol II occupancy profiles show higher expression level. This observation also suggests that highly expressed genes must have more efficient termination processes with shorter Pol II residency time downstream from their EAGs (**Figure 2C from publ. 1**). However, from these genome-wide dataset, we could not distinguish whether one gene has "allosteric" or "torpedo model" type of transcription termination (see section 7).

According to a recent publication, highly expressed genes are transcribed by specialized transcriptional factories, which consist of multiple active Pol II complexes (Eskiw & Fraser, 2011). Thus, it is conceivable that the Pol II accumulation downstream from EAGs

could be a part of these “factories”. In addition Pol IIs downstream from genes might be stably associated to the DNA in order to keep the open chromatin for further rounds of transcription.

In summary, our study was the first to characterize genome-wide the increased Pol II occupancy downstream from 3’ end of genes. Our findings further support that Pol II does not release immediately the DNA template during transcription termination and 3’ end processing. However, additional investigations are needed to decide whether slowed down or arrested Pol II complexes are composing the broad occupancy profiles. A study reported that a subset of genes accumulate Ser5-P Pol II at their 3’ end (Grosso et al, 2012). The authors suggest that these genes are the most highly expressed genes, and the Ser5-P accumulation 3’ is due to gene-looping as they also detected TBP and NELF 3’ of these genes. The authors also suggest that higher histone occupancy contributes to increased Pol II signals at EAGs. However, these observations are not complete, since the authors used “peak finder” tools prior the analyses, thus they excluded a set of potential genes. Therefore it is advised to carry out the same analyses on every non-neighboring expressed gene. Further GW analyses, such as comparing our Pol II results with distribution of elongation factors, nucleosomes, histone-variants and 3’ processing factors might give a better understanding of the mechanism responsible for the increased Pol II residency time downstream from EAGs of expressed genes.

Chapter two: The fate of RNA polymerase II transcription upon UVB irradiation

VII. Transcription is blocked by alterations in the DNA structure

Transcription of DNA encoded instructions results in proteins and regulatory RNAs which coordinate a myriad of chemical reactions. These processes can be threatened by alterations in the DNA structure that interfere with both transcription and DNA replication. DNA damage arises from both environmental and endogenous sources, including genotoxic chemicals, radiation and reactive oxygen species. In addition, DNA molecule has an intrinsic instability. These naturally occurring altered DNA structures can block replication and transcription, giving rise to gross chromosomal aberrations and mutations.

Even at the most basic biochemical level, transcript elongation by Pol II does not occur in a smooth, continuous process. During elongation, several factors such as TFIIS, Elongin, TFIIF, and ELL have to function to reduce transcriptional pausing or to reverse poised Pol II.

Transcription elongation can be disturbed in several ways (Svejstrup, 2007). At the most basic level, it is well known that certain DNA sequences are more difficult to transcribe than others (such as GC rich regions). Moreover, the DNA template is packed into chromatin, and passing through these obstacles creates a new set of challenges for Pol II. In addition, other DNA-binding proteins represent potential obstacles for efficient transcription elongation. Finally, a variety of DNA lesions are known to bias transcription. For example, UV-induced DNA lesions can completely prevent, or reduce, transcription elongation by Pol II both *in vivo* and *in vitro*.

9. Ultraviolet light

The natural source of UV light is the electromagnetic radiation coming from the Sun. As well as other electromagnetic radiations the UV light can be characterized by its wavelength, frequency, and energy (**Figure 14**). Based on the wavelength, the UV light can be divided into UVA (315–400 nm), UVB (280–315 nm) and UVC (<280 nm). The energy and frequency are inversely proportional to the wavelength, therefore UV light carries more energy than visible light (400 nm-700 nm spectrum). UV light can ionize molecules and induce chemical reactions, hence it is a strong environmental mutagen, which can cause immune-mediated diseases and cancer in humans. Fortunately the majority of the extraterrestrial UV radiation is absorbed by the Earth's atmosphere. The ozone layer (O3), oxygen molecules (O2) and water vapor selectively filter out both UVC and UVB radiation. Due to this, almost no UVC radiation penetrates the atmosphere, approximately 90% of the UVB is absorbed, and therefore UVA makes up about 95% of the UV radiation that reaches the Earth. (The mentioned ratios are depending on geographic locations and the time of day.)

In the last 50 years due to the increased amount atmospheric pollutants such as halogens, chlorofluorocarbons, organobromides and chloro-carbons, the ozone layer is continuously depleted which will give less protection against the harmful radiation. By time, more UVB and even some UVC radiation will penetrate the atmosphere, which will result in the increase of UV radiation incidence on the Earth's surface.

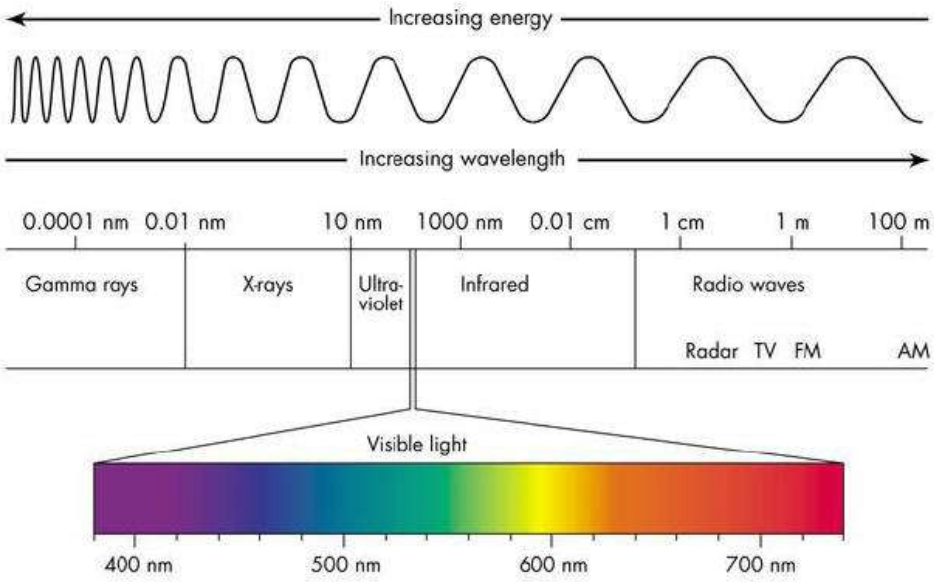


Figure 14: Electro-magnetic spectrum. (Figure adopted from: http://quarknet.fnal.gov/fnal-uc/quarknet-summer-research/QNET2010/Astronomy/)

S. UV-induced DNA damage

The ionizing energy of UV radiation can be absorbed by a variety of molecules in the cells, one of them is the DNA itself.

Despite the fact that relatively little UVB radiation reaches the Earth's surface, and a great proportion is absorbed superficially in the epidermis, it is still more genotoxic than UVA. The reason is that UVB's wavelength can be absorbed by the native DNA. However, UVA can penetrate deeper into the dermis due to its longer wavelength.

UVA has indirect harmful effect on cells. It creates reactive oxygen species (ROS) in the cell by activating some small molecules such as riboflavin, tryptophan and porphyrin. ROS then attacks DNA causing strand breaks or producing oxidative base damage like 8-hydroxyguanine in the DNA (Ikehata & Ono, 2011). ROS also attacks cellular nucleotide pools, producing oxidized nucleotides, which can still be used as nucleotide precursors for DNA synthesis.

These types of oxidative DNA and nucleotide damage are known to be mutagenic. It is also known that at high concentration, ROS can induce damage to cell structure, lipids and membranes, as well as proteins and result in oxidative stress which has been implicated in a number of human diseases (Sekiguchi & Tsuzuki, 2002).

UVB is directly absorbed by the DNA and can cause three major classes of lesions: cyclobutane pyrimidine dimers (CPDs), pyrimidine 6-4 pyrimidone photoproducts (6-4PPs), and their Dewar isomers (Rastogi et al, 2010)(**Figure 15**).

During CPD formation a four-member ring structure is generated between the neighboring bases involving their C5 and C6. 6-4PPs are formed by a non-cyclic bond between C6 (of the 5'-end) and C4 (of the 3'-end) of the involved pyrimidines. In addition, 6-4PPs show tendency to convert into their less stable Dewar isomers. It has been found that T-T and T-C sequences are more photo-reactive than C-T and C-C sequences. However, both thymidines and cytosines are considered as mutation hotspots of UVB and UVC radiations.

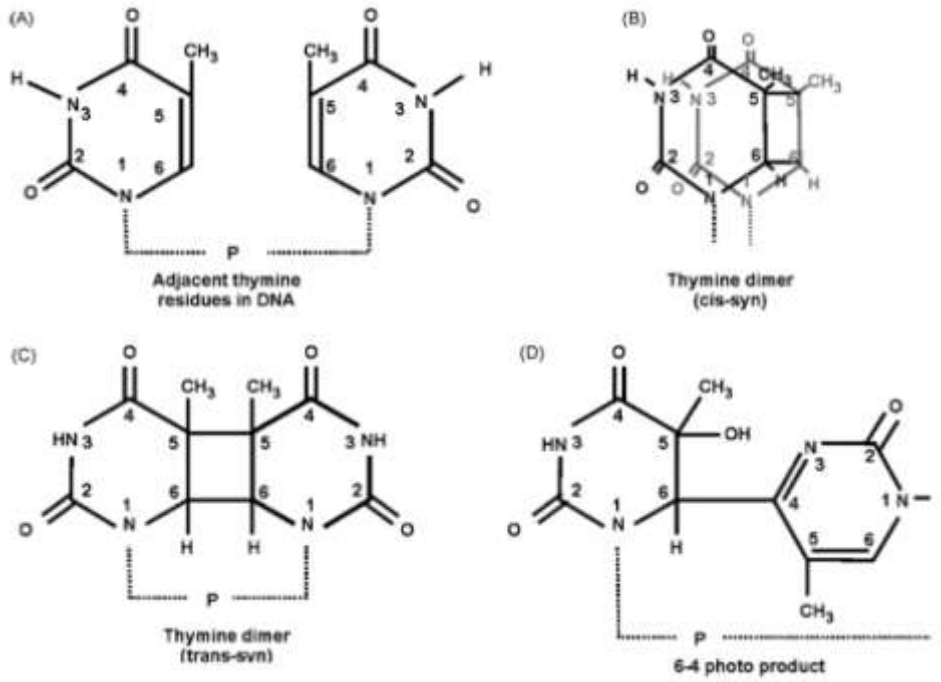


Figure 15: Structures of two main photoproducts. **(A)** Normal adjacent thymine residues in DNA (without UV). **(B)** Cyclobutane pyrimidine dimer between adjacent thymine bases in the DNA (in "cis-syn" form). **(C)** Cyclobutane pyrimidine dimer between adjacent thymine bases in the DNA (in "trans-syn" form). **(D)** Pyrimidine [6-4] pyrimidine photoproduct between adjacent thymine bases in the DNA. Figure adopted from (Tuteja et al, 2009)

Although pyrimidine photoproducts are the most frequent outcome of UVB radiation, the biological importance of UV-induced modifications of purine bases has also been recognized. Upon exposure to UVB radiation, photoproducts may appear, that involve, at least, one adenine residue that undergoes photocyclo-addition reactions with contiguous adenine or thymine (Koning et al, 1990). Overall, CPDs, 6-4PPs, strand breaks, and oxidative products are the predominant and most persistent lesions.

UVC has higher energy level than UVA and UVB due to its shorter wavelength, and can cause all of the above mentioned lesions. However as it is mentioned above, UVC is fully absorbed by the ozone layer, therefore UV related lesions in living organisms are induced by only UVA and UVB.

T. DNA double strand breaks

DNA double strand breaks (DSB) are observed in cells under UVC and UVB irradiation. UV-induced ROS as well as CPD and 6-4 PPs lesions may cause primary, as well as secondary breaks, respectively. These lesions can cause transcription and replication blockage that may lead to production of DSBs due to collapsed replication forks at the CPDs-lesion sites (Limoli et al, 2002). It has been also observed, that UVC-induced DNA DSBs, arising from replication of damaged DNA. Interestingly, a significantly low amount of DSBs was found in the cell where replication was inhibited. It gave the idea that initial photoproducts are converted into DSBs during DNA replication. Experiments showed that increased number of DSBs was observed in DNA damage repair deficient cells in comparison to wild type cells. These results further supported the idea that DSBs are produced during the replication of unrepaired DNA lesions (Dunkern & Kaina, 2002). Overall, it seems that UV-irradiation does not directly produce DSBs, but rather produces pyrimidine dimers and photoproducts which will lead to replication arrest and DSBs.

Several pathways have been considered for the formation of DSBs at a stalled replication fork. It was shown that when the DNA replication machinery encounters a lesion which block replication, DNA polymerase (DP) enzyme is stalled at the blocked site resulting in the formation of a Y shaped DNA structure. This structure is recognized by a specific endonuclease, which makes a nick in the template strand resulting in the induction of a DSB close to the replication-blocking lesion. In addition, replication stress and ROS may trap topoisomerase I and II (Top1, 2) cleavage complexes leading to generation of DSBs by preventing DNA re-ligation processes (Ohnishi et al, 2009). However, the exact DSB generation process is not entirely known and more scientific evidences are needed.

The mentioned lesions above are the predominant and most persistent lesions in the genome. The lack of an adequate repair mechanism and the subsequent accumulation of

lesions will cause severe structural distortions in the DNA molecule, thereby affecting the important cellular processes such as DNA replication and transcription.

10. Persistent blocks initiate transcription coupled repair processes

The UV-induced transcription-blocking lesions cause a particularly challenging problem. In this situation, the action of elongation factors, such as TFIIS is futile, because re-activation of Pol II does not enable it to bypass the damage. Instead, cells have evolved specialized mechanisms that enable lesions encountered by Pol II to be repaired.

U. Nucleotide excision repair (NER)

NER removes the lesions that distort the DNA double helix, interfere in base pairing and block DNA replication and transcription. It can recognize and remove CPDs and 6-4PPs besides several kinds of bulky adducts induced by chemical agents. In eukaryotes, most of NER studies were performed with cells mutated at different steps of the pathway, including cells from human patients with genetic syndromes directly related to DNA repair, such as xeroderma pigmentosum (XP), Cockayne's syndrome (CS) and trichothiodystrophy (TTD). Seven complementation groups were identified for XP syndrome (XPA–XPG) and two for CS (CSA and CSB) (Berneburg & Lehmann, 2001; Lehmann, 2003) (**Figure 16**).

The basic NER mechanisms have been strongly conserved throughout evolution, although the factors involved differ from prokaryotes to eukaryotes. During NER, after DNA damage recognition, the sequential action of helicases and endonucleases open the DNA helix and cleaves the damaged around the lesion. This is followed by the removal of the DNA segment containing the lesion and gap polymerization using the intact strand as template (Compe & Egly, 2012; Costa et al, 2003).

The damage-stalled Pol II molecule is the initiator of the transcription coupled NER, (TC-NER, a subpathway of NER) in which two additional proteins CSA and CSB play an essential role.

m) CSB is a key element of TCR / damage recognition

The CSB (ERCC6) protein plays a crucial role in TCR as the transcription-repair coupling factor. CSB contains seven conserved helicase-like ATPase motifs, which are a feature of the SWI2/SNF2 family of ATP-dependent chromatin remodelers. It has been shown that CSB has intrinsic chromatin remodeling activities *in vitro* that require the hydrolysis of ATP and can directly interact with core histones. The ability of CSB protein to alter chromatin structure could result in disruption of DNA-protein interactions. Based on this, an ATP-dependent chromatin remodeling function of CSB makes it as an ideal elongation factor as well as an important factor in TCR since the stalled Pol II might prevent the assembly of repair machinery at the lesion site (Beerens et al, 2005; Citterio et al, 1998).

When Pol II encounters a DNA lesion, a stable Pol II/CSB complex is formed which indicates the damage-site. CSB promotes the recruitment of NER pre-incision factors. This pre-incision TCNER complex creates a stable bubble structure surrounding the lesion and excises the damaged DNA fragment. CSB is also required for the recruitment of the p300 HAT and the CSA containing E3-ubiquitin (UB) ligase complex to the UV-lesion site (**Figure 16**).

Assembly of the CSA complex is needed to recruit additional TC-NER specific factors: the nucleosomal binding protein HMGN1, TFIIS and the pre-mRNA splicing factor XAB2 to the vicinity of the blocked Pol II/CSB complex. Recruitment of p300 and HMGN1 might facilitate chromatin remodeling and reverse translocation of Pol II allowing the removal of the blocking lesion by the repair machinery and resumption of transcription (Fousteri et al, 2006; Hanawalt & Spivak, 2008) (**Figure 16**).

In vitro studies show that dephosphorylation of CSB in response to UV-irradiation results in increased ATPase activity, indicating a phosphorylation-controlled mechanism of CSB function (Christiansen et al, 2003). After UV-irradiation CSB protein becomes more stably associated with chromatin, but might undergo degradation (Groisman et al, 2006). Interestingly, CSA and proteasome-dependent ubiquitination and degradation of the CSB protein was observed within 4 h after treatment of cells with high UV doses. These results might mean that CSB has to be removed from the DNA template in order to facilitate repair or to allow transcription to resume at a normal rate. However, another study reveals that

under the conditions of high UV dose CSB re-localizes from the chromatin to the soluble fraction (Lake et al, 2010)

n) TCR factor recruitment / damage verification

When a stable Pol II/CSB complex is formed, CSB promotes the recruitment of NER pre-incision factors: CSA, XPA-binding protein-2 (XAb2), TFIIS, high-mobility group nucleosome-binding domain-containing protein-1 (HmGN1), p300, XPG, replication protein A (RPA) and finally TFIIH (**Figure 16**).

- As mentioned above, CSA is part of a multiprotein complex with E3 UB-ligase activity (Kamiuchi et al, 2002).
- XAb2 seems to have a pivotal role in mRNA splicing and in TCR by its specific interaction with XPA. XPA is a protein which is essential for both Global Genome Repair (GGR, second subpathway of NER) and TCR, suggesting that it might be the key protein at the junction of these two subpathways.(Kuraoka et al, 2007).
- TFIIS might be an important player in TCR, as it activates the cryptic nuclease activity of Pol II that cleaves the nascent transcript after backtracking and allows resumed transcription (Lagerwerf et al, 2011).
- HmGN1 is a nucleosome binding protein. During TCR HmGN1 is recruited to the lesion site and facilitates the recruitment of p300, which will increase the acetylation level of residue lys14 in histone H3 and reduces compaction of chromatin (Lagerwerf et al, 2011). HmGN1 might also function to displace nucleosomes that have become re-established behind the Pol II, providing a way for backtracking (Lim et al, 2005).
- p300 is implicated in chromatin remodeling during transcription and repair (Cazzalini et al, 2008).
- RPA binds to single stranded DNA and prevents renaturation of the DNA strands if the transcription bubble collapses.

ATP-dependent remodeling of the arrested Pol II can occur in association with the recruitment of XPG, together with TFIIH and XPF (ERCC1), making the damaged DNA

template accessible to nicking by XPG and XPF-complex. Both XPF and XPG are structure-specific endonucleases (Sarker et al, 2005).

o) Role of TFIIH complex in TCR

The main role of TFIIH in NER is damage verification and opening the DNA around the lesion, which also allows the excision of the damaged oligonucleotide and its replacement by a new DNA fragment.

When bound to the damage-site, two TFIIH subunits XPB and XPD helicases act on both sides of the lesion to unwind the damaged DNA. The two helicases might move on individual DNA strands, so that blockage of either helicase could discriminate between the damaged strand and the undamaged strand (Bootsma & Hoeijmakers, 1993). In a model, XPB-mediated opening of the damaged DNA would allow the binding of XPD to the DNA, and XPD would then utilize its helicase activity to verify the DNA damage and ensure that the backbone distortion is not the result of an unusual DNA sequence (Mathieu et al, 2010). This process is called “enzymatic proofreading”. This model suggests that XPD carries out the verification of DNA damage through its helicase activity. XPB’s additional role in the NER pathway is anchoring TFIIH to the damaged chromatin. XPB ATPase activity is regulated by p8 and p52 subunit of TFIIH (Coin et al, 2006; Reardon & Sancar, 2003).

After verification step, the dual incision around a lesion site is mediated by XPF and XPG. These endonucleases show opposite polarities of substrate specificity: XPF and XPG make incisions 5’ and 3’ to the lesion respectively. The relative cleavage sites vary depending on the lesion type, the length of excised nucleotides is rather constant (24-29 nt). After incision 5’ of the damage-site, XPA and TFIIH are released from the DNA template and recycled, making space for the subsequent DNA synthesis step.

p) Late steps of TCR / DNA synthesis

Following the removal of the damage-containing section, the single-stranded gap left behind is filled by DNA repair synthesis. This step depends on the presence of proliferating cell nuclear antigen (PCNA). PCNA is a homotrimeric clap which is a processivity factor for

eukaryotic DNA polymerases. It is loaded onto the DNA primer end with the help of replication factor-C that is a DNA-dependent ATPase complex. Then, the DNA-gap is filled by DNA polymerase κ and the remaining nick is rejoined by DNA ligase I or III (Araujo et al, 2000; Ogi & Lehmann, 2006) (**Figure 16**).

V. Global genome repair (GGR)

GGR is the second subpathway of the NER process. While TCR removes lesions specifically from ORFs during transcription, GGR operates through the entire genome. GGR differs from TCR only in the damage recognition processes. In GGR the damage recognition complex is formed by XPC, centrin-2 and RAD23B/hHR23B protein. The XPC–hHR23B complex has affinity for a variety of lesions however, it shows lower affinity towards CPDs (Kusumoto et al, 2001). Damage recognition of the XPC complex is enhanced by the activity of UV-Damaged DNA binding protein complex (UV-DDB) which contains DDB1 and DDB2 (**Figure16**) (Sugasawa, 2010)

After recognition and binding of XPC-DDB complex to the lesion both TCR and GGR pathways utilize identical components for the assembly of the repair complex. As in TCR, the removal of the damage is accomplished by excising a short DNA fragment encompassing the lesion. Then the remaining gap is filled by DNA repair patch synthesis using the undamaged strand as template. It has been reported that the repair of heterocromatin and/or non-coding regions are less efficient that the repair of coding region (Lagerwerf et al, 2011).

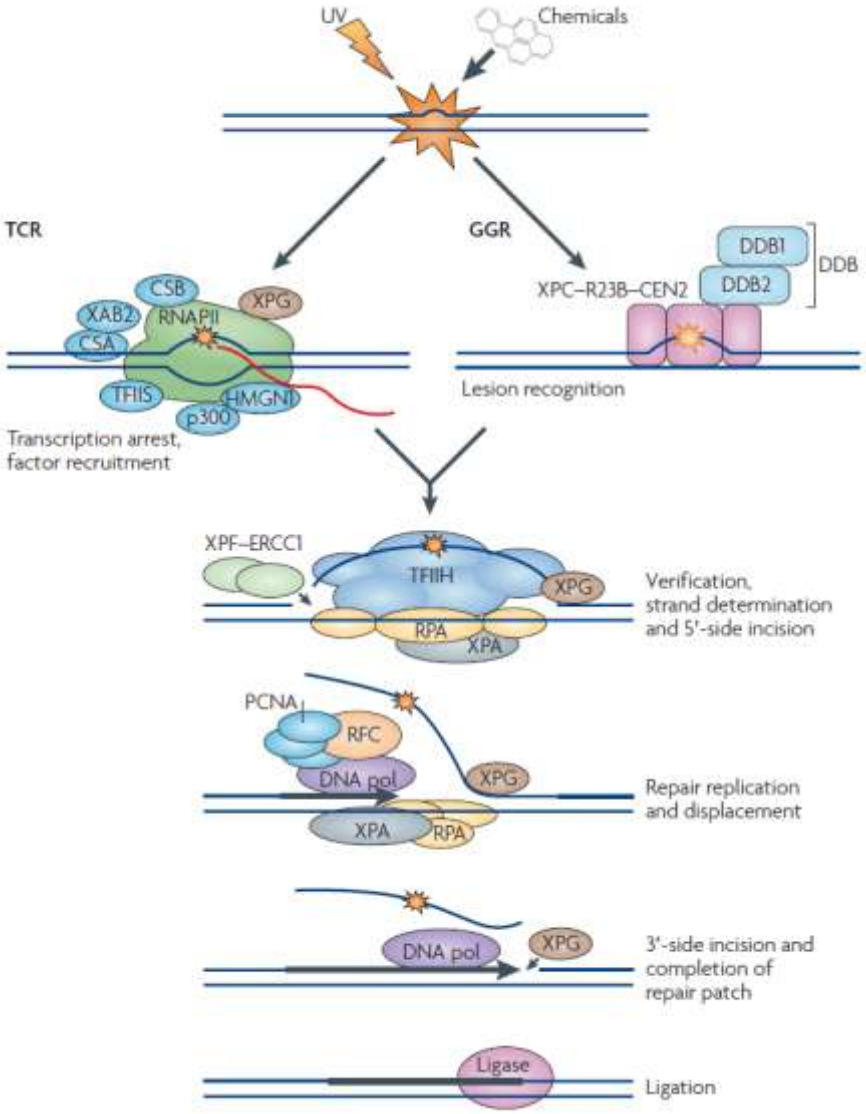


Figure 16: Sequence of events in mammalian TCR and GGR. Lesions are recognized, either by an arrested Pol II/CSB complex (**for TCR, left**) or through the lesion sensor DNA damage-binding-2 (DDB2) which forms a heterodimer with DDB1 to constitute the DDB complex and/or XPC in complex with RAD23B and centrin-2 (CEN2) (for global genomic repair (**for GGR, right**). The subpathways converge at the step when TFIIH is recruited with XPG. Replication protein A (RPA) and XPA might be present before and/or after the appearance of TFIIH. The XPF–ERCC1 endonuclease complex is recruited and incises the damaged DNA strand at the 5' of the bubble, whereas XPG incises on the 3' side. Replication factor C loads PCNA to accommodate DNA polymerase κ that will complete the single stranded template. The remaining gap is filled by a DNA ligase. Figure adopted from (Hanawalt & Spivak, 2008).

A. Single strand break repair

As it was discussed in section 9/S, one of the most common sources of single strand breaks (SSB) is oxidative attack by endogenous reactive oxygen species.

SSBs are primarily detected by PARP1 that rapidly binds to and is activated by DNA strand breaks. This step is followed by the binding of Poly(ADP-ribose) glycohydrolase (PARG), which restores PARP1 to its pre-activated state in preparation for subsequent rounds of SSB detection. The presence of PARP1 and PARG will facilitate the recruitment of the DNA end processing factors (Gao et al, 2007).

The 3'- and/or 5'-termini of SSBs must be restored to conventional 3'-hydroxyl and 5'-phosphate termini for gap filling and DNA ligation. As a wide variety of damaged termini that can arise, this is the most enzymatically diverse step of SSBR, as indicated by the large number of enzymes that are available for this process. The enzymes to repair the damaged termini can be: apurinic–apyrimidinic endonuclease I (APE1), tyrosyl-DNA phosphodiesterase 1 (TDP1), polynucleotide kinase 3'-phosphatase (PNKP), Pol β, aprataxin APTX, and flap endonuclease 1 (FEN1) (Caldecott, 2008). Additionally, XRCC1 binds to and stimulates the end repair enzymes. After the damaged 3'-termini at SSBs have been restored one of the DNA polymerases Pol β, Pol δ, or Pol ε carries out the gap filling process (Caldecott, 2007).

The last step of SSB repair is the DNA ligation step, which is carried out by DNA ligases LIG1 and/or LIG3α and its mitochondrial isoform (Cotner-Gohara et al, 2008).

B. Repair of double strand breaks

In addition to transcriptional arrest, when DSBs are not repaired, they can cause a plethora of chromosomal aberrations that often result in cell death or mutations that can lead to cancer phenotypes.

In eukaryotic cells there are two major repair pathways, which are activated upon DSBs: non-homologous end-joining (NHEJ) and homologous recombination (HR).

During NHEJ, the key proteins are the Ku70–Ku80 heterodimer that binds to DNA ends with high affinity, as well as XRCC4-like factor (XLF)/Cernunnos and DNA ligase IV (Lees-Miller & Meek, 2003). DNA ends are recognized and brought together by Ku70–Ku80. The Ku

heterodimer recruits a plethora of factors, like nucleases: Artemis with DNA-dependent protein kinase catalytic subunit (DNA-PKcs), polymerases: μ and λ, and the ligase complex (XLF with DNA ligase IV). After end processing by the nucleases and polymerases, the ends are directly ligated (Lieber, 2008; Pitcher et al, 2005). NHEJ is highly efficient however its imprecise nature makes it prone to mutations.

HR initiates when the MRN complex recognizes and binds the DNA end. MRN complex consist of Mre11, Rad50 and Nbs1/Xrs2 factors. The DNA ends on both sides of the DSB undergo 5'-3' resection. RPA binds the resulting overhanging 3' single-stranded DNA which can be recognized by RAD51 The Rad51-covered filament initiates the homology search and catalyses strand exchange. This allows the priming of DNA replication to repair the DSB. HR is considered to be an error-free pathway, as it mainly uses the homologous sequence of the sister chromatid as a template for repair (Langerak & Russell, 2011).

NHEJ act mainly in G1 phase and HR in S and G2 phases. This implies that the modes of DSB repair are regulated during the cell cycle. At the same time, cell cycle checkpoints play a critical role in delaying the onset of mitosis until DNA repair is complete.

11. DNA damage signaling

Recognition of non-canonical DNA structures by cellular surveillance proteins initiates DNA damage signaling. As it takes time to repair damaged DNA, it is essential that cell cycle progression can be temporarily stalled. Therefore cells have evolved surveillance mechanisms that are capable of delaying the cell cycle in the presence of DNA damage. These mechanisms, "checkpoints" acts at Gap1/Synthesis, during synthesis (intra-S) and at Gap2/Mitotic phase of cell cycle.

Two kinases, ATR (Rad3) together with ATM (Ataxia-telangiectasia-mutated) are considered to be the key proteins that orchestrate DNA damage signaling and the different repair processes, and each is activated by a distinct DNA structure (Shiloh, 2003). ATR is activated by single strand breaks, forms ATR/ATRIP complex and can bind to RPA-coated ssDNA, thus initiating damage signaling. ATM is activated by the appearance of double strand breaks. ATR and ATM activate additional downstream checkpoint kinases such as Chk1 and Chk2. These kinases can phosphorylate p53 tumor suppressor protein and/or

further regulate cell cycle through cyclin-dependent kinase-cyclin complexes (Poehlmann & Roessner, 2010).

It is long known that p53 phosphorylated at Ser15 rapidly accumulates in response to DNA damaging agents such as UV-light. In addition, it was also shown that the presence of arrested Pol II also positively influences the p53 level (Ljungman et al, 2001). p53 can act as a transcription factor, which up and/or downregulates hundreds of UV-responsive genes. Furthermore, p53 will decide the fate of the cells (inhibit cell cycle, or trigger apoptosis) in a dose dependent manner (Koch-Paiz et al, 2004).

Around the lesion-site DNA-damage specific histone variants (H2AX) and marks will appear. H2AX phosphorylation (γH2AX) at Ser-139 is mediated by ATR, ATM and DNAPKcs kinases. Ubiquitination of histone H2A as well as H2AX is carried out by E3 ligase RNF8 (Bergink et al, 2006; Higashi et al, 2010; Matsumoto et al, 2007). These modifications generate repressive chromatin state, which might silence transcription until the repair is finished.

12. Fate of arrested Pol II transcription

C. Transcriptional pause

A number of possible events can follow arrested Pol II and the binding of CSB. The half-life of an arrested Pol II at a CPD *in vitro* can be ~20 hours, and it extends 10 nt upstream of the CPD and 25 nt downstream (Selby et al, 1997; Tornaletti & Hanawalt, 1999). This creates a roadblock not only for all the transcription on the given ORF, even the blocking lesion cannot be accessed by repair factors. Persistent Pol II arrest, and subsequent transcriptional arrest can initiate checkpoint signaling which can lead to cellular apoptosis (Ljungman & Zhang, 1996). It has been shown, that DNA damage on ORFs leads to transiently repressed transcription. This transcription block is partly mediated by the ATM and DNAPKcs-dependent signaling and by the spreading of the repressive histone marks around the lesion. In addition, Poly (ADP-ribose) polymerase 1 (PARP1) was shown to play a

role in removing nascent RNA and elongating Pol II from the damage-site, by further generating repressive chromatin landscape (Chou et al, 2010; Pankotai et al, 2012).

D. Fate of Pol II during TCR

Pol II stalled at helix-distorting DNA damage is dangerous, because it blocks access of the nucleotide excision repair factors to the lesion (Donahue et al, 1994). Cells have evolved several solutions to deal with persistently stopped Pol II.

For example, Pol II as a damage sensor initiates TCR. During TCR factor assembly, Pol II can backtrack to provide access to the lesion.

When the TCR pathway is unable to allow continued transcription, an alternative pathway is required to remove the stalled Pol II and allow repair factors access to the lesion.

There are several proposed model to describe the different mechanisms which might displace Pol II from the lesion site:

- One mechanism might be for Pol II to bypass the lesions, but this process is slow and extremely inefficient (Walmacq et al, 2012).
- CSB and the TCR factors might cause conformational changes of the Pol II that might be required to allow accessibility to repair proteins.
- CSB might remove Pol II from the lesion site through its Swi/Snf-like activity (Svejstrup, 2002).
- Backtracking of Pol II might be another key mechanism to allow repair and/or transcription restart. Pol II and the transcription-associated DNA bubble might shift backward/backtrack along the RNA from the lesion. The restart of transcription depends on the cleavage and reposition the 3' end of the RNA to the active center of Pol II, which might me mediated by TFIIS. Indeed, there are several evidences showing that TFIIS can stimulate the cleavage activity of Pol II during TCR (Kalogeraki et al, 2005).

However, a second school of thoughts suggest that removal of Pol II is not necessary for TCR (Lagerwerf et al, 2011).

Persistent DNA damage can also result in the poly-ubiquitination and degradation of the largest subunit of Pol II. Such removal would allow repair processes at the lesion. It has been suggested that Pol II degradation pathway is activated when the blocked Pol II cannot be restored, and this pathway is an alternative to TC-NER. Accordingly, TCR impairment can lead to increased Rpb1 poly-ubiquitination and degradation (Chen et al, 2007; Woudstra et al, 2002).

E. Ubiquitination of Pol II

In general, protein ubiquitination was originally discovered as a signal for the degradation of proteins, but it has later revealed that it is a multi-functional post-translational modification (Ikeda & Dikic, 2008). The substrate of ubiquitination becomes conjugated on a lysine, via a triple enzyme cascade (E1: activating, E2: conjugating, E3: ligating). Ubiquitin itself can be extended to create chains of differing topology and promoting different outcomes. Only lysine-48 linked chains targeting proteins directly to the proteasome for degradation.

Mono-ubiquitinated Rpb1 can be found at low levels even in unstressed cell (Harreman et al, 2009). One possibility is that this modification shows stalling Pol II complexes or arrested ones during normal elongation, without always being degraded. When transcriptional arrest is permanent, poly-ubiquitination of the stopped Pol II can occur if needed. However, it is also possible that Pol II mono-ubiquitination has an additional and separate functional role during transcription, many different ubiquitination sites have been identified on Rpb1 (Beltrao et al, 2012; Wagner et al, 2011).

In mammalian cells, several distinct UB ligases can mediate Pol II ubiquitination and degradation. CSA, BRCA1–BARD1, and NEDD4 were shown to mediate mono-ubiquitination of Rpb1 (Anindya et al, 2007; Starita et al, 2005). To generate Pol II poly-ubiquitination, cooperative action of the mammalian Elongin complex together with NEDD4 and an additional UB ligase Cul5/Rbx2 is needed (Kamura et al, 2001).

Ubiquitinated Pol II can be removed from the DNA by Cdc48 ATPase and/or degraded by the 26S proteasome (Verma et al, 2011).

While TCR is the preferred pathway to handle persistently blocked Pol IIs, and Pol II degradation is an independent pathway, they might be interconnected. However, the “decision mechanism” which activates either TCR or degradation is not clear.

RESULTS II.

VIII. Genome-wide characterization of RNA polymerase II behavior upon UVB stress (Manuscript under preparation)

Introduction of the scientific question

The elongating Pol II is very stably associated to the DNA template due to its structural characteristics and the associated elongation factors. However, transcription is not a smooth and continuous process, Pol II frequently pauses and even backtracks, which can cause temporary or persistent transcriptional arrest. Conditions leading to arrested Pol II can be special sequences, non-canonical DNA structures, topological constrains and DNA lesions.

UV light is one of the strongest naturally occurring genotoxic agents. The different wavelengths of UV radiation produce a wide range of DNA lesions (see section 9), which can block DNA-dependent nuclear processes such as replication and transcription. Upon UV-irradiation even double strand breaks will appear throughout the entire genome.

Therefore cells have evolved a large number of factors and mechanisms to maintain and restore genome integrity upon genotoxic stress. These repair pathways can be induced by persistently blocked Pol II elongation complexes on the transcriptional units.

As it takes time to repair damaged DNA, it is essential that the cell cycle progression and transcription are temporary stalled. Mammalian cells have three major DNA repair checkpoints (see section 11), which can arrest cell cycle by the activation of different groups of proteins. These groups of proteins are damage sensors, signal mediators, signal transducers and effectors. The cell cycle can be arrested by the activities of downstream effector kinases: checkpoint kinase 1 and 2 (Chk1, 2; in yeast: Chk1 and Cds1). Studies in yeast showed that Cdc25 is a key target for Chk1. Phosphorylation of active Cdc25 by Cds1 and Chk1 was shown to inhibit Cdc25 activity, which is needed to remove the inhibitory phosphorylation from Cdc2, which leads to arrested cell cycle (Furnari et al, 1999; Sancar et al, 2004).

In addition, Chk2 can phosphorylate p53, thereby inhibiting the binding of Mdm2 to p53 and as a result, p53 protein is not targeted for degradation and is therefore stabilized (Bartek & Lukas, 2003). Arrested Pol II and transcription can invoke checkpoint signaling, which can end up in apoptosis, if transcription is not resumed quickly (Ljungman et al, 1999).

When transcription faces a persistent DNA lesion, Pol II can also be polyubiquitinated and therefore removed from the DNA template in order to provide access for the repair factors to the lesion (Wilson et al, 2012). During the repair process it is also essential to arrest the new rounds of transcriptions until the lesion is removed and the template strand is restored. It was reported that during repair, transcription is blocked on the corresponding ORF by the spreading of repressive chromatin marks together with PARP1 and DNAPKcs activity at the promoter (Pankotai et al, 2012). However, the exact mechanisms which remove Pol II from the template strand and temporary turn off transcription is not fully understood. In addition, most of the studies, which characterized the repair processes and the subsequent transcriptional events, investigated only a handful of model genes. Therefore, it is not known how cells maintain Pol II transcription when multiple genes are affected by DNA damage and subsequent repair process.

In order to investigate the Pol II transcription machinery and its regulation during DNA damage repair, I carried out a genome wide analysis using ChIP-seq technique on UVB irradiated cells.

By following Pol II redistribution genome-wide after UVB irradiation over time, we wanted to see how Pol II transcription alters in single and multiple gene contexts, when DNA lesions appear throughout the entire genome. Our additional goal was to determine the mechanism used by the cell to maintain and to recover Pol II transcription following TCR.

IX. UNPUBLISHED RESULTS

13. Low dose of UVB is sufficient to induce DNA damage response, but is not lethal for cells

F. Testing cell lethality

For our experiments MCF7 human cell line was chosen, because it has wild type (wt) p53, therefore it shows wt stress-induced cell cycle checkpoint activation, and proficient DNA repair (Koch-Paiz et al, 2004). It is known that upon high dose of UVB and UVC irradiation, cells induce apoptosis and degradation of Pol II (Anindya et al, 2007; Koch-Paiz et al, 2004). Four our experiments we had to find a condition, which induces TCR and transcriptional block, but not lethal for cells. Based on the mentioned publications we have chosen and tested the effects of 55 J/m² dose of UVB.

In order to test whether the chosen 55 J/m² dose of UVB is lethal for cells, I carried out Crystal violet (CV) based survival assay. CV is used as nontoxic DNA stain to indicate living cells.

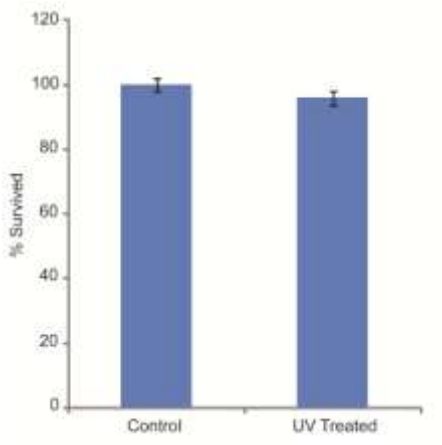


Figure 17: 55 J/m² UVB dose is not lethal for cells. Crystal violet survival assay was carried out to determine whether the selected low dose of UVB is lethal for cells. Percentage of survived cells are calculated from the measured DNA content, from control (left bar) and UV-treated (right bar) sample.

UVB treated cells and untreated cells were incubated for seven days to form colonies, then their DNA content was stained with CV and quantified with spectrophotometer.

The results (**Figure 17**) show that approximately 95% of the UV treated initial cells survived the selected dose, which also suggests that the DNA damage signaling was not inducing apoptotic processes upon this sub-lethal dose.

G. Detection of DNA lesions

To demonstrate the genotoxicity and to quantify the DNA lesions caused by UVB, I used Slot-blot coupled Western blot technique with anti-CPD antibody which recognizes thymine dimers. Human MCF7 cells were treated with 55 J/m² of UVB irradiation, and then DNA was extracted at different time points: immediately, 2 hours, 4hrs, 6hrs, 8hrs, 16hrs and 24hrs after treatment.

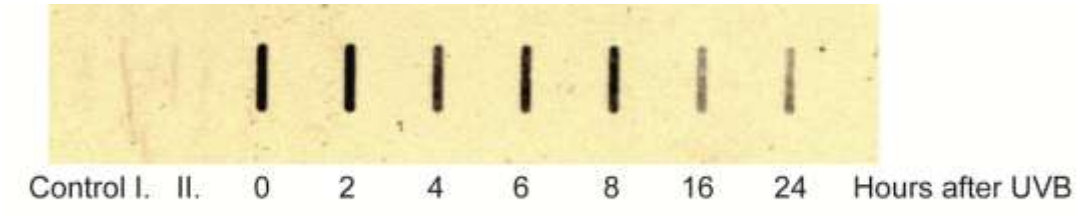


Figure 18: CPDs are formed upon 55j/m² dose of UVB. Slot blot coupled western blot analysis to detect CPDs formed upon 55J/m² UVB treatment. With a CPD specific antibody (TDM-2), we detected a massive accumulation of DNA lesions in the DNA of the treated cells. CPDs are still present 24 hours after irradiation.

The results (**Figure 18**) show that compared to the untreated controls, there is an accumulation of lesions in the genome of the UVB treated cells. Interestingly, the CPD lesions are still present in the genome 24 hours after the irradiation. Compared with the survival assay, it is surprising that cells are still viable with a massive accumulation of CPDs in their genome, which might suggest a difference in the speed of the repair processes on transcriptional units, on coding strands and on intergenic regions. However, this technique is able to detect only CPDs, and other type of DNA lesions might be present as well.

H. Activated DNA damage signaling

To test whether 55 J/m² dose is sufficient to initiate DNA damage signaling, I carried out Western blot analyses to detect if the levels of phosphorylated Chk1 and p53 are increased. For this, I used specific antibodies recognizing phospho-p53 and phospho-Chk1.

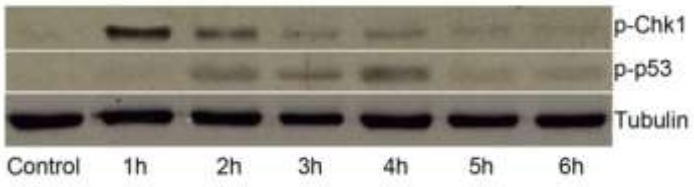


Figure 19: DNA damage signaling is activated. Western blots show that upon low dose of UVB irradiation the level of activated Chk1 (first row) together with phospho-p53 (second row) are increased. Chk1 shows a quick upregulation (maximum at 1h), which slowly diminishes. p53 is activated later, showing a maximum level at 4h. Both proteins are close to the non-treated level at 6h. Tubulin β was used as a loading control.

Our results show that even 55 J/m² dose UVB irradiation is enough to activate the checkpoint signaling (**Figure 19)**. Following UVB irradiation, the level of phospho-Chk1 increases rapidly, showing a maximum level at 1h, and then it slowly diminishes. The activation of p53 is slightly slower, it shows increased levels at 2h after treatment, a maximum at 4h and a fast decrease from 5h. The phosphorylation levels of both proteins are restored to normal at 6h.

I. Detection of p21 activation (I.)

CDKN1a/p21 protein is a cyclin-dependent kinase inhibitor which is also UV and p53-responsive gene. The expression of p21 is tightly controlled by p53. p21 binds to and inhibits the activity of CDK2 and thus functions as a regulator of cell cycle progression.

To test whether DNA damage signaling is activated by p53, I carried out ChIP qPCR assays at different time-points (4h and 8h after UVB) to follow the binding of p53 to its responsive element upstream from p21 TSS.

Our ChIP carried out with p53 AB, shows increased p53 occupancy at the upstream p21-p53 responsive element 4 hours after UVB treatment compared to the control, which declines at 8h time point (**Figure 20**). On the negative control region, as expected, we could not see increased p53 binding.

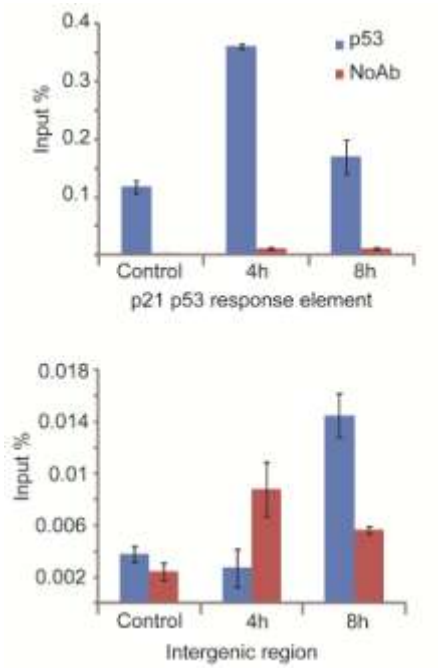


Figure 20: p53 binds and activates p21. ChIP was carried out with p53 AB (blue bars) and with No antibody control (red bars) at 4h and 8h after UVB treatment. p53 occupancy compared to the No AB negative control at different time points are represented in input%. 4 hours after irradiation p53 is bound to the upstream p21-p53 responsive element (upper panel). On the Intergenic region, which is a negative control region, there is no increased occupancy of p53 as it is shown by the low input % (lower panel).

J. Detection of p21 activation (II.)

CDKN1a/p21 is one of the best studied UV responsive model genes. It shows an increased mRNA level in response to irradiation with any wavelength of UV (Koch-Paiz et al, 2004).

To follow the mRNA changes of p21 gene, I carried out reverse transcription coupled quantitative PCR (RT-qPCR). Oligo pairs were designed to measure both pre- and total mRNA level. By measuring pre-mRNA amounts of given transcripts, we can gain information whether the transcription is active at the investigated time-point. RNA samples were extracted in every hour up to six hours after treatment. Our results show that 55 J/m² UVB induces the level of both pre and total p21 mRNAs. Interestingly the p21 pre-mRNA has a maximum at 3h time-point then rapidly decreases and normalizes at 6h. In contrast, the total mRNA level of p21 shows a slow increase until 4h then slowly decreases (**Figure 21**). This suggests that p21 transcription is active only for a short period of time and peaks at

3hours and the increased total-mRNA level at 4h is due to mRNA stabilization and not to active transcription.

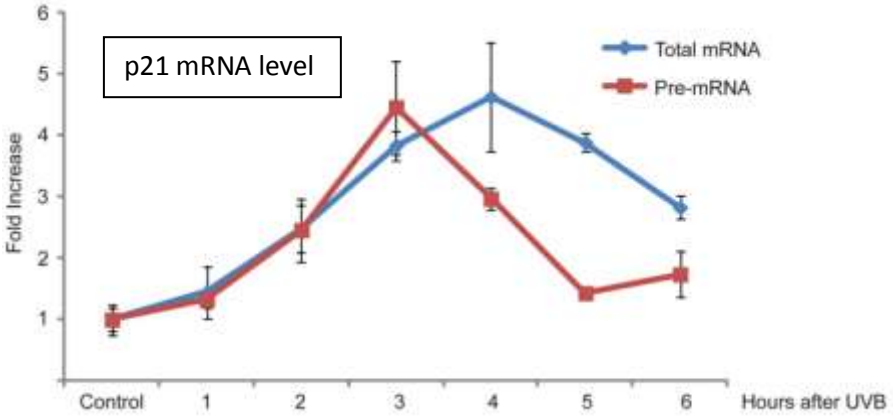


Figure 21 Transcription of p21 gene is active only for a short period. Pre- (red line) and total-mRNA (blue line) level of CDKN1a/p21 was measured with RT-QPCR at the indicated time-points after UVB treatment. Pre-mRNA of p21 shows constant increase until 3hours after irradiation then rapidly declines, showing a short period of active transcription. Total-mRNA has a maximum at 4h time-point and declines slowly which can be due to increased RNA stability.

14. Genome-wide analysis of Pol II behavior upon UVB irradiation

We have demonstrated that 55 J/m² UVB can induce DNA damage and the subsequent signaling response. In order to investigate the effect of sublethal dose of UVB on the global Pol II transcription machinery, I carried out ChIP-seq analyses at different time points after UVB treatment. ChIP was carried out on a non-treated sample and on samples treated with UVB and incubated for 1h, 2h, 3h, 4h, 5h and 6h with an antibody which recognizes the N-terminal end of human RPB1 subunit of Pol II (N-20; Santa Cruz, H-224X). This antibody allows the detection of Pol II independently of the phosphorylation status of the CTD. To test for non-specific binding, we carried out a control ChIP-seq using an antibody raised against a yeast factor that does not recognize any human proteins.

The high throughput sequencing of the precipitated DNA samples was done by Hi-Seq2000 platform.

ChIP-seq experiment can be divided in to the following steps:

1) ChIP;

2) Library preparation from chipped DNA fragments (end repair; addition of an 'A' base to the 3'-end of DNA fragments and ligation of adapters to DNA fragments)
3) Amplification of adapter-modified DNA fragments and gel purification
4) Library sequencing with Hi-seq 2000 platform

From the sequencing results identical reads were removed to avoid artificial signals (that may come from the library preparation step). This was done by mapping the sequences to the human genome (version hg19). Only uniquely mapped reads were considered for further analyses (**Table 2**).

Sample	Total number of unique reads
Control	119,548,359
1h	135,240,416
2h	131,965,772
3h	117,791,061
4h	109,408,317
5h	127,410,622
6h	129,616,209

Table 2.: List of samples/datasets and the number of uniquely mapped reads.

For the comparative analyses, the samples/datasets were normalized based on the tag densities on the intergenic region (see Materials and Methods). To generate average profiles, annotated genomic regions (reported and validated gene coordinates) and coordinates were extracted from the refseq genome database.

K. Genome-wide average Pol II occupancy profile

Pol II tag densities were collected and counted for every human refseq annotated gene (43070 annotated ORFs). Occupancy profiles were calculated from -2kb upstream from TSS until EAG+4kb around the genes (**Figure 22**).

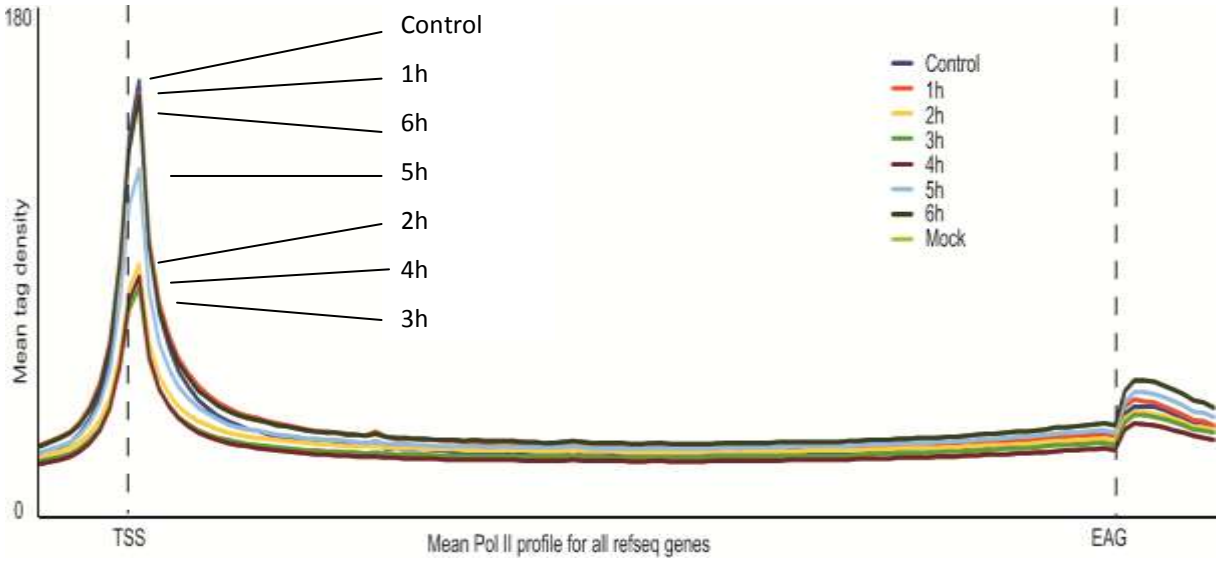


Figure 22: Mean Pol II profile for every refseq genes. Chip-seq data using an anti-Pol II antibody (N-20) was carried out in human MCF7 cell line at the indicated time points after UVB treatment. Pol II tags were calculated from ChIP-seq datasets in every time points on every annotated gene (43070 human ORFs) from -2kb from TSS through the gene body until EAG+4kb region.

Interestingly, the mean occupancy profiles show a massive Pol II clearance from the promoters in the 2h, 3h and 4h samples. Pol II occupancy starts to recover in the 5h sample and seems to be fully recovered in the 6h sample. On the gene body and downstream from EAG we found only a slightly Pol II clearance. The loss of Pol II from promoters in the 2h, 3h and 4h samples suggests that upon UVB irradiation Pol II reinitiation rate is slowed down or blocked.

The datasets from the different samples were also uploaded to the UCSC genome browser to visualize them (**Figure 23 and supplementary Figure 1**). By selecting randomly a genomic region we observed the same Pol II signal clearance from promoters.

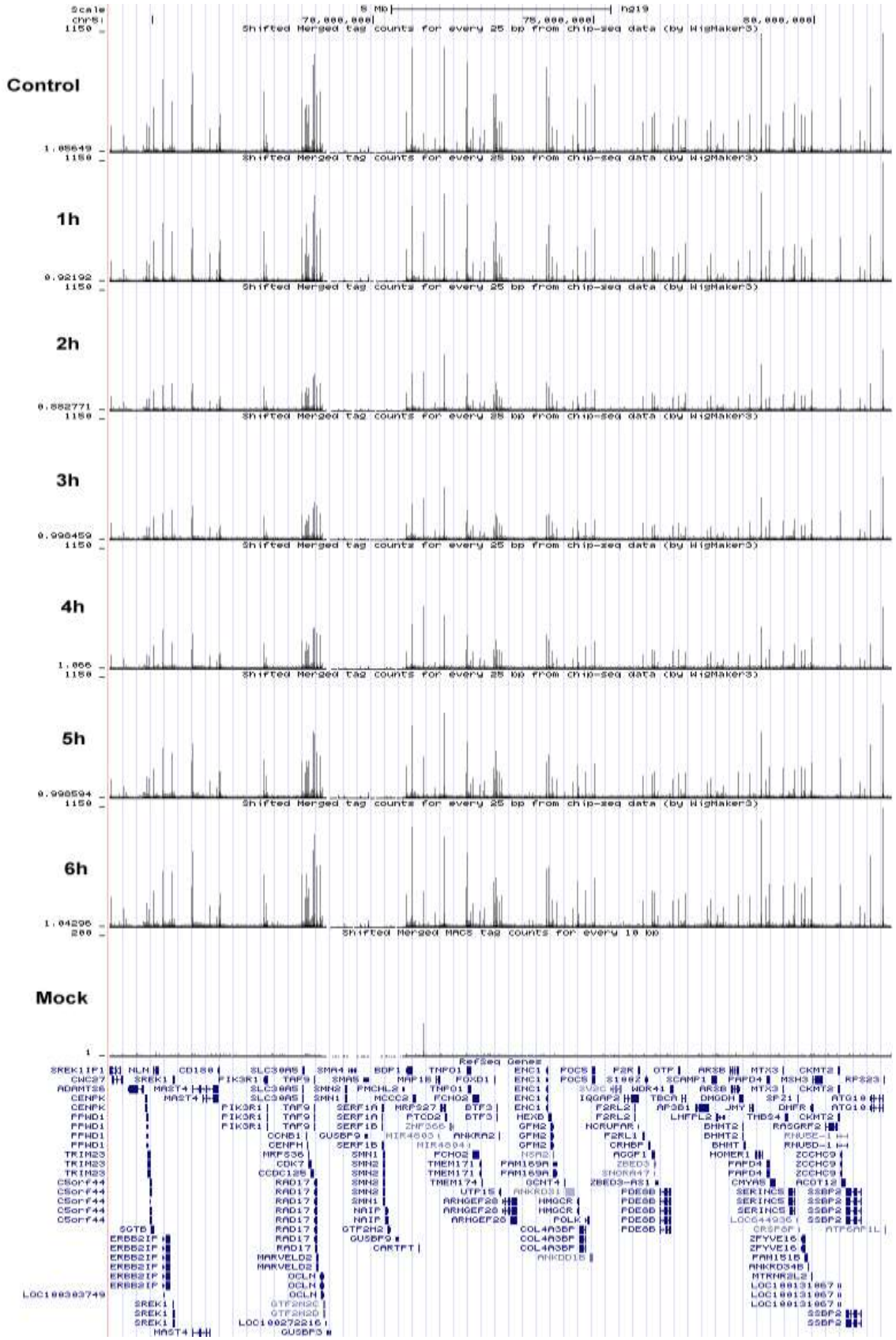


Figure 23: Pol II clearance from promoters. The ChIP-seq datasets are visualized by UCSC genome browser. Pol II signals are represented as black bars for every time point (Control - 6H). The ORFs and their location are represented as blue boxes.

L. Average Pol II occupancy profile of 4500 expressed genes

As it was discussed earlier (see section 12) during TCR, the new rounds of transcription can be blocked. To investigate the effect of UVB treatment on the active transcription, Pol II tag densities were extracted from the datasets from the following regions: TSS (+/- 300bp around TSS), on the gene body (-100bp from TSS until EAG) and downstream from EAG (from EAG to EAG+4kb (genomic regions are shown on **Figure 24**) of 4500 expressed genes.

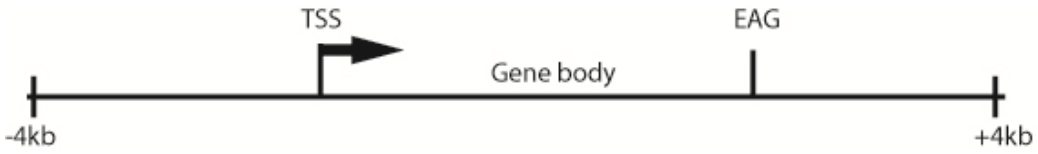


Figure 24: Schematic representation of a coding region.

From the collected tag numbers heat map and mean density profiles were generated to visualize and follow the changes Pol II distribution on expressed genes in time upon UVB treatment (**Figure 25 and 26 respectively**). The genes on figure 25 are organized according to their expression level.

Our results show that the UVB treatment has a dramatic effect on the transcription of active genes. Compared to the control sample, we observed a gradual decrease of Pol II occupancy from the TSSs of genes up to 3h time point. Pol II occupancy slowly regenerates from 4h up to 6h time point. Interestingly, on the gene body, at the 1h time point there is a slight increase in the Pol II density. These observations suggest a quick rearrangement of factors bound to promoters, which might end up in transcription inhibition. The increased Pol II density on the gene body at the 1h time point might suggest that the "last" initiating Pol II molecules move into the coding region (before the complete block of transcription). Later at 3h and 4h the Pol II signal also decreases from the gene body, then normalizes at 5h and even increases at 6h. The increased Pol II occupancy on the gene bodies at 6h suggests that the transcription apparatus might compensate for the reduced gene expression associated with TCR. This idea is supported by the alterations of Pol II signal downstream from EAGs. The slow and minor Pol II clearance from the gene bodies at 3-4h suggests a decrease in the number of terminating Pol IIs. This is followed by a strong increase of Pol II

density at 5-6h which might be a signal of increased transcription. Interestingly, the gene expression level does not influence the speed of Pol II clearance and regeneration.

After K-mean clustering analyses (which sorted the genes into distinct groups based on Pol II occupancy pattern and density), we found a set of genes which do not show dramatic Pol II loss upon UVB treatment. Moreover, they seem to be upregulated as they have increased Pol II densities on their ORF and downstream from their EAG (**Figure 27**, indicated clusters, the list of genes can be found in Supplement results section). Interestingly, Gene Ontology analyses suggest that these genes are involved in DNA damage response and signaling (Table 3).

Identified cellular process	P-Value
Transcription factor binding	1.3E-9
DNA damage response, signal transduction by p53 class mediator	6.3E-7
Regulation of cell cycle	8.7E-6
DNA damage response, signal transduction	4.6E-6
DNA integrity checkpoint	1.5E-4
G1 DNA damage checkpoint	2.2E-4
Response to UV	9.5E-6

Table 3.: The identified biological processes (at significant P-values) by the Gene Ontology analyses of the detected upregulated clusters.

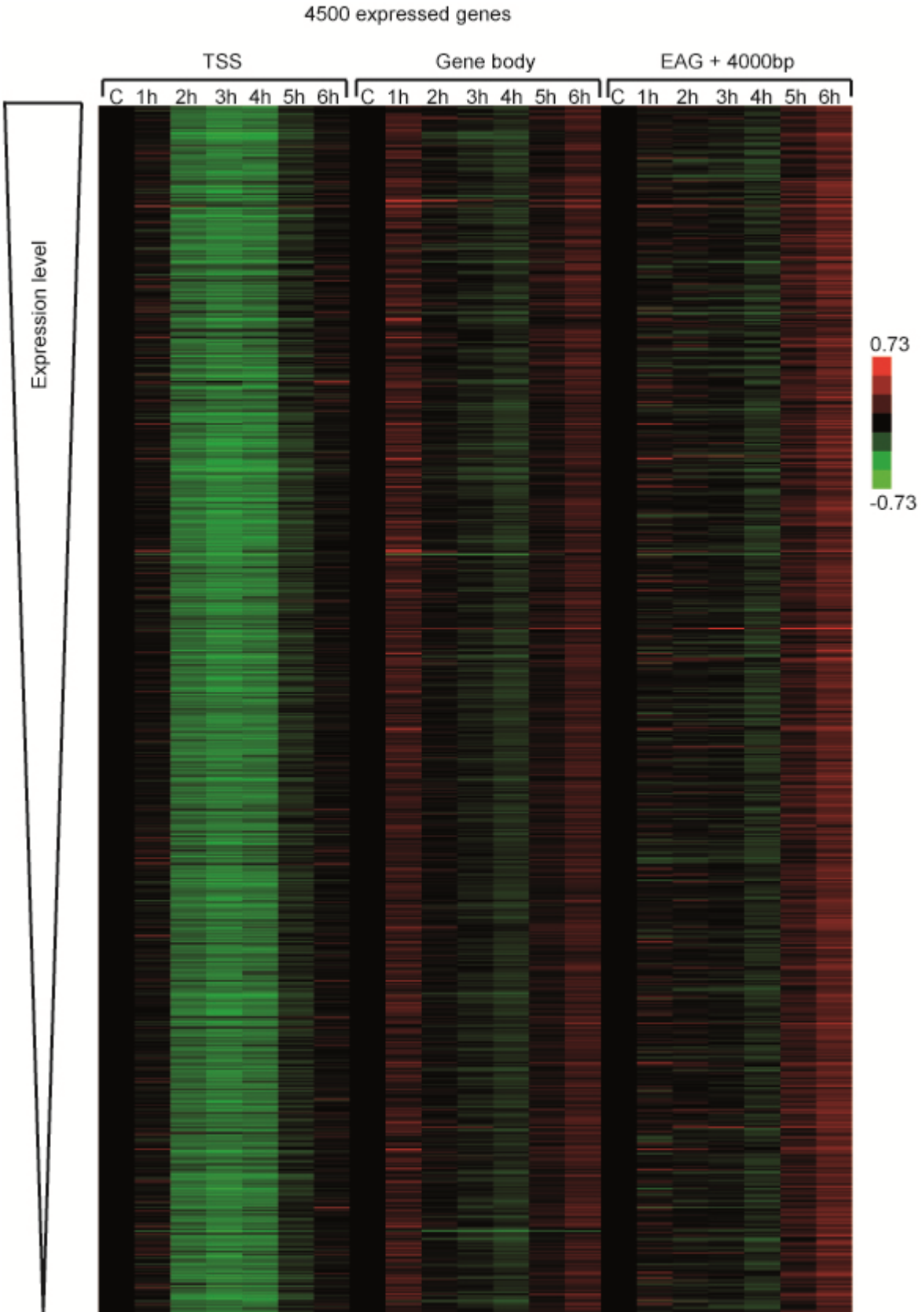


Figure 25: Pol II density changes in time on 4500 expressed genes after UVB irradiation I. Heat map was generated to follow Pol II behavior in time on 4500 expressed genes at different genomic regions after UVB treatment. Tag numbers were calculated around the TSS (+/- 300bp around TSS), on the gene body (-100bp from TSS until EAG) and downstream from EAG (from EAG to EAG+4kb). Each line represents one gene. Color code represent fold change (red shows increase; green shows decrease, black is the control).

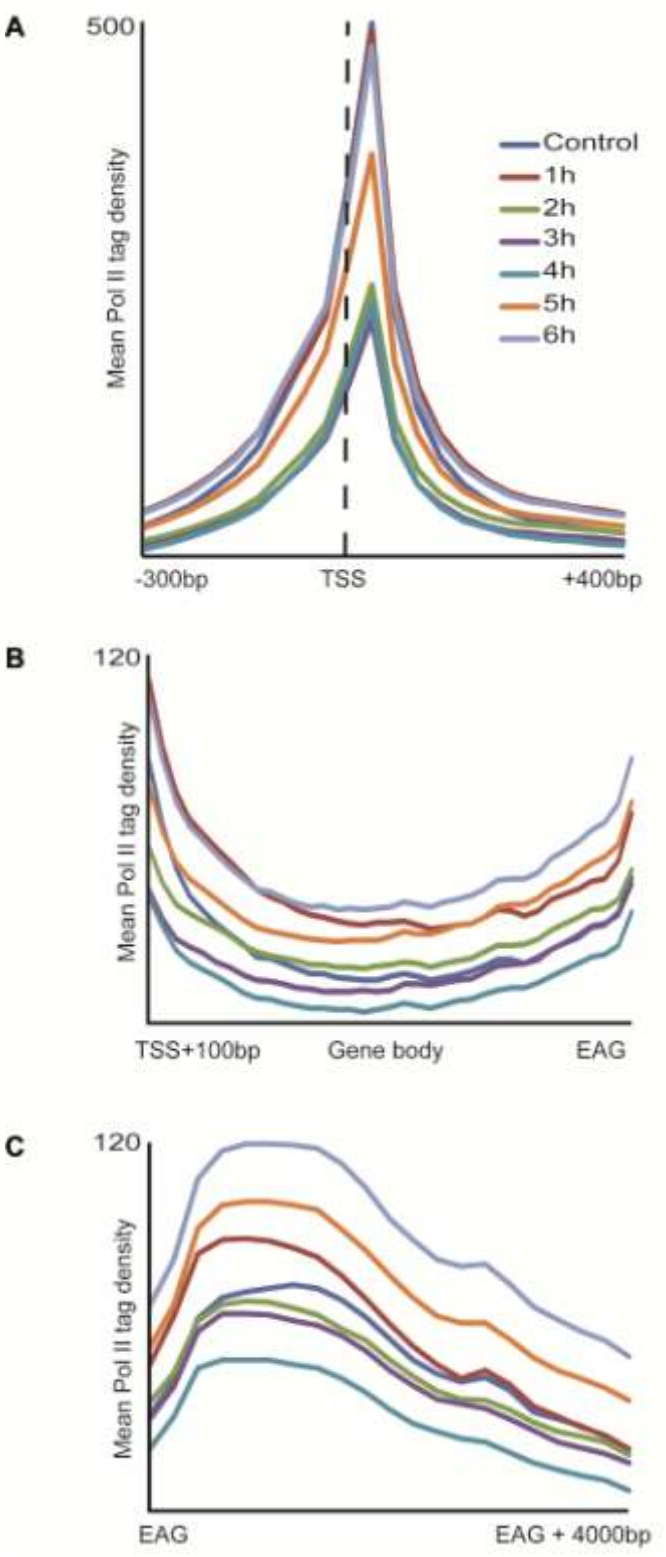


Figure 26: Pol II density changes in time on 4500 expressed genes after UVB irradiation I Mean Pol II density profile was generated for the top 4500 expressed genes at different genomic regions after UVB treatment. Tag numbers were calculated around the (**A**) TSS (+/- 300bp around TSS), (**B**) on the gene body (-100bp from TSS until EAG) (**C**) and downstream from EAG (from EAG to EAG+4kb).

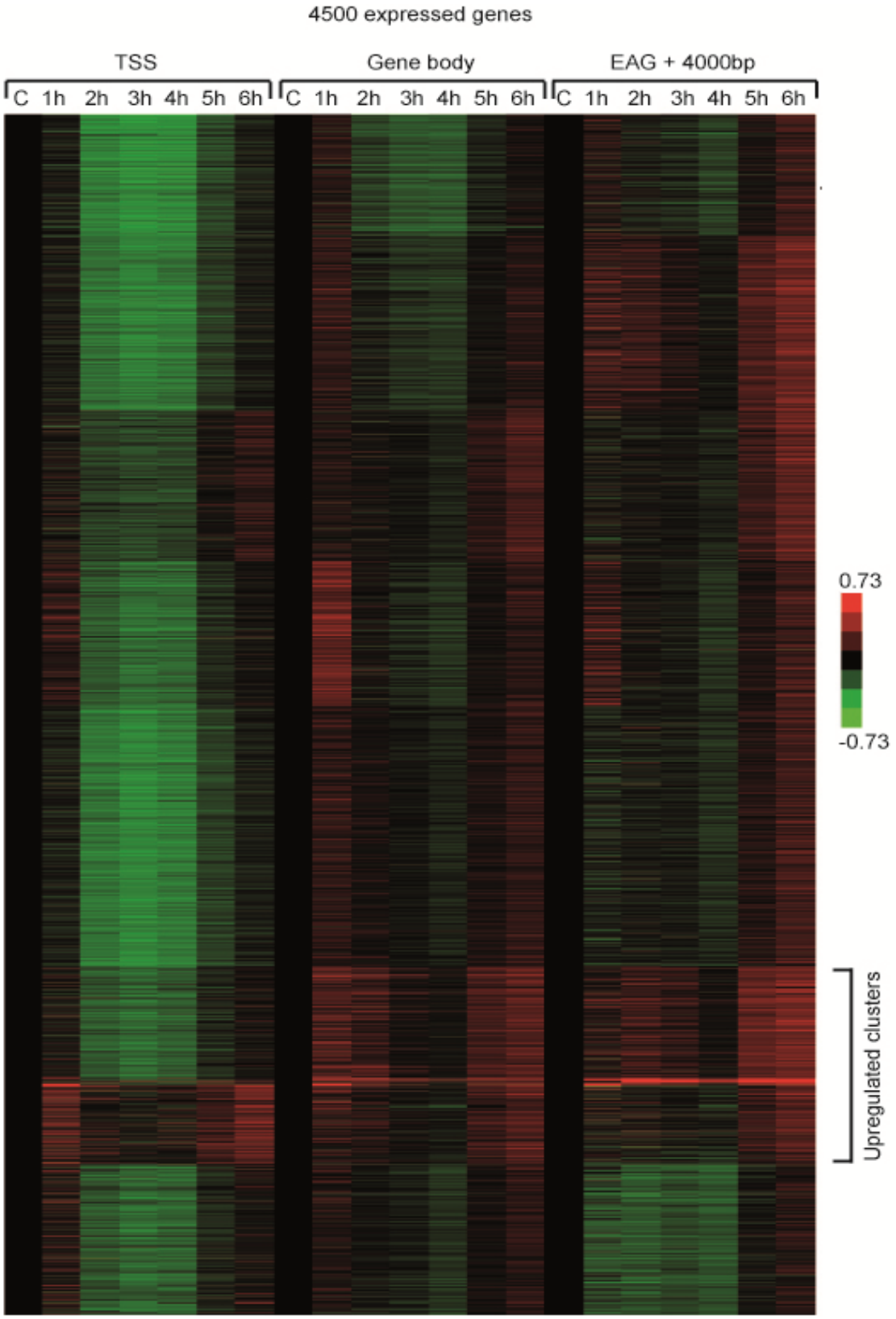


Figure 27: A set of genes show increased Pol II density at the TSS, on the gene body and downstream from the 3' ends upon UVB irradiation. Heat map was generated from the ChIP seq results of 4500 expressed genes after K-mean cluster generation. Tag numbers were calculated as described at figure 25. Each line represents one gene. Color code represent fold change (red shows increase; green shows decrease, black is the control). The identified clusters are indicated as "Upregulated clusters".

To validate the bioinformatically detected Pol II clearance from promoters, I carried out ChIP coupled qPCR detection. ChIP was carried out with Pol II and NELF-A (as it is an indicator of paused Pol II; see section 5) antibodies in non-treated samples and in samples incubated for 3h and 6h after 55 J/m² UVB irradiation. Oligos were designed to amplify Pol II and NELF-A signals on the promoters of three randomly selected genes from the list of the previously used 4500 expressed genes. To have a negative/mock control, ChIP was carried out with empty G Sepharose beads (NoAb), and for a control region, additional oligos were designed, which target an intergenic region (**Figure 28**).

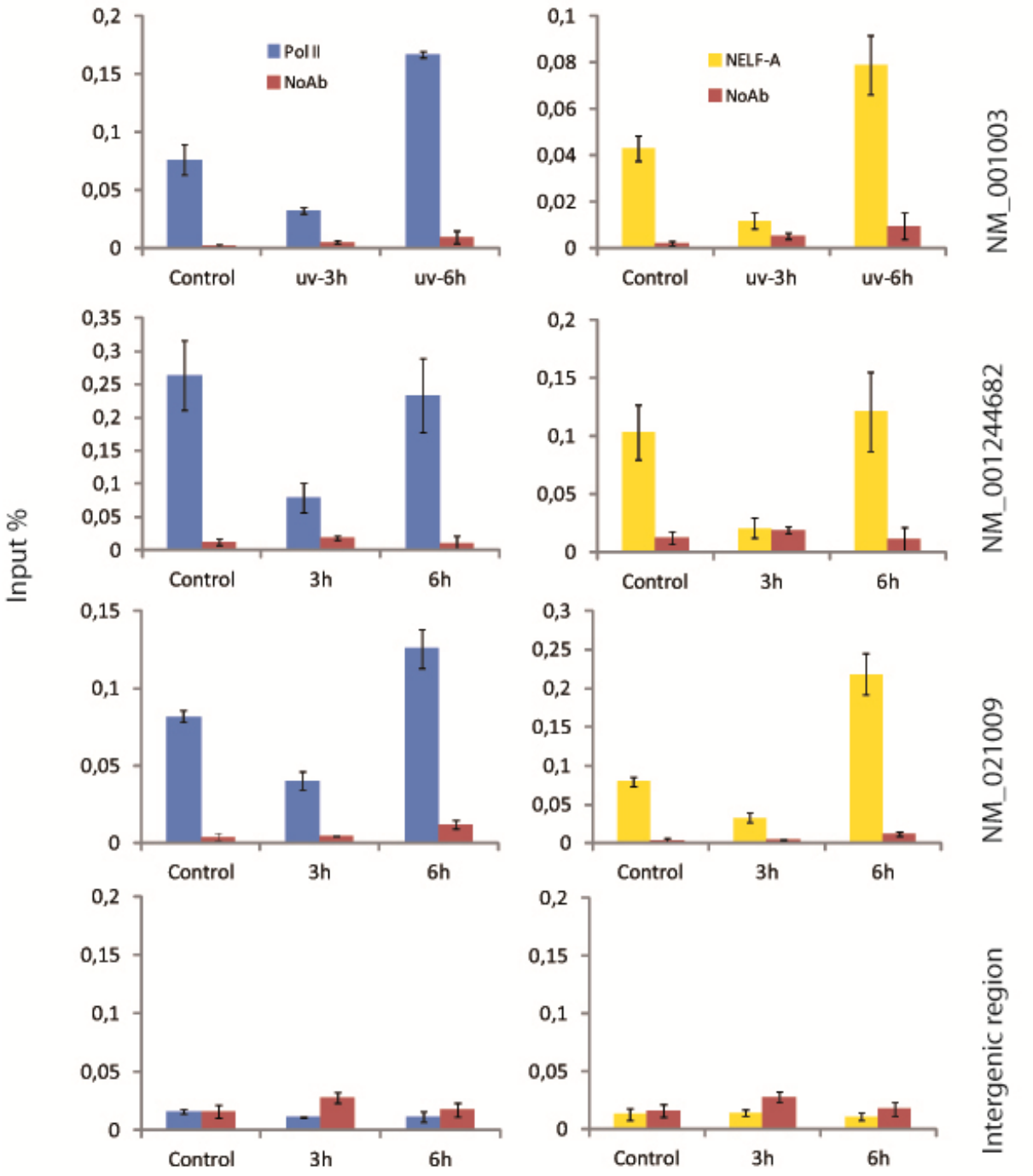


Figure 28: ChIP-qPCR validation of Pol II clearance from promoters. ChIP was carried out with Pol II, NELF-A antibodies and with empty G Sepharose beads. Pol II occupancy (Blue bars) and NELF-A occupancy (Yellow bars) compared to the No-antibody control (Red bars) at the promoters of randomly selected genes are represented in input %. Error bars represent +/− standard deviations. (Gene codes: Ubc/NM_021009, Pou2f3/NM_001244682 and Rplp1/NM_001003)

These validation experiments confirmed the detected Pol II clearance from the promoters of expressed genes. All three candidate genes show decreased Pol II occupancy in the 3h sample compared to the control. As expected, all candidate genes show an increased Pol II occupancy on their promoter region in the 6h sample. These results are in good

agreement with our bioinformatical results. Interestingly, NELF-A shows same changes in its occupancy profile as Pol II at the promoters. This observation suggests a rearrangement and/or dissociation of transcription factors from promoters and the lack of promoter proximal pause of Pol II.

Global level of Pol II

To test whether the Pol II promoter clearance at 3-4h time points is due to Pol II degradation (discussed in section 12) and/or different Pol II CTD phosphorylation status, I carried out western blot analyses to measure the global level of Pol II (RPB1) and the different phosphorylated forms of Pol II CTD. I used antibodies specifically recognizing Ser2-P, Ser5-P and Ser7-P. To measure the total level of Pol II (Pol IIa and Pol IIo together) I used the N-terminal specific antibody. Tubulin-β was used as loading control (**Figure 29**).

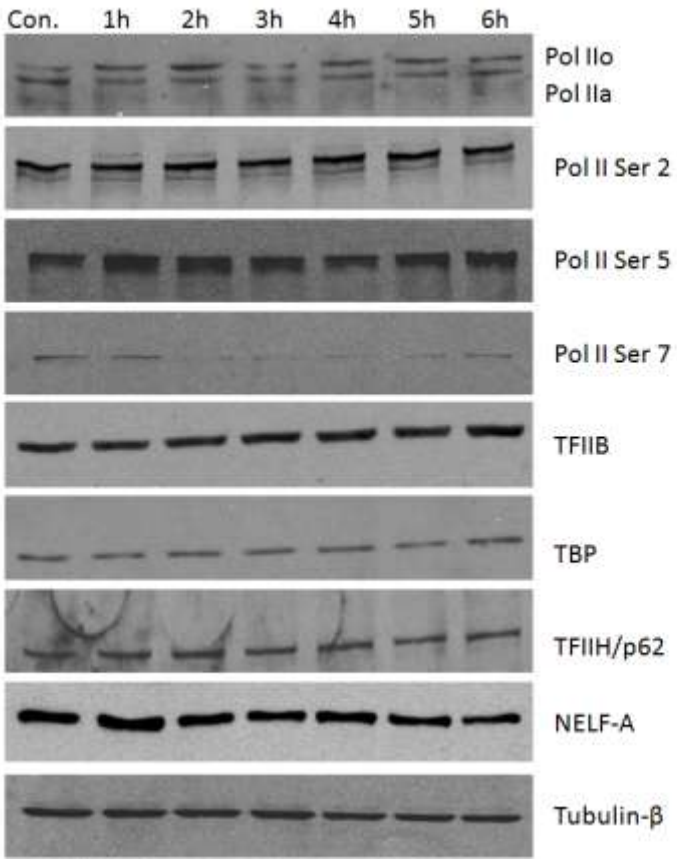


Figure 29: Global level of Pol II and PIC subunits after UVB treatment. Western blots for different Pol II phospho-CTD forms (Ser2, 5, 7-P and total Pol II), PIC subunits (TFIIH/p62, TFIIB and TBP), and negative elongation factor (NELF-A) at different time points after 55 J/m² UVB irradiation. Tubulin was used as a loading control.

The analyses with the N-terminal specific AB [1st panel] suggest that upon UVB irradiation, the level of the Pol IIo form immediately increases compared to control [compare lane 1 to 2 and 3 (1h, 2h)]. Later (5h-6h, lane 6 and 7) the level of the Pol IIo decreased and the level of the two form is balanced. This result is in good agreement with previous studies showing that upon DNA damage Pol II is hyperphosphorylated (Munoz et al, 2009). However, the Ser2-P of Pol II CTD does not show any alteration upon UVB irradiation [2nd panel]. Phosphorylated Ser5 residue of Pol II CTD [3rd panel] shows a slight increase immediately after UV irradiation (lane 2 and 3) which might be due the increased level of Pol IIo form. Surprisingly, we detected decrease (from 2h time point) in the level of phospho-Ser7 signal upon UV treatment [4th panel] which reappears at the 6h time point. Ser7-P CTD mark was suggested to be a marker of highly expressed genes (Egloff et al, 2007). Taking this into consideration together with the WB results, we suggest that upon low dose of UVB irradiation the loss of Pol II from promoters might be facilitated by hyperphosphorylation (reduction in the number of Pol IIs capable to reinitiate) and not by RPB1 degradation. Our results show that the increased level of Pol IIo [1st panel, at 2-3h in] might be related to the increased Ser5-P [3rd panel, 1-2h](**Figure 29**). Moreover, Ser7-P mark is indeed an indicator of active transcription, and its decreased level (from 2h up to 5h) upon DNA damage reflects the transcriptional pause during DNA repair.

To test whether the degradation of certain PIC subunits facilitates the Pol II clearance from the promoters upon UVB irradiation, I tested the global protein level of p62 that is a TFIIH subunit, TBP and TFIIB. Our results do not show changes in the levels of the indicated proteins. In addition, we did not see changes in the level of NELF-A upon UV treatment either. These results suggest that the loss of Pol II signal from promoters might be due to the depletion of Pol IIa form that is capable to incorporate into the PIC.

X. DISCUSSION II.

The elongating Pol II can be arrested by special sequences, non-canonical DNA structures, topological constrain and DNA lesions, which can be caused by the UV irradiation (Svejstrup, 2007). Cells have developed several DNA damage and repair pathways, which can be induced by persistently blocked Pol II elongation complexes on the transcriptional units. It has been shown, that cotemporary with the DNA repair process on ORF, the new round of transcription is also blocked (Pankotai et al, 2012). However our knowledge about Pol II transcription inhibition after UV irradiation is based only on a few model genes and conditions, therefore it is not known how the global transcription is affected upon UV irradiation.

To investigate the fate of Pol II transcription machinery during DNA repair processes at the global level, I carried out ChIP-seq experiment following UV treated human MCF7 cells to follow the alteration of Pol II occupancy patterns over time. Our aim was to study the fate of Pol II transcription in a condition, when cells are able to repair the DNA and rescue transcription, without inducing apoptosis. Therefore, I used low dose of UVB (55 J/m²) irradiation to generate DNA lesions genome wide. This dose was not lethal for the cells (**Figure 18**), however, it caused a massive accumulation of CPDs in the genome (**Figure 19**). The comparison of our CPD-slot blot results with the survival assay further support the idea that the coding regions are repaired faster than the heterocromatin, as we detected CPDs 24 hours after UV irradiation (Lagerwerf et al, 2011).

By mapping the sequencing tags from the ChIP-seq experiments to the annotated genomic regions, we were able to generate average Pol II occupancy profiles for every Pol II transcribed gene. Surprisingly, the generated average Pol II profiles showed a gradual Pol II clearance from promoters over time after UVB irradiation (**Figure 22, 23**). This observation let us to propose the existence of a coordinated genome-wide mechanism to remove Pol IIs from coding regions, mainly from promoters upon DNA damage.

To follow the Pol II behavior on active genes after UV irradiation, I generated a heat map showing Pol II density changes over time on the TSS, the gene body and 3' end region of 4500 expressed genes. Our results show a massive Pol II clearance from promoters at 3-4h time points independently from gene expression level (**Figure 25**), and a slow regeneration

at 5-6h time points. These observations propose that almost every expressed gene is targeted by a mechanism to remove Pol II from promoters during the DNA repair processes, which acts at the same time on every promoter. The detected synchronized Pol II regeneration on the TSSs at 6h time point further supports this idea. Interestingly, we also detected a set of genes, which show increased Pol II density throughout their ORF after UVB irradiation (**Figure 27**). Gene Ontology analyses suggested that these genes are participating in stress response processes (**Table 3**). However, the identified stress responsive genes, which show increased Pol II tag density upon UVB treatment, suggest that these genes might undergo a different regulation than the majority of genes that show Pol II clearance during DNA repair.

Interestingly, we observed an increase of Pol II occupancy at 1h on the gene body of those genes, which show Pol II clearance from their promoters. This increased Pol II presence might reflect “the last” round of transcription, which got arrested by lesions. In general, the Pol II residency time on the gene body and downstream from EAGs does not show notable decrease at 3-4h time points. Nevertheless at 5h and 6h time points we detected a gradual increase of Pol II presence on the coding regions, most of all on the gene bodies and downstream from EAGs, suggesting that the transcriptional machinery tries to compensate for the loss of transcription during the repair processes.

We carried out western blot analyses to determine whether Pol II degradation or CTD hyperphosphorylation (Pol IIo form) would contribute to the Pol II clearance from promoters (**Figure 29**). We detected a slight shift from Pol IIa to Pol IIo form together with increased level of CTD Ser5 phosphorylation at 1-2h time point, which suggests the depletion of initiation-competent Pol II. Surprisingly, we detected a quick drop in the Ser7 CTD phosphorylation cotemporary with the promoter clearance of Pol II. The loss of the CTD Ser7-P mark might reflect the transcriptional gap, and further supports the idea, that Ser7-P is required for gene expression (Egloff et al, 2007). Nevertheless, we did not detect changes in the global level of Pol II with the N-terminal specific AB, which suggest that the Pol II clearance from promoters is not mediated by degradation.

To test whether PIC subunits are degraded as a mechanism to facilitate the detected Pol II clearance from promoters, we also analyzed the protein levels of p62/TFIIH, TFIIB and

TBP (**Figure 29**). Additionally, we tested the level of NELF-A. As we did not detect any changes in the global level of these transcription factors, we suggest that PIC formation might be also blocked by the spreading of γH2AX and/or other repressive chromatin marks and altered nuclear architecture due to the interplay of DNA damage signaling factors (Banath et al, 2009; Misteli & Soutoglou, 2009; Pankotai et al, 2012).

Our study is the first genome-wide characterization of the effect of UVB irradiation on the Pol II transcription machinery. The results within raise additional questions:

1; What is the mechanism of Pol II clearance from promoters?

It is known that the formation of DNA damage on ORFs will silence the corresponding genes (Lagerwerf et al, 2011). This transcription repression is due to cooperative actions of multiple factors. For example ATM kinase phosphorylates H2AX and mediates the ubiquitination of H2A. These repressive chromatin marks can spread through thousands of base pairs (Massip et al, 2010). Additionally PARP1 was shown to play a role in removing pre-mRNAs and Pol IIs from DNA damage sites (Chou et al, 2010). Moreover, HP1 that promotes chromatin compaction was also reported to associate with DNA damage sites (Baldeyron et al, 2011). However, most of these observations were done during the examinations of double strand beak repair processes. As the mechanism of SSB repair, TC-NER and the DSB repair differs, it is uncertain whether we can apply these models to explain the Pol II loss from the promoters upon UVB irradiation in our case. Therefore, experiments are needed to investigate the genome-wide appearance and localization of γH2AX and UB-H2A, and/or the binding site of ATM kinase, HP1 and PARP1 upon UVB irradiation, whether they show colocalization with the genes showing Pol II clearance.

An additional explanation for the Pol II clearance from promoters could be that the number of initiation-competent Pol IIa is decreased upon UVB irradiation, as it is suggested by our results (**Figure 29**). Thus it is possible, that the activity of certain kinases which might be activated during DNA damage signaling, increase the phosphorylation level of CTD residues, such as Ser5 Tyr1 and Thr4. DNAPKcs, Chk1, Chk2 and ATM kinase are promising candidate kinases (see section 11), which might contribute to the CTD phosphorylation. An important experiment would be the individual knock-down of these kinases that is followed

by western blot analyses to detect whether Pol IIo form is increased after UV irradiation. Also, *in vitro* kinase assay could determine whether these kinases target Po II CTD.

Additionally, during the repair processes phosphatases have to be deactivated to maintain the increased level of Pol IIo form. Therefore the western blot analyses of FCP1, SCP1, PRAP2 and Ssu72 (which target Pol II CTD, see table 1 in section 1) could give a hint, whether the expression of these phosphatases are downregulated after UVB irradiation.

Finally experiments should be carried out to analyze whether the PIC composition is changed or certain subunits are modified or degraded upon UVB irradiation.

2; Is Pol II clearance from promoters governed by a global mechanism?

A study demonstrated that, when only one gene is targeted by a double strand break, the following transcriptional silencing affects the expression of the targeted gene only (Kim et al, 2007).

This observation is contradictory with our idea that suggest the existence of a global mechanism responsible to remove Pol II from promoters. Probably, there is "decision mechanism", which chooses between local and global gene silencing and it is influenced by the number of genes with arrested Pol II, or the dose of UV irradiation. In the future, different doses of UVB should be tested to see if the hypothetical global machinery is activated in a dose-dependent manner.

3; Is Pol II clearance from promoters also representing silenced transcription?

A study proposed that the increased Pol II occupancy on the gene body reflects increased/activated transcription (Welboren et al, 2009). Additionally, we detected a decreased level of phospho-Ser7 residue of Pol II CTD that is suggested to be a mark of active transcription (Egloff et al, 2007). By taking these observations into consideration, we assume that the Pol II clearance from promoters and from gene bodies will be linked to transcription arrest, while the increased Pol II signal on the identified stress responsive genes might reflect increased transcription. To further support this idea, a detailed gene expression analysis upon UVB treatment (which identifies nascent transcripts, such as GRO-seq) is needed.

4; Are arrested Pol IIs removed from the DNA template during the repair processes?

There are several models, which describe the possible mechanism, which displace arrested Pol IIs during DNA repair (see section 12/bb): **1**) Pol II might bypass the lesions (Walmacq et al, 2012), **2**) CSB and other TCR factors might cause conformational changes of the Pol II, which would facilitate its dissociation (Svejstrup, 2002), **3)** Pol II might backtrack from the lesion with the help of TFIIS (Kalogeraki et al, 2005) and finally **5**) Pol II can be polyubiquitinated and degraded from the template (Wilson et al, 2012). However, a second school of thoughts suggest that Pol II is not removed during TCR (Fousteri et al, 2006).

In our study, the observed Pol II accumulation on the gene bodies at the 1h time point might reflect Pol IIs arrested by DNA lesions. The subsequent slight Pol II signal loss at 2h-3h-4h time points might represent the dissociation of Pol II from the template. As our WB analyses did not show decrease in the global level of Pol II, we assume that in our conditions, Pol II is not degraded after UVB irradiation. However, our study did not allow us to determine how Pol II is removed from the template.

An additional candidate, which might remove arrested Pol II could be the repair factor UVSSA (UV-stimulated scaffold protein A, (Zhang et al, 2012b) which was recently found to be associated with the ubiquitinated Pol II during TCR .

5; Why the identified stress genes do not undergo Pol II clearance?

A study demonstrated that UV responsive genes are short genes in general, and because of the short size, the chance to acquire lesion during UV irradiation is decreased (McKay et al, 2004).

An additional explanation could be that these genes might have special PIC composition and/or promoter architecture, which facilitates a quick repair and activation, or simply these genes are repaired prior to the “housekeeping genes” during the DNA damage repair process. Another possibility is that these genes might have different Pol II elongation complexes which can transcribe through certain DNA lesions (i.e. CPDs).

However, further investigations are needed to determine the reason why the identified stress response genes do not show Pol II clearance from promoters upon UVB irradiation. I would test, whether these genes have special promoter architecture, nucleosome composition, and/or a special transcription activator which facilitates a rapid activation.

XI. Materials and methods

Cell culture

The MCF7 human breast cancer cell line (obtained from American Type Culture Collection; ATCC; reference number HTB-22) was grown in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) supplemented with 10% foetal calf serum (FCS). The medium contained insulin (0.6µg/ml) and gentamicin (40µg/ml).

UV-B treatment

55 Joules /m² were used for each experiment. For the UV treatment Vilber T-15M lamp and Bioblock UV-radiometer device was used to measure the UV-B dose at the 312nm wavelength. For every experiment, majority of the medium was removed before UV-treatment. Around 5% medium was left on the cells to prevent them from drying out during irradiation. After treatment, fresh and pre-warmed medium was added for the incubation period.

Crystal-violet survival assay

MCF7 cell lines were plate into a 6 well plate (300 cells / well) one day prior to UV treatment. In the following days, the survived cells will form colonies. 7 days after UV irradiation, the cells were washed with PBS then stained with Crystal violet solution (0.2% crystal violet, 2% ethanol). The staining was removed from the cells with 1% sodium dodecyl sulfate (SDS) containing MQ water. The crystal violet amount in both UV treated and in control samples were measured with spectrophotometer at 595nm. Results were calculated from biological triplicates.

Slot-blot technique to detect DNA-lesions

Cells were plated in 6 cm plates and UV-treated at 80-90% confluence. After incubation for 1min, 2, 4, 6, 8, 16, 24 hours DNA was extracted from samples as well as from

non-treated control with SIGMA GenElute Mammalian Genomic DNA Miniprep Kit. DNA content was quantified with Nanodrop. 250ng of DNA was diluted in 2x SSC (0.3M NaCl, 0.03M Sodium citrate, pH: 7) buffer from each sample and was transferred to Amersham Hybond-N+ membrane with Slot-blot vacuum chamber. The membrane was treated with Denaturizing buffer (0.5M NaOH, 1.5M NaCl) then Neutralizing buffer (0.5M Tris-HCl, 1.5M NaCl). The membrane was then treated according to Western blot protocol: blocked with PBS + 2% milk, and then incubated with mouse IgG monoclonal anti-CPD (TDM2) (MBL international corp.) primary antibody overnight in 4 °C.

Measurement of gene expression

MCF7 cell lines were plated in 6cm plates prior treatment. At 80-90% confluence, cell were irradiated then Tri-reagent was used to extract total RNA after 1, 2, 3, 4, 5, 6, hours of incubation and from non-treated control as well. Immediately after RNA extraction, DNase I treatment was carried out on the samples to fully reduce DNA contamination (DNase I, ROCHE). Phenol (pH: 4.5)-chloroform RNA extraction was performed to remove DNase. RNA samples were resuspended in MQ containing RNazin then quantified by Nanodrop. Reverse-transcription from 3000ng of each RNA samples was carried out with SuperScript II (Invitrogen) on the same day with the RNA extraction. Oligo pairs were designed to quantify total mRNA (intron-intron border) and pre-mRNA (first intron-exon border) levels.

Measurement of protein levels

MCF7 cell lines were plated in 6cm plates prior treatment. At 80-90% confluence, cells were irradiated. The irradiated and control cells 1, 2, 3, 4, 5, 6, hours later after incubation were washed twice with ice-cold PBS containing complete protease inhibitor cocktail (PIC) (1x), phenylmethylsulfonyl fluoride (PMFS)(0.5mM), β-glycerophosphate (10mM), sodium orthovanadate (1mM), sodium fluoride (20mM) inhibitors. Cells were scraped in PBS + inhibitors and frozen-thawed three times. Before loading on 6-10% SDS-PAGE samples were sonicated for 10 cycles (30 second ON, 30 seconds OFF) (Bioruptor) then boiled for 5 min. The samples were transferred on Whatman Nitrocellulose Transfer

Membrane, and then blocked in PBS + 3% milk for 1 hour on room Temperature. The following antibodies were used:

Phospho-p53 (Ser15) Antibody #9284 (Cell Signaling Technology)

Phospho-Chk1 (Ser345) (133D3) #2348 (Cell Signaling Technology)

RNA polymerase II N-terminal H-224n (Santa Cruz),

Pol II Ser2, AB: Covance, (MMS-129R)

Pol II Ser5 AB: Abcam, (ab5131)

Pol II Ser 7 AB: (Chapman et al, 2007)

NELF-A: sc-23599 (Santa Cruz)

TBP: 3G3 (Lescure et al, 1994)

TFIIH subunit (p62): 3C9MAB (Marinoni et al, 1997)

TFIIB: (Moncollin et al, 1992)

Tubulin-β: (D-10): sc-5274 (Santa Cruz)

Chromatin Immunoprecipitation and QPCR

MCF7 cells were plated in 15cm dishes. At 80–90% confluence cells were UV treated. After 1, 2, 3, 4, 5, 6 hours of incubation the cells were washed with PBS, and cross-linked with 1% formaldehyde for 20 minutes at room temperature. The reaction was stopped with 0.5 M glycine, and cells were washed three times with ice-cold PBS supplemented with 0.5 mM PMSF, scraped, and resuspended in swelling buffer (25 mM HEPES, pH 7.8, 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% NP-40, 1 mM dithiothreitol (DTT), 0.5 mM PMSF, PIC. Cells were broken with a Dounce homogenizer, and the nuclear fraction was resuspended in sonication buffer (50 mM HEPES, pH 7.8, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 1% SDS, 0.5 mM PMSF, PIC). The chromatin was sonicated with Covaris sonicator into 100–300-bp fragments and centrifuged to avoid any remaining cell debris.From the supernatant, 30 µg chromatin diluted with sonication buffer (without SDS) up to 1 ml (0.05% SDS concentration) was used for one IP. Protein G Sepharose beads were washed and blocked with sonication buffer containing cold-water fish skin gelatin (SIGMA) and yeast tRNA. Chromatin samples were pre-cleared with unblocked beads at 4°C, for 2 hours. The precleared chromatin samples were rotated overnight at 4°C with the Pol II antibody (Santa Cruz, H-224X), p53 antibody (clone DO-7, Abcam) or with NELF-A antibody

(A-20, sc-23599). The blocked beads were added for 2 hours to the samples to pull down specific protein-DNA complexes. After immunoprecipitation, samples were washed two times at 4°C with the following buffers: twice with Sonication buffer (without SDS), twice with buffer A (50 mM HEPES, pH 7.8, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS, 0.5 mM PMSF, PIC), twice with buffer B (20 mM Tris, pH 8, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 0.5 mM PMSF, PIC), and finally twice with Tris-EDTA buffer (10 mM Tris, pH 8, 1 mM EDTA). Bound fraction of the chromatin was eluted with 2×100 µl of elution buffer (50 mM Tris, pH 8, 1 mM EDTA, 1% SDS) at 65°C for 2×10 min and elutions were pooled. RNase A treatment (5 µg/ml), and reverse cross-linking (125 mM NaCl) was carried out at 65°C overnight. Elutions were finally incubated with proteinase K. DNA was phenol-chloroform extracted and precipitated by ethanol.

Validation of the ChIP was performed by quantitative PCR (qPCR) analysis using a Roche LightCycler 480 with Sybr green (Roche) master mix. Oligos were designed to amplify the promoter regions of the human Ubc/NM_021009, Pou2f3/NM_001244682 and Rplp1/NM_001003 genes. Intergenic region was selected as a negative control region. The ChIP experiments were repeated at least twice, and all the qPCR reactions were done in triplicates.

High throughput sequencing with Hi-seq 2000

The sample preparation for ChIP-seq was the same as described above. To have enough material for sequencing, 5 ChIP samples were added together per time points.

Library preparation for sequencing was performed as described by the manufacturer. The 32 base pair tags generated from Hi-seq 2000 were mapped to the human genome Build 36.1 (UCSC hg18) using the eland program allowing two mismatches. Only sequences that mapped uniquely to the genome with maximum of two mismatches were used for further analysis. Using the liftover tool from UCSC (www.ucsc.org), tags were mapped onto the human genome hg19.

To have a negative control, we carried out ChIP-seq with a non-specific yeast antibody.

Genome Annotations

Genome annotations were downloaded from the UCSC Genome Browser (www.ucsc.org), human genome Build 37 (hg19 assembly). Gene definitions were given by the refseq genes track (Pruitt et al, 2009).

Bioinformatic tools and data-analysis methods

Intergenic regions were selected that are far away from the genes about 20kb (around 8000 regions) and tag numbers were counted on 2kb interval in the middle of them for all the 7 samples. These values were used as an input for DESeq Bioconductor package, which normalizes the samples based on their median values. (http://www.bioconductor.org/packages/2.9/bioc/vignettes/DESeq/inst/doc/DESeq.pdf)

We carried out analyses on all genes found in the refseq database and on 4500 expressed genes which were selected randomly, from a recently published RNA-seq dataset for MCF-7 cell line (Kamieniarz et al, 2012).

The sequenced ChIP-seq reads represent only 36 base pair fragments instead of the precise protein-DNA binding sites. To illustrate the entire DNA fragment bound to Pol II, basically before analysis, 3' end of each ChIP-seq read was extended to 200 bp in the direction of the reads.

To generate an average gene profile from ChIP-seq results, Pol II tags were counted on the selected genes with seqMINER software (Ye et al, 2011) from their -1000bp from TSS until EAG +4000 bp. Pol II tags were averaged in a 50 bp window by the software. While doing this analysis, the strand orientation is taken in account in order to orientate all analyzed features in the same direction.

For heat map generation, Pol II read/tag numbers for the 4500 expressed genes were counted from the seven datasets around the promoter (+/-300 bp around TSS), on the annotated gene body (TSS+100 until the EAG) and downstream of EAG (EAG+4000bp). During heat map generation, Pol II tag densities were subjected to K-means clustering in order to organize or cluster genes in a same group based upon similar tag enrichment within

a defined region. In K-means clustering, number of clusters is fixed and hence the samples are sorted in the clusters based upon the tag enrichment and patterns of Pol II.

Cluster and heat map generation was carried out with FLORA, Cluster 3 and TREEview softwares.

Flora: http://www-microarrays.u-strasbg.fr/base.php?page=dataAnalysisClusterE.php

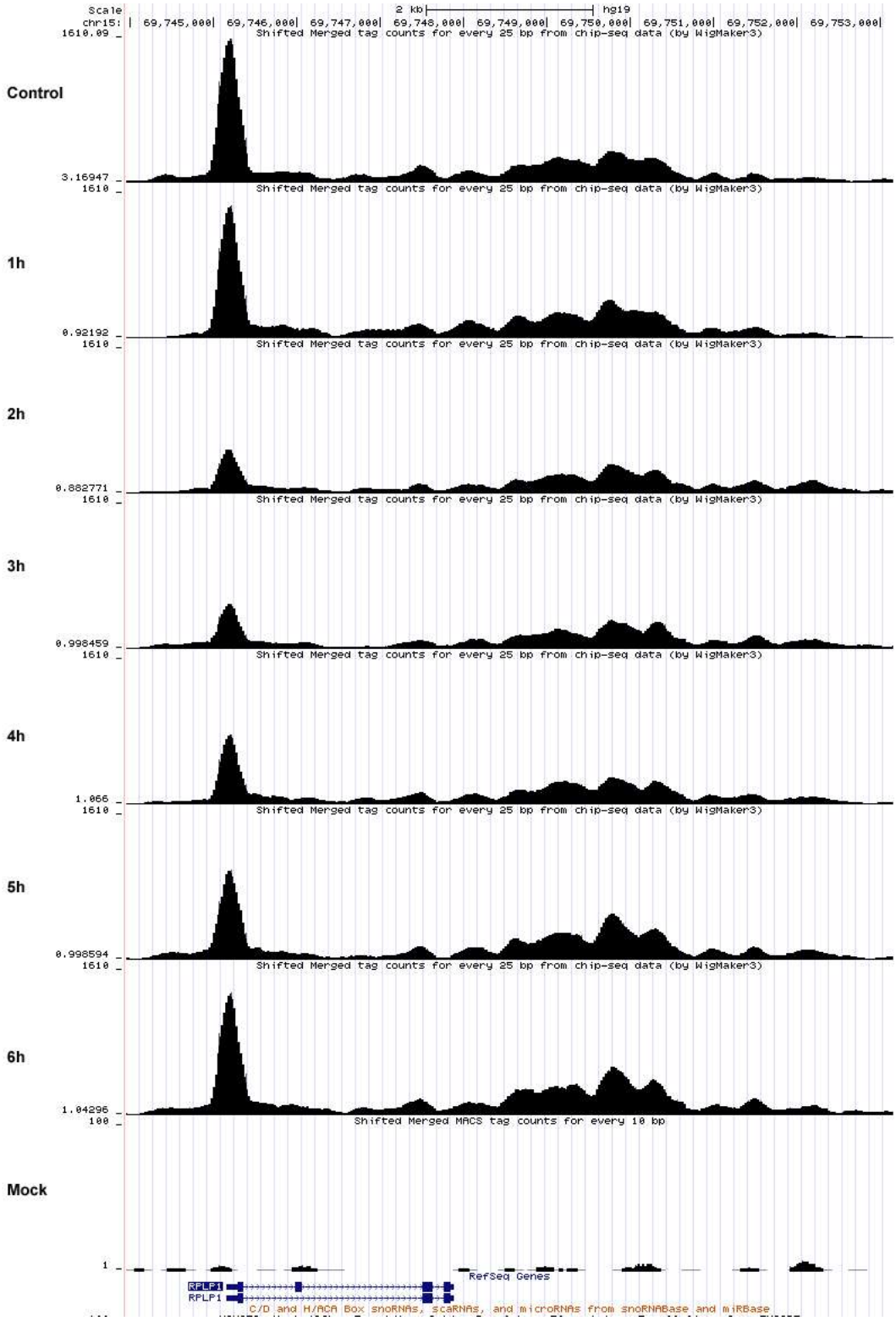
Cluster 3: http://bonsai.hgc.jp/~mdehoon/software/cluster/manual/index.html#Top

Treeview: http://sourceforge.net/projects/jtreeview/

Gene ontology (GO) analyses and Gene name conversion

GO analyses and gene ID conversion were carried out by DAVID Bioinformatics Resources 6.7. (http://david.abcc.ncifcrf.gov/home.jsp) During the analyses only the GO categories with the lowest P-Values (p-value < 0.01) were considered as positive results.

Supplementary results



Supplementary figure 1: Pol II promoter clearance on the human RPLP1 gene. ChIP-seq datasets were visualized in UCSC genome browser. Gene is indicated in blue. Pol II density is black. (The sharp peak on the left represents Pol II on the TSS)

List of "upregulated genes" after UVB treatment based on the clustering results from the list of 4500 expressed genes:

NM_052988	cdk10	cyclin-dependent kinase 10
NM_001014431	akt1	v-akt murine thymoma viral oncogene homolog 1
NM_017514	Plxna3	plexin A3
NM_001145145	SLC1A5	solute carrier family 1 (neutral amino acid transporter), member 5
NM_001014432	akt1	v-akt murine thymoma viral oncogene homolog 1
NM_001924	GADD45A	growth arrest and DNA-damage-inducible, alpha
NM_006483	DYRK1B	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B
NM_213595	iscU	iron-sulfur cluster scaffold homolog (E. coli)
NM_006484	DYRK1B	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B
NM_007056	SFRS16	splicing factor, arginine/serine-rich 16
NM_052987	cdk10	cyclin-dependent kinase 10
NM_000366	TPM1	tropomyosin 1 (alpha)
NM_078480	PUF60	poly-U binding splicing factor 60KDa
NM_138373	Myadm	myeloid-associated differentiation marker
NM_032527	zgpat	zinc finger, CCCH-type with G patch domain
NM_032525	TUBB6	tubulin, beta 6
NM_002391	Mdk	midkine (neurite growth-promoting factor 2)
NM_005343	hras	v-Ha-ras Harvey rat sarcoma viral oncogene homolog
NM_001130107	KLC1	kinesin light chain 1
NM_006341	mad2l2	MAD2 mitotic arrest deficient-like 2 (yeast)
NM_001004431	metrnl	meteorin, glial cell differentiation regulator-like; similar to meteorin, glial cell differentiation regulator-like
NM_001004431	LOC653506	meteorin, glial cell differentiation regulator-like; similar to meteorin, glial cell differentiation regulator-like
NM_015898	ZBTB7A	zinc finger and BTB domain containing 7A
NM_080861	Spsb3	splA/ryanodine receptor domain and SOCS box containing 3
NM_005481	Med16	mediator complex subunit 16
NM_001099784	fbxl19	F-box and leucine-rich repeat protein 19
NM_007346	Ogfr	opioid growth factor receptor
NM_080670	SLC35A4	solute carrier family 35, member A4
NM_001916	cyc1	cytochrome c-1
NM_004907	ier2	immediate early response 2
NM_001135054	SIGIRR	single immunoglobulin and toll-interleukin 1 receptor (TIR) domain
NM_001145138	relA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)
NM_016547	SDF4	stromal cell derived factor 4
NM_001913	CUX1	cut-like homeobox 1
NM_017528	Wbscr22	Williams Beuren syndrome chromosome region 22
NM_001040146	chtf8	CTF8, chromosome transmission fidelity factor 8 homolog (S. cerevisiae)
NM_024098	CCDC86	coiled-coil domain containing 86
NM_007040	HNRNPUL1	heterogeneous nuclear ribonucleoprotein U-like 1
NM_024881	SLC35E1	solute carrier family 35, member E1
NM_001008709	Ppp1ca	protein phosphatase 1, catalytic subunit, alpha isoform
NM_013403	STRN4	striatin, calmodulin binding protein 4
NM_175875	SIX5	SIX homeobox 5
NM_004357	cd151	CD151 molecule (Raph blood group)
NM_001136033	PUF60	poly-U binding splicing factor 60KDa
NM_016454	TMEM85	transmembrane protein 85
NM_078467	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
NM_004356	CD81	CD81 molecule
NM_004359	CDC34	cell division cycle 34 homolog (S. cerevisiae)
NM_006332	ifi30	interferon, gamma-inducible protein 30
NM_001919	dci	dodecenoyl-Coenzyme A delta isomerase (3,2 trans-enoyl-Coenzyme A isomerase)
NM_001130702	BSCL2	Bernardinelli-Seip congenital lipodystrophy 2 (seipin)
NM_000389	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
NM_016024	Rbmx2	RNA binding motif protein, X-linked 2
NM_001033549	c19orf62	chromosome 19 open reading frame 62

NM_004324	BAX	BCL2-associated X protein
NM_016639	TNFRSF12A	tumor necrosis factor receptor superfamily, member 12A
NM_181442	ADNP	activity-dependent neuroprotector homeobox
NM_024845	NAT15	N-acetyltransferase 15 (GCN5-related, putative)
NM_001080453	INTS1	integrator complex subunit 1
NM_182958	Myst1	MYST histone acetyltransferase 1
NM_016823	crk	v-crk sarcoma virus CT10 oncogene homolog (avian)
NM_015584	POLDIP2	polymerase (DNA-directed), delta interacting protein 2
NM_000107	DDB2	damage-specific DNA binding protein 2, 48kDa
NM_175847	Ptbp1	polypyrimidine tract binding protein 1
NM_001009	Rps5	ribosomal protein S5
NM_080796	DIDO1	death inducer-obliterator 1
NM_006494	erF	Ets2 repressor factor
NM_005796	NUTF2	nuclear transport factor 2
NM_025082	CENPT	centromere protein T
NM_000113	TOR1A	torsin family 1, member A (torsin A)
NM_016399	TRIAP1	TP53 regulated inhibitor of apoptosis 1
NM_024757	EHMT1	euchromatic histone-lysine N-methyltransferase 1
NM_012401	plxnb2	plexin B2
NM_015659	Rsl1d1	ribosomal L1 domain containing 1
NM_001005920	JMJD8	jumonji domain containing 8
NM_016047	SF3B14	splicing factor 3B, 14 kDa subunit
NM_013433	Tnpo2	transportin 2
NM_180982	mrpl52	mitochondrial ribosomal protein L52
NM_005762	TRIM28	tripartite motif-containing 28
NM_006443	C6orf108	chromosome 6 open reading frame 108
NM_001136203	CCDC124	coiled-coil domain containing 124
NM_005861	STUB1	STIP1 homology and U-box containing protein 1
NM_182972	IRF2BP2	interferon regulatory factor 2 binding protein 2
NM_001120	MFSD10	major facilitator superfamily domain containing 10
NM_007103	Ndufv1	NADH dehydrogenase (ubiquinone) flavoprotein 1, 51kDa
NM_013333	epn1	epsin 1
NM_004309	ARHGDIA	Rho GDP dissociation inhibitor (GDI) alpha
NM_153200	EDF1	endothelial differentiation-related factor 1
NM_007215	POLG2	polymerase (DNA directed), gamma 2, accessory subunit
NM_017432	PTOV1	prostate tumor overexpressed 1
NM_016172	ubac1	UBA domain containing 1
NM_015666	GTPBP5	GTP binding protein 5 (putative)
NM_001081560	DMPK	dystrophia myotonica-protein kinase
NM_016175	C5orf45	chromosome 5 open reading frame 45
NM_005870	SAP18	Sin3A-associated protein, 18kDa
NM_016176	SDF4	stromal cell derived factor 4
NM_001081562	DMPK	dystrophia myotonica-protein kinase
NM_016368	isyna1	inositol-3-phosphate synthase 1
NM_001081563	DMPK	dystrophia myotonica-protein kinase
NM_173853	KRTCAP3	keratinocyte associated protein 3
NM_001135700	YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
NM_001135701	YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
NM_001081559	CPSF4	cleavage and polyadenylation specific factor 4, 30kDa
NM_005189	cbx2	chromobox homolog 2 (Pc class homolog, Drosophila)
NM_006428	mrpl28	mitochondrial ribosomal protein L28
NM_006429	Cct7	chaperonin containing TCP1, subunit 7 (eta)
NM_004317	Asna1	arsA arsenite transporter, ATP-binding, homolog 1 (bacterial)
NM_000152	GAA	glucosidase, alpha; acid
NM_001040664	PPAN-P2RY11	PPAN-P2RY11 readthrough transcript
NM_007221	BGLAP	bone gamma-carboxyglutamate (gla) protein; polyamine-modulated factor 1
NM_007221	pmf1	bone gamma-carboxyglutamate (gla) protein; polyamine-modulated factor 1
NM_033375	Myo1C	myosin IC
NM_016069	TIMM16	mitochondria-associated protein involved in granulocyte-macrophage colony-stimulating factor signal transduction

NM_001080950	Myo1C	myosin IC
NM_017798	YTHDF1	YTH domain family, member 1
NM_176818	gatC	glutamyl-tRNA(Gln) amidotransferase, subunit C homolog (bacterial)
NM_007033	RER1	RER1 retention in endoplasmic reticulum 1 homolog (S. cerevisiae)
NM_212492	GPS1	G protein pathway suppressor 1
NM_015537	nelf	nasal embryonic LHRH factor
NM_006462	RBCK1	RanBP-type and C3HC4-type zinc finger containing 1
NM_001135040	DCTN1	dynactin 1 (p150, glued homolog, Drosophila)
NM_001135041	DCTN1	dynactin 1 (p150, glued homolog, Drosophila)
NM_030629	cmip	c-Maf-inducing protein
NM_001083600	NAT15	N-acetyltransferase 15 (GCN5-related, putative)
NM_001076684	UBTF	upstream binding transcription factor, RNA polymerase I
NM_006266	RALGDS	ral guanine nucleotide dissociation stimulator
NM_181485	zgpat	zinc finger, CCCH-type with G patch domain
NM_001083113	zgpat	zinc finger, CCCH-type with G patch domain
NM_080881	DBN1	drebrin 1
NM_006367	CAP1	CAP, adenylate cyclase-associated protein 1 (yeast)
NM_016400	SERF2	chromosome 15 open reading frame 63; small EDRK-rich factor 2
NM_016400	c15orf63	chromosome 15 open reading frame 63; small EDRK-rich factor 2
NM_001142285	RPS24	ribosomal protein S24
NM_032885	atg4d	ATG4 autophagy related 4 homolog D (S. cerevisiae)
NM_001020819	Myadm	myeloid-associated differentiation marker
NM_001020818	Myadm	myeloid-associated differentiation marker
NM_001136018	ephX1	epoxide hydrolase 1, microsomal (xenobiotic)
NM_006349	Znhit1	zinc finger, HIT type 1
NM_181358	hipk1	homeodomain interacting protein kinase 1
NM_001134473	KIAA0182	KIAA0182
NM_006456	St6galnac2	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2
NM_017503	SURF2	surfeit 2
NM_006454	MXD4	MAX dimerization protein 4
NM_016151	taok2	TAO kinase 2
NM_001076674	tmub2	transmembrane and ubiquitin-like domain containing 2
NM_020719	PRR12	proline rich 12
NM_001032367	Spint1	serine peptidase inhibitor, Kunitz type 1
NM_001145527	EHMT1	euchromatic histone-lysine N-methyltransferase 1
NM_001101	ACTB	actin, beta
NM_001020821	Myadm	myeloid-associated differentiation marker
NM_001005271	chd3	chromodomain helicase DNA binding protein 3
NM_001040456	RHBDD2	rhomboid domain containing 2
NM_001040457	RHBDD2	rhomboid domain containing 2
NM_001020820	Myadm	myeloid-associated differentiation marker
NM_005745	BCAP31	B-cell receptor-associated protein 31
NM_024718	C9orf86	chromosome 9 open reading frame 86
NM_001042470	sumf2	sulfatase modifying factor 2
NM_001141936	C4orf48	chromosome 4 open reading frame 48
NM_199290	NACA2	nascent polypeptide-associated complex alpha subunit 2
NM_000169	Gla	galactosidase, alpha
NM_199002	Arhgef1	Rho guanine nucleotide exchange factor (GEF) 1
NM_139029	cd151	CD151 molecule (Raph blood group)
NM_019096	gtpbp2	GTP binding protein 2
NM_014727	MLL4	myeloid/lymphoid or mixed-lineage leukemia 4
NM_002116	HLA-A	major histocompatibility complex, class I, A
NM_002201	ISG20	interferon stimulated exonuclease gene 20kDa
NM_022489	INF2	inverted formin, FH2 and WH2 domain containing
NM_006284	taf10	TAF10 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 30kDa
NM_139030	cd151	CD151 molecule (Raph blood group)
NM_001862	COX5B	cytochrome c oxidase subunit Vb
NM_013379	dpp7	dipeptidyl-peptidase 7
NM_002708	Ppp1ca	protein phosphatase 1, catalytic subunit, alpha isoform
NM_015516	TSKU	tsukushin

NM_000071	Cbs	cystathionine-beta-synthase
NM_020243	Tomm22	translocase of outer mitochondrial membrane 22 homolog (yeast)
NM_001144945	MYL12B	myosin, light chain 12B, regulatory
NM_002406	Mgat1	mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
NM_001144944	MYL12B	myosin, light chain 12B, regulatory
NM_138462	zmynd19	zinc finger, MYND-type containing 19
NM_199287	Ccdc137	coiled-coil domain containing 137
NM_032268	znrf1	zinc and ring finger 1
NM_014730	Mlec	malectin
NM_198799	BCAS4	breast carcinoma amplified sequence 4
NM_001042483	nupr1	nuclear protein, transcriptional regulator, 1
NM_000177	Gsn	gelsolin (amyloidosis, Finnish type)
NM_054013	MGAT4B	mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme B
NM_198155	C21orf33	chromosome 21 open reading frame 33
NM_019082	ddx56	DEAD (Asp-Glu-Ala-Asp) box polypeptide 56
NM_138570	SLC38A10	solute carrier family 38, member 10
NM_001130442	hras	v-Ha-ras Harvey rat sarcoma viral oncogene homolog
NM_015523	Rexo2	REX2, RNA exonuclease 2 homolog (S. cerevisiae)
NM_015414	RPL36P14	ribosomal protein L36; ribosomal protein L36 pseudogene 14
NM_015414	rpl36	ribosomal protein L36; ribosomal protein L36 pseudogene 14
NM_013387	Uqcr10	ubiquinol-cytochrome c reductase complex (7.2 kD)
NM_015629	PRPF31	PRP31 pre-mRNA processing factor 31 homolog (S. cerevisiae)
NM_015411	sumf2	sulfatase modifying factor 2
NM_152832	FAM89B	family with sequence similarity 89, member B
NM_144772	APOA1BP	apolipoprotein A-I binding protein
NM_001098833	ATXN7L3	ataxin 7-like 3
NM_018715	RCC2	regulator of chromosome condensation 2
NM_138471	C11orf84	chromosome 11 open reading frame 84
NM_001012300	MCRS1	microspherule protein 1
NM_001008897	TCP1	hypothetical gene supported by BC000665; t-complex 1
NM_001008897	TCP1P3	hypothetical gene supported by BC000665; t-complex 1
NM_033251	RPL13	ribosomal protein L13 pseudogene 12; ribosomal protein L13
NM_033251	RPL13P12	ribosomal protein L13 pseudogene 12; ribosomal protein L13
NM_178511	Prr24	hypothetical protein LOC255783
NM_005273	gnb2	guanine nucleotide binding protein (G protein), beta polypeptide 2
NM_001037984	SLC38A10	solute carrier family 38, member 10
NM_001077397	IRF2BP2	interferon regulatory factor 2 binding protein 2
NM_033250	LOC652346	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_033250	PML	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_138439	FLYWCH2	FLYWCH family member 2
NM_019070	DDX49	DEAD (Asp-Glu-Ala-Asp) box polypeptide 49
NM_001134423	CDV3	CDV3 homolog (mouse)
NM_032188	Myst1	MYST histone acetyltransferase 1
NM_181307	mrpl52	mitochondrial ribosomal protein L52
NM_181500	CUX1	cut-like homeobox 1
NM_181304	mrpl52	mitochondrial ribosomal protein L52
NM_058190	c21orf70	chromosome 21 open reading frame 70; hypothetical LOC729774; hypothetical LOC729535
NM_058190	LOC729774	chromosome 21 open reading frame 70; hypothetical LOC729774; hypothetical LOC729535
NM_058190	LOC729535	chromosome 21 open reading frame 70; hypothetical LOC729774; hypothetical LOC729535
NM_181306	mrpl52	mitochondrial ribosomal protein L52
NM_181305	mrpl52	mitochondrial ribosomal protein L52
NM_022156	DUS1L	dihydrouridine synthase 1-like (S. cerevisiae)
NM_032389	ARFGAP2	ADP-ribosylation factor GTPase activating protein 2
NM_018067	MAP7D1	MAP7 domain containing 1
NM_006693	CPSF4	cleavage and polyadenylation specific factor 4, 30kDa
NM_001105079	FBRS	fibrosin
NM_033244	LOC652346	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_033244	PML	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_001144925	mx1	myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)
NM_007263	cope	coatomer protein complex, subunit epsilon
NM_002824	PTMS	parathymosin

NM_033249	LOC652346	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_033249	PML	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_001283	AP1S1	adaptor-related protein complex 1, sigma 1 subunit
NM_000455	STK11	serine/threonine kinase 11
NM_033247	LOC652346	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_033247	PML	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_033246	LOC652346	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_033246	PML	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_002931	RING1	ring finger protein 1
NM_002335	LRP5	low density lipoprotein receptor-related protein 5
NM_001287	CLCN7	chloride channel 7
NM_033240	LOC652346	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_033240	PML	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_005262	Gfer	growth factor, augmenter of liver regeneration
NM_006396	SSSCA1	Sjogren syndrome/scleroderma autoantigen 1
NM_000593	tap1	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)
NM_199184	C6orf108	chromosome 6 open reading frame 108
NM_005163	akt1	v-akt murine thymoma viral oncogene homolog 1
NM_001018137	NME2	non-metastatic cells 1, protein (NM23A) expressed in; NME1-NME2 readthrough transcript; non-metastatic cells 2, protein (NM23B) expressed in
NM_001018137	NME1-NME2	non-metastatic cells 1, protein (NM23A) expressed in; NME1-NME2 readthrough transcript; non-metastatic cells 2, protein (NM23B) expressed in
NM_001018137	Nme1	non-metastatic cells 1, protein (NM23A) expressed in; NME1-NME2 readthrough transcript; non-metastatic cells 2, protein (NM23B) expressed in
NM_014615	KIAA0182	KIAA0182
NM_001018020	TPM1	tropomyosin 1 (alpha)
NM_013365	GGA1	golgi associated, gamma adaptin ear containing, ARF binding protein 1
NM_033239	LOC652346	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_033239	PML	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_178863	Kctd13	potassium channel tetramerisation domain containing 13
NM_020132	AGPAT3	1-acylglycerol-3-phosphate O-acyltransferase 3
NM_032595	Ppp1r9b	protein phosphatase 1, regulatory (inhibitor) subunit 9B
NM_021806	FAM3A	family with sequence similarity 3, member A
NM_001086521	C17orf89	chromosome 17 open reading frame 89
NM_007278	GABARAP	GABA(A) receptor-associated protein
NM_002342	LTBR	lymphotoxin beta receptor (TNFR superfamily, member 3)
NM_001878	crabp2	cellular retinoic acid binding protein 2
NM_002730	PRKACA	protein kinase, cAMP-dependent, catalytic, alpha
NM_005112	wdr1	WD repeat domain 1
NM_001037335	PRIC285	peroxisomal proliferator-activated receptor A interacting complex 285
NM_001145303	TMC4	transmembrane channel-like 4
NM_020385	rexo4	REX4, RNA exonuclease 4 homolog (S. cerevisiae)
NM_014064	mettl11a	methyltransferase like 11A
NM_002949	mRpL12	mitochondrial ribosomal protein L12
NM_001085454	GIT1	G protein-coupled receptor kinase interacting ArfGAP 1
NM_005253	Fosl2	FOS-like antigen 2
NM_001022	rps19	ribosomal protein S19 pseudogene 3; ribosomal protein S19
NM_001022	RPS19P3	ribosomal protein S19 pseudogene 3; ribosomal protein S19
NM_058246	DNAJB6	DnaJ (Hsp40) homolog, subfamily B, member 6
NM_001012334	Mdk	midkine (neurite growth-promoting factor 2)
NM_199054	MKNK2	MAP kinase interacting serine/threonine kinase 2
NM_001964	Egr1	early growth response 1
NM_001012333	Mdk	midkine (neurite growth-promoting factor 2)
NM_002067	GNA11	guanine nucleotide binding protein (G protein), alpha 11 (Gq class)
NM_007285	GABARAPL2	GABA(A) receptor-associated protein-like 2
NM_001102564	c14orf179	chromosome 14 open reading frame 179
NM_001960	EEF1D	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)
NM_080822	ovca2	candidate tumor suppressor in ovarian cancer 2
NM_014866	SEC16A	SEC16 homolog A (S. cerevisiae)
NM_181527	naa20	N-acetyltransferase 5 (GCN5-related, putative)
NM_018045	BSDC1	BSD domain containing 1
NM_032667	BSCL2	Bernardinelli-Seip congenital lipodystrophy 2 (seipin)

NM_005101	ISG15	ISG15 ubiquitin-like modifier
NM_015037	KIAA0913	KIAA0913
NM_080426	Gnas	GNAS complex locus
NM_001018004	TPM1	tropomyosin 1 (alpha)
NM_001018005	TPM1	tropomyosin 1 (alpha)
NM_003195	tcea2	transcription elongation factor A (SII), 2
NM_003290	TPM4	tropomyosin 4
NM_002462	mx1	myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)
NM_001103175	ccdc64b	coiled-coil domain containing 64B
NM_032038	spns1	spinster homolog 1 (Drosophila)
NM_032039	ITFG3	integrin alpha FG-GAP repeat containing 3
NM_032960	MAPKAPK2	mitogen-activated protein kinase-activated protein kinase 2
NM_002819	Ptbp1	polypyrimidine tract binding protein 1
NM_015356	scrib	scribbled homolog (Drosophila)
NM_003915	Rbm12	RNA binding motif protein 12; copine I
NM_003915	CPNE1	RNA binding motif protein 12; copine I
NM_001950	e2f4	E2F transcription factor 4, p107/p130-binding
NM_032673	PCGF1	polycomb group ring finger 1
NM_003917	ap1g2	adaptor-related protein complex 1, gamma 2 subunit
NM_005967	NAB2	NGFI-A binding protein 2 (EGR1 binding protein 2)
NM_014780	CUL7	cullin 7
NM_003959	Hip1r	huntingtin interacting protein 1 related
NM_031297	RNF208	ring finger protein 208
NM_001018008	TPM1	tropomyosin 1 (alpha)
NM_001018007	TPM1	tropomyosin 1 (alpha)
NM_001018006	TPM1	tropomyosin 1 (alpha)
NM_020218	ATXN7L3	ataxin 7-like 3
NM_003910	bud31	BUD31 homolog (S. cerevisiae)
NM_001810	Cenpb	centromere protein B, 80kDa
NM_001040100	C3orf57	chromosome 3 open reading frame 57
NM_003952	RPS6KB2	ribosomal protein S6 kinase, 70kDa, polypeptide 2
NM_014047	C19orf53	chromosome 19 open reading frame 53
NM_031946	AGAP3	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3
NM_001114331	CLCN7	chloride channel 7
NM_001081492	KRT80	keratin 80
NM_001127664	Gsn	gelsolin (amyloidosis, Finnish type)
NM_194460	RNF126	ring finger protein 126
NM_001127663	Gsn	gelsolin (amyloidosis, Finnish type)
NM_001142499	FLYWCH2	FLYWCH family member 2
NM_001127662	Gsn	gelsolin (amyloidosis, Finnish type)
NM_005234	NR2F6	nuclear receptor subfamily 2, group F, member 6
NM_198976	Th1l	TH1-like (Drosophila)
NM_001142350	THOC6	THO complex 6 homolog (Drosophila)
NM_002038	IFI6	interferon, alpha-inducible protein 6
NM_005028	PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha
NM_001131018	Ciz1	CDKN1A interacting zinc finger protein 1
NM_001131017	Ciz1	CDKN1A interacting zinc finger protein 1
NM_002566	P2RY11	purinergic receptor P2Y, G-protein coupled, 11
NM_003801	gpaa1	glycosylphosphatidylinositol anchor attachment protein 1 homolog (yeast)
NM_001127667	Gsn	gelsolin (amyloidosis, Finnish type)
NM_001127665	Gsn	gelsolin (amyloidosis, Finnish type)
NM_176795	hras	v-Ha-ras Harvey rat sarcoma viral oncogene homolog
NM_001127666	Gsn	gelsolin (amyloidosis, Finnish type)
NM_001131016	Ciz1	CDKN1A interacting zinc finger protein 1
NM_007165	SF3A2	splicing factor 3a, subunit 2, 66kDa
NM_022117	TSPYL2	TSPY-like 2
NM_001009184	grina	glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1 (glutamate binding)
NM_145685	BRF1	BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB (S. cerevisiae)
NM_005803	flot1	flotillin 1
NM_005125	Ccs	copper chaperone for superoxide dismutase

NM_152285	Arrdc1	arrestin domain containing 1
NM_020376	PNPLA2	patatin-like phospholipase domain containing 2
NM_001042469	sumf2	sulfatase modifying factor 2
NM_181552	CUX1	cut-like homeobox 1
NM_014077	Fam32a	family with sequence similarity 32, member A
NM_020378	nat14	N-acetyltransferase 14 (GCN5-related, putative)
NM_032940	Polr2c	polymerase (RNA) II (DNA directed) polypeptide C, 33kDa
NM_001005619	ITGB4	integrin, beta 4
NM_000683	ADRA2C	adrenergic, alpha-2C-, receptor
NM_007371	BRD3	bromodomain containing 3
NM_014173	c19orf62	chromosome 19 open reading frame 62
NM_002970	Sat1	spermidine/spermine N1-acetyltransferase 1
NM_080548	PTPN6	protein tyrosine phosphatase, non-receptor type 6
NM_182923	KLC1	kinesin light chain 1
NM_201380	LOC652460	similar to Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1); plectin 1, intermediate filament binding protein 500kDa
NM_201380	PLEC	similar to Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1); plectin 1, intermediate filament binding protein 500kDa
NM_001042368	RALGDS	ral guanine nucleotide dissociation stimulator
NM_201381	LOC652460	similar to Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1); plectin 1, intermediate filament binding protein 500kDa
NM_201381	PLEC	similar to Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1); plectin 1, intermediate filament binding protein 500kDa
NM_013993	DDR1	discoidin domain receptor tyrosine kinase 1
NM_007173	prss23	protease, serine, 23
NM_138496	CYHR1	cysteine/histidine-rich 1
NM_013994	DDR1	discoidin domain receptor tyrosine kinase 1
NM_015339	ADNP	activity-dependent neuroprotector homeobox
NM_198576	AGRN	agrin
NM_201384	LOC652460	similar to Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1); plectin 1, intermediate filament binding protein 500kDa
NM_201384	PLEC	similar to Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1); plectin 1, intermediate filament binding protein 500kDa
NM_015335	Med13l	mediator complex subunit 13-like
NM_002435	MPI	mannose phosphate isomerase
NM_005115	MVP	major vault protein
NM_001933	DLST	dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex); dihydrolipoamide S-succinyltransferase pseudogene (E2 component of 2-oxo-glutarate complex)
NM_001933	DLSTP1	dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex); dihydrolipoamide S-succinyltransferase pseudogene (E2 component of 2-oxo-glutarate complex)
NM_001077489	Gnas	GNAS complex locus
NM_014856	dennd4b	DENN/MADD domain containing 4B
NM_001077488	Gnas	GNAS complex locus
NM_001130524	ap1m1	adaptor-related protein complex 1, mu 1 subunit
NM_002434	mpg	N-methylpurine-DNA glycosylase
NM_001005336	DNM1	dynamin 1
NM_003977	aip	aryl hydrocarbon receptor interacting protein
NM_017702	DEF8	differentially expressed in FDCP 8 homolog (mouse)
NM_138769	RHOT2	ras homolog gene family, member T2
NM_001142646	tpra1	G protein-coupled receptor 175
NM_001142641	FBRSL1	fibrosin-like 1
NM_002191	inhA	inhibin, alpha
NM_002675	LOC652346	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_002675	PML	promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1
NM_006099	PIAS3	protein inhibitor of activated STAT, 3
NM_014417	BBC3	BCL2 binding component 3
NM_001009996	dalrd3	DALR anticodon binding domain containing 3
NM_017900	AURKAIP1	aurora kinase A interacting protein 1
NM_021259	tmem8a	transmembrane protein 8A
NM_006736	Dnajb2	DnaJ (Hsp40) homolog, subfamily B, member 2
NM_001139457	BCAP31	B-cell receptor-associated protein 31
NM_001142449	spns1	spinster homolog 1 (Drosophila)

NM_001142448	spns1	spinster homolog 1 (Drosophila)
NM_173211	Tgif1	TGFB-induced factor homeobox 1
NM_000516	Gnas	GNAS complex locus
NM_014941	MORC2	MORC family CW-type zinc finger 2
NM_198488	Fam83h	family with sequence similarity 83, member H
NM_032304	haghl	hydroxyacylglutathione hydrolase-like
NM_001015881	TSC22D3	TSC22 domain family, member 3; GRAM domain containing 4
NM_001015881	GRAMD4	TSC22 domain family, member 3; GRAM domain containing 4
NM_201525	GPR56	G protein-coupled receptor 56
NM_005206	crk	v-crk sarcoma virus CT10 oncogene homolog (avian)
NM_201524	GPR56	G protein-coupled receptor 56
NM_006869	ADAP1	ArfGAP with dual PH domains 1
NM_021174	KIAA1967	KIAA1967
NM_153824	Pycr1	pyrroline-5-carboxylate reductase 1
NM_001129995	KCTD15	potassium channel tetramerisation domain containing 15
NM_014405	Cacng4	calcium channel, voltage-dependent, gamma subunit 4
NM_001142451	spns1	spinster homolog 1 (Drosophila)
NM_014161	MRPL18	mitochondrial ribosomal protein L18
NM_001142450	spns1	spinster homolog 1 (Drosophila)
NM_001129994	KCTD15	potassium channel tetramerisation domain containing 15
NM_021168	RAB40C	RAB40C, member RAS oncogene family
NM_001114619	Mgat1	mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
NM_001142554	Trmt1	TRM1 tRNA methyltransferase 1 homolog (S. cerevisiae)
NM_198205	MLX	MAX-like protein X
NM_015005	kiaa0284	KIAA0284
NM_198204	MLX	MAX-like protein X
NM_014931	SAPS1	SAPS domain family, member 1
NM_203352	Pdlim7	PDZ and LIM domain 7 (enigma)
NM_182706	scrib	scribbled homolog (Drosophila)
NM_002798	PSMB6	proteasome (prosome, macropain) subunit, beta type, 6
NM_001079803	GAA	glucosidase, alpha; acid
NM_001079804	GAA	glucosidase, alpha; acid
NM_018270	c20orf20	chromosome 20 open reading frame 20
NM_003124	spr	sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase)
NM_003470	Usp7	ubiquitin specific peptidase 7 (herpes virus-associated)
NM_015379	BRI3P1	brain protein I3; brain protein I3 pseudogene 1
NM_015379	Bri3	brain protein I3; brain protein I3 pseudogene 1
NM_001033026	c19orf6	chromosome 19 open reading frame 6
NM_014935	PLEKHA6	pleckstrin homology domain containing, family A member 6
NM_006899	IDH3B	isocitrate dehydrogenase 3 (NAD+) beta
NM_004714	DYRK1B	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B
NM_001142421	Morf4l2	mortality factor 4 like 2
NM_001114620	Mgat1	mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
NM_001142420	Morf4l2	mortality factor 4 like 2
NM_006612	Kif1c	kinesin family member 1C
NM_021134	MRPL23	mitochondrial ribosomal protein L23
NM_001142422	Morf4l2	mortality factor 4 like 2
NM_182507	KRT80	keratin 80
NM_203346	HDLBP	high density lipoprotein binding protein
NM_003655	cbx4	chromobox homolog 4 (Pc class homolog, Drosophila)
NM_013291	CPSF1	cleavage and polyadenylation specific factor 1, 160kDa
NM_018683	RNF114	ring finger protein 114
NM_001142427	Morf4l2	mortality factor 4 like 2
NM_002997	sdc1	syndecan 1
NM_000964	rarA	retinoic acid receptor, alpha
NM_001142428	Morf4l2	mortality factor 4 like 2
NM_032902	PPP1R16A	protein phosphatase 1, regulatory (inhibitor) subunit 16A
NM_031459	SESN2	sestrin 2
NM_001122955	BSCL2	Bernardinelli-Seip congenital lipodystrophy 2 (seipin)
NM_014030	GIT1	G protein-coupled receptor kinase interacting ArfGAP 1
NM_138778	WDR85	WD repeat domain 85

NM_001031714	INF2	inverted formin, FH2 and WH2 domain containing
NM_024339	THOC6	THO complex 6 homolog (Drosophila)
NM_031991	Ptbp1	polypyrimidine tract binding protein 1
NM_031990	Ptbp1	polypyrimidine tract binding protein 1
NM_021975	relA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)
NM_206873	Ppp1ca	protein phosphatase 1, catalytic subunit, alpha isoform
NM_003461	ZYX	zyxin
NM_002882	RanBP1	similar to RAN binding protein 1; RAN binding protein 1
NM_002882	LOC389842	similar to RAN binding protein 1; RAN binding protein 1
NM_002882	LOC727803	similar to RAN binding protein 1; RAN binding protein 1
NM_022450	Rhbdf1	rhomboid 5 homolog 1 (Drosophila)
NM_001042679	RHOC	ras homolog gene family, member C
NM_174855	IDH3B	isocitrate dehydrogenase 3 (NAD+) beta
NM_001042678	RHOC	ras homolog gene family, member C
NM_006086	MC1R	tubulin, beta 3; melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)
NM_006086	tubb3	tubulin, beta 3; melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)
NM_032125	TMEM222	transmembrane protein 222
NM_001142432	Morf4l2	mortality factor 4 like 2
NM_000977	RPL13	ribosomal protein L13 pseudogene 12; ribosomal protein L13
NM_000977	RPL13P12	ribosomal protein L13 pseudogene 12; ribosomal protein L13
NM_001142431	Morf4l2	mortality factor 4 like 2
NM_006602	Tcfl5	transcription factor-like 5 (basic helix-loop-helix)
NM_001001560	GGA1	golgi associated, gamma adaptin ear containing, ARF binding protein 1
NM_201555	FHL2	four and a half LIM domains 2
NM_020963	MOV10	Mov10, Moloney leukemia virus 10, homolog (mouse)
NM_198335	GANAB	glucosidase, alpha; neutral AB
NM_198334	GANAB	glucosidase, alpha; neutral AB
NM_003779	B4GALT3	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 3
NM_001010972	ZYX	zyxin
NM_032353	vps25	vacuolar protein sorting 25 homolog (S. cerevisiae)
NM_001010974	BCAS4	breast carcinoma amplified sequence 4
NM_015260	Sin3b	SIN3 homolog B, transcription regulator (yeast)
NM_177441	tmub2	transmembrane and ubiquitin-like domain containing 2
NM_001002018	HCFC1R1	host cell factor C1 regulator 1 (XPO1 dependent)
NM_012396	Phlda3	pleckstrin homology-like domain, family A, member 3
NM_172195	EIF2B4	eukaryotic translation initiation factor 2B, subunit 4 delta, 67kDa
NM_012394	PFDN2	prefoldin subunit 2
NM_001130969	nelf	nasal embryonic LHRH factor
NM_001002017	HCFC1R1	host cell factor C1 regulator 1 (XPO1 dependent)
NM_001105530	CAP1	CAP, adenylate cyclase-associated protein 1 (yeast)
NM_024107	tmub2	transmembrane and ubiquitin-like domain containing 2
NM_001017987	C5orf45	chromosome 5 open reading frame 45
NM_013274	Poll	polymerase (DNA directed), lambda
NM_022574	gigyf1	GRB10 interacting GYF protein 1
NM_001142500	FLYWCH2	FLYWCH family member 2
NM_020184	Cnnm4	cyclin M4
NM_014811	KIAA0649	KIAA0649
NM_017949	CUEDC1	CUE domain containing 1
NM_003780	b4galt2	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 2
NM_178167	ZNF598	zinc finger protein 598
NM_021959	ppp1r11	protein phosphatase 1, regulatory (inhibitor) subunit 11
NM_003486	SLC7A5	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5
NM_004628	XPC	xeroderma pigmentosum, complementation group C
NM_001383	DPH1	DPH1 homolog (S. cerevisiae)
NM_001130971	nelf	nasal embryonic LHRH factor
NM_001130970	nelf	nasal embryonic LHRH factor
NM_001127190	CSK	c-src tyrosine kinase
NM_001110556	FLNA	filamin A, alpha (actin binding protein 280)
NM_001077350	NPRL3	chromosome 16 open reading frame 35
NM_018890	rac1	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
NM_014301	iscU	iron-sulfur cluster scaffold homolog (E. coli)

NM_198252	Gsn	gelsolin (amyloidosis, Finnish type)
NM_001003684	Uqcr10	ubiquinol-cytochrome c reductase complex (7.2 kD)
NM_015276	USP22	ubiquitin specific peptidase 22
NM_001142419	Morf4l2	mortality factor 4 like 2
NM_001025243	Irak1	interleukin-1 receptor-associated kinase 1
NM_001025242	Irak1	interleukin-1 receptor-associated kinase 1
NM_032493	ap1m1	adaptor-related protein complex 1, mu 1 subunit
NM_024112	C9orf16	chromosome 9 open reading frame 16
NM_001143944	LEMD2	LEM domain containing 2
NM_001098785	FAM89B	family with sequence similarity 89, member B
NM_003792	EDF1	endothelial differentiation-related factor 1
NM_001098784	FAM89B	family with sequence similarity 89, member B
NM_006830	Uqcr11	ubiquinol-cytochrome c reductase, 6.4kDa subunit
NM_001035235	Sra1	steroid receptor RNA activator 1
NM_053056	CCND1	cyclin D1
NM_003755	eif3g	eukaryotic translation initiation factor 3, subunit G
NM_032378	EEF1D	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)
NM_004759	MAPKAPK2	mitogen-activated protein kinase-activated protein kinase 2
NM_003897	Ier3	immediate early response 3
NM_021035	Znfx1	zinc finger, NFX1-type containing 1
NM_013245	vps4a	vacuolar protein sorting 4 homolog A (S. cerevisiae)
NM_006763	BTG2	BTG family, member 2
NM_181715	CRTC2	CREB regulated transcription coactivator 2
NM_020062	SLC2A4RG	SLC2A4 regulator
NM_022552	DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha
NM_006755	taldo1	transaldolase 1
NM_001014841	Ncdn	neurochondrin
NM_013974	ddah2	dimethylarginine dimethylaminohydrolase 2
NM_003367	USF2	upstream transcription factor 2, c-fos interacting
NM_006035	CDC42BPB	CDC42 binding protein kinase beta (DMPK-like)
NM_015963	THAP4	THAP domain containing 4
NM_015920	Rps27l	ribosomal protein S27-like
NM_004761	Rgl2	ral guanine nucleotide dissociation stimulator-like 2
NM_144686	TMC4	transmembrane channel-like 4
NM_004604	STX4	syntaxin 4
NM_178582	hm13	histocompatibility (minor) 13
NM_031229	RBCK1	RanBP-type and C3HC4-type zinc finger containing 1
NM_012473	TXN2	thioredoxin 2
NM_006753	SURF6	surfeit 6
NM_003620	PPM1D	protein phosphatase 1D magnesium-dependent, delta isoform
NM_198269	hipk1	homeodomain interacting protein kinase 1
NM_145030	C7orf47	chromosome 7 open reading frame 47
NM_006187	Oas3	2'-5'-oligoadenylate synthetase 3, 100kDa
NM_000875	IGF1R	insulin-like growth factor 1 receptor
NM_020686	ABAT	4-aminobutyrate aminotransferase
NM_001014839	Ncdn	neurochondrin
NM_033643	RPL36P14	ribosomal protein L36; ribosomal protein L36 pseudogene 14
NM_033643	rpl36	ribosomal protein L36; ribosomal protein L36 pseudogene 14
NM_001348	dapk3	death-associated protein kinase 3
NM_001006946	sdc1	syndecan 1
NM_017896	c20orf11	chromosome 20 open reading frame 11
NM_213636	Pdlim7	PDZ and LIM domain 7 (enigma)
NM_006014	Lage3	L antigen family, member 3
NM_004168	SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)
NM_025157	PXN	paxillin
NM_006526	ZNF217	zinc finger protein 217
NM_019034	Rhof	ras homolog gene family, member F (in filopodia)
NM_012245	SNW1	SNW domain containing 1
NM_023019	DCTN1	dynactin 1 (p150, glued homolog, Drosophila)
NM_001130079	MOV10	Mov10, Moloney leukemia virus 10, homolog (mouse)
NM_013232	pdcd6	aryl-hydrocarbon receptor repressor; programmed cell death 6

NM_013232	AHRR	aryl-hydrocarbon receptor repressor; programmed cell death 6
NM_000709	BCKDHA	branched chain keto acid dehydrogenase E1, alpha polypeptide
NM_001130072	epn1	epsin 1
NM_133436	asnS	asparagine synthetase
NM_001039476	NPRL3	chromosome 16 open reading frame 35
NM_030978	ARPC5L	actin related protein 2/3 complex, subunit 5-like
NM_145905	LOC100130009	hypothetical LOC100130009; high mobility group AT-hook 1
NM_145905	HMGA1	hypothetical LOC100130009; high mobility group AT-hook 1
NM_004864	Gdf15	growth differentiation factor 15
NM_004584	RAD9A	RAD9 homolog A (S. pombe)
NM_004862	LITAF	lipopolysaccharide-induced TNF factor
NM_001110792	mecp2	methyl CpG binding protein 2 (Rett syndrome)
NM_001079863	DBI	diazepam binding inhibitor (GABA receptor modulator, acyl-Coenzyme A binding protein)
NM_139315	TAF6	TAF6 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 80kDa
NM_030973	MED25	mediator complex subunit 25
NM_005533	IFI35	interferon-induced protein 35
NM_017885	HCFC1R1	host cell factor C1 regulator 1 (XPO1 dependent)
NM_175744	RHOC	ras homolog gene family, member C
NM_017458	MVP	major vault protein
NM_004152	oaz1	ornithine decarboxylase antizyme 1
NM_016201	Amotl2	angiomotin like 2
NM_001042610	Dbndd1	dysbindin (dystrobrevin binding protein 1) domain containing 1
NM_213646	wars	tryptophanyl-tRNA synthetase
NM_001130069	sumf2	sulfatase modifying factor 2
NM_001010938	TNK2	tyrosine kinase, non-receptor, 2
NM_006122	MAN2A2	mannosidase, alpha, class 2A, member 2
NM_207291	USF2	upstream transcription factor 2, c-fos interacting
NM_001098533	cdk10	cyclin-dependent kinase 10
NM_001673	asnS	asparagine synthetase
NM_018321	BRIX1	brix domain containing 2
NM_031206	LAS1L	LAS1-like (S. cerevisiae)
NM_001537	hsbp1	heat shock factor binding protein 1
NM_006400	DCTN2	dynactin 2 (p50)
NM_031209	Qtrt1	queuine tRNA-ribosyltransferase 1
NM_133328	DEDD2	death effector domain containing 2
NM_175609	ARFGAP1	ADP-ribosylation factor GTPase activating protein 1
NM_207115	ZNF580	zinc finger protein 580
NM_004295	TRAF4	TNF receptor-associated factor 4
NM_001539	Dnaja1	DnaJ (Hsp40) homolog, subfamily A, member 1
NM_004148	Ninj1	ninjurin 1
NM_012264	TMEM184B	transmembrane protein 184B
NM_001143889	BSDC1	BSD domain containing 1
NM_024407	NDUFS7	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)
NM_001143888	BSDC1	BSD domain containing 1
NM_017877	C2orf18	chromosome 2 open reading frame 18
NM_004766	copb2	coatomer protein complex, subunit beta 2 (beta prime)
NM_019058	ddit4	DNA-damage-inducible transcript 4
NM_178336	mrpl52	mitochondrial ribosomal protein L52
NM_004146	Ndufb7	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7, 18kDa
NM_001572	IRF7	interferon regulatory factor 7
NM_012225	NUBP2	nucleotide binding protein 2 (MinD homolog, E. coli)
NM_001130053	EEF1D	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)
NM_001130055	EEF1D	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)
NM_004047	Atp6v0b	ATPase, H+ transporting, lysosomal 21kDa, V0 subunit b
NM_001130056	EEF1D	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)
NM_017770	Elovl2	elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 2
NM_014335	EID1	EP300 interacting inhibitor of differentiation 1
NM_001039492	FHL2	four and a half LIM domains 2
NM_001143890	BSDC1	BSD domain containing 1
NM_030818	CCDC130	coiled-coil domain containing 130
NM_016208	VPS28	vacuolar protein sorting 28 homolog (S. cerevisiae)

NM_006813	PNRC1	proline-rich nuclear receptor coactivator 1
NM_183425	rbm38	RNA binding motif protein 38
NM_001039490	cd151	CD151 molecule (Raph blood group)
NM_001002249	anapc11	anaphase promoting complex subunit 11
NM_000435	NOTCH3	Notch homolog 3 (Drosophila)
NM_001569	Irak1	interleukin-1 receptor-associated kinase 1
NM_014235	UBL4A	ubiquitin-like 4A
NM_004860	FXR2	fragile X mental retardation, autosomal homolog 2
NM_006513	SARS	seryl-tRNA synthetase
NM_014233	UBTF	upstream binding transcription factor, RNA polymerase I
NM_183057	VPS28	vacuolar protein sorting 28 homolog (S. cerevisiae)
NM_000837	grina	glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1 (glutamate binding)
NM_178150	FBXO18	F-box protein, helicase, 18
NM_005558	LAD1	ladinin 1
NM_024036	LRFN4	leucine rich repeat and fibronectin type III domain containing 4
NM_001456	FLNA	filamin A, alpha (actin binding protein 280)
NM_030665	RAI1	retinoic acid induced 1
NM_004031	IRF7	interferon regulatory factor 7
NM_023008	KRI1	KRI1 homolog (S. cerevisiae)
NM_175629	DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha
NM_033557	YIF1B	Yip1 interacting factor homolog B (S. cerevisiae)
NM_012385	nupr1	nuclear protein, transcriptional regulator, 1
NM_004649	C21orf33	chromosome 21 open reading frame 33
NM_004029	IRF7	interferon regulatory factor 7
NM_052848	ccdc97	coiled-coil domain containing 97
NM_001112726	kiaa0284	KIAA0284
NM_001015052	mpg	N-methylpurine-DNA glycosylase
NM_004642	Cdk2ap1	cyclin-dependent kinase 2 associated protein 1
NM_030981	RAB1B	RAB1B, member RAS oncogene family
NM_001042728	rarg	retinoic acid receptor, gamma
NM_030980	ISG20L2	interferon stimulated exonuclease gene 20kDa-like 2
NM_014329	EDC4	enhancer of mRNA decapping 4
NM_006907	Pycr1	pyrroline-5-carboxylate reductase 1
NM_001128848	SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
NM_006908	rac1	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
NM_004785	slc9a3r2	solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulator 2
NM_024043	Dbndd1	dysbindin (dystrobrevin binding protein 1) domain containing 1
NM_001128846	SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
NM_016219	Man1b1	mannosidase, alpha, class 1B, member 1
NM_024042	metrn	meteorin, glial cell differentiation regulator
NM_001128847	SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
NM_178148	SLC35B2	solute carrier family 35, member B2
NM_001128845	SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
NM_012242	DKK1	dickkopf homolog 1 (Xenopus laevis)
NM_006801	KDELR1	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 1
NM_198426	CHMP2A	chromatin modifying protein 2A
NM_004127	GPS1	G protein pathway suppressor 1
NM_183356	asnS	asparagine synthetase
NM_202001	Ercc1	excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)
NM_004435	endog	endonuclease G
NM_016491	MRPL37	mitochondrial ribosomal protein L37
NM_005452	wdr46	WD repeat domain 46
NM_024653	prkrip1	PRKR interacting protein 1 (IL11 inducible)
NM_005451	Pdlim7	PDZ and LIM domain 7 (enigma)
NM_017572	MKNK2	MAP kinase interacting serine/threonine kinase 2
NM_022767	AEN	apoptosis enhancing nuclease
NM_207346	TSEN54	tRNA splicing endonuclease 54 homolog (S. cerevisiae)

NM_003403	YY1	YY1 transcription factor
NM_003575	ZNF282	zinc finger protein 282
NM_024798	Snx22	sorting nexin 22
NM_180990	ZACN	zinc activated ligand-gated ion channel
NM_014260	pfdn6	prefoldin subunit 6
NM_018209	ARFGAP1	ADP-ribosylation factor GTPase activating protein 1
NM_030937	CCR6	cyclin L2; chemokine (C-C motif) receptor 6
NM_030937	Ccnl2	cyclin L2; chemokine (C-C motif) receptor 6
NM_001039577	CCR6	cyclin L2; chemokine (C-C motif) receptor 6
NM_001039577	Ccnl2	cyclin L2; chemokine (C-C motif) receptor 6
NM_016494	Rnf181	ring finger protein 181
NM_030935	TSC22D4	TSC22 domain family, member 4
NM_052873	c14orf179	chromosome 14 open reading frame 179
NM_017843	BCAS4	breast carcinoma amplified sequence 4
NM_001128631	Dcakd	dephospho-CoA kinase domain containing
NM_001098208	HNRNPF	heterogeneous nuclear ribonucleoprotein F
NM_003683	rrp1	ribosomal RNA processing 1 homolog (S. cerevisiae)
NM_033405	PRIC285	peroxisomal proliferator-activated receptor A interacting complex 285
NM_001450	FHL2	four and a half LIM domains 2
NM_001134231	NT5DC2	5'-nucleotidase domain containing 2
NM_052876	Nacc1	nucleus accumbens associated 1, BEN and BTB (POZ) domain containing
NM_022872	IFI6	interferon, alpha-inducible protein 6
NM_031449	Zmiz2	zinc finger, MIZ-type containing 2
NM_004424	E4F1	E4F transcription factor 1
NM_022873	IFI6	interferon, alpha-inducible protein 6
NM_005632	SOLH	small optic lobes homolog (Drosophila)
NM_025128	MUS81	MUS81 endonuclease homolog (S. cerevisiae)
NM_133373	PLCD3	phospholipase C, delta 3
NM_001017917	CYB561	cytochrome b-561
NM_012212	PTGR1	prostaglandin reductase 1
NM_030928	Cdt1	chromatin licensing and DNA replication factor 1
NM_022772	EPS8L2	EPS8-like 2
NM_004395	DBN1	drebrin 1
NM_024067	C7orf26	chromosome 7 open reading frame 26
NM_014453	CHMP2A	chromatin modifying protein 2A
NM_001009552	PPP2CB	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform
NM_005641	TAF6	TAF6 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 80kDa
NM_004390	CTSH	cathepsin H
NM_001039457	Atp6v0b	ATPase, H+ transporting, lysosomal 21kDa, V0 subunit b
NM_001130012	slc9a3r2	solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulator 2
NM_030752	TCP1	hypothetical gene supported by BC000665; t-complex 1
NM_030752	TCP1P3	hypothetical gene supported by BC000665; t-complex 1
NM_005526	hsf1	heat shock transcription factor 1
NM_004082	DCTN1	dynactin 1 (p150, glued homolog, Drosophila)
NM_022744	C16orf58	chromosome 16 open reading frame 58
NM_014281	PUF60	poly-U binding splicing factor 60KDa
NM_017590	ZC3H7B	zinc finger CCCH-type containing 7B
NM_004184	wars	tryptophanyl-tRNA synthetase
NM_199444	cope	coatomer protein complex, subunit epsilon
NM_033420	c19orf6	chromosome 19 open reading frame 6
NM_018426	Tmem63b	transmembrane protein 63B
NM_001039550	Dnajb2	DnaJ (Hsp40) homolog, subfamily B, member 2
NM_199442	cope	coatomer protein complex, subunit epsilon
NM_017729	EPS8L1	EPS8-like 1
NM_004383	CSK	c-src tyrosine kinase
NM_017495	rbm38	RNA binding motif protein 38
NM_004380	CREBBP	CREB binding protein
NM_001039877	STRN4	striatin, calmodulin binding protein 4
NM_017491	wdr1	WD repeat domain 1
NM_001079846	CREBBP	CREB binding protein
NM_004699	Fam50a	family with sequence similarity 50, member A

NM_017722	Trmt1	TRM1 tRNA methyltransferase 1 homolog (S. cerevisiae)
NM_001136473	LITAF	lipopolysaccharide-induced TNF factor
NM_001136472	LITAF	lipopolysaccharide-induced TNF factor
NM_012140	slc25a10	solute carrier family 25 (mitochondrial carrier; dicarboxylate transporter), member 10
NM_014567	LOC646079	similar to breast cancer anti-estrogen resistance 1; breast cancer anti-estrogen resistance 1
NM_014567	bcar1	similar to breast cancer anti-estrogen resistance 1; breast cancer anti-estrogen resistance 1
NM_017582	Ube2q1	ubiquitin-conjugating enzyme E2Q family member 1
NM_017588	WDR5	WD repeat domain 5
NM_004078	CSRP1	cysteine and glycine-rich protein 1
NM_018337	ZNF444	zinc finger protein 444
NM_024663	npepl1	aminopeptidase-like 1
NM_000937	POLR2A	polymerase (RNA) II (DNA directed) polypeptide A, 220kDa
NM_001007074	SNORA7B	small nucleolar RNA, H/ACA box 7A; small nucleolar RNA, H/ACA box 7B; ribosomal protein L32
NM_001007074	rpl32	small nucleolar RNA, H/ACA box 7A; small nucleolar RNA, H/ACA box 7B; ribosomal protein L32
NM_001007074	Snora7a	small nucleolar RNA, H/ACA box 7A; small nucleolar RNA, H/ACA box 7B; ribosomal protein L32
NM_004409	DMPK	dystrophia myotonica-protein kinase
NM_004547	LOC402175	hypothetical gene supported by AF044957; NM_004547; NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa
NM_004547	Ndufb4	hypothetical gene supported by AF044957; NM_004547; NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa
NM_004408	DNM1	dynamin 1
NM_024571	SNRNP25	small nuclear ribonucleoprotein 25kDa (U11/U12)
NM_001009570	Cct7	chaperonin containing TCP1, subunit 7 (eta)
NM_170607	MLX	MAX-like protein X
NM_004373	Cox6a1	cytochrome c oxidase subunit VIa polypeptide 1
NM_004553	NDUFS6	NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13kDa (NADH-coenzyme Q reductase)
NM_001139441	BCAP31	B-cell receptor-associated protein 31
NM_004953	Eif4g1	eukaryotic translation initiation factor 4 gamma, 1
NM_032807	FBXO18	F-box protein, helicase, 18

REFERENCES

(2009) Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. *Nature* **457:** 1028-1032

Adelman K, Lis JT (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13:** 720-731

Adelman K, Marr MT, Werner J, Saunders A, Ni Z, Andrulis ED, Lis JT (2005) Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol Cell* **17:** 103-112

Aiyar SE, Sun JL, Blair AL, Moskaluk CA, Lu YZ, Ye QN, Yamaguchi Y, Mukherjee A, Ren DM, Handa H, Li R (2004) Attenuation of estrogen receptor alpha-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor. *Genes Dev* **18:** 2134-2146

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

Alen C, Kent NA, Jones HS, O'Sullivan J, Aranda A, Proudfoot NJ (2002) A role for chromatin remodeling in transcriptional termination by RNA polymerase II. *Mol Cell* **10:** 1441-1452

Andrulis ED, Guzman E, Doring P, Werner J, Lis JT (2000) High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes Dev* **14:** 2635-2649

Anindya R, Aygun O, Svejstrup JQ (2007) Damage-induced ubiquitylation of human RNA polymerase II by the ubiquitin ligase Nedd4, but not Cockayne syndrome proteins or BRCA1. *Mol Cell* **28:** 386-397

Araujo SJ, Tirode F, Coin F, Pospiech H, Syvaoja JE, Stucki M, Hubscher U, Egly JM, Wood RD (2000) Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK. *Genes Dev* **14:** 349-359

Armache KJ, Kettenberger H, Cramer P (2003) Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A* **100:** 6964-6968

Armache KJ, Mitterweger S, Meinhart A, Cramer P (2005) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* **280:** 7131-7134

Armstrong JA, Papoulas O, Daubresse G, Sperling AS, Lis JT, Scott MP, Tamkun JW (2002) The Drosophila BRM complex facilitates global transcription by RNA polymerase II. *EMBO J* **21:** 5245-5254

Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* **97:** 7090-7095

Auble DT, Hahn S (1993) An ATP-dependent inhibitor of TBP binding to DNA. *Genes Dev* **7:** 844-856

Baillat D, Hakimi MA, Naar AM, Shilatifard A, Cooch N, Shiekhattar R (2005) Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123:** 265-276

Baldeyron C, Soria G, Roche D, Cook AJ, Almouzni G (2011) HP1alpha recruitment to DNA damage by p150CAF-1 promotes homologous recombination repair. *J Cell Biol* **193:** 81-95

Banath JP, Banuelos CA, Klokov D, MacPhail SM, Lansdorp PM, Olive PL (2009) Explanation for excessive DNA single-strand breaks and endogenous repair foci in pluripotent mouse embryonic stem cells. *Exp Cell Res* **315:** 1505-1520

Barboric M, Nissen RM, Kanazawa S, Jabrane-Ferrat N, Peterlin BM (2001) NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol Cell* **8:** 327-337

Bartek J, Lukas J (2003) Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell* **3:** 421-429

Baskaran R, Dahmus ME, Wang JY (1993) Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc Natl Acad Sci U S A* **90:** 11167-11171

Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F (2012) A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* **45:** 158-170

Beerens N, Hoeijmakers JH, Kanaar R, Vermeulen W, Wyman C (2005) The CSB protein actively wraps DNA. *J Biol Chem* **280:** 4722-4729

Beltrao P, Albanese V, Kenner LR, Swaney DL, Burlingame A, Villen J, Lim WA, Fraser JS, Frydman J, Krogan NJ (2012) Systematic functional prioritization of protein posttranslational modifications. *Cell* **150:** 413-425

Bengal E, Flores O, Krauskopf A, Reinberg D, Aloni Y (1991) Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. *Mol Cell Biol* **11:** 1195-1206

Bergink S, Salomons FA, Hoogstraten D, Groothuis TA, de Waard H, Wu J, Yuan L, Citterio E, Houtsmuller AB, Neefjes J, Hoeijmakers JH, Vermeulen W, Dantuma NP (2006) DNA damage triggers nucleotide excision repair-dependent monoubiquitylation of histone H2A. *Genes Dev* **20:** 1343-1352

Berneburg M, Lehmann AR (2001) Xeroderma pigmentosum and related disorders: defects in DNA repair and transcription. *Adv Genet* **43:** 71-102

Bhaumik SR (2011) Distinct regulatory mechanisms of eukaryotic transcriptional activation by SAGA and TFIID. *Biochim Biophys Acta* **1809:** 97-108

Bhaumik SR, Green MR (2001) SAGA is an essential in vivo target of the yeast acidic activator Gal4p. *Genes Dev* **15:** 1935-1945

Bird G, Zorio DA, Bentley DL (2004) RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol* **24:** 8963-8969

Boeger H, Bushnell DA, Davis R, Griesenbeck J, Lorch Y, Strattan JS, Westover KD, Kornberg RD (2005) Structural basis of eukaryotic gene transcription. *FEBS Lett* **579:** 899-903

Boehm AK, Saunders A, Werner J, Lis JT (2003) Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. *Mol Cell Biol* **23:** 7628-7637

Boeing S, Rigault C, Heidemann M, Eick D, Meisterernst M (2010) RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. *J Biol Chem* **285:** 188-196

Bootsma D, Hoeijmakers JH (1993) DNA repair. Engagement with transcription. *Nature* **363:** 114-115

Bortvin A, Winston F (1996) Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science* **272:** 1473-1476

Briggs SD, Xiao T, Sun ZW, Caldwell JA, Shabanowitz J, Hunt DF, Allis CD, Strahl BD (2002) Gene silencing: trans-histone regulatory pathway in chromatin. *Nature* **418:** 498

Brookes E, de Santiago I, Hebenstreit D, Morris KJ, Carroll T, Xie SQ, Stock JK, Heidemann M, Eick D, Nozaki N, Kimura H, Ragoussis J, Teichmann SA, Pombo A (2012) Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell* **10:** 157-170

Brown CE, Lechner T, Howe L, Workman JL (2000) The many HATs of transcription coactivators. *Trends Biochem Sci* **25:** 15-19

Buratowski S (2005) Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol* **17:** 257-261

Buratowski S, Hahn S, Guarente L, Sharp PA (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56:** 549-561

Buratowski S, Sharp PA (1990) Transcription initiation complexes and upstream activation with RNA polymerase II lacking the C-terminal domain of the largest subunit. *Mol Cell Biol* **10:** 5562-5564

Bushnell DA, Kornberg RD (2003) Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* **100:** 6969-6973

Caldecott KW (2007) Mammalian single-strand break repair: mechanisms and links with chromatin. *DNA Repair (Amst)* **6:** 443-453

Caldecott KW (2008) Single-strand break repair and genetic disease. *Nat Rev Genet* **9:** 619-631

Calvo O, Manley JL (2003) Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev* **17:** 1321-1327

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38:** 626-635

Cazzalini O, Perucca P, Savio M, Necchi D, Bianchi L, Stivala LA, Ducommun B, Scovassi AI, Prosperi E (2008) Interaction of p21(CDKN1A) with PCNA regulates the histone acetyltransferase activity of p300 in nucleotide excision repair. *Nucleic Acids Res* **36:** 1713-1722

Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E, Eick D (2007) Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318:** 1780-1782

Chapman RD, Heidemann M, Hintermair C, Eick D (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet* **24:** 289-296

Chen W, Roeder RG (2011) Mediator-dependent nuclear receptor function. *Semin Cell Dev Biol* **22:** 749-758

Chen X, Ruggiero C, Li S (2007) Yeast Rpb9 plays an important role in ubiquitylation and degradation of Rpb1 in response to UV-induced DNA damage. *Mol Cell Biol* **27:** 4617-4625

Cheng B, Price DH (2007) Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J Biol Chem* **282:** 21901-21912

Cheng B, Price DH (2008) Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Res* **36:** e135

Chodchoy N, Pandey NB, Marzluff WF (1991) An intact histone 3'-processing site is required for transcription termination in a mouse histone H2a gene. *Mol Cell Biol* **11:** 497-509

Chou DM, Adamson B, Dephoure NE, Tan X, Nottke AC, Hurov KE, Gygi SP, Colaiacovo MP, Elledge SJ (2010) A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc Natl Acad Sci U S A* **107:** 18475-18480

Christiansen M, Stevnsner T, Modin C, Martensen PM, Brosh RM, Jr., Bohr VA (2003) Functional consequences of mutations in the conserved SF2 motifs and post-translational phosphorylation of the CSB protein. *Nucleic Acids Res* **31:** 963-973

Citterio E, Rademakers S, van der Horst GT, van Gool AJ, Hoeijmakers JH, Vermeulen W (1998) Biochemical and biological characterization of wild-type and ATPase-deficient Cockayne syndrome B repair protein. *J Biol Chem* **273:** 11844-11851

Coin F, Proietti De Santis L, Nardo T, Zlobinskaya O, Stefanini M, Egly JM (2006) p8/TTD-A as a repair-specific TFIIH subunit. *Mol Cell* **21:** 215-226

Coleman RA, Taggart AK, Burma S, Chicca JJ, 2nd, Pugh BF (1999) TFIIA regulates TBP and TFIID dimers. *Mol Cell* **4:** 451-457

Comer FI, Hart GW (2001) Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* **40:** 7845-7852

Compe E, Egly JM (2012) TFIIH: when transcription met DNA repair. *Nat Rev Mol Cell Biol* **13:** 343-354

Conaway JW, Conaway RC (1989) A multisubunit transcription factor essential for accurate initiation by RNA polymerase II. *J Biol Chem* **264:** 2357-2362

Conaway RC, Kong SE, Conaway JW (2003) TFIIS and GreB: two like-minded transcription elongation factors with sticky fingers. *Cell* **114:** 272-274

Concino MF, Lee RF, Merryweather JP, Weinmann R (1984) The adenovirus major late promoter TATA box and initiation site are both necessary for transcription in vitro. *Nucleic Acids Res* **12:** 7423-7433

Connelly S, Manley JL (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* **2:** 440-452

Corden JL (1990) Tails of RNA polymerase II. *Trends Biochem Sci* **15:** 383-387

Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322:** 1845-1848

Costa RM, Chigancas V, Galhardo Rda S, Carvalho H, Menck CF (2003) The eukaryotic nucleotide excision repair pathway. *Biochimie* **85:** 1083-1099

Cotner-Gohara E, Kim IK, Tomkinson AE, Ellenberger T (2008) Two DNA-binding and nick recognition modules in human DNA ligase III. *J Biol Chem* **283:** 10764-10772

Cramer P (2004a) RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev* **14:** 218-226

Cramer P (2004b) Structure and function of RNA polymerase II. *Adv Protein Chem* **67:** 1-42

Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292:** 1863-1876

Czudnochowski N, Bosken CA, Geyer M (2012) Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun* **3:** 842

Dantonel JC, Murthy KG, Manley JL, Tora L (1997) Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* **389:** 399-402

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

de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* **12:** 525-532

de Vries H, Ruegsegger U, Hubner W, Friedlein A, Langen H, Keller W (2000) Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J* **19:** 5895-5904

Deng W, Roberts SG (2007) TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma* **116:** 417-429

Diribarne G, Bensaude O (2009) 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. *RNA Biol* **6:** 122-128

Donahue BA, Yin S, Taylor JS, Reines D, Hanawalt PC (1994) Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template. *Proc Natl Acad Sci U S A* **91:** 8502-8506

Dunkern TR, Kaina B (2002) Cell proliferation and DNA breaks are involved in ultraviolet light-induced apoptosis in nucleotide excision repair-deficient Chinese hamster cells. *Mol Biol Cell* **13:** 348-361

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

Dvir A, Conaway RC, Conaway JW (1997) A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. *Proc Natl Acad Sci U S A* **94:** 9006-9010

Eberhardy SR, Farnham PJ (2002) Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J Biol Chem* **277:** 40156-40162

Egloff S, Murphy S (2008) Cracking the RNA polymerase II CTD code. *Trends Genet* **24:** 280-288

Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318:** 1777-1779

Egloff S, O'Reilly D, Murphy S (2008) Expression of human snRNA genes from beginning to end. *Biochem Soc Trans* **36:** 590-594

Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S (2010) The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem* **285:** 20564-20569

Eskiw CH, Fraser P (2011) Ultrastructural study of transcription factories in mouse erythroblasts. *J Cell Sci* **124:** 3676-3683

Evans R, Fairley JA, Roberts SG (2001) Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes Dev* **15:** 2945-2949

Fire A, Samuels M, Sharp PA (1984) Interactions between RNA polymerase II, factors, and template leading to accurate transcription. *J Biol Chem* **259:** 2509-2516

Fish RN, Kane CM (2002) Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta* **1577:** 287-307

Flores O, Ha I, Reinberg D (1990) Factors involved in specific transcription by mammalian RNA polymerase II. Purification and subunit composition of transcription factor IIF. *J Biol Chem* **265:** 5629-5634

Fousteri M, Vermeulen W, van Zeeland AA, Mullenders LH (2006) Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Mol Cell* **23:** 471-482

Furnari B, Blasina A, Boddy MN, McGowan CH, Russell P (1999) Cdc25 inhibited in vivo and in vitro by checkpoint kinases Cds1 and Chk1. *Mol Biol Cell* **10:** 833-845

Gao H, Coyle DL, Meyer-Ficca ML, Meyer RG, Jacobson EL, Wang ZQ, Jacobson MK (2007) Altered poly(ADP-ribose) metabolism impairs cellular responses to genotoxic stress in a hypomorphic mutant of poly(ADP-ribose) glycohydrolase. *Exp Cell Res* **313:** 984-996

Gerber M, Ma J, Dean K, Eissenberg JC, Shilatifard A (2001) Drosophila ELL is associated with actively elongating RNA polymerase II on transcriptionally active sites in vivo. *EMBO J* **20:** 6104-6114

Gilchrist DA, Adelman K (2012) Coupling polymerase pausing and chromatin landscapes for precise regulation of transcription. *Biochim Biophys Acta* **1819:** 700-706

Gilchrist DA, Dos Santos G, Fargo DC, Xie B, Gao Y, Li L, Adelman K (2010) Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143:** 540-551

Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, Bentley DL (2009) TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol Cell Biol* **29:** 5455-5464

Gomes NP, Bjerke G, Llorente B, Szostek SA, Emerson BM, Espinosa JM (2006) Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. *Genes Dev* **20:** 601-612

Goppelt A, Stelzer G, Lottspeich F, Meisterernst M (1996) A mechanism for repression of class II gene transcription through specific binding of NC2 to TBP-promoter complexes via heterodimeric histone fold domains. *EMBO J* **15:** 3105-3116

Groisman R, Kuraoka I, Chevallier O, Gaye N, Magnaldo T, Tanaka K, Kisselev AF, Harel-Bellan A, Nakatani Y (2006) CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway

establishes a link between complementation factors of the Cockayne syndrome. *Genes Dev* **20:** 1429-1434

Grosso AR, de Almeida SF, Braga J, Carmo-Fonseca M (2012) Dynamic transitions in RNA polymerase II density profiles during transcription termination. *Genome Res* **22:** 1447-1456

Gu W, Reines D (1995) Identification of a decay in transcription potential that results in elongation factor dependence of RNA polymerase II. *J Biol Chem* **270:** 11238-11244

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

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130:** 77-88

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

Hampsey M (1998) Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* **62:** 465-503

Hanawalt PC, Spivak G (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* **9:** 958-970

Hanyu-Nakamura K, Sonobe-Nojima H, Tanigawa A, Lasko P, Nakamura A (2008) Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature* **451:** 730-733

Hargreaves DC, Horng T, Medzhitov R (2009) Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138:** 129-145

Harreman M, Taschner M, Sigurdsson S, Anindya R, Reid J, Somesh B, Kong SE, Banks CA, Conaway RC, Conaway JW, Svejstrup JQ (2009) Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proc Natl Acad Sci U S A* **106:** 20705-20710

Heidemann M, Hintermair C, Voss K, Eick D (2012) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta*

Hendrix DA, Hong JW, Zeitlinger J, Rokhsar DS, Levine MS (2008) Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. *Proc Natl Acad Sci U S A* **105:** 7762-7767

Higashi M, Inoue S, Ito T (2010) Core histone H2A ubiquitylation and transcriptional regulation. *Exp Cell Res* **316:** 2707-2712

Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R, Ferrier P, Flatley A, Kremmer E, Chapman RD, Andrau JC, Eick D (2012) Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J* **31:** 2784-2797

Hirose Y, Manley JL (1998) RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395:** 93-96

Hirose Y, Tacke R, Manley JL (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev* **13:** 1234-1239

Hsin JP, Sheth A, Manley JL (2011) RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* **334:** 683-686

Ikeda F, Dikic I (2008) Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series. *EMBO Rep* **9:** 536-542

Ikehata H, Ono T (2011) The mechanisms of UV mutagenesis. *J Radiat Res* **52:** 115-125

Ivan M, Kaelin WG, Jr. (2001) The von Hippel-Lindau tumor suppressor protein. *Curr Opin Genet Dev* **11:** 27-34

Ivanov D, Kwak YT, Guo J, Gaynor RB (2000) Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol Cell Biol* **20:** 2970-2983

Izban MG, Luse DS (1992) Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Biol Chem* **267:** 13647-13655

Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT (2008) The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol* **20:** 253-259

Kaiser K, Meisterernst M (1996) The human general co-factors. *Trends Biochem Sci* **21:** 342-345

Kalogeraki VS, Tornaletti S, Cooper PK, Hanawalt PC (2005) Comparative TFIIS-mediated transcript cleavage by mammalian RNA polymerase II arrested at a lesion in different transcription systems. *DNA Repair (Amst)* **4:** 1075-1087

Kamieniarz K, Izzo A, Dundr M, Tropberger P, Ozretic L, Kirfel J, Scheer E, Tropel P, Wisniewski JR, Tora L, Viville S, Buettner R, Schneider R (2012) A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation. *Genes Dev* **26:** 797-802

Kamiuchi S, Saijo M, Citterio E, de Jager M, Hoeijmakers JH, Tanaka K (2002) Translocation of Cockayne syndrome group A protein to the nuclear matrix: possible relevance to transcription-coupled DNA repair. *Proc Natl Acad Sci U S A* **99:** 201-206

Kamura T, Burian D, Yan Q, Schmidt SL, Lane WS, Querido E, Branton PE, Shilatifard A, Conaway RC, Conaway JW (2001) Muf1, a novel Elongin BC-interacting leucine-rich repeat protein that can assemble with Cul5 and Rbx1 to reconstitute a ubiquitin ligase. *J Biol Chem* **276:** 29748-29753

Kaplan CD, Morris JR, Wu C, Winston F (2000) Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. *Genes Dev* **14:** 2623-2634

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

Kaufmann I, Martin G, Friedlein A, Langen H, Keller W (2004) Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J* **23:** 616-626

Kettenberger H, Armache KJ, Cramer P (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114:** 347-357

Kim JA, Kruhlak M, Dotiwala F, Nussenzweig A, Haber JE (2007) Heterochromatin is refractory to gamma-H2AX modification in yeast and mammals. *J Cell Biol* **178:** 209-218

Kim JB, Sharp PA (2001) Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *J Biol Chem* **276:** 12317-12323

Kim JL, Burley SK (1994) 1.9 A resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat Struct Biol* **1:** 638-653

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

Kim M, Suh H, Cho EJ, Buratowski S (2009) Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J Biol Chem* **284:** 26421-26426

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

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

Koch-Paiz CA, Amundson SA, Bittner ML, Meltzer PS, Fornace AJ, Jr. (2004) Functional genomics of UV radiation responses in human cells. *Mutat Res* **549:** 65-78

Kokubo T, Gong DW, Yamashita S, Horikoshi M, Roeder RG, Nakatani Y (1993) Drosophila 230-kD TFIID subunit, a functional homolog of the human cell cycle gene product, negatively regulates DNA binding of the TATA box-binding subunit of TFIID. *Genes Dev* **7:** 1033-1046

Komarnitsky P, Cho EJ, Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14:** 2452-2460

Koning TM, Davies RJ, Kaptein R (1990) The solution structure of the intramolecular photoproduct of d(TpA) derived with the use of NMR and a combination of distance geometry and molecular dynamics. *Nucleic Acids Res* **18:** 277-284

Kornberg RD (2005) Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* **30:** 235-239

Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M, Greenblatt JF, Shilatifard A (2003) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* **11:** 721-729

Krogan NJ, Kim M, Ahn SH, Zhong G, Kobor MS, Cagney G, Emili A, Shilatifard A, Buratowski S, Greenblatt JF (2002) RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol Cell Biol* **22:** 6979-6992

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

Kugel JF, Goodrich JA (1998) Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA. *Proc Natl Acad Sci U S A* **95:** 9232-9237

Kugel JF, Goodrich JA (2002) Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape. *Mol Cell Biol* **22:** 762-773

Kumar KP, Akoulitchev S, Reinberg D (1998) Promoter-proximal stalling results from the inability to recruit transcription factor IIH to the transcription complex and is a regulated event. *Proc Natl Acad Sci U S A* **95:** 9767-9772

Kuraoka I, Suzuki K, Ito S, Hayashida M, Kwei JS, Ikegami T, Handa H, Nakabeppu Y, Tanaka K (2007) RNA polymerase II bypasses 8-oxoguanine in the presence of transcription elongation factor TFIIS. *DNA Repair (Amst)* **6:** 841-851

Kusumoto R, Masutani C, Sugasawa K, Iwai S, Araki M, Uchida A, Mizukoshi T, Hanaoka F (2001) Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro. *Mutat Res* **485:** 219-227

Kutach AK, Kadonaga JT (2000) The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol Cell Biol* **20:** 4754-4764

Lagerwerf S, Vrouwe MG, Overmeer RM, Fousteri MI, Mullenders LH (2011) DNA damage response and transcription. *DNA Repair (Amst)* **10:** 743-750

Lake RJ, Geyko A, Hemashettar G, Zhao Y, Fan HY (2010) UV-induced association of the CSB remodeling protein with chromatin requires ATP-dependent relief of N-terminal autorepression. *Mol Cell* **37:** 235-246

Landick R (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* **34:** 1062-1066

Langerak P, Russell P (2011) Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair. *Philos Trans R Soc Lond B Biol Sci* **366:** 3562-3571

Lee TI, Young RA (2000) Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* **34:** 77-137

Lees-Miller SP, Meek K (2003) Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* **85:** 1161-1173

Lehmann AR (2003) DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *Biochimie* **85:** 1101-1111

Lescure A, Lutz Y, Eberhard D, Jacq X, Krol A, Grummt I, Davidson I, Chambon P, Tora L (1994) The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters. *EMBO J* **13:** 1166-1175

Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. *Cell* **128:** 707-719

Li B, Pattenden SG, Lee D, Gutierrez J, Chen J, Seidel C, Gerton J, Workman JL (2005) Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc Natl Acad Sci U S A* **102:** 18385-18390

Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell* **9:** 1101-1111

Licht CL, Stevnsner T, Bohr VA (2003) Cockayne syndrome group B cellular and biochemical functions. *Am J Hum Genet* **73:** 1217-1239

Lieber MR (2008) The mechanism of human nonhomologous DNA end joining. *J Biol Chem* **283:** 1-5

Lim JH, West KL, Rubinstein Y, Bergel M, Postnikov YV, Bustin M (2005) Chromosomal protein HMGN1 enhances the acetylation of lysine 14 in histone H3. *EMBO J* **24:** 3038-3048

Limoli CL, Giedzinski E, Bonner WM, Cleaver JE (2002) UV-induced replication arrest in the xeroderma pigmentosum variant leads to DNA double-strand breaks, gamma -H2AX formation, and Mre11 relocalization. *Proc Natl Acad Sci U S A* **99:** 233-238

Lin C, Smith ER, Takahashi H, Lai KC, Martin-Brown S, Florens L, Washburn MP, Conaway JW, Conaway RC, Shilatifard A (2010) AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol Cell* **37:** 429-437

Lindstrom DL, Squazzo SL, Muster N, Burckin TA, Wachter KC, Emigh CA, McCleery JA, Yates JR, 3rd, Hartzog GA (2003) Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol Cell Biol* **23:** 1368-1378

Lis J (1998) Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. *Cold Spring Harb Symp Quant Biol* **63:** 347-356

Lis JT, Mason P, Peng J, Price DH, Werner J (2000) P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev* **14:** 792-803

Ljungman M, O'Hagan HM, Paulsen MT (2001) Induction of ser15 and lys382 modifications of p53 by blockage of transcription elongation. *Oncogene* **20:** 5964-5971

Ljungman M, Zhang F (1996) Blockage of RNA polymerase as a possible trigger for u.v. light-induced apoptosis. *Oncogene* **13:** 823-831

Ljungman M, Zhang F, Chen F, Rainbow AJ, McKay BC (1999) Inhibition of RNA polymerase II as a trigger for the p53 response. *Oncogene* **18:** 583-592

Logan J, Falck-Pedersen E, Darnell JE, Jr., Shenk T (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A* **84:** 8306-8310

Lu H, Zawel L, Fisher L, Egly JM, Reinberg D (1992) Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* **358:** 641-645

Luo Z, Lin C, Shilatifard A (2012) The super elongation complex (SEC) family in transcriptional control. *Nat Rev Mol Cell Biol* **13:** 543-547

Malik S, Roeder RG (2005) Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci* **30:** 256-263

Mandal SS, Chu C, Wada T, Handa H, Shatkin AJ, Reinberg D (2004) Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc Natl Acad Sci U S A* **101:** 7572-7577

Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444:** 953-956

Margaritis T, Holstege FC (2008) Poised RNA polymerase II gives pause for thought. *Cell* **133:** 581-584

Marinoni JC, Roy R, Vermeulen W, Miniou P, Lutz Y, Weeda G, Seroz T, Gomez DM, Hoeijmakers JH, Egly JM (1997) Cloning and characterization of p52, the fifth subunit of the core of the transcription/DNA repair factor TFIIH. *EMBO J* **16:** 1093-1102

Marshall NF, Price DH (1992) Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol* **12:** 2078-2090

Marshall NF, Price DH (1995) Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem* **270:** 12335-12338

Marzluff WF, Wagner EJ, Duronio RJ (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* **9:** 843-854

Mason PB, Struhl K (2003) The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Mol Cell Biol* **23:** 8323-8333

Massip L, Caron P, Iacovoni JS, Trouche D, Legube G (2010) Deciphering the chromatin landscape induced around DNA double strand breaks. *Cell Cycle* **9:** 2963-2972

Mathieu N, Kaczmarek N, Naegeli H (2010) Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. *Proc Natl Acad Sci U S A* **107:** 17545-17550

Matsumoto M, Yaginuma K, Igarashi A, Imura M, Hasegawa M, Iwabuchi K, Date T, Mori T, Ishizaki K, Yamashita K, Inobe M, Matsunaga T (2007) Perturbed gap-filling synthesis in nucleotide excision repair causes histone H2AX phosphorylation in human quiescent cells. *J Cell Sci* **120:** 1104-1112

Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, Gilmour DS, Albert I, Pugh BF (2008) Nucleosome organization in the Drosophila genome. *Nature* **453:** 358-362

Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336:** 1723-1725

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

McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385:** 357-361

McKay BC, Stubbert LJ, Fowler CC, Smith JM, Cardamore RA, Spronck JC (2004) Regulation of ultraviolet light-induced gene expression by gene size. *Proc Natl Acad Sci U S A* **101:** 6582-6586

Medlin J, Scurry A, Taylor A, Zhang F, Peterlin BM, Murphy S (2005) P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *EMBO J* **24:** 4154-4165

Meijer HA, Bushell M, Hill K, Gant TW, Willis AE, Jones P, de Moor CH (2007) A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells. *Nucleic Acids Res* **35:** e132

Meinhart A, Cramer P (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430:** 223-226

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. *Genes Dev* **19:** 1401-1415

Miller T, Williams K, Johnstone RW, Shilatifard A (2000) Identification, cloning, expression, and biochemical characterization of the testis-specific RNA polymerase II elongation factor ELL3. *J Biol Chem* **275:** 32052-32056

Minakhin L, Bhagat S, Brunning A, Campbell EA, Darst SA, Ebright RH, Severinov K (2001) Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* **98:** 892-897

Minvielle-Sebastia L, Preker PJ, Wiederkehr T, Strahm Y, Keller W (1997) The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. *Proc Natl Acad Sci U S A* **94:** 7897-7902

Missra A, Gilmour DS (2010) Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc Natl Acad Sci U S A* **107:** 11301-11306

Misteli T, Soutoglou E (2009) The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nat Rev Mol Cell Biol* **10:** 243-254

Mitsui A, Sharp PA (1999) Ubiquitination of RNA polymerase II large subunit signaled by phosphorylation of carboxyl-terminal domain. *Proc Natl Acad Sci U S A* **96:** 6054-6059

Mitsuzawa H, Ishihama A (2004) RNA polymerase II transcription apparatus in Schizosaccharomyces pombe. *Curr Genet* **44:** 287-294

Moncollin V, Schaeffer L, Chalut C, Egly JM (1992) Expression in Escherichia coli: purification and properties of the recombinant human general transcription factor rTFIIB. *Protein Expr Purif* **3:** 374-379

Moore MJ, Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136:** 688-700

Moreland RJ, Tirode F, Yan Q, Conaway JW, Egly JM, Conaway RC (1999) A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. *J Biol Chem* **274:** 22127-22130

Morillon A, Karabetsou N, O'Sullivan J, Kent N, Proudfoot N, Mellor J (2003) Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. *Cell* **115:** 425-435

Muller F, Zaucker A, Tora L (2010) Developmental regulation of transcription initiation: more than just changing the actors. *Curr Opin Genet Dev* **20:** 533-540

Munoz MJ, Perez Santangelo MS, Paronetto MP, de la Mata M, Pelisch F, Boireau S, Glover-Cutter K, Ben-Dov C, Blaustein M, Lozano JJ, Bird G, Bentley D, Bertrand E, Kornblihtt AR (2009) DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* **137:** 708-720

Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K (2007) RNA polymerase is poised for activation across the genome. *Nat Genet* **39:** 1507-1511

Nag A, Narsinh K, Kazerouninia A, Martinson HG (2006) The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity. *RNA* **12:** 1534-1544

Nag A, Narsinh K, Martinson HG (2007) The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nat Struct Mol Biol* **14:** 662-669

Nagy Z, Riss A, Fujiyama S, Krebs A, Orpinell M, Jansen P, Cohen A, Stunnenberg HG, Kato S, Tora L (2010) The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. *Cell Mol Life Sci* **67:** 611-628

Narita T, Yamaguchi Y, Yano K, Sugimoto S, Chanarat S, Wada T, Kim DK, Hasegawa J, Omori M, Inukai N, Endoh M, Yamada T, Handa H (2003) Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Mol Cell Biol* **23:** 1863-1873

Narita T, Yung TM, Yamamoto J, Tsuboi Y, Tanabe H, Tanaka K, Yamaguchi Y, Handa H (2007) NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Mol Cell* **26:** 349-365

Narlikar GJ, Fan HY, Kingston RE (2002) Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108:** 475-487

Nechaev S, Adelman K (2008) Promoter-proximal Pol II: when stalling speeds things up. *Cell Cycle* **7:** 1539-1544

Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K (2010) Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* **327:** 335-338

Nikolov DB, Burley SK (1994) 2.1 A resolution refined structure of a TATA box-binding protein (TBP). *Nat Struct Biol* **1:** 621-637

Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A (2005) Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* **12:** 144-151

Ogi T, Lehmann AR (2006) The Y-family DNA polymerase kappa (pol kappa) functions in mammalian nucleotide-excision repair. *Nat Cell Biol* **8:** 640-642

Ohkuma Y, Roeder RG (1994) Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* **368:** 160-163

Ohnishi T, Mori E, Takahashi A (2009) DNA double-strand breaks: their production, recognition, and repair in eukaryotes. *Mutat Res* **669:** 8-12

Orlova M, Newlands J, Das A, Goldfarb A, Borukhov S (1995) Intrinsic transcript cleavage activity of RNA polymerase. *Proc Natl Acad Sci U S A* **92:** 4596-4600

Orphanides G, Wu WH, Lane WS, Hampsey M, Reinberg D (1999) The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* **400:** 284-288

Pal M, Luse DS (2003) The initiation-elongation transition: lateral mobility of RNA in RNA polymerase II complexes is greatly reduced at +8/+9 and absent by +23. *Proc Natl Acad Sci U S A* **100:** 5700-5705

Palancade B, Bensaude O (2003) Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur J Biochem* **270:** 3859-3870

Pankotai T, Bonhomme C, Chen D, Soutoglou E (2012) DNAPKcs-dependent arrest of RNA polymerase II transcription in the presence of DNA breaks. *Nat Struct Mol Biol* **19:** 276-282

Pankotai T, Ujfaludi Z, Vamos E, Suri K, Boros IM (2010) The dissociable RPB4 subunit of RNA Pol II has vital functions in Drosophila. *Mol Genet Genomics* **283:** 89-97

Pei Y, Schwer B, Shuman S (2003) Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *J Biol Chem* **278:** 7180-7188

Pei Y, Shuman S (2002) Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J Biol Chem* **277:** 19639-19648

Peterlin BM, Price DH (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23:** 297-305

Phatnani HP, Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20:** 2922-2936

Pitcher RS, Wilson TE, Doherty AJ (2005) New insights into NHEJ repair processes in prokaryotes. *Cell Cycle* **4:** 675-678

Poehlmann A, Roessner A (2010) Importance of DNA damage checkpoints in the pathogenesis of human cancers. *Pathol Res Pract* **206:** 591-601

Pokholok DK, Hannett NM, Young RA (2002) Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol Cell* **9:** 799-809

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122:** 517-527

Price DH (2000) P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol Cell Biol* **20:** 2629-2634

Proudfoot NJ (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14:** 105-110

Proudfoot NJ (2011) Ending the message: poly(A) signals then and now. *Genes Dev* **25:** 1770-1782

Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* **108:** 501-512

Pruitt KD, Tatusova T, Klimke W, Maglott DR (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res* **37:** D32-36

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA (2010) c-Myc regulates transcriptional pause release. *Cell* **141:** 432-445

Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD (2005) Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123:** 233-248

Ranuncolo SM, Ghosh S, Hanover JA, Hart GW, Lewis BA (2012) Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. *J Biol Chem* **287:** 23549-23561

Rasmussen EB, Lis JT (1993) In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc Natl Acad Sci U S A* **90:** 7923-7927

Rastogi RP, Richa, Kumar A, Tyagi MB, Sinha RP (2010) Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J Nucleic Acids* **2010:** 592980

Ratner JN, Balasubramanian B, Corden J, Warren SL, Bregman DB (1998) Ultraviolet radiation-induced ubiquitination and proteasomal degradation of the large subunit of RNA polymerase II. Implications for transcription-coupled DNA repair. *J Biol Chem* **273:** 5184-5189

Reardon JT, Sancar A (2003) Recognition and repair of the cyclobutane thymine dimer, a major cause of skin cancers, by the human excision nuclease. *Genes Dev* **17:** 2539-2551

Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH (2001) A highly purified RNA polymerase II elongation control system. *J Biol Chem* **276:** 42601-42609

Rodriguez-Navarro S (2009) Insights into SAGA function during gene expression. *EMBO Rep* **10:** 843-850

Rosonina E, Kaneko S, Manley JL (2006) Terminating the transcript: breaking up is hard to do. *Genes Dev* **20:** 1050-1056

Rougvie AE, Lis JT (1988) The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell* **54:** 795-804

Rudd MD, Izban MG, Luse DS (1994) The active site of RNA polymerase II participates in transcript cleavage within arrested ternary complexes. *Proc Natl Acad Sci U S A* **91:** 8057-8061

Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* **73:** 39-85

Sarker AH, Tsutakawa SE, Kostek S, Ng C, Shin DS, Peris M, Campeau E, Tainer JA, Nogales E, Cooper PK (2005) Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: insights for transcription-coupled repair and Cockayne Syndrome. *Mol Cell* **20:** 187-198

Saunders A, Core LJ, Lis JT (2006) Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* **7:** 557-567

Schaeffer L, Moncollin V, Roy R, Staub A, Mezzina M, Sarasin A, Weeda G, Hoeijmakers JH, Egly JM (1994) The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. *EMBO J* **13:** 2388-2392

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132:** 887-898

Schroeder SC, Schwer B, Shuman S, Bentley D (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev* **14:** 2435-2440

Schwartz BE, Ahmad K (2005) Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev* **19:** 804-814

Schwer B, Shuman S (2011) Deciphering the RNA polymerase II CTD code in fission yeast. *Mol Cell* **43:** 311-318

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA (2008) Divergent transcription from active promoters. *Science* **322:** 1849-1851

Sekiguchi M, Tsuzuki T (2002) Oxidative nucleotide damage: consequences and prevention. *Oncogene* **21:** 8895-8904

Selby CP, Drapkin R, Reinberg D, Sancar A (1997) RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair. *Nucleic Acids Res* **25:** 787-793

Shao H, Revach M, Moshonov S, Tzuman Y, Gazit K, Albeck S, Unger T, Dikstein R (2005) Core promoter binding by histone-like TAF complexes. *Mol Cell Biol* **25:** 206-219

Shilatifard A, Conaway RC, Conaway JW (2003) The RNA polymerase II elongation complex. *Annu Rev Biochem* **72:** 693-715

Shilatifard A, Lane WS, Jackson KW, Conaway RC, Conaway JW (1996) An RNA polymerase II elongation factor encoded by the human ELL gene. *Science* **271:** 1873-1876

Shiloh Y (2003) ATM and related protein kinases: safeguarding genome integrity. *Nat Rev Cancer* **3:** 155-168

Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311:** 844-847

Sims RJ, 3rd, Rojas LA, Beck D, Bonasio R, Schuller R, Drury WJ, 3rd, Eick D, Reinberg D (2011) The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* **332:** 99-103

Singh J, Padgett RA (2009) Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* **16:** 1128-1133

Sluder AE, Greenleaf AL, Price DH (1989) Properties of a Drosophila RNA polymerase II elongation factor. *J Biol Chem* **264:** 8963-8969

Smale ST (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim Biophys Acta* **1351:** 73-88

Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. *Annu Rev Biochem* **72:** 449-479

Smith E, Lin C, Shilatifard A (2011a) The super elongation complex (SEC) and MLL in development and disease. *Genes Dev* **25:** 661-672

Smith ER, Lin C, Garrett AS, Thornton J, Mohaghegh N, Hu D, Jackson J, Saraf A, Swanson SK, Seidel C, Florens L, Washburn MP, Eissenberg JC, Shilatifard A (2011b) The little elongation complex regulates small nuclear RNA transcription. *Mol Cell* **44:** 954-965

Smolle M, Venkatesh S, Gogol MM, Li H, Zhang Y, Florens L, Washburn MP, Workman JL (2012) Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol* **19:** 884-892

Sobhian B, Laguette N, Yatim A, Nakamura M, Levy Y, Kiernan R, Benkirane M (2010) HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. *Mol Cell* **38:** 439-451

Squazzo SL, Costa PJ, Lindstrom DL, Kumer KE, Simic R, Jennings JL, Link AJ, Arndt KM, Hartzog GA (2002) The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J* **21:** 1764-1774

St Amour CV, Sanso M, Bosken CA, Lee KM, Larochelle S, Zhang C, Shokat KM, Geyer M, Fisher RP (2012) Separate domains of fission yeast Cdk9 (P-TEFb) are required for capping enzyme recruitment and primed (Ser7-phosphorylated) Rpb1 carboxyl-terminal domain substrate recognition. *Mol Cell Biol* **32:** 2372-2383

Starita LM, Horwitz AA, Keogh MC, Ishioka C, Parvin JD, Chiba N (2005) BRCA1/BARD1 ubiquitinate phosphorylated RNA polymerase II. *J Biol Chem* **280:** 24498-24505

Steinmetz EJ, Brow DA (2003) Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Mol Cell Biol* **23:** 6339-6349

Stiller JW, Cook MS (2004) Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot Cell* **3:** 735-740

Stiller JW, McConaughy BL, Hall BD (2000) Evolutionary complementation for polymerase II CTD function. *Yeast* **16:** 57-64

Stokes DG, Tartof KD, Perry RP (1996) CHD1 is concentrated in interbands and puffed regions of Drosophila polytene chromosomes. *Proc Natl Acad Sci U S A* **93:** 7137-7142

Subtil-Rodriguez A, Reyes JC (2010) BRG1 helps RNA polymerase II to overcome a nucleosomal barrier during elongation, in vivo. *EMBO Rep* **11:** 751-757

Subtil-Rodriguez A, Reyes JC (2011) To cross or not to cross the nucleosome, that is the elongation question. *RNA Biol* **8:** 389-393

Sugasawa K (2010) Regulation of damage recognition in mammalian global genomic nucleotide excision repair. *Mutat Res* **685:** 29-37

Svejstrup JQ (2002) Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol* **3:** 21-29

Svejstrup JQ (2007) Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem Sci* **32:** 165-171

Taft RJ, Kaplan CD, Simons C, Mattick JS (2009) Evolution, biogenesis and function of promoter-associated RNAs. *Cell Cycle* **8:** 2332-2338

Takagaki Y, Manley JL (2000) Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* **20:** 1515-1525

Takagi Y, Conaway RC, Conaway JW (1996) Characterization of elongin C functional domains required for interaction with elongin B and activation of elongin A. *J Biol Chem* **271:** 25562-25568

Takagi Y, Kornberg RD (2006) Mediator as a general transcription factor. *J Biol Chem* **281:** 80-89

Takagi Y, Masuda CA, Chang WH, Komori H, Wang D, Hunter T, Joazeiro CA, Kornberg RD (2005) Ubiquitin ligase activity of TFIIH and the transcriptional response to DNA damage. *Mol Cell* **18:** 237-243

Tan S, Aso T, Conaway RC, Conaway JW (1994) Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. *J Biol Chem* **269:** 25684-25691

Thirman MJ, Levitan DA, Kobayashi H, Simon MC, Rowley JD (1994) Cloning of ELL, a gene that fuses to MLL in a t(11;19)(q23;p13.1) in acute myeloid leukemia. *Proc Natl Acad Sci U S A* **91:** 12110-12114

Tietjen JR, Zhang DW, Rodriguez-Molina JB, White BE, Akhtar MS, Heidemann M, Li X, Chapman RD, Shokat K, Keles S, Eick D, Ansari AZ (2010) Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* **17:** 1154-1161

Timmers HT, Meyers RE, Sharp PA (1992) Composition of transcription factor B-TFIID. *Proc Natl Acad Sci U S A* **89:** 8140-8144

Tornaletti S, Hanawalt PC (1999) Effect of DNA lesions on transcription elongation. *Biochimie* **81:** 139-146

Tran HG, Steger DJ, Iyer VR, Johnson AD (2000) The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor. *EMBO J* **19:** 2323-2331

Tuteja N, Ahmad P, Panda BB, Tuteja R (2009) Genotoxic stress in plants: shedding light on DNA damage, repair and DNA repair helicases. *Mutat Res* **681:** 134-149

Uptain SM, Kane CM, Chamberlin MJ (1997) Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem* **66:** 117-172

Verma R, Oania R, Fang R, Smith GT, Deshaies RJ (2011) Cdc48/p97 mediates UV-dependent turnover of RNA Pol II. *Mol Cell* **41:** 82-92

Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog GA, Winston F, Buratowski S, Handa H (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* **12:** 343-356

Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, Choudhary C (2011) A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol Cell Proteomics* **10:** M111 013284

Walmacq C, Cheung AC, Kireeva ML, Lubkowska L, Ye C, Gotte D, Strathern JN, Carell T, Cramer P, Kashlev M (2012) Mechanism of translesion transcription by RNA polymerase II and its role in cellular resistance to DNA damage. *Mol Cell* **46:** 18-29

Watanabe T, Hayashi K, Tanaka A, Furumoto T, Hanaoka F, Ohkuma Y (2003) The carboxy terminus of the small subunit of TFIIE regulates the transition from transcription initiation to elongation by RNA polymerase II. *Mol Cell Biol* **23:** 2914-2926

Welboren WJ, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FC, Span PN, Stunnenberg HG (2009) ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO J* **28:** 1418-1428

West S, Gromak N, Proudfoot NJ (2004) Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432:** 522-525

Wild T, Cramer P (2012) Biogenesis of multisubunit RNA polymerases. *Trends Biochem Sci* **37:** 99-105

Wilson MD, Harreman M, Svejstrup JQ (2012) Ubiquitylation and degradation of elongating RNA polymerase II: The last resort. *Biochim Biophys Acta*

Wittschieben BO, Fellows J, Du W, Stillman DJ, Svejstrup JQ (2000) Overlapping roles for the histone acetyltransferase activities of SAGA and elongator in vivo. *EMBO J* **19:** 3060-3068

Wood A, Schneider J, Dover J, Johnston M, Shilatifard A (2003) The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J Biol Chem* **278:** 34739-34742

Workman JL (2006) Nucleosome displacement in transcription. *Genes Dev* **20:** 2009-2017

Workman JL, Kingston RE (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem* **67:** 545-579

Woudstra EC, Gilbert C, Fellows J, Jansen L, Brouwer J, Erdjument-Bromage H, Tempst P, Svejstrup JQ (2002) A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. *Nature* **415:** 929-933

Wu CH, Yamaguchi Y, Benjamin LR, Horvat-Gordon M, Washinsky J, Enerly E, Larsson J, Lambertsson A, Handa H, Gilmour D (2003) NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev* **17:** 1402-1414

Yamaguchi Y, Filipovska J, Yano K, Furuya A, Inukai N, Narita T, Wada T, Sugimoto S, Konarska MM, Handa H (2001) Stimulation of RNA polymerase II elongation by hepatitis delta antigen. *Science* **293:** 124-127

Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97:** 41-51

Yan Q, Moreland RJ, Conaway JW, Conaway RC (1999) Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J Biol Chem* **274:** 35668-35675

Yang XC, Sullivan KD, Marzluff WF, Dominski Z (2009) Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. *Mol Cell Biol* **29:** 31-42

Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, Tora L (2011) seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res* **39:** e35

Young RA (1991) RNA polymerase II. *Annu Rev Biochem* **60:** 689-715

Zanton SJ, Pugh BF (2006) Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock. *Genes Dev* **20:** 2250-2265

Zawel L, Kumar KP, Reinberg D (1995) Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev* **9:** 1479-1490

Zehring WA, Lee JM, Weeks JR, Jokerst RS, Greenleaf AL (1988) The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. *Proc Natl Acad Sci U S A* **85:** 3698-3702

Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ (2012a) Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *J Biol Chem* **287:** 8541-8551

Zhang H, Roberts DN, Cairns BR (2005) Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123:** 219-231

Zhang J, Corden JL (1991) Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J Biol Chem* **266:** 2290-2296

Zhang X, Horibata K, Saijo M, Ishigami C, Ukai A, Kanno S, Tahara H, Neilan EG, Honma M, Nohmi T, Yasui A, Tanaka K (2012b) Mutations in UVSSA cause UV-sensitive syndrome and destabilize ERCC6 in transcription-coupled DNA repair. *Nat Genet* **44:** 593-597

Zhang Y, Smith CL, Saha A, Grill SW, Mihardja S, Smith SB, Cairns BR, Peterson CL, Bustamante C (2006) DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC. *Mol Cell* **24:** 559-568

Zhong H, Zhu J, Zhang H, Ding L, Sun Y, Huang C, Ye Q (2004) COBRA1 inhibits AP-1 transcriptional activity in transfected cells. *Biochem Biophys Res Commun* **325:** 568-573

Summary:

Living cells are continuously exposed to stimuli from internal (basic metabolic processes) and external (environmental or chemical stress) sources. Therefore a very accurate and tightly regulated cellular process is needed to access and express the DNA-encoded information in order to maintain the normal physiological state of a cell. In eukaryotes this is carried out by the RNA polymerase II (Pol II) machinery, which is responsible for the expression of thousands of genes coding for proteins and non-coding regulatory RNAs. The Pol II transcription cycle can be divided into three main phases: transcription initiation, elongation and termination. Each phase requires sequential activities of different protein complexes, which independently represent a possibility for the regulation of gene expression. Polymerase II pause was first described as a regulatory step between initiation and early elongation at the promoters of certain heat shock genes. Recently, genome-wide studies demonstrated that Pol II pausing is an important regulatory step that is present at almost every eukaryotic Pol II promoter. Surprisingly, paused or slowed down polymerases were also discovered downstream of 3' end of genes, of which the exact role is still not fully understood.

During my Ph.D. I carried out projects using chromatin immunoprecipitation assay coupled to high-throughput sequencing techniques to analyse genome-wide Pol II behaviour in two aspects:

First, we analysed Pol II occupancy downstream of 3' end of transcription units. Our analyses suggest that accumulation of Pol II downstream of genes is a genome-wide feature of active transcription. We found broad, often up to 6kb long Pol II occupancy signals at genes coding for polyadenylated transcripts. In contrast, Pol II occupancy shows a narrow profile at the annotated end of core histone genes. We also found a link between RNA 3' end processing and Pol II accumulation at the end of transcription units.

Second, we were following the genome-wide response and alteration of Pol II transcription upon genotoxic stress. We found that even a sub-lethal dose of ultraviolet light (UV-B) can temporary interrupt transcription genome-wide. Following UV-B treatment we observed a progressive and massive Pol II signal loss from the promoters of expressed genes, which will then extend through the entire transcription unit, up to four hours after irradiation. In contrast, we found a subset of genes with increased Pol II signal upon irradiation, which were identified as DNA damage response genes. This is in good agreement with the observation that after UV irradiation transcription is arrested during the period of transcription-coupled repair (TCR). Interestingly, six hours after UV-B treatment, genes that were negatively influenced, showed increased Pol II signal compared to the control state, suggesting that the Pol II transcription apparatus may compensate for the reduced gene expression associated with TCR.

RNA polymerase II transcription has been extensively studied in the past decades, but lots of details are still missing to establish a model, which covers every regulatory mechanism during gene expression. With my Ph.D. projects, we managed to expand the current knowledge about Pol II transcription machinery. First we characterized the Pol II accumulation around the 3' end of genes that was not yet investigated at a genome wide level. By studying the effect of genotoxic stress on Pol II mediated transcription genome-wide, we describe a global mechanism that may facilitate the dissociation of Pol II complex from the promoters of expressed genes.