Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Cécilia STUDNIAREK

Le 4 octobre 2019

Contrôle de l'expression génique par la RNP 7SK

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **GENETIQUE MOLECULAIRE**

Unité de recherche :
LBME - Laboratoire de Biologie Moléculaire des Eucaryotes

Thèse dirigée par
Tamas KISS et Sylvain EGLOFF

Jury

Mme Rosemary Kiernan, Rapporteure
Mme Patricia Uguen, Rapporteure
M. Frédéric Coin, Rapporteur
M. Jean-Christophe Andrau, Examinateur
M. Pierre-Emmanuel Gleizes, Examinateur
M. Tamas KISS, Directeur de thèse
M. Sylvain Egloff, Co-directeur de thèse

Remerciements

Si je devais résumer ma thèse aujourd'hui, je dirais que c'est d'abord des rencontres, des gens qui m'ont tendu la main. Et c'est assez curieux de se dire que les hasards, les rencontres, forgent une destinée…

Je tiens en tout premier lieu à remercier chaleureusement les membres de mon jury de thèse, Patricia Uguen, Rosemary Kiernan, Frédéric Coin et Jean-Christophe Andrau. J'ai été très fière de vous avoir dans mon jury, et heureuse de vous rencontrer. Merci d'avoir pris le temps de lire et juger mon manuscrit de thèse, d'être venus pour la soutenance, merci pour vos commentaires et nos discussions. Et je vous souhaite à tous une bonne continuation dans vos recherches, qui je l'espère nous permettront de nous rencontrer à nouveau. Un très grand merci à Pierre-Emmanuel Gleizes pour avoir présidé mon jury de thèse, ainsi que pour son suivi bienveillant de longue date. Je remercie aussi très chaleureusement Béatrice Clouet d'Orval et Estelle Nicolas pour avoir fait partie de mon comité de thèse et avoir suivi assidument mon avancée au cours de ces quelques années. Enfin, je remercie le corps enseignant de biologie moléculaire pour m'avoir permis d'être DCE pendant 3 ans, et en particulier Pascale Belenguer et Estelle Espinos pour m'avoir fait confiance et donné autant d'heures de leurs UE. Par-dessus tout, je remercie mon tuteur, David Umlauf, pour m'avoir fait partager ses qualités d'enseignant, pour sa patience, son aide et son soutien. David, tu as accompli ton rôle de tuteur au-delà de mes espérances. J'ai adoré t'assister pour les premières séances de TP, préparer les séances avec toi et discuter de la vie, de tout, et de rien.

Then I will switch to english to thank you Tamás (whereas I am 99 % sure you secretly understand french pretty well…). Thank you for allowing me to perform my M2 internship in your lab, which, unfortunately, turned into a PhD study. Unfortunately, yes, because it meant you have to put up with me for 3… no, wait… 4 years. Thank you for allowing me to perform all these expensive experiments, always accepting my antibody orders, even on Sundays holidays barbecue sessions. Oh and also not complaining too much about the loud and not classical music we listened in the lab, that I know sometimes – maybe quite often – almost hurt your ears.. Hopefully you had many worse reasons to complain about! Such as the weather, morning traffic jam or noisy building work. I will miss you Tamás!!

Sylvain, quel plaisir, et le mot est faible, d'avoir été sous ta supervision et d'avoir travaillé à tes côtés toutes ces années. Merci de m'avoir choisie pour le M2, je sais que tu ne l'as pas – souvent – regretté. Scientifique de talent, j'ai tellement appris grâce à toi. Réaliser de magnifiques western et des ChIP reproductibles, ça, je te le dois. L'expérience mNET-seq fut également un moment inoubliable, je dirais même dingo ! On aura surtout appris que les data, c'est les data. Cela dit, je t'ai aussi découvert personnellement, et je ne suis pas déçue du voyage. Tant de bons moments, de fous-rire et de délires entre et pendant les manips ont jalonné ces 5 années passées au labo. On aura bien progressé tous les deux en lancer de boules d'alu, en retournement de bac à glace (plein), en calage d'expressions farfelues (à ce propos, on s'en est aussi mis plein la lampe à de très nombreuses reprises), en racontage de blagues, le tout dans une ambiance musicale plutôt très éclectique (merci de ne pas voir mis trop souvent du Noir Désir, tu vas pouvoir maintenant t'en donner à cœur joie). Et puis tu as été là – et plus que là – tout du long, à tous les moments (sauf un, mais je l'avais bien mérité :p), je te dois un immense merci. J'ai vraiment eu de la chance de t'avoir en encadrant et tout simplement de te connaître et je me ravie d'ores et déjà de nos futures collaborations.

La team Kiss, je ne l'oublierais jamais. Même si la journée s'annonçait difficile et lourde de manips, l'optique de vous retrouver m'a bien souvent donné la force de me lever le matin. Fluctuellement présent en fonction des périodes d'enseignement, mais toujours le mot pour rire, je mets

alternativement ma sonde ou mon cours sur moodle avant de partir, je suis je suis je suis ??? Paaaat !!! Tu as été, au cours de ma thèse, comme dirait le sage Anaël, mon soleil. Ton déhanché restera dans ma mémoire et j'espère que tu la rafraîchiras régulièrement !! Béa, ma superwoman, qui m'a porté secours à tellement de reprises !!! Que ce soit pour lancer mon transfert, pour m'ouvrir la salle de culture un samedi, pour un problème de vélo ou pour retrouver des clés égarées, j'ai toujours pu compter sur toi. Merci pour ton grand cœur et ta bienveillance !!! Aaaah et Amandine, madame à tout faire !!! Pilier des Kiss, sans conteste !!! Amandiiiine !! Ou est-ce qu'on met ça ?? Tu sais où je peux trouver ça ?? Et ça, ça se jette où ?? Je ne sais pas ce qu'on ferait sans toi dans l'équipe, et surtout on n'aurait pas eu tous ces chocolats avec la promo Marcherey Nagel, ni l'enceinte maintenant indétrônable. Merci aussi pour le lègue des légumes en excès de ton potager, et puis surtout, pour ta joie de vivre communicative quotidienne. Je vais aussi tâcher de m'inspirer de ton sens de l'organisation pour ma future carrière... Je ne peux pas ne pas citer Aline aussi. Malheureusement pour moi, tu as quitté l'équipe peu de temps après mon arrivée… Mais tu n'as pas oublié les Kiss pour autant et te revoir régulièrement pour apporter à Pat son magazine était toujours un plaisir ! Je te remercie beaucoup pour les conseils et retours que tu m'as donnés sur mes talks (ceux auxquels tu as pu assister ^^). Je n'oublierai certainement pas non plus les étudiants qui sont passés par là, et en particulier Juju, Lucie, Olivia et Sari. Vous avez amené de la fraîcheur, du baume au cœur, et tant d'autres choses, c'était si bon de vous avoir dans l'équipe !!

Ah et les Cava, avec qui nous formions dans le cœur une seule et même équipe !! Jérôme et ses célèbres éternuements à réveiller un loir – ou, en l'occurrence, Jade –, Marie, toujours présente pour manger de la galette, ou autre, Virginie, malheureusement beaucoup trop souvent avec les souris, Raph, qui m'a préparé comme personne aux castings pour jouer quelqu'un qui se fait tirer dessus, et Jade, mon achate, mon acolyte comme personne. Cette formation Galaxy a tout changé pour moi. Merci pour ton soutien sans faille, pour m'avoir accueillie chez toi, et pour m'avoir accordé ton amitié, tellement précieuse pour moi. Alors merci aux atypiques Cava pour ces moments (avant dispersion) que nous avons partagés. Les CRISPR-burgers clubs, les repas de Noël, les galettes, crêpes, beer sessions... j'en passe et des meilleurs !!

Qu'aurait été la thèse sans les thésards de l'IBCG ? Je ne préfère pas y penser, mais plutôt me remémorer tous ces bons moments qui nous ont rassemblés, bercés par le doux ronron du rire de Lise. Lise, Sylvain, Ramtin, Duy Khanh, Thomas, Maral, Clément, Manon, Maleaume (le seul non-IBCG de la bande), Steph (bon, pas thésarde, mais aussi déjantée que nous), les parties de tarot avec vous, sous le soleil de l'upsi, à potiner et rigoler, en sont le parfait exemple. On aura vraiment passé des moments de folie au labo et en dehors, et quelque-chose me dit que ce n'est pas fini.. Lud et Alex (avec qui je dois encore une revanche au Twister), Emma, qui squatte toujours l'IBCG pour nous amener sa bonne humeur, Emilie, Ali, Sophie, David, Yiying, Sevan, Olivier, Coline et Jérémy, j'ai adoré vous croiser, discuter et rigoler avec vous tout au long de ma thèse, sur les marches de l'IBCG, au détour d'un couloir, sur les tables dehors… tous les endroits étaient bons pour échanger quelques papotages. Merci aussi aux thésards sportifs, Tommy, Clément, Maxime, Duy Khanh et Manon avec qui j'ai pu partager quelques courses au bord du canal ou, pour les plus téméraires, à Pech David, et Lise et Sylvain qui m'ont souvent motivée pour aller à la piscine ! Je retiendrai aussi les séminaires passionnés et passionnants au Dubliners avec Luis, Micka, le parrain, Cyril, Charlotte, Alix, Laura, Valentin, Côme, Ludo, Moïse, Marie etc, toujours bien animés autour de quelques pichets... Et pour finir en beauté, les soirées et week-end avec les hédonistes, qui se reconnaîtront, auront été du plaisir à l'état pur, et sous toutes ses coutures… Ce qui est sûr, en plus du fait qu'on ait développé des goûts de luxe, c'est que vous êtes entrés dans ma vie et je ne vous en laisserai pas sortir. Bref, des thésards de qualité tout ça, vous allez terriblement me manquer !!! Et je vous souhaite toute la réussite que vous méritez dans vos thèses respectives et dans votre futur !!!

La jeunesse, c'est bien, mais les aînés, c'est bien aussi. Autour de la machine à café ou dans les couloirs, j'ai eu la chance d'en côtoyer bon nombre. Les papotages en tout genre avec Odile, Régine, Yves, Martha, Raph, Safi, les deux Christophe, Bertrand, pour n'en citer que quelques-uns, égayaient systématiquement ma journée. Alors merci pour cela, vous allez beaucoup me manquer.

Il faisait bon aussi de temps en temps de se détacher du laboratoire, et pour cela, j'ai toujours pu compter sur mes amis de longue date, mes copines Manue (ma plus vieille pote !!), Clem, Lisa et Mary, mes petits vieux Pasca et Loïs, Pierre Cab, Clémence, Zoulou, Olivia, Dani et Emilie. On n'a pas eu l'occasion de se voir ni souvent ni assez, mais à chaque fois c'était pour moi un vrai bonheur de vous retrouver, et vous me faisiez oublier tous mes malheurs. Merci d'avoir toujours été et d'être encore là pour moi.

Pour finir, je dois les remerciements les plus tendres et les plus chers à mon cœur à ma famille. Papa, Maman, Marina, Eva et Amanda. Vous êtes mon bonheur, ma fierté, ma force. Merci.

Titre : Etude de la fonction de la particule ribonucléoprotéique 7SK dans le contrôle de la transcription par l'ARN Polymérase II

Auteur : Cécilia Studniarek

Directeurs de thèse : Dr. Tamás Kiss et Dr. Sylvain Egloff

Résumé :

Le facteur positif de l'élongation P-TEFb stimule l'élongation de la transcription par l'ARN Polymérase II (ARN Pol II) en phosphorylant les facteurs négatifs de l'élongation et le domaine carboxy-terminal (CTD) de l'ARN Pol II. L'activité nucléaire de P-TEFb est contrôlée par la particule ribonucléoprotéique (RNP) 7SK, composée du petit ARN nucléaire 7SK et des protéines Larp7 et MePCE. Avec l'aide d'un dimère de protéines HEXIM1/2, la RNP 7SK séquestre de manière dynamique et réversible une fraction de P-TEFb dans la RNP 7SK/P-TEFb catalytiquement inactive. En dehors de la RNP 7SK, P-TEFb s'associe avec différents facteurs de transcription et cofacteurs avec lesquels il forme des complexes actifs recrutés sur les gènes pour activer leur expression. De nombreux stimuli et stress cellulaires affectant la croissance cellulaire, dont les dommages à l'ADN causés par les UV, induisent la dissociation rapide de la RNP 7SK/P-TEFb, libérant ainsi une fraction conséquente de P-TEFb actif. En plus de son rôle de réservoir nucléaire de P-TEFb, la RNP 7SK/P-TEFb occupe également les régions promotrices de certains gènes pour permettre une libération ciblée de la forme active de P-TEFb et ainsi stimuler la transcription de gènes spécifiques. En utilisant des lignées cellulaires dans lesquelles l'expression de l'ARN 7SK est abolie (cellules KO-7SK), nous avons évalué les conséquences fonctionnelles de la perte de l'ARN 7SK sur la transcription par l'ARN Pol II avant et après induction de dommages à l'ADN par les UV. L'analyse de la transcription naissante par Mammalian Native Elongating Transcript-sequencing (mNET-seq) a révélé que la perte de l'ARN 7SK a peu d'impact sur la transcription par l'ARN Pol II en conditions normales. Cependant, en réponse aux UV, l'activation transcriptionnelle de gènes clé de la réponse aux dommages à l'ADN est compromise dans les cellules KO-7SK alors qu'ils sont rapidement induits dans les cellules contrôles. Etant donné que les niveaux d'activité de P-TEFb dans les lignées contrôle et KO-7SK sont comparables, ces données indiquent que l'activation correcte de ces gènes après irradiation repose sur la fraction de P-TEFb séquestrée au sein de la RNP 7SK. Des expériences supplémentaires suggèrent qu'après irradiation aux UV, P-TEFb est spécifiquement extrait de la RNP 7SK et transféré dans le Super Elongation Complex (SEC), au sein duquel il est recruté sur ces gènes cibles pour stimuler l'entrée de l'ARN Pol II en élongation productive et le recrutement de facteurs d'élongation. En conclusion, nous avons démontré que l'ARN 7SK n'est pas requis pour la viabilité cellulaire en conditions de croissance normales, mais s'avère essentiel pour orchestrer la réponse transcriptionnelle aux dommages à l'ADN générés par les UV.

Liste des abréviations

ADN : Acide désoxyribonucléique
ARN : Acide ribonucléique
ARNm : ARN messager
ARN Pol : ARN polymérase
ARNr : ARN ribosomique
ATP : Adénosine Tri Phosphate
CAK : CDK-Activating Kinase
CDK : Cyclin-dependant kinase
ChIP : Chromatine immunoprecipitation
ChIRP : Chromatine isolation by RNA purification
CPSF : Cleavage and polyadenylation specificity factor
CTD : Carboxyl-terminal domain
CTR : C-terminal repeat
DRB : 5,6-dichloro-1-b-D-ribofuranosylbenzimidazole
DSIF : DRB sensitivity inducing factor
EGF : Epidermal growth factor
eRNA : enhancer RNA
GB : gene body
GTF : General Transcription Factors
H2A, H2B, H3, H4 : histone H2A, histone H2B, histone H3, histone H4
HEXIM : Hexamethylene bisacetamine-inducible protein
hnRNP : heterogeneous nuclear riborucleoprotein particule
Hsp : Heat Shock Protein
Larp7 : La-related protein 7
LTR : Long Terminal Repeat
MePCE : Methylphosphate capping enzyme
mNET-seq : mammalian native elongating transcript sequencing
NELF : Negative elongation factor
NER : Nucleotide excision repair
nt : nucléotide
N-TEF : Negative Transcription Elongation Factor
PAF: Polymerase Associated Factor
PAS : Polyadenylation signal
PIC : PreInitiation complex
PseudoU : Pseudouridine
P-TEFb : Positive transcription elongation factor b
pb : paires de bases
RNase : Ribonucléase
RNP : Ribonucleoprotein particule
SEC : Super Elongation Complex
siRNA : small interfering RNA
snARN : “small nuclear” ARN ou petit ARN nucléaire
snoARN : “small nucleolar” ARN ou petit ARN nucleolaire

TAF : TBP associated factor
TAR : Transactivating-responsive region
TBP : TATA box binding protein
TES : Transcription end site
TFII : Transcription factor II
TRM : Tat/TAR Recognition Motif
TSS : Transcription Start Site
UV : ultraviolet
VIH : Virus de l'immunodéficience humaine

Sommaire

.Introduction 7
I. La transcription par l'ARN Pol II chez les mammifères 8
i. L'initiation de la transcription par l'ARN Pol II 9
a. Configuration chromatinienne aux promoteurs 10
b. Mécanismes moléculaires de l'initiation 10
c. Régulation de l'initiation 13
d. Transcription antisens 14
ii. L'élongation de la transcription 15
a. La pause de l'ARN Pol II à proximité du promoteur : de la découverte à la généralisation 16
b. Fonctions de la pause de l'ARN Pol II à proximité du promoteur 19
c. Synergie pause transcriptionnelle/initiation 21
d. Mise en place de la pause de l'ARN Pol II 21
e. Libération de l'ARN Pol II en élongation productive 26
f. Un objectif, de multiples options : le recrutement de P-TEFb 28
g. Stabilité du complexe d'élongation en pause à proximité du promoteur 31
h. L'élongation productive 32
iii. Terminaison 34
a. Clivage et polyadénylation de l'ARN 35
b. Décrochage et recyclage de l'ARN Pol II 36
c. Elongation abortive à l'étape de pause à proximité du promoteur 37
d. Terminaison prématurée causée par l'arrêt de la polymérase 37
iv. Le CTD de l'ARN Pol II 38
a. Les phosphorylations du CTD 39
b. Les autres modifications du CTD 41
c. Modifications des répétitions non-consensus du CTD 42
II. Régulation de P-TEFb par la particule ribonucléoprotéique (RNP) 7SK 43
i. La RNP 7SK cœur 44
a. L'ARN 7SK 44
b. MePCE 46
c. Larp7 47
ii. La RNP 7SK/P-TEFb 48
a. Le complexe P-TEFb 49
b. Les protéines HEXIM 52
c. Pathologies associées à P-TEFb 55
iii. Libération de P-TEFb de la RNP 7SK/P-TEFb 56

a. Voies de signalisation et modifications post-traductionnelles de la RNP 7SK/P-TEFb 56
b. Extraction de P-TEFb de la RNP 7SK/P-TEFb 59
c. Dissociation de la RNP 7SK/P-TEFb et changement conformationnel.....60
d. Remodelage de la RNP 7SK/P-TEFb 60
iv. Régulation de l'activité de P-TEFb à la chromatine.....61
a. Association de la RNP 7SK/P-TEFb à la chromatine.....62
b. Dissociation de la RNP 7SK/P-TEFb à la chromatine.....63
v. Les autres RNP 7SK.....64
a. La RNP 7SK/LEC 64
b. L'ARN 7SK et le complexe BAF 65
c. La RNP 7SK/Ars2.....65
III. Réponse transcriptionnelle aux dommages à l'ADN causés par les UV 66
i. Lésions de l'ADN provoquées par les UV et leur signalisation.....66
a. Détection des lésions par le TC-NER.....67
b. Détection des lésions par le GG-NER.....68
c. Réparation des lésions par excision de nucléotide 68
d. Cinétique de réparation.....70
e. Voies de signalisation activées en réponse aux UV.....70
ii. Reprogrammation transcriptionnelle lors de la réponse aux UV 71
a. Répression de l'initiation de la transcription par l'ARN Pol II 72
b. Libération de l'ARN Pol II en pause à proximité du promoteur 72
c. Restriction du taux d'élongation par l'ARN Pol II 73
d. Activation des gènes de réponse 74
e. Reprise générale de la transcription.....74
.Résultats.....77
Préambule 78
Résultats complémentaires.....81
I. Rôle de la RNP 7SK/P-TEFb dans la régulation de la transcription par l'ARN Pol II 81
i. Rôle spécifique de l'ARN 7SK dans la régulation des gènes 82
ii. Rôle spécifique de la protéine Larp7 dans la régulation de l'expression des gènes.....85
I. Rôle de la RNP 7SK/P-TEFb dans la réponse transcriptionnelle aux UV.....88
i. Libération de l'ARN Pol II en pause par le réservoir inactif de P-TEFb 90
ii. Activation transcriptionnelle dépendante de l'ARN 7SK en réponse aux UV 92
.Discussion.....96
I. Fonction de la RNP 7SK/P-TEFb dans la régulation de l'expression génique en conditions normales 97

II. Contrôle de la pause de l'ARN Pol II à proximité du promoteur par la RNP 7SK/P-TEFb.....99
III. Rôle déterminant du réservoir inactif de P-TEFb dans la réponse aux UV.....101
i. Libération massive des ARN Pol II en pause par le réservoir inactif de P-TEFb.....101
ii. Rôle de la RNP 7SK/P-TEFb dans l'activation des gènes de réponse aux UV.....102
a. Libération spécifique de la pause au niveau des gènes activés.....102
b. Activation transcriptionnelle dépendante du réservoir inactif de P-TEFb.....103
c. Conclusion et modèle.....104
Références bibliographiques.....107
Annexe.....153

.Introduction

L'expression des gènes doit être méticuleusement régulée afin d'assurer le bon déroulement de processus biologiques fondamentaux tels que la prolifération et la différentiation cellulaires ou encore la réponse à divers stimuli. La transcription constitue une étape clé de l'expression des gènes en permettant la synthèse des ARN cellulaires. Chez les eucaryotes, trois ARN polymérases (Pol) se partagent cette lourde tâche (cinq chez les plantes). Tandis que les ARN Pol I et III permettent la production d'ARN non codants uniquement (notamment les ARN ribosomiques et de transfert, respectivement), c'est l'ARN Pol II qui se charge de la synthèse des ARN messagers, matrices pour la synthèse des protéines, mais aussi de nombreux types d'ARN non codants (sn(o)ARN, microARN, longs ARN non codants…). Nous allons ici nous concentrer sur la régulation complexe de la transcription par l'ARN Pol II, processus encore bien méconnu faisant intervenir de manière orchestrée toute une pléiade de facteurs aux rôles multiples.

I. La transcription par l'ARN Pol II chez les mammifères

- *L'ARN Pol II*

L'ARN Pol II est un complexe macromoléculaire de 516 kDa composé de 12 sous-unités, Rpb1 à 12, très conservées chez les eucaryotes (Sainsbury et al., 2015). La résolution de la structure de l'ARN Pol II, seule et en élongation, a permis d'établir l'agencement des différentes sous-unités la constituant et d'éclaircir les mécanismes moléculaires de la transcription (Cramer et al., 2001; Gnatt et al., 2001). De ces 12 sous-unités, 10 forment le cœur enzymatique et Rpb4/7 constituent un sous-complexe modulaire (figure 1) (Armache et al., 2005; Bernecky et al., 2016). Les deux plus grandes sous-unités, Rpb1 et Rpb2, sont situées au centre du complexe et forment un sillon riche en charges positives au fond duquel se trouve le site catalytique. Le brin d'ADN matrice traverse le sillon dans lequel il s'hybride sur 8-9 nt avec l'ARN naissant. Les sous-unités constituant le cœur enzymatique établissent des interactions complexes avec le duplexe ADN/ARN, contribuant à la stabilité du complexe d'élongation tout en permettant la translocation de l'ADN (Gnatt et al., 2001). L'ARN naissant émerge du complexe par un canal de sortie à proximité du sous-domaine Rpb4/7, avec lequel il s'associe co-transcriptionnellement (Schulz et al., 2014; Ujvari and Luse, 2006). Le canal de sortie de l'ARN jouxte également le domaine carboxy-terminal (CTD) de Rpb1 (Chen et al., 2009; Suh et al., 2013). Le CTD de Rpb1 est un domaine unique de faible complexité s'étendant en dehors du cœur du complexe. Il est composé de répétitions des résidus Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, cibles de nombreuses modifications particulièrement importantes au cours du cycle de transcription (voir I.iv. Le CTD de l'ARN Pol II).

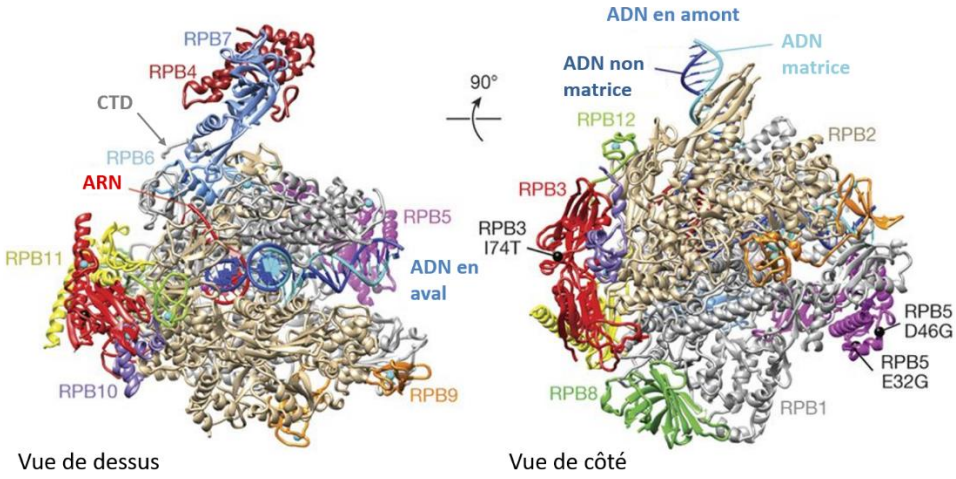


Figure 1 : Structure de l'ARN Pol II bovine en complexe avec l'ADN et l'ARN naissant (Bernecky et al., 2016). La vue de dessus (gauche) permet de distinguer le cœur catalytique et le sous-complexe Rpb4/7. L'ADN matrice et l'ARN s'hybrident au niveau du site catalytique de l'ARN Pol II, au cœur du complexe. En raison de sa faible structuration, le CTD de Rpb1 n'est pas représenté mais sa localisation est indiquée par la flèche. Les points représentent les résidus qui diffèrent entre l'ARN Pol II bovine et humaine.

- *Le cycle de transcription*

Le cycle de transcription par l'ARN Pol II peut être assimilé à un véritable parcours du combattant jalonné d'obstacles et de points de contrôle. Le cycle commence avec le recrutement de l'ARN Pol II au niveau du promoteur du gène et s'en suit l'initiation de la transcription. L'ARN Pol II quitte les régions promotrices, marque une pause au début de l'élongation avant d'entrer en élongation productive. Le cycle de transcription se termine après le passage du site de polyadénylation (pA). L'ARN nouvellement synthétisé est clivé et l'ARN Pol II libérée de la matrice ADN peut éventuellement démarrer un nouveau cycle. Un nombre incalculable de facteurs entre en jeu pour permettre à l'ARN Pol II de franchir chacune de ces étapes de manière régulée, dans un contexte chromatinien et pour réaliser au fur et à mesure diverses maturations co-transcriptionnelles de l'ARN (ajout de la coiffe, épissage, modifications de bases, polyadénylation).

i. L'initiation de la transcription par l'ARN Pol II

L'initiation de la transcription par l'ARN Pol II est un processus extrêmement régulé et complexe. Elle commence par le recrutement de l'ARN Pol II et des facteurs d'initiation au niveau des régions promotrices des gènes, permis par un contexte chromatinien favorable. La double hélice d'ADN est ensuite ouverte pour permettre à l'ARN Pol II d'accéder au brin matrice et de catalyser la synthèse de l'ARN complémentaire. Ces étapes sont le fruit de réarrangements structuraux complexes accompagnés et stimulés par de nombreux facteurs.

a. Configuration chromatinienne aux promoteurs

Chez les eucaryotes, l'ADN est condensé sous forme de chromatine, dont l'unité de base, le nucléosome, est composée de 147 pb d'ADN enroulées autour de huit protéines histones (H2A, H2B, H3 et H4 présents en deux exemplaires). Une telle compaction restreint *per se* l'accessibilité de l'ADN et constitue une barrière à différentes étapes de la transcription. Au niveau des promoteurs, la présence de nucléosomes empêche le recrutement et l'assemblage de la machinerie de transcription (Lorch et al., 1987). Cependant, des études à l'échelle du génome de la structure de la chromatine ont révélé que les régions promotrices des gènes actifs étaient dépourvues de nucléosomes ou occupées par des nucléosomes fortement dynamiques constitués de variants d'histones tels que H3.3 et H2A.Z (Jin et al., 2009; Mavrich et al., 2008; Schones et al., 2008; Venkatesh and Workman, 2015). Certains facteurs de transcription, tels que les facteurs FoxA et GATA, ont la capacité de se lier aux régions promotrices dans un contexte chromatinien et de recruter des facteurs de remodelage de la chromatine permettant une ouverture chromatinienne locale (Ansari et al., 2014; Fuda et al., 2015; Ganguli et al., 2014; Hartley and Madhani, 2009; Kremer et al., 2012; Shimojima et al., 2003; Zaret and Carroll, 2011). Les modifications post-traductionnelles des histones favorisent également l'accessibilité des promoteurs. La tri-méthylation de l'histones H3 sur la lysine 4 (H3K4me3) et l'acétylation de divers résidus de H3 sont par exemple des marques de chromatine ouverte et sont enrichies au niveau des promoteurs actifs et des régions régulatrices transcriptionnellement actives de type enhancer (Barski et al., 2007; Creyghton et al., 2010; Di Cerbo et al., 2014; Kouzarides, 2007; Luebben et al., 2010; Rada-Iglesias et al., 2011). Ces marques d'histones sont déposées par des méthyl- et acétyl-transférases spécifiques des histones (HMT et HAT respectivement), telles que les HMT de la famille SET et les HAT p300/CBP. L'ajout d'un groupement me3 ou Ac diminue l'affinité de l'ADN pour le nucléosome, provoquant un relâchement de la chromatine (Tessarz and Kouzarides, 2014; Tropberger and Schneider, 2013). De plus, ces marques d'histones sont spécifiquement reconnues par des cofacteurs qui participent à l'activation de la transcription, comme le cofacteur Brd4 qui reconnait les queues acétylées des histones grâce à son bromodomaine (BR) (Dey et al., 2003; Dey et al., 2009). L'établissement d'un contexte chromatinien favorable au niveau des promoteurs à travers ces différents mécanismes est un prérequis essentiel pour permettre l'assemblage de la machinerie de transcription et ainsi d'initier la transcription.

b. Mécanismes moléculaires de l'initiation

La transcription débute à une position précise dans la région 5' du gène appelé site d'initiation de la transcription (TSS, Transcription Start Site). Environ 40 pb en amont et en aval du TSS s'étend le promoteur minimal, une région composée d'une séquence particulière permettant la liaison de la

machinerie transcriptionnelle de base ou complexe de pré-initiation (PIC, Pre-Initiation Complexe) comprenant l'ARN Pol II et les facteurs généraux de la transcription (GTF, General Transcription Factors) TFIID, TFIIA, TFIIB, TFIIE, TFIIF et TFIIH (Hampsey, 1998). Le promoteur minimal, comme son nom l'indique, est suffisant pour que l'initiation de la transcription ait lieu à un niveau basal, généralement faible (Smale and Kadonaga, 2003). Des régions régulatrices plus éloignées et l'environnement chromatinien permettent alors de moduler le taux d'initiation via le recrutement de facteurs de transcription, de cofacteurs et via des modifications de la chromatine pour faciliter ou empêcher l'assemblage du PIC (Haberle and Stark, 2018).

Des études biochimiques et structurales ont établi un modèle séquentiel pour l'assemblage du PIC et l'initiation de la transcription : Le facteur TFIID, composé de TBP (TATA-binding protein) et de ses protéines accessoires (TAFs, TBP associated proteins), reconnait le promoteur minimal et engendre une distorsion de l'ADN (figure 2 a-b) (Buratowski et al., 1989; Kim et al., 1993; Louder et al., 2016; Thomas and Chiang, 2006). TFIIA et TFIIB rejoignent alors le complexe (figure 2b). TFIIA stabilise l'association de TBP avec l'ADN (Hieb et al., 2007). TFIIB contacte TBP et permet le recrutement de l'ARN Pol II, avec laquelle il interagit fortement, accompagnée de TFIIF (figure 2c) (He et al., 2013; Sainsbury et al., 2013). TFIIF joue un rôle clé dans la stabilisation du PIC, et notamment de TFIIB (Luse, 2012). Il a été proposé que TFIIB et TFIIF participent également à la sélection du TSS (Ghazy et al., 2004; Hawkes and Roberts, 1999; Pardee et al., 1998). Le PIC est alors complété par TFIIE qui permet le recrutement de TFIIH au niveau de l'ADN en aval du complexe d'initiation (figure 2d) (Compe et al., 2019; He et al., 2013; Maxon et al., 1994). A ce stade, le complexe est dit « fermé » car l'ADN est encore sous forme double brin. En présence d'ATP, l'activité ATPase de la sous-unité XPB de TFIIH permet d'ouvrir la double hélice d'ADN pour former la bulle transcriptionnelle (figure 2e) (Fishburn et al., 2015; Grunberg et al., 2012; Guzman and Lis, 1999; Ohkuma and Roeder, 1994). Le complexe est alors dit ouvert et la matrice d'ADN simple brin est insérée dans le site actif de l'ARN Pol II (Giardina and Lis, 1993; Plaschka et al., 2016). L'ARN peut alors commencer à être synthétisé (figure 2f) (Sainsbury et al., 2015). Une fois incorporé dans le PIC, TFIIH catalyse également la phosphorylation du CTD de l'ARN Pol II au niveau des serines 5 (Ser5P) via son module kinase cdk7/cycline H (Compe et al., 2019; Schwartz et al., 2003). Cette étape s'avère nécessaire pour la libération de l'ARN Pol II du promoteur et des GTF après stabilisation du complexe ARN Pol II/ADN, environ 10 nt en aval du TSS (figure 2g) (Luse, 2013; Wong et al., 2014).

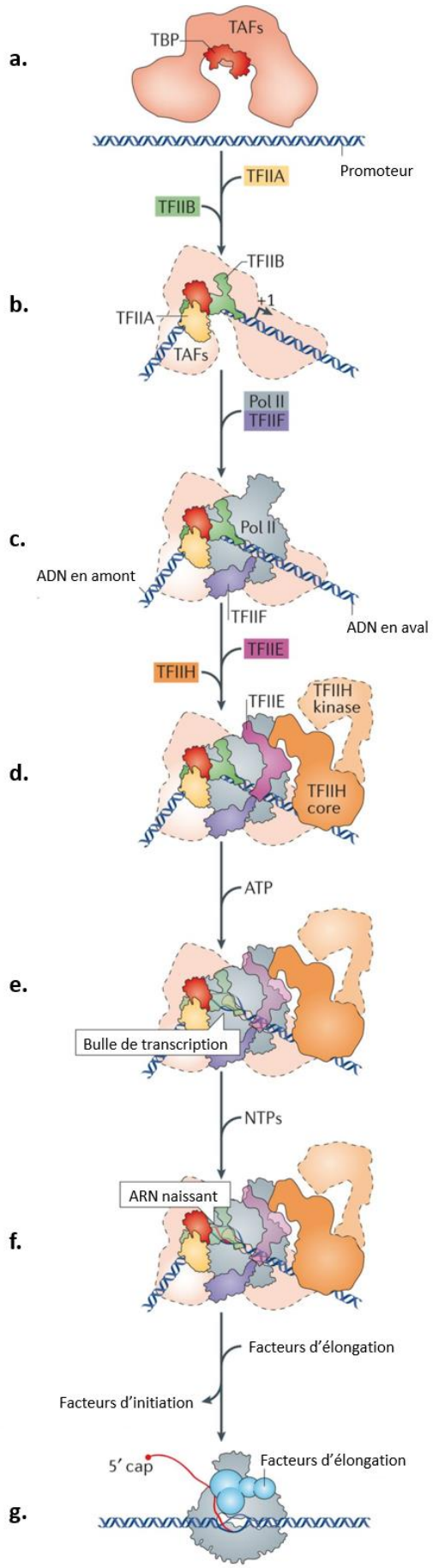


Figure 2 : Assemblage du complexe de pré-initiation (PIC) et initiation de la transcription par l'ARN Pol II (Sainsbury et al., 2015). (a) Le facteur TFIID se lie au promoteur et induit une distorsion de l'ADN. (b) TFIIA et TFIIB sont recrutés et stabilisent le complexe. (c) TFIIB permet le recrutement de l'ARN Pol II et TFIIF. (d)TFIIE et TFIIH, composé d'un cœur enzymatique (core) et d'un complexe kinase, sont recrutés pour compléter le PIC. (e) En présence d'ATP, la sous-unité à activité ATPase de TFIIH, XPB, permet l'ouverture de la double hélice d'ADN (bulle de transcription) et la formation d'un complexe ouvert. (f) Le brin d'ADN matrice s'insère dans le site catalytique de l'ARN Pol II et la synthèse de l'ARN commence. (g) Quand l'ARN atteint une taille d'environ 10 nt, la bulle de transcription s'effondre et l'ARN Pol II est libérée du promoteur. Cet événement est associé à la permutation des facteurs d'initiation avec les facteurs d'élongation.

c. Régulation de l'initiation

Au cours de l'initiation, la multiplicité des étapes aboutissant au début de la transcription sont autant d'opportunités de contrôler le niveau d'expression d'un gène donné (Michel and Cramer, 2013). Ainsi, la reconnaissance du promoteur, le recrutement et l'assemblage du PIC, l'ouverture de la double hélice et la libération de l'ARN Pol II du promoteur peuvent être régulés par la liaison de facteurs de transcription et de cofacteurs au niveau d'éléments génomiques régulateurs situés à une distance plus ou moins grande des promoteurs minimaux (Haberle and Stark, 2018). L'analyse approfondie des séquences génomiques a permis d'identifier un grand nombre de sites de liaison pour des facteurs de transcription spécifiques (Vaquerizas et al., 2009). Ces sites de liaisons sont retrouvés au niveau de régions promotrices proximales, situées, comme leur nom l'indique, proches des promoteurs, mais surtout au niveau de régions plus distales appelées enhancers.

D'un point de vue mécanistique, des études de l'organisation du noyau en trois dimension (chromosome conformation capture (3C) et ses dérivés) ont permis de mettre en évidence des interactions longues distances entre diverses régions génomiques, et notamment entre enhancers et promoteurs (Hughes et al., 2014). Ainsi, la liaison de facteurs de transcription et de cofacteurs au niveau des enhancers permet de mettre en contact la machinerie transcriptionnelle avec le promoteur et augmenter l'initiation de la transcription. Le complexe Mediator, composé de 30 sous-unités chez l'homme, est un acteur clé de ce processus d'activation. Le complexe Mediator interagit à la fois avec les facteurs de transcription recrutés au niveau des enhancers et avec les composants du PIC. Il permet alors d'intégrer les signaux d'activation reçus par les enhancers et d'augmenter le recrutement et l'assemblage du PIC sur les régions promotrices (figure 3a) (Allen and Taatjes, 2015; Eychenne et al., 2016; Eychenne et al., 2017; Jeronimo and Robert, 2017; Soutourina, 2018; Soutourina et al., 2011). Le complexe Mediator participe à la libération de l'ARN Pol II du promoteur en régulant la phosphorylation du CTD de l'ARN Pol II (figure 3b). Il est également impliqué dans les étapes précoces de l'élongation qui seront évoquées dans la partie suivante (voir I.ii.f. Un objectif, de multiples options : le recrutement de P-TEFb). Bien que le complexe Mediator ne soit pas requis pour la transcription *in vitro*, son activité s'avère cruciale pour l'expression globale des gènes *in vivo* (Soutourina, 2018). La déplétion de diverses sous-unités du Mediator est d'ailleurs létale de la levure aux mammifères (Ito et al., 2000; Soutourina et al., 2011; Thompson and Young, 1995; Tudor et al., 1999; Westerling et al., 2007).

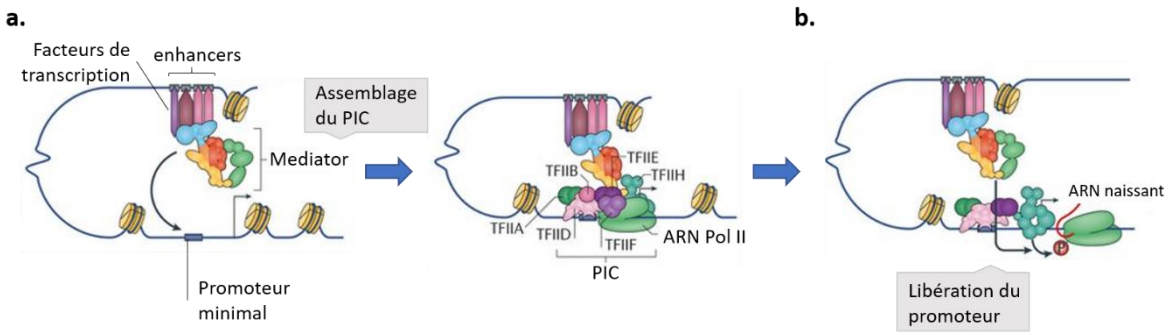


Figure 3 : Modèle d'action du complexe Médiator dans les étapes précoces de la transcription (Soutourina, 2018). (a) Le Mediator contacte les facteurs de transcription situés aux enhancers et facilite l'assemblage du PIC et son recrutement au promoteur. (b) Le Mediator participe également à la libération de l'ARN Pol II des régions promotrices

Des études génomiques et transcriptomiques ont révélé qu'en plus de faciliter le contact entre la machinerie transcriptionnelle et les promoteurs, les facteurs de transcription qui se lient aux enhancers y recrutent également l'ARN Pol II et les GTF et de courts ARN non-codants appelés enhancer RNAs (eRNAs) sont ainsi produits (De Santa et al., 2010; Kim et al., 2010c; Koch et al., 2011). Mais contrairement aux promoteurs de gènes qui favorisent la production d'ARN stables, les eRNA sont généralement peu abondants et rapidement dégradés par l'exosome nucléaire (Andersson et al., 2014b; Mousavi et al., 2013; Pefanis et al., 2015). La transcription des enhancers peut être activée comme celle des gènes en réponse à divers stimuli, et corrèle avec leur fonction d'activateurs transcriptionnels (Andersson et al., 2014a; Kim et al., 2010c; Li et al., 2013b; Li et al., 2016). Ainsi, la transcription des enhancers pourrait stimuler la transcription des gènes par transfert de machinerie transcriptionnelle. Le rôle des eRNAs dans l'activité stimulatrice des enhancers reste néanmoins élusif, tout comme l'origine de la transcription des enhancers : participe-t-elle à leur fonction ou est-elle simplement une conséquence de l'accessibilité de ces régions ? (Li et al., 2016).

d. Transcription antisens

L'ère du séquençage haut-débit a révélé des aspects insoupçonnés de la transcription du génome. Ainsi, il a émergé que plus de 70 % du génome de différentes espèces eucaryotes était transcrit, et parmi eux seulement 2 % correspondaient à la transcription de gènes codants les protéines (Consortium et al., 2007). Une partie de cette transcription non-codante provient de la synthèse d'ARN antisens, générés en amont de promoteurs de gènes et dans le sens opposé (figure 4) (Beck and Warren, 1988; Trinklein et al., 2004; Wei et al., 2011). Similairement à la transcription des enhancers, la transcription antisens génère généralement des ARN non-codants courts (< 2 kb) et instables appelés PROMPT (promoter upstream transcripts) ou uaRNA (upstream antisense RNA) (Flynn et al., 2011; Ntini et al., 2013; Preker et al., 2008). Les transcrits sens et antisens sont produits par des complexes

ARN Pol II distincts recrutés sur des éléments promoteurs généralement différents et à partir de TSS situés aux deux extrémités de la région déplétée en nucléosomes, en orientation opposée (Duttke et al., 2015; Ibrahim et al., 2018; Scruggs et al., 2015; Weingarten-Gabbay et al., 2019). A noter que quelques promoteurs permettent la production de transcrits stables en sens et antisens (Scruggs et al., 2015). Dans ce cas, les nucléosomes -1 et +1 portent les marques classiques de promoteurs de gènes actifs, H3K4me3 et H3K27ac (figure 4). Lorsque les transcrits antisens sont instables, les nucléosomes en position -1 sont enrichis en H3K4me2 en plus de H3K4me3 et H3K27ac (figure 4). Il semblerait que l'instabilité des transcrits antisens soit due à un enrichissement en motifs de type signal de polyadénylation (riches en A-T) dans les séquences en aval du promoteur antisens, entraînant une terminaison rapide et la dégradation par l'exosome nucléaire (Almada et al., 2013; Ntini et al., 2013). Bien que la fonction, s'il y a, de cette transcription divergente reste encore méconnue, elle pourrait participer à l'activation du gène apparenté en fonctionnant comme un enhancer, i.e. en optimisant le recrutement de facteurs de transcription et de la machinerie transcriptionnelle (Scruggs et al., 2015). De plus, la transcription bidirectionnelle pourrait contribuer au maintien d'une région dépourvue en nucléosomes, d'autant plus large que les TSS divergents sont éloignés (Scruggs et al., 2015).

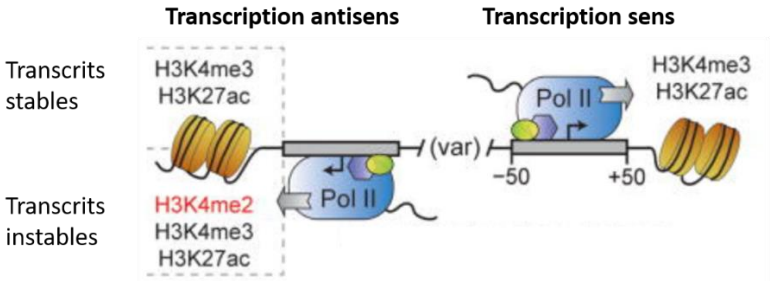


Figure 4 : Représentation schématique de la transcription antisens (Scruggs et al., 2015). La transcription des gènes par l'ARN Pol II (transcription sens) est souvent couplée à la production de courts ARN non-codants instables en orientation inverse (transcription antisens). Le nucléosome en position +1 des promoteurs antisens est caractérisé par la présence des marques H3K4me2/me3 et H3K27ac. L'enrichissement en H3K4me3 et H3K27ac au niveau du nucléosome +1 est associé à la production de transcrits stables, sens et antisens.

ii. L'élongation de la transcription

Une fois l'étape d'initiation achevée, avec la libération du promoteur, l'ARN Pol II rencontre de nouvelles étapes limitantes. Notamment, une pause transcriptionnelle majeure a lieu au début de l'élongation, à proximité du promoteur, pour la quasi-totalité des gènes transcrits par l'ARN Pol II (Adelman and Lis, 2012; Guo and Price, 2013; Jonkers and Lis, 2015; Mayer et al., 2017; Zhou et al., 2012). Cette étape limitante ne semble pas exister chez la levure (Churchman and Weissman, 2011; Lidschreiber et al., 2013). En revanche, la pause à proximité du promoteur constitue un mécanisme de régulation de l'entrée en élongation productive conservé chez les métazoaires, et particulièrement important pour l'intégration des voies de signalisation cellulaires et le développement (Levine, 2011;

Liu et al., 2015; Williams et al., 2015; Zeitlinger et al., 2007). La pause de l'ARN Pol II à proximité du promoteur est imposée par des facteurs négatifs de l'élongation et sa libération repose sur le recrutement du facteur positif de l'élongation P-TEFb. Durant la phase d'élongation productive, l'ARN Pol II marque aussi fréquemment des pauses. Cependant, ces pauses dans le corps du gène diffèrent de la pause à proximité du promoteur. Elles sont provoquées par la séquence d'ADN traversée ou par les nucléosomes et sont généralement facilement surmontées grâce aux facteurs d'élongation qui accompagnent l'ARN Pol II durant l'élongation, stimulant sa progression dans le gène et remodelant la chromatine sur son passage.

a. La pause de l'ARN Pol II à proximité du promoteur : de la découverte à la généralisation

Une accumulation de l'ARN Pol II a initialement été détectée en 5' de gènes particulièrement étudiés, comme les gènes de réponse au choc thermique *hsp70* chez la drosophile ou les gènes de réponse précoce *c-myc* et *c-fos* chez les mammifères (Bentley and Groudine, 1986; Collart et al., 1991; Gilmour and Lis, 1986; Spencer and Groudine, 1990). Des expériences de transcription *in vitro* sur des matrices ADN immobilisées et incubées avec différents extraits nucléaires totaux ou fractionnés ont ensuite permis d'identifier les facteurs nécessaires pour bloquer l'élongation à proximité du promoteur, appelés N-TEFs (Negative Transcriptional Elongation Factors) : NELF (negative elongation factor) et DSIF (DRB-sensitivity inducing factor), et un facteur requis pour générer des complexes d'élongation productifs : P-TEFb (Positive Transcription Elongation Factor b) (Marshall and Price, 1992, 1995; Wada et al., 1998; Yamaguchi et al., 1999). Ainsi, selon le modèle prévalent, l'initiation de la transcription ne semble pas limitante sur ces gènes, qui sont plutôt régulés au stade d'élongation précoce : à proximité du promoteur, les N-TEFs s'associent à la machinerie transcriptionnelle et empêchent sa progression (figure 5b). L'activité de P-TEFb est alors nécessaire pour la transition vers l'élongation productive (figure 5c).

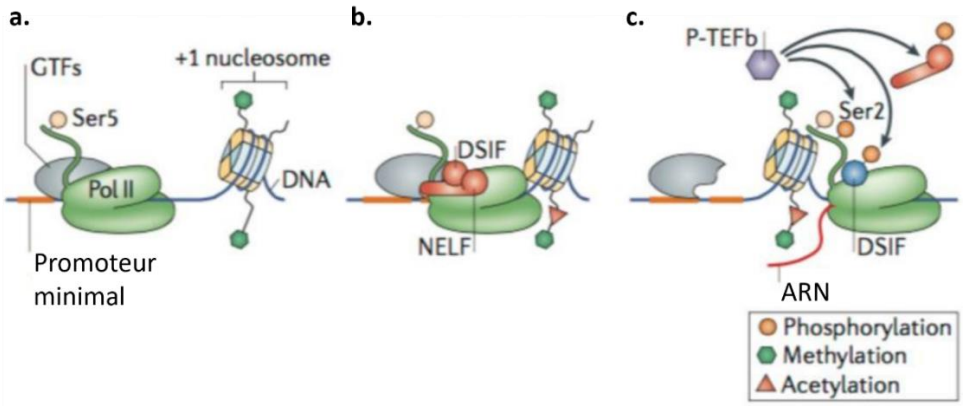


Figure 5 : Pause de la transcription par l'ARN Pol II à proximité du promoteur (Jonkers and Lis, 2015). (a) L'ARN Pol II est recrutée au niveau du promoteur minimal et initie la transcription. (b) Peu après le début de l'élongation, la liaison des facteurs négatifs de l'élongation, NELF et DSIF, empêche la progression de l'ARN Pol II dans le corps du gène. (c) La reprise de l'élongation est permise par le facteur P-TEFb qui phosphoryle DSIF, NELF et le CTD de l'ARN Pol II au niveau des sérines 2. Phospho-DSIF est convertit en facteur positif de l'élongation tandis que NELF est relargué du complexe d'élongation.

Grâce aux profondes avancées techniques ayant vu le jour au cours des deux dernières décennies, l'occurrence de cette pause à proximité du promoteur a pu être généralisée et son mécanisme approfondi (Chen et al., 2018; Guo and Price, 2013; Jonkers and Lis, 2015). La plupart des gènes transcrits par l'ARN Pol II chez les métazoaires et la drosophile, y compris les gènes de ménages fortement exprimés, sont soumis à cette régulation post-initiation (Core et al., 2008; Muse et al., 2007; Zeitlinger et al., 2007). Différentes techniques pour étudier l'élongation de la transcription ont permis de détecter et caractériser la pause de l'ARN Pol II à proximité du promoteur (tableau 1). Ainsi, une accumulation d'ARN Pol II, mais aussi des N-TEFs NELF et DSIF, est détectée par immunoprécipitation de chromatine (ChIP) dans une région localisée 20 à 120 nt en aval du TSS sur 40-70 % des gènes actifs (Core et al., 2008; Kwak et al., 2013; Min et al., 2011; Nechaev et al., 2010; Rahl et al., 2010). Le séquençage des ARN naissants montre également une transcription accrue à proximité du promoteur, comparé au niveau de transcription dans le reste du gène pour la plupart des gènes. Les gènes présentant un fort niveau d'ARN Pol II ou de transcrits à proximité du promoteur sont alors dits en pause, ou fortement régulés par la pause à proximité du promoteur.

Technique	***Principe***	Références
ChIP-seq *(Chromatine immunoprecipitation and sequencing)*	Immunoprécipitation de l'ARN Pol II et séquençage des fragments d'ADN (~200 pb) co-immunoprécipités.	(Chen et al., 2015b; Guenther et al., 2007; Muse et al., 2007)
GRO-seq *(Global nuclear run-on and sequencing) et* ***PRO-seq*** *(Precision nuclear run-on and sequencing)*	Purification des noyaux et courte transcription *in vitro* avec incorporation de bromo-UTP ; immunoprécipitations des ARN néosynthétisés-bromo-UTP et séquençage.	(Core et al., 2008; Kwak et al., 2013; Mahat et al., 2016)
BruDRB-seq *et* ***4sUDRB-seq***	Synchronisation de la transcription par inhibition de la libération de la pause au DRB puis GRO-seq.	(Bugai et al., 2019; Jonkers et al., 2014)
Start-seq *(start site-associated RNA sequencing)*	Purification des ARN courts (< 150 nt) coiffés associés à la chromatine et séquençage par les deux extrémités.	(Henriques et al., 2018; Nechaev et al., 2010; Scruggs et al., 2015)
mNET-seq *(mammalian native elongating transcript sequencing)*	Immunoprécipitation de l'ARN Pol II à la chromatine et séquençage des ARN associés.	(Nojima et al., 2015; Nojima et al., 2018)

Tableau 1 : Principales techniques de détection de la pause de l'ARN Pol II à l'échelle du génome.

L'activité kinase de P-TEFb s'avère cruciale pour transiter vers l'élongation productive pour tous les gènes actifs, qu'ils présentent une accumulation de l'ARN Pol II à proximité du promoteur ou pas (Gressel et al., 2017; Henriques et al., 2013; Jonkers et al., 2014). En effet, l'inhibition de P-TEFb par le Flavopiridol ou le DRB, un analogue de l'ATP, engendre un blocage général de la transcription concomitant à une augmentation du niveau d'ARN Pol II à proximité du promoteur. De la même manière, la durée de la pause varie d'un gène à l'autre, de quelques minutes pour la plupart des gènes jusqu'à une heure pour quelques gènes très stablement en pause (Chen et al., 2015a; Erickson et al., 2018; Henriques et al., 2013; Jonkers et al., 2014). Ces données impliquent un mécanisme d'action ou de recrutement de P-TEFb sur les complexes d'ARN Pol II en pause différent selon les gènes. L'activité de P-TEFb semble ainsi limitante pour certains gènes (ceux présentant une forte accumulation de l'ARN Pol II au TSS, indiquant un temps de résidence du complexe d'élongation plus long) et non limitante pour d'autres (ceux ne présentant pas d'accumulation d'ARN Pol II au TSS dans la condition étudiée). Pour conforter cette assomption, de nombreux facteurs peuvent interagir de manière régulée avec P-

TEFb pour l'activer, le cibler sur les gènes ou au contraire l'inhiber (voir I.ii.e. Libération de l'ARN Pol II en élongation productive).

Globalement, la pause de l'ARN Pol II à proximité du promoteur peut être vue comme la balance entre le recrutement de facteurs négatifs et de facteurs positifs favorisant la transition vers l'élongation productive. Mais d'autres éléments participent à la complexité et diversité de ce processus de régulation, tels que le taux d'initiation, la nature de la séquence d'ADN au niveau du promoteur et du site de pause et la composition locale de la chromatine.

La pause transcriptionnelle est donc un mécanisme complexe dépendant principalement de trois paramètres qui seront détaillés dans les prochains paragraphes (figure 6) :

- Le taux de recrutement de l'ARN Pol II
- L'entrée en pause et son maintien
- La levée de la pause ou terminaison précoce

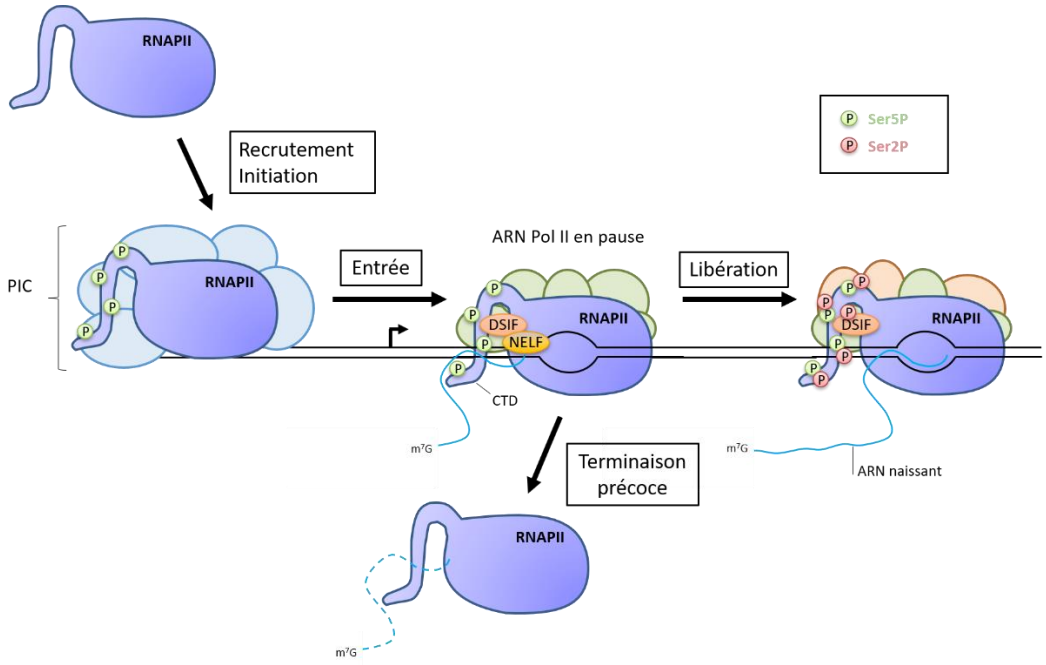


Figure 6 : Modulation du niveau d'ARN Pol II en pause à proximité du promoteur. PIC : preinitiation complex.

b. Fonctions de la pause de l'ARN Pol II à proximité du promoteur

A première vue, le contrôle de la transcription à une étape post-initiation présente un avantage cinétique pour l'induction rapide des gènes en réponse à un stimulus. L'entrée en élongation productive est directement permise par le recrutement de P-TEFb au niveau de l'ARN Pol II en pause, en s'affranchissant donc des étapes précoces de la transcription. L'accumulation de l'ARN Pol II à proximité du promoteur est en effet détectée sur de nombreux gènes rapidement induits en réponse

à diverses voies de signalisation, tels que les gènes *c-jun*, *c-myc*, *c-fos*, les gènes de réponse à l'hypoxie (Galbraith et al., 2013), aux chocs thermiques (Mahat et al., 2016), aux signaux inflammatoires (Adelman et al., 2009; Escoubet-Lozach et al., 2011; Luo et al., 2014) et aux œstrogènes (Hah et al., 2011; Kininis et al., 2007; Kininis et al., 2009). Cependant, l'induction rapide en réponse à un stimulus ne repose pas uniquement sur la libération de la pause. Le contrôle de la pause à proximité du promoteur peut en effet agir de concert ou en opposition à la régulation de l'initiation de la transcription, permettant un contrôle précis du taux de transcription. Par exemple, les gènes de réponse aux œstrogènes présentent une accumulation de l'ARN Pol II à proximité du promoteur en conditions non stimulées. Suite au traitement aux œstrogènes, une augmentation de l'accumulation de l'ARN Pol II aux promoteurs et dans les régions transcrites des gènes de réponse est observée. Ainsi, la pause de l'ARN Pol II contrôle la transcription de ces gènes en condition normale, mais leur induction repose principalement sur l'augmentation de l'initiation de la transcription (Danko et al., 2013). La pause à proximité du promoteur apparait d'ailleurs particulièrement importante pour la régulation des acteurs des voies de signalisation (Gilchrist et al., 2012; Williams et al., 2015). Ainsi, l'altération de la pause à proximité du promoteur par déplétion de NELF affecte majoritairement l'expression des gènes impliqués dans la transduction des signaux. L'induction des gènes de réponse après stimulation est donc compromise en raison du défaut d'activation de la voie de signalisation, comme par exemple la voie FGF/Erk, requise pour la différenciation des cellules embryonnaires de souris (Gilchrist et al., 2012; Williams et al., 2015).

Il est important de noter que la pause de l'ARN Pol II à proximité du promoteur n'est pas un mécanisme associé à la répression transcriptionnelle. Elle a lieu sur des gènes de niveaux d'expression variés, des fortement transcrits à ceux faiblement exprimés, mais est quasiment toujours associée à une activité transcriptionnelle. Notamment, la pause de l'ARN Pol II pourrait permette de maintenir un état chromatinien permissif aux promoteurs (Core et al., 2012; Gilchrist et al., 2010; Gilchrist et al., 2012; Gilchrist et al., 2008). Les gènes fortement en pause présentent en effet une faible occupation en nucléosomes au promoteur. La déplétion de NELF a ainsi été associée à une réorganisation des nucléosomes aux promoteurs de ces gènes, couplée à une diminution de leur expression (Gilchrist et al., 2010). Dans le même ordre d'idée, l'altération de la pause à proximité du promoteur sur le gène hsp70, par mutation de la séquence d'ADN, restreint l'accessibilité du promoteur aux facteurs de transcription et compromet l'induction transcriptionnelle en réponse à un choc thermique (Lee et al., 1992; Shopland et al., 1995). Le contrôle de la pause de l'ARN Pol II à proximité du promoteur semble toutefois participer à la répression des gènes en réponse à certains stimuli (Aiyar et al., 2004; Gupte et al., 2013; Mahat et al., 2016). Après activation, le récepteur aux œstrogènes ERα permet le recrutement de NELF sur certains de ses gènes cibles, induisant ainsi leur répression transcriptionnelle

(Aiyar et al., 2004). Le même mécanisme est employé par le récepteur aux glucocorticoïdes (Gupte et al., 2013). En réponse à un choc thermique, l'expression de nombreux gènes impliqués dans le métabolisme et le cycle cellulaires est diminuée par une baisse de la libération de l'ARN Pol II en élongation productive, mais le mécanisme n'est pas connu (Mahat et al., 2016).

La pause de l'ARN Pol II à proximité du promoteur permet donc une régulation fine de l'expression des gènes. Elle permet de maintenir les promoteurs dans une configuration ouverte et accessible aux facteurs de transcription, et de maintenir une activité basale, assurant ainsi une activation efficace après stimulation.

c. Synergie pause transcriptionnelle/initiation

Le niveau d'ARN Pol II en pause à proximité du promoteur est par essence dépendant de la fréquence d'initiation (figure 6). En effet, l'entrée de l'ARN Pol II en pause implique que la machinerie transcriptionnelle doit avoir au préalable été recrutée sur le promoteur et avoir initié la transcription. Cependant, il a récemment été montré que la présence d'une ARN Pol II en pause à proximité du promoteur limitait le recrutement de nouvelles ARN Pol II (Gressel et al., 2017; Shao and Zeitlinger, 2017). Des modélisations moléculaires suggèrent qu'une ARN Pol II en pause génère un encombrement stérique empêchant le recrutement de la machinerie transcriptionnelle au même promoteur (Ehrensberger et al., 2013). L'étude résolutive de la distribution de l'ARN Pol II à l'échelle du génome par ChIP-nexus a mis en évidence une corrélation inverse entre la présence de l'ARN Pol II au sein du PIC aux promoteurs et celle de l'ARN Pol II en pause. De plus, l'augmentation de la stabilité des complexes en pause, par inhibition de P-TEFb, induit une diminution du signal correspondant aux PIC au niveau des promoteurs. En combinant plusieurs méthodes de séquençage des ARN naissants à l'inhibition de la libération de la pause, une limite pause-initiation a d'ailleurs été mise en évidence. En dessous de cette limite, la durée de la pause n'est pas limitante et la production d'ARN peut s'accroitre par augmentation du taux d'initiation. Si en revanche la pause est plus longue, la libération de l'ARN Pol II en élongation est nécessaire pour augmenter le taux de transcription. Ainsi, la stabilité des complexes d'élongation en pause, et donc le recrutement de P-TEFb, semble pouvoir déterminer les capacités d'initiation d'un gène donné.

d. Mise en place de la pause de l'ARN Pol II

L'établissement de la pause à proximité du promoteur repose essentiellement sur la liaison des facteurs négatifs de l'élongation, NELF et DSIF. Cependant, les mécanismes induisant leur association aux complexes d'élongation au niveau d'une région précise du gène sont encore un domaine de

recherche actif. Encore une fois, des études à l'échelle du génome ont permis d'identifier de nouveaux éléments impliqués dans le blocage de l'ARN Pol II à proximité du promoteur, comme la séquence d'ADN ou la chromatine. Ces éléments pourraient ainsi induire le ralentissement ou l'arrêt de l'ARN Pol II à proximité du promoteur, et ouvrir ainsi une fenêtre temporelle pour faciliter le recrutement des facteurs négatifs sur le complexe d'élongation qui finalisent la mise en place de la pause de l'ARN Pol II (Li et al., 2013a).

- *Rôle de la séquence d'ADN des régions promotrices proximales :*

Une étude métagénomique a récemment mis en évidence que le niveau de pause de l'ARN Pol II à proximité du promoteur était comparable entre différents types cellulaires, y compris cancéreux, pour de nombreux gènes (Day et al., 2016). Ces données suggèrent que la séquence de la région promotrice pourrait jouer un rôle dans l'établissement de la pause. En accord avec cette hypothèse, les séquences promotrices de gènes fortement en pause ont tendance à être enrichies en GC, îlots CpG et dépourvues de boîte TATA, bien que ce ne soit pas une caractéristique générale (Core et al., 2008; Day et al., 2016; Hendrix et al., 2008; Kellner et al., 2015; Li and Gilmour, 2011). Le gène c-myc est par exemple caractérisé par la présence d'une boîte TATA dans son promoteur, nécessaire pour que l'ARN Pol II entre en pause (Krumm et al., 1995). Chez la drosophile, des éléments promoteurs appelés Downstream Promoter Element (DPE), Pause Button (PB) et Initiator (Inr) ainsi que des motifs de type GAGA sont souvent associés à une pause marquée de l'ARN Pol II à proximité du promoteur (Hendrix et al., 2008; Lagha et al., 2013; Li et al., 2013a; Shao et al., 2019). Notamment, les DPE et PB, riches en GC, sont précisément localisés au niveau du site de pause et suivis de séquences riches en AT (Nechaev et al., 2010). Or, la transcription de régions riches en AT provoque une diminution de la stabilité de l'hybride ADN/ARN naissant, et l'ARN Pol II a alors tendance à glisser vers l'arrière et à se stabiliser au niveau de régions plus stables thermodynamiquement, menant à l'arrêt de la synthèse d'ARN du fait de son désalignement avec le site actif de l'ARN Pol II (Komissarova and Kashlev, 1997). En outre, durant la transcription, l'ADN non matrice engagé dans la bulle transcriptionnelle est sous forme simple brin et peut ainsi potentiellement adopter des structures secondaires stables (Bochman et al., 2012). Par exemple, l'enrichissement en G, en particulier dans la séquence du brin non matrice, retrouvé au niveau des régions 5' des gènes en pause peut induire la formation d'une structure en G-quadruplex (repliement de l'ADN ou de l'ARN provoqué par la formation de liaisons hydrogène entre guanines) empêchant la progression de l'ARN Pol II (Eddy et al., 2011). Le repliement de l'ADN non matrice dans la bulle transcriptionnelle peut également former des motifs structuraux reconnus par des facteurs influençant la pause. Cela a par exemple été suggéré pour le facteur TRIM28 (également appelé KAP1) qui reconnait le motif GCCGCG exposé sur le brin non matrice du gène *hsp70* et semble stabiliser la pause de l'ARN Pol II (Bunch et al., 2014). Il a d'ailleurs été très récemment montré que les

régions d'ADN au niveau des sites de pause étaient particulièrement propices à la formation de structures secondaires. Trois modèles ont été proposés pour l'implication des structures secondaires de l'ADN dans l'établissement de la pause : (1) par la formation de motifs structuraux permettant le recrutement de facteurs spécifiques, comme TRIM28, (2) par la formation d'hybrides ADN-ARN stables, autrement appelés R-loops, en amont de la pause, empêchant ainsi la réhybridation des deux brins d'ADN, (3) par la structuration du brin non matrice et la forte stabilité de l'hybride ADN-ARN au sein même de la bulle transcriptionnelle (Szlachta et al., 2018). Ces données ouvrent de nouvelles perspectives concernant les mécanismes gouvernant la pause de l'ARN Pol II à proximité du promoteur.

- *Rôle de la chromatine :*

Les gènes en pause présentent une plus forte déplétion en nucléosome au promoteur que les gènes non en pause (Core et al., 2012; Day et al., 2016; Gilchrist et al., 2010). Il a donc été suggéré que la pause de l'ARN Pol II à proximité du promoteur permettrait de garder au promoteur un contexte chromatinien favorable à l'activation transcriptionnelle. En revanche, le nucléosome +1 est fortement positionné juste en aval du site de pause et pourrait participer à l'établissement de la pause en générant une barrière hostile à la poursuite de l'élongation (Egloff et al., 2009; Jimeno-Gonzalez et al., 2015; Mavrich et al., 2008; Voong et al., 2016; Weber et al., 2014). Chez les mammifères, un enrichissement du variant d'histone H2A.Z a été observé dans les nucléosomes aux positions -1 et +1 des gènes particulièrement en pause (Day et al., 2016). Cependant, la présence de ce variant d'histone est plutôt associé à une diminution de la pause chez la drosophile (Weber et al., 2010). Ces données suggèrent des mécanismes de pause variés et probablement espèce-spécifiques.

- *Les facteurs négatifs de l'élongation DSIF et NELF (Cf. figure 5) :*

DSIF est un hétérodimère composé des sous-unités Spt4 et Spt5. Spt5 possède un domaine C terminal à régions répétées, de manière similaire au CTD de l'ARN Pol II, et lui aussi cible de phosphorylations, notamment par P-TEFb. Contrairement à NELF, DSIF est requis pour l'établissement de la pause mais aussi pour l'élongation de la transcription. La longueur de l'ARN naissant semble être un déterminant pour la liaison de DSIF, dont l'affinité pour le complexe d'élongation augmente substantiellement lorsque le transcrit atteint une taille supérieure à 18 nt (Missra and Gilmour, 2010), et est perdue si le transcrit est dégradé par traitement à la RNase A (Cheng and Price, 2008). DSIF émet en effet des contacts multiples avec l'ARN Pol II et l'ARN naissant, notamment via les motifs KOW (Kyrpides, Ouzounis, Woese) de Spt5, requis pour l'établissement de la pause (Bernecky et al., 2017; Cheng and Price, 2008; Ehara et al., 2017; Missra and Gilmour, 2010; Qiu and Gilmour, 2017). Curieusement, les régions de Spt5 impliquées dans la liaison à l'ARN sont également importantes pour la stabilité et la

processivité du complexe d'élongation en participant au maintien de l'ARN dans le canal de sortie (Bernecky et al., 2016; Bernecky et al., 2017). Il a récemment été montré que ces régions de Spt5 pouvaient être phosphorylées par P-TEFb, expliquant potentiellement la conversion répression/activation de la transcription par DSIF (Sanso et al., 2016).

NELF est un complexe de 4 sous-unités, NELF-A, -B, -C/D et -E conservé chez les métazoaires. Les sous-unités C et D proviennent de la traduction différentielle du même ARNm, mais des rôles spécifiques n'ont pas été identifiés. NELF-A et NELF-C/D forment le sous-complexe cœur auquel s'associent les sous-unités B et E (Vos et al., 2016) vos 2018. Les sous-unités -B, -C/D et -E de NELF peuvent contacter l'ARN *in vivo*, et NELF-E possède notamment un motif de reconnaissance d'ARN (RRM, RNA Recognition Motif), et pourrait participer à la liaison de NELF à l'ARN Pol II alors que l'ARN naissant commence à émerger du complexe d'élongation (Vos et al., 2016). La délétion du domaine RRM de la sous-unité NELF-E empêche en effet la liaison de NELF sur des promoteurs de gènes de réponse primaires de cellules neuronales de souris *in vivo* (Schaukowitch et al., 2014). De même, le domaine RRM de NELF-E, bien que dispensable pour l'intégrité du complexe NELF, est nécessaire pour inhiber la transcription de gènes rapporteurs en cellules HeLa (Yamaguchi et al., 2002). La liaison de NELF au complexe d'élongation requiert également la liaison préalable de DSIF à l'ARN Pol II (Missra and Gilmour, 2010; Vos et al., 2018b; Yamaguchi et al., 2002; Yamaguchi et al., 1999). NELF contacte en effet directement DSIF via deux extensions générées par les sous-unités NELF-A et NELF-E, dont la première apparait nécessaire pour la stabilisation de la pause. Au sein du complexe ARN Pol II-DSIF, NELF émet des contacts clés avec l'ARN Pol II permettant d'expliquer son rôle dans l'inhibition de la transcription (Vos et al., 2018b). Notamment, les sous-unités NELF-A et -C restreignent la mobilité de l'ARN Pol II, son activité catalytique et stabilisent une conformation particulière, inclinée (tilted), de l'hybride ADN/ARN dans le site actif de l'ARN Pol II qui empêche la poursuite de la synthèse du transcrit. De plus, la liaison de NELF est incompatible avec la liaison de TFIIS (Vos et al., 2018b), facteur qui est nécessaire pour la réactivation de l'ARN Pol II et la reprise de l'élongation (Adelman et al., 2005; Sheridan et al., 2019).

A noter que la liaison de NELF-A à l'ARN Pol II est incompatible avec la liaison de TFIIF, et celle de DSIF avec TFIIB et TFIIE (Bernecky et al., 2017; Grohmann et al., 2011; He et al., 2016; Vos et al., 2018a). Leur liaison à l'ARN Pol II ne peut ainsi se faire qu'après le début de l'élongation marqué par l'éviction des facteurs d'initiation.

Ainsi, ces différentes données soutiennent un modèle dans lequel l'association de DSIF est permise lorsque l'ARN atteint une taille suffisante pour émerger du complexe d'élongation, après dissociation des facteurs d'initiation, et est stabilisée par ses contacts avec l'ARN Pol II. La liaison de DSIF à la

machinerie transcriptionnelle induit un contexte favorable pour le recrutement de NELF qui vient alors stabiliser le complexe et empêcher la poursuite de l'élongation.

- *Autres facteurs intervenant dans le maintien de la pause :*

In vitro, NELF et DSIF sont suffisants pour établir la pause. Cependant, *in vivo,* d'autres facteurs apparaissent participer à ce processus (Cheng and Price, 2007; Missra and Gilmour, 2010; Renner et al., 2001) :

Gdown1 : Environ 30-50 % des complexes ARN Pol II contiennent une sous-unité facultative appelée Gdown1. Gdown1 s'associe après l'étape d'initiation à l'ARN Pol II, avec laquelle elle co-localise fortement à proximité du promoteur (Cheng et al., 2012a; Jishage et al., 2012; Mullen Davis et al., 2014). En empêchant la liaison du facteur de terminaison de la transcription TTF2 et celle de TFIIF, aussi facteur d'élongation, Gdown1 joue un rôle clé dans la stabilisation de la pause de l'ARN Pol II (DeLaney and Luse, 2016; Guo et al., 2014).

GAF : Chez la drosophile, le facteur GAGA (GAF) reconnait les motifs GAGA au niveau des régions promotrices, enrichis sur les gènes en pause. GAF participe au contrôle de la localisation de la pause en favorisant le recrutement de NELF (Li et al., 2013a).

Les récepteurs nucléaires : De la même manière, NELF est recruté par le récepteur aux œstrogènes ERα sur les régions promotrices de ses gènes cibles pour en atténuer l'expression (Aiyar et al., 2004). Le récepteur aux glucocorticoïdes (GR) agit à différentes étapes de la transcription pour réprimer l'expression des gènes codant les cytokines (Chinenov et al., 2012). Avec l'aide de son co-répresseur GRIP1 (GR-interacting protein), le GR recrute NELF pour diminuer l'entrée en élongation d'une partie de ses gènes cibles (Gupte et al., 2013).

PAF1C : Le complexe PAF1 (Polymerase Associated Factor, PAF1c) est récemment ressorti d'un crible visant à identifier des facteurs requis pour la pause de l'ARN Pol II à proximité du promoteur (Chen et al., 2015b). La déplétion de la sous-unité PAF1 de ce complexe dans différentes lignées cellulaires humaines et de drosophile induit une importante augmentation de la transcription de nombreux gènes, que les auteurs de cette étude expliquent par une augmentation de la libération de la pause de l'ARN Pol II. Le complexe PAF1 semble en effet contraindre le recrutement de P-TEFb, au sein du Super Elongation Complex (SEC), aux promoteurs, limitant ainsi le passage de l'ARN Pol II en élongation productive. Plus récemment, le même groupe a trouvé que PAF1 était recruté sur une fraction d'enhancers et inhibait leur expression (Chen et al., 2017). Ainsi, la déplétion de PAF1 par shRNA stimule leur activité transcriptionnelle (augmentation du recrutement de l'ARN Pol II et de la production d'eRNA), mais aussi celle des gènes avoisinants par libération de la pause

transcriptionnelle. Cependant, alors que ces études établissent un rôle du PAF1C dans la stabilisation de la pause, ce complexe était jusque-là reconnu pour son activité stimulatrice de l'élongation. Deux études récentes rapportent également un effet positif du PAF1C dans la libération de la pause, qui sera décrit dans la section suivante. Le PAF1C semble donc être un facteur polyvalent intervenant par divers mécanismes dans le contrôle de l'élongation de l'ARN Pol II. Reste à déterminer dans quels contextes PAF1C exerce-t-il plutôt une activité répressive ou stimulatrice sur la libération de la pause de l'ARN Pol II.

De nombreux paramètres et facteurs peuvent favoriser l'entrée en pause et son maintien. Les multiples combinaisons possibles offrent un immense potentiel de régulation et l'interaction entre ces différents éléments pourrait participer aux variations gènes-spécifiques observées. Une concentration d'éléments favorisant l'établissement de la pause pourrait augmenter la difficulté à outrepasser ce point de contrôle par les facteurs promouvant l'élongation de la transcription.

e. Libération de l'ARN Pol II en élongation productive

De nombreuses études ont montré que le facteur clé pour la libération de la pause de l'ARN Pol II est P-TEFb (Positive Transcription Elongation Factor b), un hétérodimère composé de la kinase Cdk9 et de la sous-unité régulatrice Cycline T1 ou T2 (AJ et al., 2016; Peng et al., 1998; Price, 2000; Zhou et al., 2012). P-TEFb est recruté au niveau de l'ARN Pol II en pause et phosphoryle le complexe d'élongation au niveau de nombreux sites (Vos et al., 2018a). Parmi eux, le CTD de l'ARN Pol II au niveau des serines 2 (Marshall et al., 1996), plusieurs résidus de la sous-unité Spt5 de DSIF (Sanso et al., 2016; Yamada et al., 2006), notamment au niveau du CTR, et les sous-unité NELF-A et NELF-E (Fujinaga et al., 2004; Vos et al., 2018a). Ces événements de phosphorylation permettent la dissociation du complexe NELF, la conversion de DSIF en un facteur positif de l'élongation (Wada et al., 1998) et le recrutement de facteurs d'élongation, de maturation de l'ARN, de remodelage/modification de la chromatine et de terminaison (Ahn et al., 2004; Bentley, 1999; Bentley, 2005, 2014; Lenasi and Barboric, 2010; Liu et al., 2009; Lunde et al., 2010; Schuller et al., 2016; Yoh et al., 2007). Ainsi, l'activité de P-TEFb s'avère non seulement nécessaire pour la levée de la pause, mais également critique pour toutes les étapes ultérieures de la transcription.

Cependant, *in vitro,* P-TEFb seul ne parvient pas à lever la pause de l'ARN Pol II imposée par NELF et DSIF et d'autres facteurs apparaissent nécessaires pour le relargage de la pause (Vos et al., 2018a; Wada et al., 2000). Ainsi, dans un système minimal de transcription *in vitro* contenant l'ARN Pol II, TFIIE, TFIIH, NELF et DSIF, le facteur d'élongation FACT agit en collaboration avec P-TEFb pour activer l'élongation transcriptionnelle (Wada et al., 2000). L'activité de ces deux facteurs ne permet toutefois

pas de recouvrer le niveau de transcription obtenu en l'absence des N-TEFs, suggérant que des facteurs additionnels peuvent stimuler la levée de la pause. De plus, très récemment, un rôle de FACT dans le maintien de l'ARN Pol II en pause a été identifié (Tettey et al., 2019). La diminution des niveaux de FACT dans des cellules de drosophiles entraîne en effet une forte réduction des niveaux d'ARN Pol II à proximité du promoteur associée à une durée de pause raccourcie.

Comme mentionné dans la section « mise en place de la pause », les séquences d'ADN dans les régions promotrices proximales peuvent favoriser le recul de l'ARN Pol II en pause, créant un désalignement de l'extrémité 3' de l'ARN naissant avec le site actif de l'ARN Pol II. Le facteur d'élongation TFIIS participe également à la reprise de la transcription après la pause à proximité du promoteur en stimulant le clivage de la portion d'ARN dépassant du site actif (Adelman et al., 2005; Nechaev et al., 2010; Palangat et al., 2005; Sheridan et al., 2019). TFIIS est également nécessaire durant l'élongation pour résoudre les pauses de l'ARN Pol II induites par différents obstacles (voir I.ii.h. L'élongation productive).

Deux études ont récemment indiqué un rôle clé du complexe PAF1 dans la libération de la pause de l'ARN Pol II par P-TEFb (Vos et al., 2018a; Yu et al., 2015). Comme cela a été entrevu dans la section précédente, PAF1C est un régulateur multifonctionnel de l'expression des gènes chez les eucaryotes, intervenant à de nombreuses étapes de la transcription par l'ARN Pol II et notamment dans l'élongation, dont le rôle sera décrit dans la section suivante (Van Oss et al., 2017). Chez l'homme, PAF1C est composé des 6 sous-unités Paf1, Ctr9, Cdc73, Leo1, Rtf1 et Ski8/Wdr61 (Zhu et al., 2005). PAF1C interagit directement avec le CTD phosphorylé de l'ARN Pol II et avec DSIF, et son recrutement nécessite l'activité kinase de P-TEFb (Liu et al., 2009; Mbogning et al., 2013; Qiu et al., 2012; Vos et al., 2018a; Yu et al., 2015). En accord avec sa localisation, il a récemment été montré que PAF1C participe au contrôle de la pause à l'étape d'élongation précoce (Chen et al., 2015b; Lu et al., 2016b; Wu et al., 2014; Yu et al., 2015). Cependant, des rôles opposés dans le maintien et la libération de la pause ont été rapportés, dépendant potentiellement du type cellulaire utilisé (Chen et al., 2015b; Chen et al., 2017; Lu et al., 2016b; Wu et al., 2014; Yu et al., 2015). En effet, la déplétion de PAF1 dans les cellules humaines de leucémie myéloïde THP1 induit une augmentation globale de la pause de la Pol II, tandis qu'elle diminue dans les cellules CCRF-CEM (cellules de leucémie lymphoblastique). L'étude structurale et moléculaire de la transition de complexes ARN Pol II en pause à des complexes en élongation a mis en évidence que NELF et PAF1C partageaient les mêmes sites de liaison à l'ARN Pol II (Vos et al., 2018a). La liaison du PAF1C au complexe d'élongation est donc incompatible avec la liaison de NELF aux complexes en pause. Cette étude a également montré que la liaison de PAF1C à l'ARN Pol II était nécessaire pour le relargage de NELF, de manière dépendante de l'activité kinase de P-TEFb : sans P-TEFb, l'incubation des facteurs négatifs NELF et DSIF et de PAF1C avec le complexe ARN Pol II induit la

formation du complexe de pause, mais PAF1C est exclu. Avec P-TEFb en revanche, la formation d'un complexe d'élongation est favorisée, et NELF est exclu. Ces données sont en accord avec des études réalisées chez la levure démontrant que la phosphorylation de la région C-ter de Spt5 par Bur1, l'homologue de Cdk9, stimule le recrutement du PAF1C sur les régions transcrites (Laribee et al., 2005; Liu et al., 2009; Qiu et al., 2012; Zhou et al., 2009). Ainsi, l'ensemble de ces données tendent vers un modèle dans lequel, au niveau des complexes ARN Pol II en pause, la phosphorylation de Spt5 par P-TEFb permet le recrutement du PAF1C. PAF1C entre alors en compétition avec NELF pour sa liaison à l'ARN Pol II et entraîne son exclusion du complexe. Des études supplémentaires sont nécessaires pour comprendre les effets négatifs du PAF1C sur la libération de la pause.

f. Un objectif, de multiples options : le recrutement de P-TEFb

La libération de la pause de l'ARN Pol II dépend principalement du recrutement de P-TEFb sur les régions promotrices des gènes. Depuis la découverte de son rôle clé dans cette étape de la transcription par l'ARN Pol II, la liste des facteurs permettant de favoriser son recrutement au niveau de la machinerie transcriptionnelle ne cesse de s'agrandir. Il apparait maintenant que ces facteurs peuvent avoir des fonctions redondantes et/ou coopératives (Huang et al., 2009; Lu et al., 2016b; Stewart et al., 2013).

- *Association de P-TEFb avec des facteurs de transcription spécifiques :*

De nombreux de facteurs de transcription gènes-spécifiques interagissent avec P-TEFb et promeuvent son recrutement sur leur gènes cibles, tels que NF-kB (Barboric et al., 2001), le facteur de réponse à l'hypoxie HIF1α (Galbraith et al., 2013), le transactivateur de la classe II du complexe majeur d'histocompatibilité CIITA (Kanazawa et al., 2000), le récepteur aux androgènes (Lee et al., 2001), le récepteurs aux hydrocarbures aryliques (Tian et al., 2003), c-myc (Gargano et al., 2007; Kanazawa et al., 2003; Rahl et al., 2010), le facteur de différenciation musculaire MyoD (Giacinti et al., 2006), STAT3 (Giraud et al., 2004), le facteur de transcription des cellules musculaires MEF2 (Nojima et al., 2008), le facteur d'activation des cellules épithéliales thymiques médullaires AIRE (Huang et al., 2018; Oven et al., 2007; Zumer et al., 2011), Sox2 et Sox10 (Arter and Wegner, 2015). Ainsi, ces interactions permettent le ciblage de P-TEFb sur des gènes spécifiques pour activer leur expression en réponse à diverses voies de signalisation cellulaires.

- *Recrutement de P-TEFb au sein du Super Elongation Complex :*

Des études biochimiques ont montré que P-TEFb faisait également partie de larges complexes nucléaires. Ainsi, une fraction de P-TEFb réside au sein du Super Elongation Complex (SEC) (Luo et al., 2012b). En plus de P-TEFb, le SEC est composé de différentes combinaisons de facteurs d'élongation

de la famille ELL (eleven-nineteen Lys-rich leukemia) (ELL1-3), de protéines de la famille AFF (AF4/FMR2) (AFF1-AFF4), de protéines à domaine YEATS (ENL ou AF9), et d'un facteur associé à ELL (EAF1 ou EAF2) (Luo et al., 2012a). Au sein du SEC, les protéines de la famille AFF peuvent s'homo- ou hétéro-dimériser et jouent le rôle de plateforme pour le recrutement des autres composants du SEC et notamment de P-TEFb (He et al., 2011; He et al., 2010; Lu et al., 2014; Luo et al., 2012a). Plusieurs sous-complexes de SEC peuvent se former autour des protéines AFF4/1, AFF2 et AFF3 respectivement (figure 7a). La composition du SEC pourrait alors spécifier son recrutement sur les gènes en permettant son association avec des facteurs distincts. Le sous-complexe SEC-AFF4/1 est par exemple requis pour l'expression des gènes de réponse précoce (IEGs, Immediate Early Genes) (Luo et al., 2012a).

Les gènes codant les protéines du SEC sont fréquemment retrouvés fusionnés au gène de la protéine de liaison à l'ADN MLL (mixed lineage leukemia) dans des cas de cancers du sang pédiatriques (Lin et al., 2010). Ces protéines chimères permettent alors de cibler le SEC sur les gènes cibles de la protéine MLL et d'en activer l'expression. Dans un contexte cellulaire sans translocation avec la protéine MLL, le SEC peut être recruté au niveau des promoteurs de différents groupes de gènes via des complexes multi-protéiques impliqués dans l'initiation, facilitant certainement la transition entre initiation et élongation productive. Ainsi, la sous-unité Med26 du complexe Mediator, spécifique des métazoaires, peut interagir via son domaine N-terminal avec le facteur d'initiation TFIID ou la protéine ELL pour recruter le SEC et activer l'expression de ses gènes cibles (Lu et al., 2016b; Takahashi et al., 2011). TFIID contacte également les sous-unités AF9 et EAF1 du SEC, indépendamment de Med26 (Yadav et al., 2019). Cette association semble requise pour l'expression de gènes impliqués dans la prolifération cellulaire, bien que l'analyse des gènes affectés par la déplétion de sous-unité de TFIID soit compliquée au vu de son implication dans l'étape d'initiation. Le complexe multirégulateur PAF1 est aussi impliqué dans le recrutement du SEC, via la sous-unité PAF1 qui contacte les domaines YEATS de ENL/AF9 (He et al., 2011; Sobhian et al., 2010). Une étude récente a également montré un recrutement coopératif et complexe de SEC de compositions différentes par PAF1C et Mediator (Lu et al., 2016b). D'un côté, PAF1C favorise le recrutement du SEC-AFF1/ENL/AF9 pour la phosphorylation de NELF-A, et du SEC-AFF4/ENL/AF9 pour la phosphorylation du CTD de l'ARN Pol II. De l'autre, Med26 stimule le recrutement du SEC-AFF4/AFF1/ENL/AF9 pour la phosphorylation de NELF-E. Ce réseau de recrutement de P-TEFb semble s'appliquer aux gènes de réponse précoce, dont l'activation après stimulation en dépend. En plus de son rôle dans la biogénèse des snARN (Baillat et al., 2005), le complexe Integrator est requis pour l'activation des gènes de réponse au facteur de croissance EGF (epidermal growth factor) en stimulant l'initiation de la transcription et l'entrée en élongation (Gardini et al., 2014). L'Integrator interagit avec les sous-unités Cdk9 et AFF4 du SEC et est nécessaire pour son recrutement sur les régions promotrices des gènes activés par l'EGF après induction. Divers modes de

recrutement du SEC ont donc été mis en évidence (figure 7b). Bien que la composition du SEC ne soit pas toujours claire, l'ensemble de ces données s'accordent sur l'interaction spécifique de certains composants du SEC avec des facteurs qui permettent de le recruter sur les gènes. Identifier les déterminants requis pour l'interaction du SEC avec ses divers facteurs de recrutement et les différents groupes de gènes qu'il régule par ce moyen reste un défi majeur.

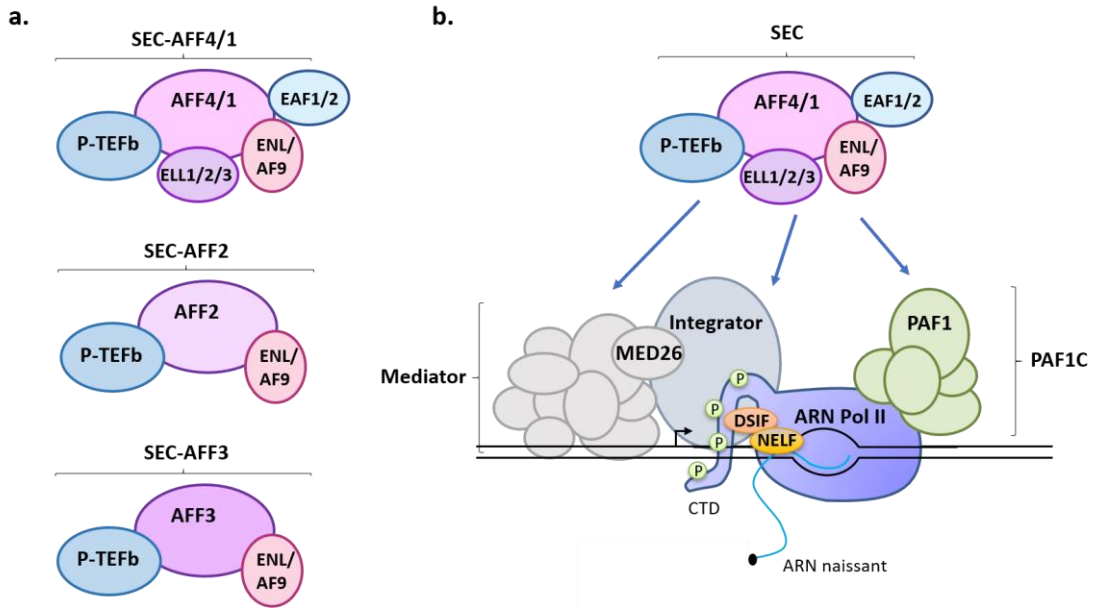


Figure 7 : Composition et recrutement du Super Elongation Complex (SEC). (a) Les protéines de la famille AFF permettent la formation de différents sous-complexes (Luo 2012). (b) Le SEC-AFF4/1 peut être recruté au niveau de l'ARN Pol II en pause via la sous-unité Med26 du complexe Mediator, via le complexe Integrator et via le complexe PAF1 (PAF1C).

- *Recrutement de P-TEFb par Brd4 :*

Brd4 également été identifié comme un partenaire majeur de P-TEFb (Jang et al., 2005; Yang et al., 2005). Brd4 est une protéine abondante appartenant à la famille des protéines à domaine BET caractérisées par la présence de bromodomaines en tandem du côté N-terminal et d'un domaine extraterminal (ET) leur permettant d'interagir avec les queues acétylées des histones. Via son domaine C-terminal, Brd4 peut interagir avec la cyclin T1 (Bisgrove et al., 2007; Jang et al., 2005; Yang et al., 2005) et cibler P-TEFb au niveau de nombreux promoteurs viraux et cellulaires présentant des histones H3 et H4 acétylés (Bisgrove et al., 2007; Dey et al., 2003). Notamment, Brd4 permet de recruter P-TEFb sur les régions promotrices de certains gènes pour leur activation transcriptionnelle après la mitose en reconnaissant les marques d'histones acétylés persistantes sur la chromatine mitotique (Behera et al., 2019; Dey et al., 2009; Mochizuki et al., 2008; Yang et al., 2008). Brd4 est également nécessaire pour l'activation P-TEFb dépendante des gènes de réponse à l'interféron après stimulation (Patel et al., 2013). Cependant, la coopération entre Brd4 et P-TEFb pour promouvoir l'élongation de la

transcription reste méconnue. Il a en effet été montré récemment que la dégradation ciblée de Brd4 par le protéasome affectait la transition de la pause de l'ARN Pol II vers l'élongation productive, mais sans affecter la localisation chromatinienne de P-TEFb (Winter et al., 2017). Après stimulation à l'acide rétinoïque, le SEC et Brd4 sont co-recrutés sur les gènes de réponse, mais la déplétion de Brd4 par siRNA n'affecte pas le recrutement du SEC (Lin et al., 2011). Brd4 pourrait alors fonctionner en stimulant l'activité de P-TEFb plutôt qu'en le recrutant sur les régions promotrices, ou en activant l'élongation transcriptionnelle par un mécanisme indépendant de P-TEFb (Itzen et al., 2014; Muhar et al., 2018; Winter et al., 2017).

- *Recrutement de P-TEFb sous sa forme inactive*

Une fraction de P-TEFb, dont la taille dépend du type cellulaire, réside dans la particule ribonucléoprotéique (RNP) 7SK, composée de l'ARN 7SK et des protéines MePCE, Larp7 et HEXIM. Dans ce complexe, P-TEFb est maintenu inactif et la dissociation de la RNP 7SK en réponse à divers stimuli représente un mécanisme majeur pour réguler la disponibilité de P-TEFb. La RNP 7SK/P-TEFb a récemment été détectée au niveau des promoteurs de certains gènes, suggérant un nouveau mode de recrutement de P-TEFb. Ce mécanisme sera détaillé dans le chapitre suivant.

g. Stabilité du complexe d'élongation en pause à proximité du promoteur

La détection d'une accumulation de l'ARN Pol II juste en aval du TSS de la plupart des gènes signifie un temps de résidence relativement long dans cette région comparé au temps passé par la polymérase dans d'autres régions. Cependant, les techniques classiques permettant de détecter la pause de l'ARN Pol II (tableau 1) n'apportent pas d'indications temporelles absolues. Pour déterminer la durée de la pause, plusieurs études se sont basées sur l'utilisation du triptolide (trp), un composé qui inhibe l'initiation de la transcription en bloquant l'activité ATPase de XPB du facteur TFIIH (Titov et al., 2011). En empêchant l'entrée de nouvelles ARN Pol II en pause, il devient alors possible de mesurer le temps nécessaire pour la disparition des complexes en pause par détection de l'ARN Pol II (ChIP-seq) ou de la transcription naissante (GRO-seq, start-seq, voir tableau 1). Ces études concordantes ont révélé une grande variabilité dans la stabilité des complexes d'élongation en pause en fonction des gènes, de quelques minutes à plus d'1h pour certains gènes avec une moyenne de 5-15 min (Chen et al., 2015a; Henriques et al., 2013; Jonkers et al., 2014; Shao and Zeitlinger, 2017). Les variations observées sont cohérentes au vu des différences de niveau d'expression entre gènes : de manière générale, la durée de la pause doit être courte pour les gènes fortement transcrits, et plus longue pour les gènes à faible niveau d'expression. Cependant, l'utilisation du trp semble induire une surestimation de la stabilité des complexes d'élongation en pause, causée par une inhibition de l'initiation incomplète (Nilson et

al., 2017). Ainsi, l'analyse de la mobilité de l'ARN Pol II par microscopie à fluorescence et photoblanchiement (FRAP, fluorescence recovery after photobleaching) a mis en évidence une durée de pause moyenne beaucoup plus courte, de 42 s, due à un important renouvellement des complexes d'élongation en pause à proximité du promoteur (Steurer et al., 2018). Cette étude suggère en effet que seulement 10 % des ARN Pol II en pause progressent dans le corps des gènes, et que la grande majorité subit une terminaison précoce. Ces résultats ne sont pas en contradiction avec une pause stable, étant donné que la durée mesurée reste largement supérieure (>10 fois) au temps estimé pour l'étape d'initiation (Darzacq et al., 2007; Steurer et al., 2018). De plus, une pause plus courte (42 s contre 5-15 min) mais fréquemment renouvelée pourrait participer au maintien d'un contexte chromatinien favorable aux promoteurs, et notamment de la marque des promoteurs actifs H3K4me3 dont la demi-vie est de 6.8 h (Price, 2018; Zheng et al., 2014b). Ces différentes études renforcent donc l'idée que la pause à proximité du promoteur est une étape clé du processus de transcription, et que le passage en élongation productive est fortement régulé. Le niveau de pause pour un gène est conditionné par le taux d'initiation, et les données récentes suggèrent que le taux de renouvellement par terminaison précoce prend une part importante dans la régulation de l'entrée en élongation productive qu'il sera intéressant d'étudier plus amplement.

h. L'élongation productive

Après la levée de la pause transcriptionnelle, l'élongation productive commence. De nombreuses études ont permis d'établir une vitesse moyenne de transcription par l'ARN Pol II de 2,5 kb/min, mais le taux d'élongation est loin d'être constant le long du gène, et est également très variable d'un gène à l'autre (Fitz et al., 2018; Fuchs et al., 2014; Jonkers et al., 2014). En effet, l'ARN Pol II s'accélère graduellement après la libération de la pause (0,5 kb/min sur le premier kb) et atteint sa vitesse maximale (2-5 kb/min) à environ 15 kb du TSS, conservant ce taux jusqu'à la fin du gène (Danko et al., 2013; Fitz et al., 2018; Jonkers et al., 2014). Ces différences de taux d'élongation reflètent certainement l'assemblage progressif d'un complexe d'élongation efficace avec le recrutement et la modification séquentiels des composants de la machinerie transcriptionnelle. Comme énoncé précédemment, les phosphorylations du CTD de l'ARN Pol II sur les serines 2 et de la sous-unité Spt5 de DSIF posent les bases du recrutement de nombreux facteurs qui assistent l'ARN Pol II durant l'étape d'élongation, en augmentant notamment sa processivité. Le facteur Spt5 lui-même est un acteur clé de l'élongation productive (Fitz et al., 2018). Il voyage avec l'ARN Pol II jusqu'à l'extrémité 3' du gène et améliore sa processivité. En se liant d'un part au canal de sortie de l'ARN sur l'ARN Pol II, et d'autre part au canal de sortie de l'ADN, DSIF facilite la réhybridation de l'ADN après le passage de l'ARN Pol II (Bernecky et al., 2017; Ehara et al., 2017; Vos et al., 2018a).

- *La progression de l'ARN Pol II à travers la chromatine*

L'élongation productive affecte et est affecté par la structure chromatinienne. Une structure chromatinienne permissive à l'élongation requiert des facteurs permettant le désassemblage des nucléosomes pour permettre le passage de l'ARN Pol II, et leur réassemblage dans son sillage afin d'éviter des initiations cryptiques et des dommages à l'ADN. Le passage de l'ARN Pol II provoque la libération d'un dimère d'histones H2A-H2B, laissant une particule sous-nucléosomale hexamèrique composée d'un tétramère d'histones H3-H4 et du dimère d'histones H2A-H2B conservé (Kireeva et al., 2002). Ce processus est notamment facilité par les complexes chaperonnes d'histones FACT (facilitates chromatin transcription) et Spt6. FACT interagit directement avec le dimère H2A-H2B à exclure et permet son réassemblage en aval du complexe d'élongation (Hondele et al., 2013; Hsieh et al., 2013; Kemble et al., 2015; Mason and Struhl, 2003; Orphanides et al., 1998). Le rôle de Spt6 est moins bien caractérisé, mais il participe également au réassemblage des nucléosomes pendant la transcription, probablement via son interaction avec les histones H3-H4, et sa déplétion impacte négativement le taux d'élongation par l'ARN Pol II (Andrulis et al., 2000; Ardehali et al., 2009; Ivanovska et al., 2011; Kaplan et al., 2003; Kaplan et al., 2000). Les complexes de remodelage des nucléosomes, tels que Chd1, permettant de déplacer et d'exclure les nucléosomes de manière ATP dépendante, jouent également un rôle important dans le passage de l'ARN Pol II à travers la chromatine (Lee et al., 2012; Skene et al., 2014). Enfin, la mono-ubiquitinylation de l'histone H2B (H2Bub1) a émergé comme un acteur clé pour moduler la barrière chromatinienne. Déposée par les ubiquitines-ligase RNF20/40 et MSL, cette marque d'histones permet notamment de stimuler l'activité des complexes FACT et Chd1, potentiellement en stabilisant leur liaison au complexe d'élongation (Lee et al., 2012; Pavri et al., 2006; Zheng et al., 2014a), et est requise pour l'ajout de méthylations sur les lysines 4 et 79 de l'histone 3, marquant les promoteurs et corps de gènes activement transcrits, respectivement (Batta et al., 2011; Kim et al., 2009; Wu et al., 2011). Le complexe PAF1 semble être requis pour la mise en place de l'ubiquitinylation de l'histone H3 en contactant directement RNF20/40 et MSL et stabilisant leur liaison au complexe d'élongation (Pavri et al., 2006; Wu et al., 2014). H2Bub1 semble aussi nécessiter la phosphorylation du CTD de l'ARN Pol II sur les serines 2 par Cdk9 pour être mis en place (Pirngruber et al., 2009), et H2Bub1 favorise en retour le recrutement de P-TEFb sur les régions promotrices, révélant une boucle de régulation positive entre ces différents facteurs (Wu et al., 2014)

- *Ralentissements de l'ARN Pol II au cours de l'élongation :*

La rencontre de l'ARN Pol II avec un nucléosome ou un facteur de liaison à l'ADN, ou bien lorsqu'une mauvaise base est incorporée, provoque une interruption de la synthèse de l'ARN et la marche arrière du complexe d'élongation (James et al., 2017; Teves et al., 2014). La reprise de la transcription

nécessite alors le facteur TFIIS, qui diminue la durée de ces pauses en stimulant l'activité intrinsèque de l'ARN Pol II pour cliver l'ARN naissant et réaligner son extrémité 3' dans le site actif (Adelman et al., 2005; Izban and Luse, 1992; Reines et al., 1989; Sheridan et al., 2019). Le facteur TFIIF et le complexe PAF1 participent aussi à ce processus, en potentialisant l'action de TFIIS et via des mécanismes indépendants (Ishibashi et al., 2014; Kim et al., 2010b; Luse et al., 2011; Schweikhard et al., 2014).

RECQL5 est une protéine de la famille des hélicases RECQ qui régule négativement le taux d'élongation en s'associant directement à l'ARN Pol II (Aygun et al., 2008; Aygun and Svejstrup, 2010; Aygun et al., 2009; Saponaro et al., 2014). Ainsi, sa déplétion induit une augmentation globale du taux d'élongation dans les cellules HEK293T, mais également une augmentation de délétions ou réarrangements chromosomiques trouvant leur origine dans les régions transcrites (Saponaro et al., 2014). Ces données suggèrent donc que le contrôle du taux d'élongation est non seulement nécessaire pour contrôler la production d'ARN, mais aussi pour préserver l'intégrité génomique.

De plus, le taux d'élongation influe directement et est contrôlé par les processus de maturation de l'ARN co-transcriptionnels tels que l'épissage. Ainsi, l'ARN Pol II est ralentie par la présence d'exons, qui possèdent un plus fort taux en GC comparé à celui des introns (Amit et al., 2012), et une plus forte occupation en nucléosomes (Schwartz et al., 2009; Tilgner et al., 2009). Ces ralentissements sont proposés favoriser la reconnaissance du site d'épissage et éviter ainsi le saut d'exons (de la Mata et al., 2003; Ip et al., 2011; Nogues et al., 2003).

iii. Terminaison

La terminaison de la transcription par l'ARN Pol II est un processus dynamique reposant sur la libération de l'ARN naissant et de l'ARN Pol II de la matrice ADN (Proudfoot, 2016). Ainsi, à la fin du cycle de terminaison, l'ARN naissant subit une maturation terminale consistant à son clivage et à l'ajout à l'extrémité 3' nouvellement formée d'une queue poly(A). L'ARN Pol II continue à transcrire la matrice après le site de terminaison, correspondant à l'extrémité 3' de l'ARN mature, avant de subir des réarrangements conformationnels déclenchant son relargage. La terminaison de la transcription est essentiel pour le contrôle de l'expression génique (Porrua and Libri, 2015; West and Proudfoot, 2009). Un défaut de terminaison à la fin d'un gène, en plus d'affecter son expression, peut interférer avec la transcription du gène en aval (Hobson et al., 2012; Shearwin et al., 2005). Un échec dans le décrochage de l'ARN Pol II peut également impacter la réplication, à travers la collision entre les machineries de transcription et de réplication, menaçant alors l'intégrité du génome (Gaillard et al., 2015). La terminaison peut également avoir lieu précocement durant la transcription, avant que l'ARN Pol II n'atteigne la fin du gène. Deux types de terminaison précoces peuvent être distinguées : (1) l'élongation abortive a lieu lorsque l'ARN Pol II est en pause, notamment à proximité du promoteur

(Brannan et al., 2012). Ainsi, si l'ARN Pol II en pause ne reçoit pas le signal pour entrer en élongation productive, elle subira une terminaison prématurée impliquant la dégradation de l'ARN naissant mais l'intégrité de l'ARN Pol II est préservée ; (2) la terminaison non productive a lieu lorsque l'ARN Pol II est arrêtée, par exemple lorsqu'elle rencontre un dommage à l'ADN (Wilson et al., 2013). L'ARN Pol II est alors dégradée par le protéasome pour permettre la réparation de la lésion et de nouveaux cycles de transcription.

a. Clivage et polyadénylation de l'ARN

La maturation de l'ARN en 3' repose sur la reconnaissance du signal de polyadénylation (PAS, Poly(A) signal) sur l'ARN naissant, alors qu'il émerge du complexe d'élongation, par le facteur spécifique de clivage et de polyadénylation (CPSF, cleavage and polyadenylation specific factor) et le facteur stimulateur du clivage (CstF) (Connelly and Manley, 1988; Proudfoot, 1989, 2011). CPSF et CstF sont des complexes multiprotéiques (figure 8). Chaque sous-unité de ces complexes participe aux réactions de clivage et de polyadénylation de l'ARN en permettant notamment la liaison à l'ARN et en stimulant les activités enzymatiques (Mandel et al., 2008). Ces deux complexes sont également assistés par les complexes accessoires CFI et CFII (cleavage factor I et II) (figure 8). La sous-unité CPSF73 de CPSF possède une activité endonucléase responsable du clivage du transcrit, qui a lieu 10-30 nt en aval du PAS (Legendre and Gautheret, 2003; Mandel et al., 2006; Proudfoot, 1991; Xiang et al., 2014). La poly(A) polymérase commence alors la synthèse de la queue poly(A) à l'extrémité 3'OH de l'ARN. La longueur de la queue poly(A), de 50-100 nt en moyenne dans les cellules humaines, peut varier d'un gène à l'autre et joue un rôle important dans la stabilité du transcrit (Chang et al., 2014). Elle est contrôlée par la liaison des protéines PABP (Poly(A) binding protein) (Kuhn et al., 2009), qui assurent également d'autres processus post-transcriptionnels (Bear et al., 2003; Bresson and Conrad, 2013).

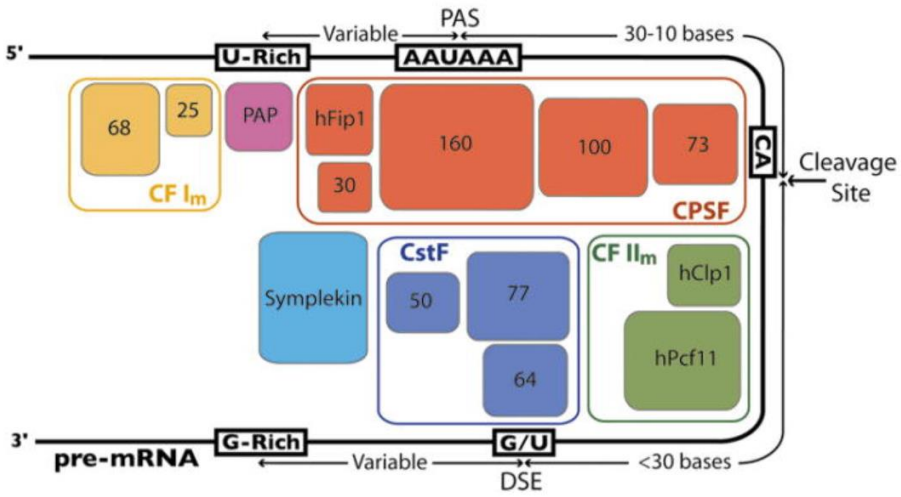


Figure 8 : La machinerie de maturation en 3' des ARNm (Mandel et al., 2008). Le facteur spécifique de clivage et de polyadénylation (CPSF) reconnait le signal de polyadénylation (PAS). Le facteur stimulateur de clivage (CstF) reconnait une séquence en aval du PAS (DES, downstream sequence element) riche en G/U. Le clivage du pré-ARNm a lieu 10-30 nt en aval du PAS, généralement après le dinucléotide CA. Une séquence riche en U (U-rich) est retrouvée en amont du PAS, tandis que la région en aval du DES est fréquemment riche en G (G-rich). Les facteurs de clivage I et II (CFI et CFII) participent à la reconnaissance du PAS et au recrutement des facteurs de clivage. Symplekin promeut l'interaction entre CstF et CPSF. PAP (poly(A)-polymerase) catalyse l'ajout de la queue poly(A) à l'extrémité 3' du transcrit après le clivage.

Cette maturation se fait co-transcriptionnellement, et les interactions entre la machinerie de clivage et de poly(A) et le complexe d'élongation semblent jouer un rôle important dans ce processus. L'interaction entre CPSF/Cstf et l'ARN naissant est facilité par son recrutement au préalable par le CTD de l'ARN Pol II phosphorylé sur les serines 2, qui atteint d'ailleurs son maximum de phosphorylation à la fin du gène (Hsin and Manley, 2012; McCracken et al., 1997).

b. Décrochage et recyclage de l'ARN Pol II

Le décrochage de l'ARN Pol II de la matrice a lieu à une distance variable pouvant aller jusqu'à 10 kb du site de terminaison, et résulte d'une compétition cinétique entre l'élongation et la dégradation de l'ARN qui continue à être produit par l'ARN Pol II par l'exonucléase 5'-3' Xrn2 (figure 9) (Eaton et al., 2018; Fong et al., 2015; Lian et al., 2008; Proudfoot, 2016). Alors qu'elle passe le PAS, l'ARN Pol II est ralentie par l'interaction entre les facteurs de clivage et le transcrit naissant (Nag et al., 2007; Nojima et al., 2015). De plus, le transcrit naissant forme fréquemment des R-loop en s'hybridant au brin d'ADN non matrice en amont du complexe d'élongation dans la région 3' du gène, ralentissant davantage l'ARN Pol II (Schwalb et al., 2016; Skourti-Stathaki et al., 2014; Skourti-Stathaki and Proudfoot, 2014). La diminution du taux d'élongation permet à Xrn2 de rattraper l'ARN Pol II, et de provoquer son décrochage selon le modèle appelé « torpedo » (Proudfoot, 2016; West et al., 2004).

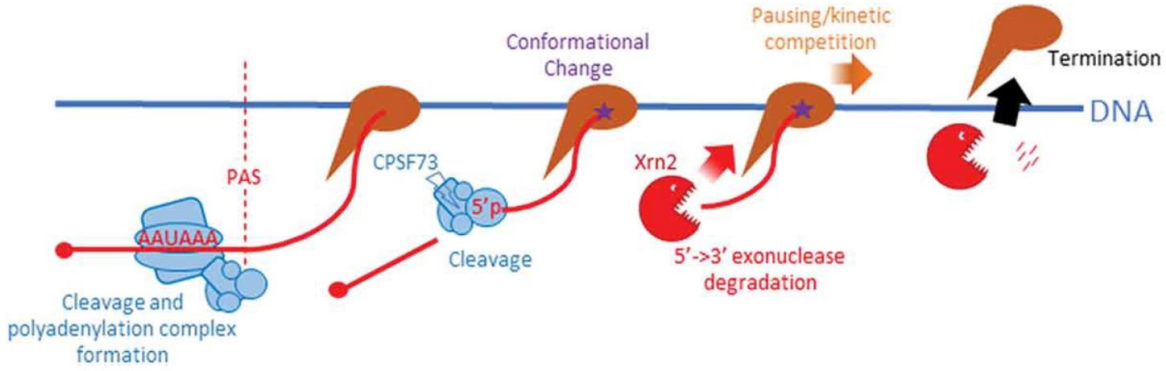


Figure 9 : Clivage de l'ARNm et terminaison de la transcription dans les cellules humaines (Eaton and West, 2018). L'ARN Pol II transcrit le signal de polyadénylation (PAS), reconnu sur l'ARNm par la machinerie de clivage et de polyadénylation. L'ARNm est clivé par l'endonucléase CPSF73, générant une extrémité 5'OH sur l'ARN toujours produit par l'ARN Pol II qui est reconnue par l'exonucléase 5'-3' Xrn2. Après le clivage, l'ARN Pol II est ralentie (étoile). Xrn2, qui dégrade l'ARN, entre alors en collision avec l'ARN Pol II et provoque son décrochage.

c. Elongation abortive à l'étape de pause à proximité du promoteur

Comme mentionné dans le chapitre sur la pause de la transcription, l'ARN Pol II réside plus ou moins longtemps en pause à proximité du promoteur. La stabilité de l'ARN Pol II en pause dépend de son relargage dans le corps du gène par P-TEFb, mais également du taux de terminaison précoce. Deux mécanismes de terminaison précoce au promoteur, ou élongation abortive, ont été proposés (Brannan et al., 2012; Wagschal et al., 2012). Le premier nécessite l'intervention d'une enzyme pour enlever la coiffe du transcrit naissant suivie de l'intervention de l'exonucléase 5'-3' Xrn2 pour le dégrader et décrocher l'ARN Pol II. Dans de modèle, l'activité de Xrn2 serait stimulée par le facteur à activité ATPase TTF2, tous deux localisés au niveau du promoteur de nombreux gènes. Ainsi, la co-déplétion de ces deux facteurs entraîne une redistribution de l'ARN Pol II dans les corps de gènes. De plus, la présence de la sous-unité Gdown1, qui empêche la liaison de TTF2, dans les complexes d'élongation corrèle avec le niveau de pause (Cheng et al., 2012a). Le second modèle repose sur la reconnaissance de structures particulières en tige-boucle formées par l'ARN naissant par le microprocessor, un complexe composé de la RNase Drosha et de la sous-unité de liaison à l'ARN Dgcr8. Drosha pourrait donc cliver le transcrit naissant, ce qui permettrait à Xrn2 d'initier la terminaison précoce. Ce mécanisme a été identifié au niveau du promoteur du gène intégré du VIH-1, mais pourrait également s'appliquer à des gènes cellulaires (Macias et al., 2012).

d. Terminaison prématurée causée par l'arrêt de la polymérase

Un mécanisme de terminaison prématurée « de dernier recours » a été mis en évidence lorsque la transcription par l'ARN Pol II est arrêtée de manière irréversible, et est notamment observé après

induction de dommages à l'ADN (Caron et al., 2019; Wilson et al., 2013). L'ARN Pol II bloquée en élongation est alors mono- puis poly-ubiquitinylée, délogée par le complexe Cdc48-Ubx4, et adressée au protéasome pour sa dégradation (Ratner et al., 1998; Verma et al., 2011). Différentes ubiquitine-ligases ont été impliquées dans ce processus et semblent dépendre de la cause de l'arrêt de l'ARN Pol II : WWP2 ubiquitinyle l'ARN Pol II au niveau des cassures double-brins de l'ADN (Caron et al., 2019), tandis que NDD4 et le complexe ElonginA/B/C-Cul5-Rbx2 mono- et poly-ubiquitinylent respectivement l'ARN Pol II en réponse aux dommages à l'ADN causés par les UV (Anindya et al., 2007; Harreman et al., 2009).

iv. Le CTD de l'ARN Pol II

Les trois ARN Pol eucaryotes sont très similaires structurellement (Cramer et al., 2008). Cependant, l'ARN Pol II se distingue de ses consœurs par la présence d'un domaine carboxy-terminal (CTD) unique non structuré sur la sous-unité Rpb1. Il est constitué de plusieurs répétitions de la séquence consensus Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (figure 10A). Le nombre de répétitions est variable selon les organismes : le CTD compte 26 répétitions chez la levure, tandis qu'il comporte 52 répétitions chez les mammifères (Chapman et al., 2008; Yang and Stiller, 2014). Cependant, sur les 52 répétitions, 21 sont très conservées et les 31 dernières divergent légèrement de la séquence consensus (figure 10A) (Harlen and Churchman, 2017). Tous les résidus qui composent l'unité de base du CTD peuvent être modifiés post-traductionnellement. Le CTD est notamment dynamiquement modifié au cours du cycle de transcription par phosphorylation, glycosylation, isomérisation, acétylation et méthylation (figure 10B) (Egloff and Murphy, 2008; Heidemann et al., 2013). La combinaison de ces différentes modifications établit le « code du CTD » et constitue une plateforme de recrutement pour de nombreux facteurs. Le CTD participe ainsi à la coordination des différentes étapes de la transcription par l'ARN Pol II, à la maturation co-transcriptionnelle des ARN et aux modifications de la chromatine (Heidemann et al., 2013; Kim et al., 2010a). Il a récemment été montré que le CTD joue également un rôle clé dans la localisation et la dynamique des ARN Pol II dans le noyau : la formation de régions nucléaires hautement actives transcriptionnellement, ou usines transcriptionnelles, est permis par l'interaction de plusieurs ARN Pol II entre elles via leur CTD (Boehning et al., 2018). Ces regroupements d'ARN Pol II à proximité de promoteurs permettraient ainsi d'augmenter le taux d'initiation lors de l'activation transcriptionnelle.

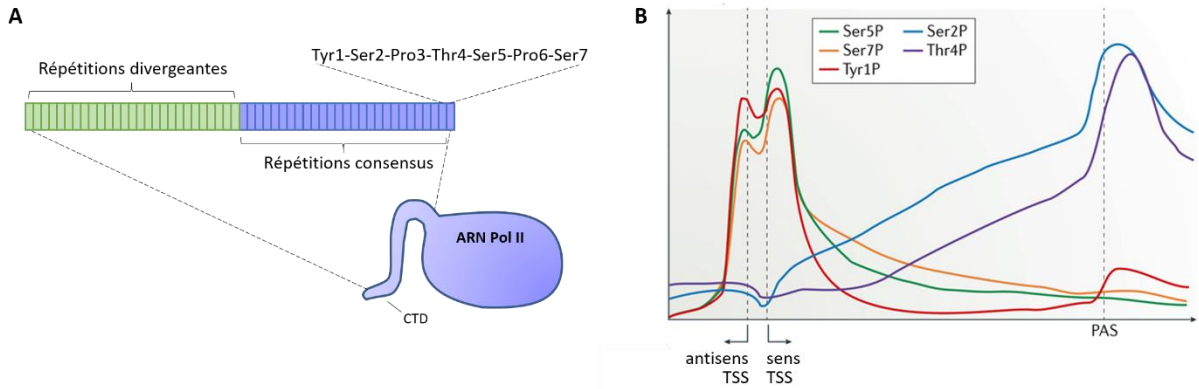


Figure 10 : Modification du CTD de l'ARN Pol II. (A) Organisation du CTD. (B) Profil des principales modifications du CTD de l'ARN Pol II sur les gènes chez les mammifères (Harlen and Churchman, 2017). TSS : transcription start site. PAS : poly(A) signal.

a. Les phosphorylations du CTD

Les phosphorylations des Ser5 et Ser2 sont les modifications majeures du CTD chez l'homme et la levure (Schuller et al., 2016; Suh et al., 2016). Le CTD de l'ARN Pol II recrutée au promoteur des gènes est non-phosphorylé (Lu et al., 1991). Au cours du processus d'initiation, la kinase cdk7 du complexe TFIIH phosphoryle la serine 5 (Ser5P), avant d'être relarguée du complexe d'initiation après ouverture de la double hélice (Compe et al., 2019). La Ser5P participe à la libération de l'ARN Pol II du promoteur en diminuant notamment son affinité pour le complexe Mediator (Robinson et al., 2016; Wong et al., 2014). Elle est importante pour le recrutement des enzymes qui ajoutent co-transcriptionnellement la coiffe à l'extrémité 5' des ARN naissants, et pourrait recruter le complexe d'épissage (Harlen et al., 2016; Nojima et al., 2015; Nojima et al., 2018). Chez la levure, la Ser5P est spécifiquement reconnue par la protéine Wdr82 qui permet le recrutement de la méthyltransférase H3K4 Set1 dans la région 5' des gènes (Lee and Skalnik, 2008; Ng et al., 2003). L'homologue humain de Set1, SETD1A/B reconnait également spécifiquement le CTD de l'ARN Pol II phosphorylé sur les Ser5 (Ebmeier et al., 2017; Fabrega et al., 2003; Ghosh et al., 2011; Nilson et al., 2015; Nojima et al., 2018; Schroeder et al., 2000). Cdk7 a également été impliquée dans l'établissement de la pause à proximité du promoteur via le recrutement de NELF (Glover-Cutter et al., 2009). Ainsi, la Ser5P caractérise l'ARN Pol II au niveau des régions promotrices, puis son niveau diminue progressivement durant l'élongation sous l'action des phosphatases RPAP2 et Ssu72 (figure 10B) (Krishnamurthy et al., 2004; Ni et al., 2014; Nojima et al., 2015; Rosado-Lugo and Hampsey, 2014).

La phosphorylation de la Ser2 est accomplie par différentes kinases, et est associée à l'élongation et à la terminaison de la transcription (Nojima et al., 2015). Ainsi, les niveaux de Ser2P augmentent le long du gène et diminuent après le site de poly(A), concourant avec le ralentissement de l'ARN Pol II lors de la terminaison (figure 10B) (Grosso et al., 2012; Hintermair et al., 2012; Nag et al., 2007; Rahl et al.,

2010). La kinase Cdk9 du complexe P-TEFb permet la phosphorylation de la Ser2 au début du gène, lorsqu'elle est recrutée pour libérer la pause de l'ARN Pol II à proximité du promoteur (voir I.ii.e. Libération de l'ARN Pol II en élongation productive). La phosphorylation de la Ser2 par Cdk9 est notamment importante pour le recrutement de la méthyltransférase SETD2, qui catalyse la tri-méthylation de la lysine 36 de l'histone H3 (H3K36me3) (Ebmeier et al., 2017), le facteur d'élongation Spt6 (Sun et al., 2010) et le complexe PAF1 (Vos et al., 2018a; Yu et al., 2015). Suite à la libération de la pause, les kinases Cdk12, et potentiellement Cdk13, relaient Cdk9 dans la phosphorylation de la Ser2 (Bartkowiak et al., 2010; Blazek et al., 2011; Cheng et al., 2012b; Yu et al., 2015). Cdk12 interagit directement avec plusieurs sous-unités du complexe PAF1 et la déplétion de PAF1 diminue grandement l'occupation de Cdk12 sur les gènes (Yu et al., 2015). Ces données suggèrent une coopération entre les Ser2P-kinases : Cdk9 stimule le recrutement du PAF1C qui lui-même participerait au recrutement de Cdk12. Brd4 semble également pouvoir phosphoryler la Ser2 du CTD, mais son implication dans cette fonction en coopération ou en substitution des autres Ser2P-kinases reste à découvrir (Devaiah et al., 2012). Durant l'élongation, la Ser2P joue un rôle important dans la maturation co-transcriptionnelle des ARN et la terminaison de la transcription à travers le recrutement des facteurs d'épissage U2AF65 et Prp19 et des facteurs de clivage et polyadénylation Pcf11, CPSF et CstF77 (David et al., 2011; Davidson et al., 2014; Gu et al., 2013).

Le CTD de l'ARN Pol II peut également être phosphorylé au niveau de la Tyrosine 1, la thréonine 4 et la Serine 7, mais ces modifications sont bien moins abondantes que les Ser5/2P (Schuller et al., 2016; Suh et al., 2016). Le profil de phosphorylation de la Tyrosine 1 (Tyr1P) au cours de la transcription est très différent chez la levure de celui de cellules de mammifères, suggérant des différences fonctionnelles (Hsin et al., 2014). Chez les mammifères, la Tyr1P est catalysée par la kinase c-Abl (Baskaran et al., 1997). Son niveau culmine aux sites de pause à proximité des promoteurs sens et antisens, et s'accumule également dans la région 3' des gènes, suivant le profil de l'ARN Pol II (figure 10B) (Descostes et al., 2014; Hsin et al., 2014). La Tyr1P s'avère requise pour la stabilité du CTD (Descostes et al., 2014; Hsin et al., 2014). La mutation de toutes les Tyr du CTD pour des résidus non phosphorylables est létale et engendre la dégradation de Rpb1 vers une forme dépourvue de CTD. La mutation partielle des Tyr du CTD a permis de mettre en évidence un rôle global de la Tyr1P dans la terminaison de la transcription. Sa diminution provoque une persistance de la transcription bien après le site de terminaison en 3' des gènes et une accumulation des PROMPTs, selon un mécanisme restant à définir (Baskaran et al., 1993; Shah et al., 2018; Yurko and Manley, 2018). La Tyr1P participe également à la maturation des snARN et des ARNm des histones, probablement en contribuant au recrutement du complexe Integrator. De plus, un rôle de la Tyr1P dans la pause de l'ARN Pol II à proximité du promoteur et au niveau des enhancers a été identifié et pourrait être en lien avec le

complexe Mediator, dont le recrutement dépend de cette modification du CTD (Descostes et al., 2014; Shah et al., 2018). Enfin, la Tyr1P a très récemment été impliquée dans la réponse aux dommages à l'ADN, dans laquelle elle permettrait la transcription aux sites de lésions, favorisant le recrutement de la machinerie de signalisation des dommages à l'ADN (Burger et al., 2019).

La phosphorylation de la thréonine 4 (Thr4P) par la kinase Plk3 est associée à l'étape d'élongation de la transcription, comme la Ser2P (Hintermair et al., 2012). Mais son profil est différent sur les gènes, avec un niveau maximal 500-2000 pb après le site de poly(A), ce qui suggère que son apparition suit celle de la Ser2P (figure 10B). Cette hypothèse est renforcée par la perte de la Thr4P après inhibition de Cdk9 ou la mutation des Ser2 par un acide-aminé non phosphorylable (Hintermair et al., 2012). De plus, la Thr4P joue un rôle spécialisé dans la transcription des gènes codant les histones dépendant de la réplication (Hsin et al., 2011). Le recrutement co-transcriptionnel du facteur SLBP (stem loop binding protein), requis pour la maturation de l'extrémité 3 ' des ARNm des histones, est en effet facilité par la Thr4P.

Cdk7 phosphoryle également la serine 7 (Ser7P) du CTD, mais à un niveau plus faible que la Ser5 (Akhtar et al., 2009; Glover-Cutter et al., 2009; Schuller et al., 2016). Chez les mammifères, la fonction majeure de la Ser7P semble être de favoriser la transcription et la maturation des sn et snoARN via le recrutement du complexe Integrator en 3' de leurs gènes en collaboration avec la Ser2P (Egloff et al., 2007; Egloff et al., 2010). De plus, la Ser7P est spécifiquement reconnu par la phosphatase RPAP2 sur les gènes des snARN, et sa mutation pour un résidu non phosphorylable, ou la déplétion de RPAP2 entraîne une phosphorylation persistante de la Ser5. En conséquence, le recrutement de l'Integrator et ainsi l'expression des snARN sont diminués (Egloff et al., 2010; Egloff et al., 2012). La Ser7P est également retrouvée sur les gènes codants les protéines selon un profile similaire à celui de la Ser5P (figure 10B). Sa fonction est encore peu connue, mais elle semble favoriser la phosphorylation des Ser2 par la kinase Cdk9 du complexe P-TEFb (Czudnochowski et al., 2012).

b. Les autres modifications du CTD

En plus d'être phosphorylées, les Ser et Thr au sein d'une répétition du CTD peuvent également être glycosylées. Cependant, cette modification, qui empêche la phosphorylation du même résidu, n'est retrouvée que chez les vertébrés (Comer and Hart, 2001). Les enzymes responsables de cette modification sont les N-acetylglycosamine transférases (O-GlcNAc transférases ou OGT) et N-acetylglycosamine aminidase (O-GlcNAc aminidase ou OGA) (Ranuncolo et al., 2012). Le rôle de la glycosylation du CTD est encore peu caractérisé, mais cette modification est retrouvée sur les Ser2, Ser5 et Ser7 du CTD au niveau des promoteurs. Notamment, la Ser5-OGlcNAc serait impliquée dans la

formation du PIC, avant d'être retirée par un mécanisme ATP-dépendant au cours de l'initiation (Lewis et al., 2016). Le niveau global de substrats glycosylés augmentant après un afflux en nutriment ou un stress cellulaire, la modification du CTD par glycosylation pourrait jouer un rôle important durant la régulation transcriptionnelle en réponse à des stimuli particuliers (Krause et al., 2018; Lu et al., 2016a).

L'heptade qui forme le CTD contient deux prolines (Pro) pouvant être isomérisées en conformation *cis* ou *trans* par la peptidyl–prolyl *cis/trans*-isomérase Pin1 (Heidemann et al., 2013). La conformation *trans* des prolines (*trans*-Pro) semble être la plus stable (Gibbs et al., 2017; Noble et al., 2005). Récemment, il a été montré par spectrométrie de masse et spectroscopie à résonnance magnétique nucléaire que la Ser5P du CTD de drosophile induit l'isomérisation en conformation *cis* de la Pro6 (Gibbs et al., 2017). Ce changement conformationnel s'accompagne d'un réarrangement structural du CTD qui participe à la spécificité de liaison de facteurs reconnaissant des résidus ou modifications particuliers du CTD, comme la phosphatase de la Ser5P Ssu72 (Gibbs et al., 2017; Werner-Allen et al., 2011). Chez la levure, la conformation *trans* de la Pro3 est impliquée en combinaison avec la Ser2P dans le recrutement de facteurs de terminaison (Meinhart and Cramer, 2004; Noble et al., 2005).

c. Modifications des répétitions non-consensus du CTD

Le CTD de Rpb1 chez les mammifères a accumulé une diversité de séquence non-consensus, en particulier dans la région C-terminale. Les résidus divergents de l'heptade originel peuvent également être modifiés post-traductionnellement, et l'étude du rôle spécifique de ces modifications dans la transcription par l'ARN Pol II est encore à ses balbutiements. Comme le CTD est très conservé chez les vertébrés, y compris au niveau des séquences non-consensus, les modifications des répétitions non-consensus pourraient jouer un rôle clé dans le développement. La substitution la plus fréquente a lieu à la position 7 de l'heptade, où la Ser7 est remplacée par une Lysine (Lys7) sur 8 répétitions (Simonti et al., 2015). La Lys7 peut être acétylée (Lys7Ac) par la lysine-acétyltransférase p300, et est requise pour l'activation spécifique de gènes de réponse précoce après stimulation (Schroder et al., 2013). De plus, des analyses par spectrométrie de masse ont révélées que la Lys7 pouvait également être mono-, di- ou tri-méthylée (Voss et al., 2015). La Ser7 peut également être remplacée par une Arginine a une position unique dans le CTD (Arg1810). Arg1810 est asymétriquement di-méthylée (Arg1810me2a) par la méthyltransférase CARM1 et contrôle spécifiquement l'expression des sn/snoARN (Sims et al., 2011). Arg1810 peut également être symétriquement di-méthylée (Arg1810me2s) par PRMT5 et promouvoir la terminaison de la transcription (Zhao et al., 2016a). Cette modification est en effet reconnue par la protéine SMN (survival of motor neuron), requise pour la résolution des R-loop générées dans la région 3' des gènes lors de la terminaison. Récemment, Arg1810 a été impliquée dans la régulation de la pause de l'ARN Pol II à proximité du promoteur, cette fois via sa citrullination

(Arg1810cit) par la peptidyl-arginine deiminase PADI2 (Sharma et al., 2019). La localisation de PADI2 sur les gènes est similaire à celle de l'ARN Pol II, en accord avec l'association de l'Arg1810cit à la forme phosphorylée de l'ARN Pol II. PADI2 est particulièrement enrichie sur les gènes fortement transcrits, sur lesquels elle faciliterait l'interaction entre P-TEFb et l'ARN Pol II.

II. Régulation de P-TEFb par la particule ribonucléoprotéique (RNP) 7SK

P-TEFb est un acteur central du cycle de transcription par l'ARN Pol II. En plus de contrôler l'entrée en élongation productive, l'activité de P-TEFb est critique pour le recrutement séquentiel de nombreux facteurs nécessaire pour qu'elle soit efficace. P-TEFb intervient également à l'étape de terminaison. De ce fait, l'activité de P-TEFb influence indirectement les processus de maturation co-transcriptionnels des ARN naissants. Comprendre comment P-TEFb est ciblé sur les gènes est donc primordial. Cependant, la multiplicité des partenaires de P-TEFb rend l'étude de son recrutement très complexe et nécessite d'être mise au clair. D'autant plus qu'au cours de ces dernières années, un nouveau modèle a émergé : la forme inactive de P-TEFb, au sein de la RNP 7SK, pourrait également être recrutée sur les gènes. Au cours de ma thèse, je me suis particulièrement intéressée à la régulation de l'activité de P-TEFb par la RNP 7SK.

L'ARN 7SK est un abondant petit ARN nucléaire (snARN), autour duquel différents sets de protéines s'assemblent, incluant de prime abord MePCE et Larp7, avec lesquelles il forme la particule cœur (RNP 7SK cœur). La fonction la plus détaillée de l'ARN 7SK est sans conteste la régulation de l'activité de P-TEFb, avec lequel il s'associe dynamiquement dans le noyau, élaborant ainsi un mécanisme inédit de régulation de l'activité d'une kinase (Nguyen et al., 2001; Yang et al., 2001). En association avec la RNP 7SK cœur et les protéines HEXIM1/2, P-TEFb est incapable d'assurer sa fonction d'activateur transcriptionnel décrite précédemment. Ainsi, la régulation dynamique de l'équilibre entre les formes actives et inactives de P-TEFb constitue un mécanisme majeur de contrôle de la disponibilité de P-TEFb et donc de la transcription par l'ARN Pol II. La dissociation de la RNP 7SK/P-TEFb peut être déclenchée par un certain nombre de signaux extracellulaires et en réponse à plusieurs voies de signalisation. Elle résulte en une augmentation temporaire de la disponibilité de P-TEFb, permettant d'adapter l'activité de P-TEFb au contexte cellulaire (Peterlin and Price, 2006; Zhou and Yik, 2006). De plus, une fraction non négligeable de la RNP 7SK/P-TEFb apparait se localiser à la chromatine, permettant de délivrer P-TEFb de manière régulée à proximité des gènes. Au cours de la dernière décennie, de nouveaux partenaires de l'ARN 7SK ont également été découverts, avec lesquels la RNP 7SK cœur s'associe et régule d'autres aspects de la transcription par l'ARN Pol II. Ces nouvelles RNP 7SK sont moins abondantes que la RNP 7SK/P-TEFb, aussi sont-elles encore peu caractérisées mais dévoilent le potentiel de régulation de ce petit ARN non-codant.

i. La RNP 7SK cœur

Les protéines MePCE (Methylphosphate capping enzyme) et Larp7 (La-related protein 7) lient et protègent les extrémités 5' et 3' de l'ARN 7SK, respectivement. Leur association à l'ARN 7SK s'avère indispensable pour sa stabilité. Larp7 et MePCE sont donc des partenaires constitutifs, ou chaperonnes de l'ARN 7SK. Ensemble, MePCE, Larp7 et l'ARN 7SK forment la RNP 7SK cœur et coopèrent pour recruter leurs multiples partenaires.

a. L'ARN 7SK

L'ARN 7SK est un ARN non-codant, court, 331 nt de long, très conservé chez les vertébrés mais des reliques sont aussi présentes chez les invertébrés (Diribarne and Bensaude, 2009; Yazbeck et al., 2018). Il est synthétisé par l'ARN Pol III, à partir d'un gène unique localisé sur le chromosome 6, bien que de nombreux pseudogènes auxquels il manque le promoteur et la région 3' existent dans le génome. Le promoteur du gène 7SK est situé exclusivement en amont du TSS, contrairement à la majorité des promoteurs Pol III. Il se compose d'un élément distal (DSE), d'un élément proximal (PSE) et d'une boîte TATA (Jawdekar and Henry, 2008; Murphy et al., 1987). Ces éléments permettent le recrutement des facteurs Oct-1, SNAPc (snRNA-activating complex) et TBP/TAF (TATA-binding protein/TBP-associated factor), respectivement. SNAPc et TBP sont absolument nécessaires au recrutement de l'ARN Pol III et à l'expression de l'ARN 7SK, tandis qu'Oct-1 permet d'augmenter le niveau de transcription en stabilisant la liaison de SNAPc. Comme tous les transcrits synthétisés par l'ARN Pol III, l'ARN 7SK se termine par une séquence riche en uridines (UUU-3'OH) à laquelle s'associe co-transcriptionnellement la protéine La, qui sera ensuite échangée avec la protéine Larp7 (figure 12) (Chambers and Dahmus, 1994; Maraia and Bayfield, 2006).

- *Expression de l'ARN 7SK*

L'ARN 7SK est très stable et fortement transcrit, ce qui en fait un des ARN cellulaires les plus abondants (Zieve and Penman, 1976; Zieve, 1981). Son expression est ubiquitaire mais des variations de ses niveaux cellulaires ont été observées dans certains organes et au cours de la différenciation et transformation cellulaires (Abasi et al., 2016; Bazi et al., 2018; Skreka et al., 2012). Les niveaux de l'ARN 7SK apparaissent de manière générale plus élevés dans les cellules différenciées que dans les cellules souches (Abasi et al., 2016). Le même constat s'applique en comparant des lignées non-cancéreuses (forte expression de l'ARN 7SK) et cancéreuses (faible expression) (Abasi et al., 2016). La surexpression de l'ARN 7SK dans des cellules cancéreuses humaines induit d'ailleurs une baisse de la viabilité cellulaire (Keramati et al., 2015). Ces données confortent le rôle de l'ARN 7SK dans la régulation de la prolifération cellulaire proposé à travers le contrôle de l'activité de P-TEFb (Peterlin and Price, 2006;

Zhou and Yik, 2006). La régulation de l'expression de l'ARN 7SK semble également jouer un rôle important dans le développement neuronal (Bazi et al., 2018; Santoro et al., 2016). L'expression de l'ARN 7SK est en effet plus élevée dans le cerveau adulte de souris que dans d'autres organes (Bazi et al., 2018). De plus, elle augmente au cours du développement neuronal chez la souris et cette augmentation s'avère cruciale pour induire la différenciation de cellules souches embryonnaires en cellules neuronales (Bazi et al., 2018; Skreka et al., 2012). En accord avec ces observations, des cellules de patients atteints de sclérose en plaque, une maladie touchant le système nerveux central, présentent des niveaux anormaux d'ARN 7SK comparés à des cellules saines (Santoro et al., 2016).

- *Structure de l'ARN 7SK :*

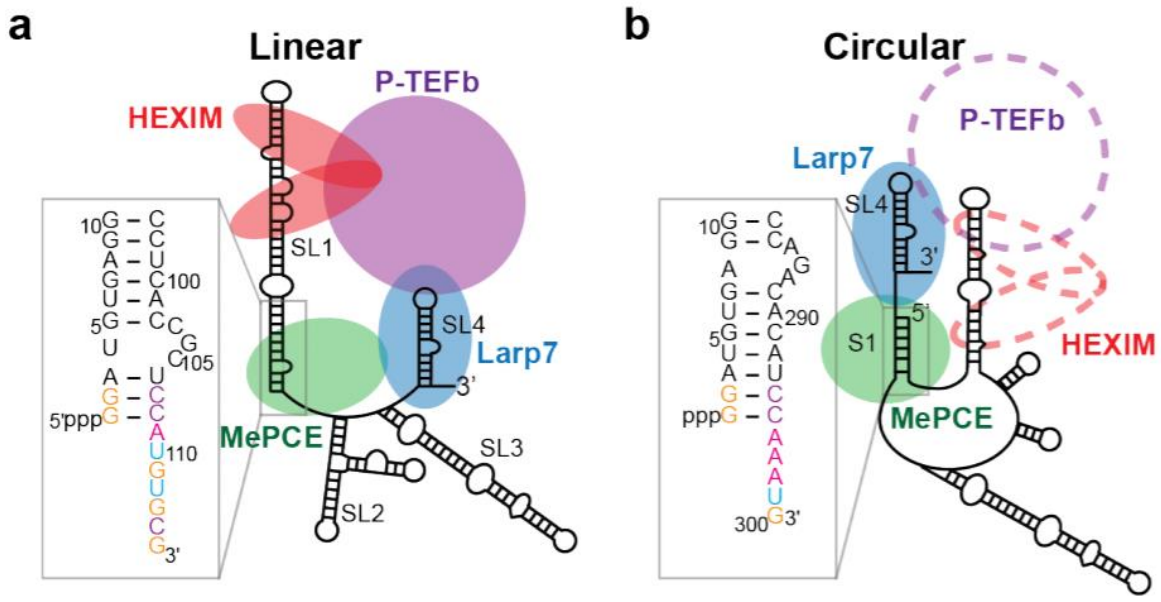


Figure 11 : Représentation schématique des deux structures secondaires proposées de l'ARN 7SK et liaison avec ses principaux partenaires protéiques (Yang et al., 2019). Ces deux structures diffèrent notamment au niveau de l'extrémité 5' (voir les encarts) (a) Structure secondaire « linéaire » déterminée par cartographie chimique et enzymatique (Wassarman and Steitz, 1991). Cette structure a été utilisée pour mettre en évidence les domaines de liaison à HEXIM et P-TEFb. (b) Structure secondaire prédite par analyse de la conservation des séquences au cours de l'évolution (Marz et al., 2009). La position hypothétique de P-TEFb et HEXIM dans cette conformation est indiquée en ligne pointillée.

La structure secondaire de l'ARN 7SK a initialement été établie grâce à des expériences de cartographie par modifications chimiques et clivages enzymatiques (Wassarman and Steitz, 1991). La structure déduite de ces expériences est composée de quatre motifs structuraux en tige-boucle séparés par de courtes régions simple-brin (figure 11a). La plupart des études réalisées sur l'ARN 7SK se sont basées sur cette structure pour établir ses domaines fonctionnels. Par analyse de la conservation phylogénétique de la séquence de l'ARN 7SK, une structure alternative a été proposée plus récemment (figure 11b) (Marz et al., 2009). Cette nouvelle structure, dite circulaire, suggère un rapprochement

des extrémités 5' et 3' par hybridation des séquences adjacentes, et une organisation plus structurée en huit tiges-boucles. Malgré ces différences, les motifs structuraux importants pour la liaison avec les principaux partenaires protéiques de l'ARN 7SK restent globalement conservés. En outre, la structure de l'ARN 7SK semble dynamique *in vivo* et dépendre de l'interaction avec ses différents partenaires (Brogie and Price, 2017; Krueger et al., 2010). Ainsi, la RNP 7SK cœur adopte la conformation circulaire, favorable à l'interaction Larp7-MePCE (figure 11b). Lorsque P-TEFb et HEXIM1/2 sont incorporés, l'ARN 7SK adopte une structure ouverte se rapprochant de celle proposée initialement par *Wassarman and Steitz,* permettant l'extension de la tige-boucle en 5' et la liaison d'HEXIM (figure 11a).

- *Modifications de l'ARN 7SK*

L'ARN 7SK est mono-méthylé sur le phosphate en position γ de la guanosine à son extrémité 5' (5'mpppG), une modification commune à peu d'ARN cellulaires, exclusivement non-codants et issus de la transcription Pol III, dont le snARN U6 (Singh and Reddy, 1989) ou l'ARN B2 chez les rongeurs (Gupta et al., 1990; Shumyatsky et al., 1990). L'ajout de cette coiffe est catalysé par la protéine MePCE, qui sera décrite dans la section suivante. Outre le 5'mpppG, d'autres nucléotides sont modifiés post-transcriptionnellement (Hussain et al., 2013b; Warda et al., 2017; Zhao et al., 2016b). Ainsi, la quasi-totalité des ARN 7SK apparaissent pseudo-uridylés à la position U250 par la RNP DKC1–box H/ACA–snoRA70. Etonnamment, alors que la U250 se trouve dans la troisième tige-boucle de l'ARN 7SK, sa conversion en pseudoU semble critique pour l'incorporation de P-TEFb. Il est possible que la pseudo-uridylation de la U250 soit nécessaire à la stabilité de la RNP 7SK/P-TEFb ou prévienne l'association de l'ARN 7SK avec des protéines incompatibles à la liaison de P-TEFb.
Des expériences de miCLIP, une version modifiée de l'iCLIP (individual-nucleotide-resolution crosslinking and immunoprecipitation) pour détecter les méthylations à l'échelle du transcriptome, ont identifié l'ARN 7SK comme cible potentielle de la C-5-méthyltransférase NSun2, qui pourrait donc le méthyler à la position C237 (Hussain et al., 2013b). De même, l'ARN 7SK a été retrouvé associé à la 6-A-methyltransférase METTL16 dans un crible visant à déterminer les cibles de cette enzyme (Warda et al., 2017). Comme cela a été montré pour la pseudo-uridylation de la U250, ces modifications pourraient jouer un rôle important dans la biogénèse des différentes RNP 7SK.

b. MePCE

MePCE est une méthyltransférase S-adénosylméthionine-dépendante qui catalyse l'ajout d'un groupement méthyl sur le phosphate en position gamma de la guanosine à l'extrémité 5' des ARN 7SK et U6 (Jeronimo et al., 2007; Yang et al., 2019). MePCE est détectée par ChIP sur leurs gènes respectifs et semble ainsi coiffer ces ARN co-transcriptionnellement (Xue et al., 2010), similairement à l'ajout de

la coiffe en 5’ des ARN messagers. En plus de « coiffer » l’extrémité 5’ de l’ARN 7SK, MePCE y reste stablement associée en interagissant avec l’extrémité 5’ et la base de la première tige-boucle de l’ARN 7SK (Muniz et al., 2013). Il a d’ailleurs été très récemment montré que l’affinité de la liaison de MePCE pour l’ARN 7SK augmentait avec l’ajout de la coiffe, expliquant la persistance de la liaison entre les deux après la réaction enzymatique (Yang et al., 2019).

La liaison de MePCE est nécessaire à la stabilité de l’ARN 7SK et sa déplétion par siARN provoque la co-déplétion de l’ARN 7SK (Barboric et al., 2009; Jeronimo et al., 2007; Xue et al., 2010). MePCE reste probablement associé à l’ARN U6 jusqu’à la formation de la tr-snRNP U4/U5/U6, mais curieusement sa déplétion n’affecte pas les niveaux de l’ARN U6 (Jeronimo et al., 2007; Yang et al., 2019).

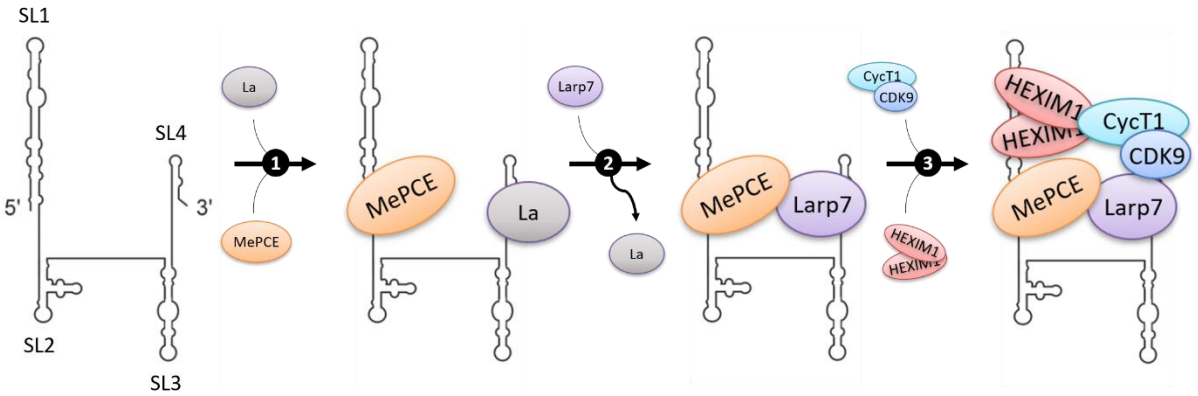


Figure 12 : Biogénèse de la RNP 7SK cœur et de la RNP 7SK/P-TEFb (AJ Quaresma 2016). SL: stem-loop. (1) MePCE s'associe cotranscriptionnellement à l'ARN 7SK au niveau de la SL1 (5') et coiffe son extrémité 5’. La protéine La se lie à l’extrémité 3' de l’ARN 7SK. (2) la protéine La est rapidement remplacée par Larp7, qui s’associe à la SL4 en plus de l’extrémité (UUU 3-OH’) et avec MePCE. (3) Un dimère de protéines HEXIM s’associe au niveau de deux motifs distincts dans les régions distale et proximale de la SL1, exposant son domaine de liaison à P-TEFb. La liaison de HEXIM à P-TEFb inhibe l’activité kinase de Cdk9.

c. Larp7

Comme mentionné lors de la description de l’ARN 7SK, la protéine La reconnait et se lie à l’extrémité UUU-3’OH de l’ARN 7SK, mais cette liaison est transitoire : la protéine La est rapidement remplacée par la protéine Larp7 (figure 12) (Bayfield et al., 2010). Comme MePCE, Larp7 est nécessaire à l’intégrité de l’ARN 7SK et sa déplétion par siARN réduit profondément les niveaux de l’ARN 7SK (he 2008, krueger 2008). Larp7 possède un motif La (LAM, La motif) dans sa région N-terminale, très similaire à celui de la protéine La, qui lui permet de reconnaitre l’extrémité UUU-3’OH de l’ARN 7SK (figure 13) (Muniz et al., 2013). Larp7 contacte aussi la tige-boucle en 3’ de l’ARN 7SK via ses domaines de liaison à l’ARN RRM1 et RRM2 (RNA recognition motif, figure 13) (Eichhorn et al., 2018; Uchikawa et al., 2015). Ces différents points d’ancrage résultent en une interaction stable et spécifique entre Larp7 et l’ARN 7SK (Brogie and Price, 2017; Muniz et al., 2013; Uchikawa et al., 2015). Alors que le LAM

de Larp7 peut être remplacé par celui de la protéine La sans impacter son association à l'ARN 7SK, les RRM de Larp7 s'avèrent critiques pour la formation de la RNP 7SK (He et al., 2008).

Au sein de la RNP 7SK, Larp7 interagit également avec MePCE via son domaine C-terminal, renforçant ainsi la stabilité de la RNP cœur (figure 12) (Brogie and Price, 2017; Muniz et al., 2013; Xue et al., 2010). Au vu des propriétés de chaperonne de Larp7 communes à toutes les protéines de la famille Larp (Hussain et al., 2013a), il a été proposé que Larp7 provoque un réarrangement structurel de l'ARN 7SK pour contacter MePCE par rapprochement des extrémités 5' et 3' (figure 11b) (Xue et al., 2010; Yang et al., 2019). De plus, la liaison de Larp7 à MePCE bloque l'activité catalytique de MePCE, bien que la fonction de cette inhibition ne soit pas connue et puisse simplement résulter de l'encombrement stérique généré (Brogie and Price, 2017; Xue et al., 2010).

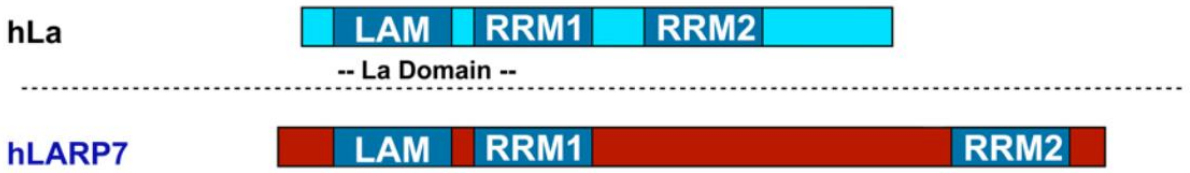


Figure 13 : domaines structuraux des protéines La et Larp7. LAM : La Motif ; RRM : RNA Recognition Motif (Bayfield et al., 2010).

Des perturbations de l'expression de Larp7 sont associées à de nombreuses pathologies. Des cellules de cancers du sein, de la thyroïde, gastriques, et cervicaux présentent une diminution de l'expression de Larp7, d'autant plus basse que le stade est avancé (Biewenga et al., 2008; Cheng et al., 2012c; Ji et al., 2014; Sui et al., 2018). La déplétion de Larp7 dans des cellules non-tumorales suffit d'ailleurs à leur transformation (Cheng et al., 2012c; He et al., 2008; Ji et al., 2014). Dans certains cas, une augmentation de l'activité de P-TEFb a également été détectée, reliant la déplétion de Larp7 à la perte du mécanisme de contrôle de P-TEFb par la RNP 7SK (He et al., 2008; Ji et al., 2014). Des mutations dans le gène de Larp7 résultant en la perte d'expression ou l'expression de protéines tronquées ont également été associés à diverses pathologies neurodéveloppementales comme le nanisme primordial, le syndrome Alazami et la microcéphalie (Alazami et al., 2012; Hollink et al., 2016; Holohan et al., 2016; Imbert-Bouteille et al., 2019; Ling and Sorrentino, 2016; Najmabadi et al., 2011; Wojcik et al., 2019).

ii. La RNP 7SK/P-TEFb

Au début du XXIè siècle, l'ARN 7SK, cet abondant transcrit Pol III de fonction encore inconnue, a été identifié comme se liant à P-TEFb et inhibant son activité kinase (Nguyen et al., 2001; Yang et al., 2001). Suite à cette révolution moléculaire, il a été découvert que l'ARN 7SK servait en réalité de charpente

pour l'interaction inactivatrice de P-TEFb avec la protéine hexamethylene bisacetamide inducible protein 1 (HEXIM1) ou son paralogue HEXIM2 (Michels et al., 2003). Dans les cellules, une fraction de P-TEFb est séquestrée par la RNP 7SK en coopération avec HEXIM1/2. L'autre fraction est active et suffit aux besoins transcriptionnels basiques de la cellule. L'interaction entre P-TEFb, HEXIM1/2 et la RNP 7SK cœur est réversible, permettant d'adapter le niveau de P-TEFb actif en fonction de la demande transcriptionnelle (figure 14). La fraction de P-TEFb associée à la RNP 7SK varie ainsi en fonction du type cellulaire : alors qu'environ la moitié des complexes P-TEFb sont séquestrés dans les cellules HeLa hautement prolifératives, cette fraction atteint presque 90 % dans les lymphocytes T stationnaires (Kim et al., 2011; Nguyen et al., 2001; Yang et al., 2001). Certains signaux intra- et extracellulaires dont de nombreux stress peuvent également affecter ce ratio (Peterlin and Price, 2006; Zhou and Yik, 2006). La RNP 7SK/P-TEFb est ainsi vue comme un réservoir de P-TEFb pouvant être rapidement mobilisé par la cellule (Nguyen et al., 2001; Peterlin and Price, 2006; Yang et al., 2001; Zhou and Yik, 2006).

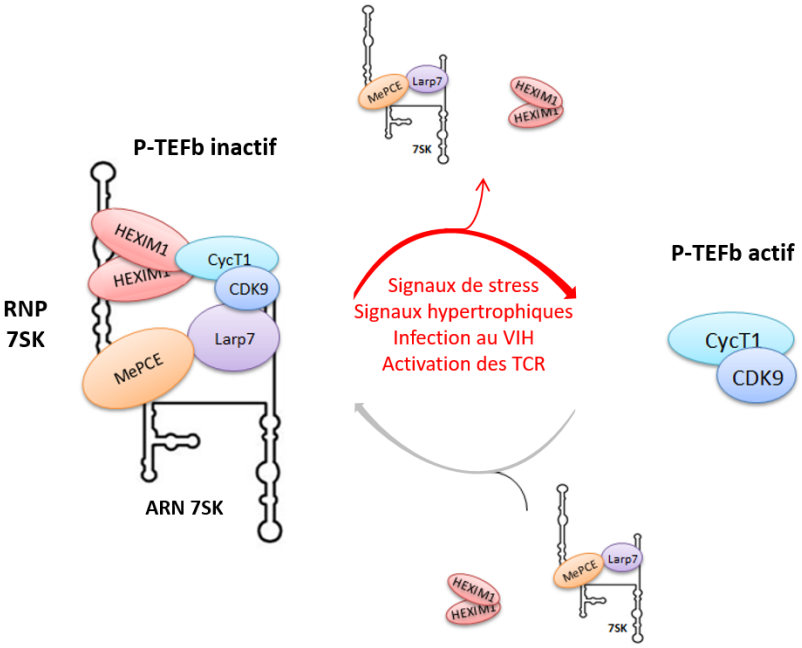


Figure 14 : Equilibre dynamique entre les formes actives et inactives de P-TEFb. Les formes actives (droite) et inactives (gauche) de P-TEFb coexistent dans les cellules. P-TEFb est incorporé au sein de la RNP 7SK cœur avec un dimère de protéines HEXIM. La dissociation de la RNP 7SK en réponse à divers stimuli repose sur la libération de P-TEFb et d'HEXIM tandis que l'intégrité de la RNP 7SK cœur est préservée.

a. Le complexe P-TEFb

P-TEFb est un hétérodimère composé d'une sous-unité kinase dépendante des cyclines, Cdk9, et de Cycline T1 ou T2. Comme les autres Cdk impliquées dans la transcription par l'ARN Pol II, Cdk7 et Cdk8, les concentrations de Cdk9, Cycline T1 et T2 ne changent pas de façon significative durant le cycle cellulaire et ces protéines sont ubiquitairement exprimées (Garriga et al., 1998). De plus, la liaison de

la cycline induit un changement conformationnel de Cdk9 nécessaire à son activation et à sa stabilité (Majello and Napolitano, 2001). Comme décrit dans le chapitre précédent, l'activité kinase de P-TEFb est requise au cours du cycle de transcription par l'ARN Pol II pour stimuler la libération de la pause à proximité du promoteur. Cette étape s'avère être un point de contrôle important permettant une fine régulation de l'expression des gènes. De ce fait, la disponibilité nucléaire de P-TEFb est elle-même scrupuleusement régulée par son association avec la RNP 7SK dans laquelle son activité kinase est inhibée par un dimère de protéines HEXIM. L'étude des liaisons entre les différents composants de la RNP 7SK/P-TEFb menée dans le laboratoire a montré que les tiges-boucles en 5' et 3' de l'ARN 7SK étaient toutes les deux requises pour son association avec P-TEFb *in vivo* (Egloff et al., 2006; Muniz et al., 2010). Cependant, il semblerait que P-TEFb n'interagisse pas directement avec l'ARN 7SK, mais soit stablement associé à la RNP à travers la liaison de HEXIM à Cycline T1/T2 et à Cdk9, et via une interaction entre le domaine C-terminal de Larp7 et Cdk9 (figure 14) (Markert et al., 2008; Muniz et al., 2013).

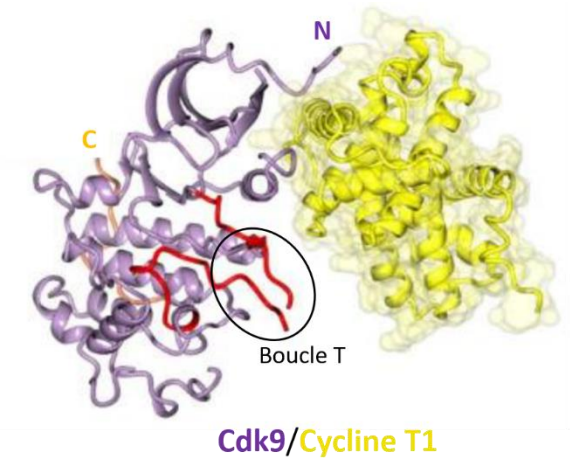


Figure 15 : Représentation en ruban de la structure du complexe P-TEFb (Wood and Endicott, 2018). Cdk9 (violet) ; Cycline T1 (jaune) ; orange : segment carboxy-terminal de Cdk9 ; rouge : segment activateur

- *Cdk9 :*

Cdk9 contient un domaine de liaison à l'ATP et une structure en boucle T, dont la phosphorylation, sur le résidu Thr186, est nécessaire pour l'accès du substrat et de l'ATP au cœur catalytique de Cdk9, et donc pour son activité kinase (figure 15) (Baumli et al., 2008). La composante kinase du facteur d'initiation TFIIH, Cdk7, s'avère requise pour la phosphorylation de la Thr186, mettant en évidence une coopération entre les deux kinases pour promouvoir l'élongation de la transcription par l'ARN Pol II (Larochelle et al., 2012; Mbonye et al., 2018). La déplétion de la kinase calmoduline-dépendante 1D (CaMK1D) par siRNA diminue la phosphorylation de Cdk9 sur la Thr186, suggérant que CaMK1D pourrait également participer à l'activation de Cdk9 (Ramakrishnan and Rice, 2012). La phosphorylation de la Thr186 de Cdk9 est aussi un prérequis pour l'incorporation de P-TEFb dans la RNP 7SK cœur (Chen et al., 2004; Li et al., 2005). Ainsi, c'est la forme active de P-TEFb qui est séquestrée au sein de la RNP 7SK/P-TEFb, directement capable d'activer l'élongation de la transcription

suite à sa libération. Cependant, la déphosphorylation de la Thr186 semble également être un mécanisme pour extraire P-TEFb de la RNP 7SK, avec la potentielle contrainte de devoir la re-phosphoryler pour que la Cdk9 soit active (Chen et al., 2008; Wang et al., 2008). Ce mécanisme sera plus amplement détaillé dans la partie suivante.

Deux isoformes de Cdk9 existent (de 42 et 55 kDa), exprimées à partir de deux promoteurs différents du même gène, et diffèrent donc par leur région N-terminale (Shore et al., 2003). Les deux isoformes sont exprimées dans différentes lignées cellulaires et se localisent majoritairement dans le noyau (Liu and Herrmann, 2005). L'isoforme courte est prépondérante et retrouvée dans le nucléoplasme, tandis que l'isoforme longue semble nucléolaire. Contrairement à l'isoforme courte, l'isoforme longue de Cdk9 a très peu été étudiée. En plus des substrats précédemment cités NELF, DSIF et le CTD de l'ARN Pol II, une étude a récemment identifié de nombreuses cibles potentielles de Cdk9 (Sanso et al., 2016). La majorité des facteurs retrouvés dans ce crible participent à la transcription et au métabolisme des ARN, et parmi eux figurent MePCE, Larp7, Brd4, un composant du SEC, AFF1, et le facteur de terminaison Xrn2. Ces cibles devront être vérifiées, mais il est intriguant de retrouver MePCE et Larp7, alors que la seule interaction connue entre ces protéines et Cdk9 se fait au sein de la RNP 7SK/P-TEFb. La phosphorylation de Xrn2 a été analysée plus en détail et validée, et permet de stimuler son activité exonucléase, indiquant un rôle de Cdk9 au-delà de la libération de la pause à proximité du promoteur.

- *Cyclines T1/T2 :*

Comme toutes les kinases dépendantes des cyclines (CDK, cyclin dependent kinase), la stabilité de Cdk9 est régulée par son association à une cycline (Chiu et al., 2004). Trois cyclines différentes peuvent s'associer à Cdk9 et former des complexes P-TEFb kinase-compétents : Cycline T1 et deux isoformes de Cycline T2, T2a et T2b, bien que Cycline T1 soit le partenaire majoritaire de Cdk9 (Tahirov et al., 2010) (Peng et al., 1998). Les trois cyclines partagent une forte homologie (> 80 %), et divergent principalement au niveau de leur domaine C-terminal (figure 16) (De Luca et al., 2003). La région N-terminale des Cycline T contient deux boîtes cycline requises pour leur liaison à Cdk9 et à HEXIM (figure 16). La sous-unité cycline est impliquée dans la liaison de P-TEFb à de nombreux partenaires, et certains semblent avoir une préférence de liaison pour Cycline T1 ou Cycline T2. L'association de Cdk9 à ses différentes cyclines permet donc de conférer à P-TEFb des fonctions spécifiques et non-redondantes (De Luca et al., 2003; Kohoutek et al., 2009; Ramakrishnan et al., 2011; Simone et al., 2002). La délétion du gène codant la cycline T2 s'avère par exemple embryonnaire-létal chez la souris (Kohoutek et al., 2009). Dans les cellules HEK293, seule la fraction de P-TEFb associée à Cycline T2 (T2a et T2b) co-précipite avec la protéine du Super Elongation Complex ENL (Mueller et al., 2009; Wimmer et al., 1999). Cycline T2 interagit également spécifiquement avec le facteur transcription MyoD impliqué dans la

différentiation des cellules musculaires (Simone et al., 2002). Cycline T1 est requise pour la transcription du gène intégré du VIH (virus de l'immunodéficience humaine) de type 1. La protéine virale Tat interagit en effet spécifiquement avec Cycline T1 en reconnaissant un domaine unique à Cycline T1 (TRM, figure 16). Ce domaine est également impliqué dans la liaison entre Cycline T1 et HEXIM, et permet à Tat d'extraire P-TEFb de la RNP 7SK (voir II.iii.b. Extraction de P-TEFb de la RNP 7SK)

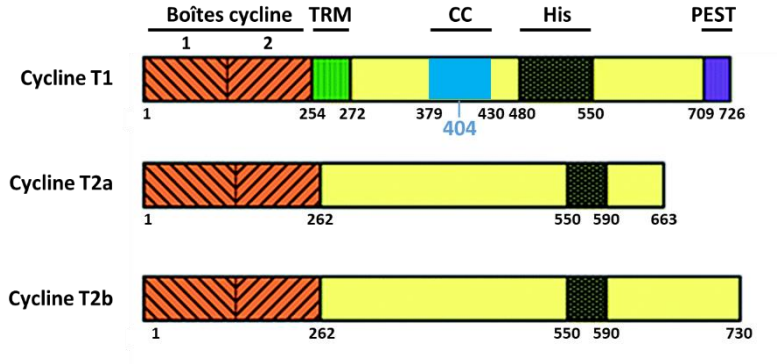


Figure 16 : Représentation schématique des domaines fonctionnels des cyclines T1, T2a et T2b (Paparidis et al., 2017).Les cycline T possèdent deux boîtes cycline à leur extrémité N-terminale requises pour l'interaction avec Cdk9 et HEXIM, et un domaine riche en histidine (His) qui favorise l'interaction de P-TEFb avec le CTD de l'ARN Pol II (taube 2002, lu 2018). Cycline T1 possède spécifiquement un motif de reconnaissance de la protéine Tat (TRM,) également impliqué dans son interaction avec HEXIM, un domaine coiled-coil, et un domaine PEST cible d'ubiquitinylation. Le résidu sur Cycline T1 modifié post-traductionnellement par acétylation est indiqué en bleu (voir II.iii.a. Voies de signalisation et modifications post-traductionnelles de la RNP 7SK/P-TEFb).

b. Les protéines HEXIM

- *HEXIM et la RNP 7SK*

HEXIM1 inhibe l'activité catalytique de P-TEFb de façon strictement dépendante de l'ARN 7SK (Barboric et al., 2005; Michels et al., 2003). Des analyses mutationnelles réalisées dans notre équipe et par d'autres laboratoires, ainsi que des analyses structurales ont permis de définir précisément les éléments de l'ARN 7SK requis pour la liaison de HEXIM1 (Belanger et al., 2009; Egloff et al., 2006; Lebars et al., 2010; Martinez-Zapien et al., 2016; Martinez-Zapien et al., 2015; Muniz et al., 2010). Ainsi, HEXIM1 s'associe sous forme de dimère à l'ARN 7SK au niveau de deux motifs distincts dans les régions distale et proximale de la première tige-boucle (5'), de manière interdépendante : la mutation d'un de ces deux motifs abolit la liaison de HEXIM à l'ARN 7SK. Notamment, deux motifs GAUC organisés en double brins dans la partie distale de la tige-boucle en 5' apparaissent critiques pour la liaison d'HEXIM et sont impliqués dans le remodelage de l'ARN 7SK suite à la libération de HEXIM et P-TEFb (Brogie and Price, 2017; Lebars et al., 2010; Martinez-Zapien et al., 2017; Martinez-Zapien et al., 2016; Martinez-Zapien et al., 2015; Muniz et al., 2010). La partie centrale de la protéine HEXIM1, composée d'une

région basique (BR) suivie d'une région acide (AR), est impliquée dans la liaison à l'ARN 7SK et dans l'inhibition de P-TEFb (figure 17) (Barboric et al., 2005). En dehors de la RNP 7SK, les BR et AR de HEXIM forment des liaisons électrostatiques provocant un rapprochement des régions N- et C-terminales. En revanche, la liaison de HEXIM à l'ARN 7SK induit un changement conformationnel de la protéine qui dévoile un motif PYNT entre la BR et l'AR et permet l'interaction avec P-TEFb (figure 17). Ce motif masque le site catalytique de Cdk9 et l'accès au substrat et est ainsi responsable de son inhibition (Kobbi et al., 2016; Krueger et al., 2010). La région C-terminale de HEXIM contient un domaine « coiled-coil » par lequel les protéines HEXIM se dimérisent et qui lui permet de contacter les Cyclines T au niveau de leur boîtes cycline (figure 16, 17) (Blazek et al., 2005; Dames et al., 2007; Kuzmina et al., 2014; Michels et al., 2003; Schulte et al., 2005).

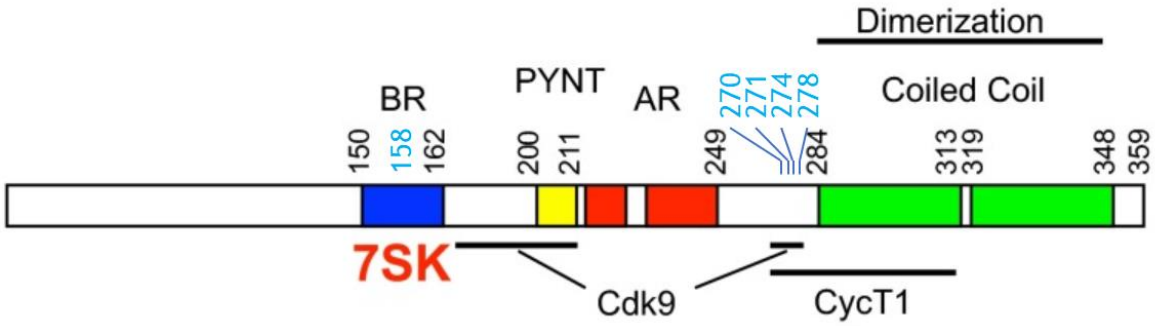


Figure 17 : Représentation schématique des domaines fonctionnels de HEXIM1 (Kobbi et al., 2016). En l'absence de l'ARN 7SK, les régions basiques (BR, boîte bleue) et acides (AR, boîtes rouges) interagissent ensemble, masquant les sites d'interaction à Cdk9 et à Cycline T1 (CycT1). La liaison à l'ARN 7SK via la BR induit un changement conformationnel permettant l'inhibition de P-TEFb. La dimérisation des protéines Hexim1 se fait via le domaine Coiled Coil dans la région C-terminale, qui chevauche le domaine de liaison à CycT1. Les résidus modifiés post-traductionnellement par phosphorylation sont indiqués en bleu (voir II.iii.a. Voies de signalisation et modifications post-traductionnelles de la RNP 7SK/P-TEFb).

Les protéines HEXIM1 et HEXIM2 présentent une forte homologie et diffèrent principalement via leur domaine N-terminal. Cependant, toutes les deux peuvent être incorporées dans la RNP 7SK sous forme d'homo- ou hétéro-dimères, interagir avec les cyclines T1/T2 et inhiber l'activité de P-TEFb (Byers et al., 2005; Yik et al., 2005). Les deux protéines sont ubiquitaires, mais les niveaux cellulaires de HEXIM1 sont généralement plus élevés que ceux de HEXIM2, et l'incorporation de HEXIM1 au sein de la RNP 7SK semble être favorisée (Byers et al., 2005; Yik et al., 2005). HEXIM1 présente une affinité de liaison pour Cycline T1 plus forte que pour Cycline T2, tandis que HEXIM2 s'associe préférentiellement à Cycline T2, en accord avec la prépondérance de la RNP 7SK/P-TEFb contenant Cycline T1 et HEXIM1 (Czudnochowski et al., 2010). Cependant, lorsque les niveaux de HEXIM1 sont réduits par siARN, les niveaux de HEXIM2 associés à la RNP 7SK/P-TEFb augmentent de façon à maintenir l'équilibre entre les formes actives et inactives de P-TEFb. Ainsi, HEXIM2 semble pouvoir compenser la perte de HEXIM1 au sein de la RNP 7SK/P-TEFb, mettant en évidence l'importance du contrôle de l'activité de P-TEFb.

HEXIM2 en revanche apparait ne pas pouvoir se substituer à HEXIM1 au cours du développement, la délétion du gène HEXIM1 chez la souris étant embryonnaire-létal (Huang et al., 2004). Ces données révèlent de potentielles fonctions spécifiques de HEXIM1 qui pourraient être indépendantes de la RNP 7SK.

- *Autres partenaires de HEXIM1 :*

La majorité des protéines HEXIM sont retrouvées en dehors de la RNP 7SK/P-TEFb. HEXIM1 s'associe au récepteur aux glucocorticoïdes (GR) indépendamment de l'ARN 7SK et inhibe sa fonction d'activateur transcriptionnel (Shimizu et al., 2005; Shimizu et al., 2008; Yeh et al., 2014; Yoshikawa et al., 2008). HEXIM1 s'associe également spécifiquement à la protéine p53 et la stabilise en masquant son site d'ubiquitinylation ciblé par la protéine human double minute-2 (HDM2) (Lew et al., 2012). Cette interaction augmente par traitement au DRB, au Flavoripidol ou aux UV, qui induisent également la dissociation de la RNP 7SK/P-TEFb. Cela suggère que HEXIM1 libéré de la RNP 7SK s'associe avec p53. Curieusement, HDM2 ubiquitinyle également HEXIM1 au niveau de sa BR (figure 17), entraînant non pas sa dégradation mais une augmentation de l'inhibition de P-TEFb (Lau et al., 2009). Nucleophosmin (NPM), une autre protéine impliquée dans le métabolisme de p53, s'associe à HEXIM1 et induit sa dégradation par le protéasome (Gurumurthy et al., 2008). Une mutation de NPM empêchant sa translocation dans le noyau, fréquemment retrouvée dans les leucémies myéloïdes aigues, entraîne la séquestration d'HEXIM1 dans le cytoplasme. Cependant, un rôle physiologique de l'interaction entre HEXIM1 et NPM n'a pas été rapporté. HEXIM1 s'associe via sa BR (figure 17) au long ARN non-codant NEAT1 et forme une RNP avec le complexe DNA-PK et plusieurs protéines des paraspeckles (Morchikh et al., 2017). L'ARN NEAT1 permet d'ancrer cette RNP aux paraspeckles, des sous-domaines nucléaires impliqués dans la réponse au stress via la régulation de la rétention nucléaire des ARNm (Nakagawa and Hirose, 2012). La RNP HEXIM1/NEAT1 est requise pour la réponse immunitaire à l'ADN exogène, lors de laquelle elle est libérée des paraspeckles et s'associe au facteur de transcription STING pour l'activation de ses gènes cibles (Morchikh et al., 2017). Les ARN NEAT1 et 7SK lient la même région de HEXIM1, et l'association de P-TEFb et des composants de la RNP HEXIM1/NEAT1 est mutuellement exclusive. Ainsi, HEXIM1 fait partie de deux RNP distinctes, toutes deux impliquées dans la réponse au stress. HEXIM1 peut également lier et stabiliser des ARNm pro-apoptotiques en réponse au stress nucléotidique (Tan et al., 2016). Curieusement, la privation en nucléotides conduit en parallèle à une augmentation de la séquestration de P-TEFb par HEXIM1 au sein de la RNP 7SK, contrairement à ce qui est observé en réponse à d'autres stress inhibant la transcription.

c. Pathologies associées à P-TEFb

Sachant son rôle capital dans la régulation de la transcription par l'ARN Pol II, il n'est pas étonnant que P-TEFb soit impliqué dans de nombreuses pathologies (Franco et al., 2018). P-TEFb a d'abord été identifié comme un acteur clé de l'expression du génome intégré du VIH. La transcription à partir du promoteur LTR (long terminal repeat) du VIH est en effet pris en charge par la machinerie de transcription de la cellule hôte du virus, et s'est avérée être un modèle de choix pour l'étude du mécanisme de pause de l'ARN Pol II à proximité du promoteur (Wang et al., 2009; Yankulov and Bentley, 1998). L'ARN Pol II en pause au promoteur du VIH est associée à un ARN naissant formant une structure en tige-boucle particulière appelée TAR (trans-activation response element), reconnue par la protéine virale Tat. Tat recrute P-TEFb au niveau de l'ARN TAR, étape essentielle pour la production de transcrits viraux pleine taille (Zhu et al., 1997). L'expression du VIH est particulièrement sensible aux quantités de P-TEFb actif (Flores et al., 1999). L'établissement de la latence virale implique d'ailleurs une restriction de l'activité de P-TEFb couplée à la mise en place de marques de chromatine inactive au promoteur du VIH (Tyagi et al., 2010).

Comme indiqué précédemment pour la protéine Larp7, certains cancers sont également associés à une augmentation de l'activité de P-TEFb résultant de l'altération de l'expression des composants de la RNP 7SK. Ainsi, l'expression de Hexim1 est plus faible dans des cellules de cancer mammaires comparées à des cellules de tissus sain, et sa surexpression inhibe la prolifération tumorale (Wittmann et al., 2003). HEXIM1 et Cycline T1 interagissent tous les deux indépendamment avec le récepteur aux œstrogènes ERα (Wittmann et al., 2005). De faibles niveaux d'HEXIM1 favoriseraient donc l'association P-TEFb-ERα et la suractivation des gènes de réponse aux œstrogènes. P-TEFb est également directement impliqué dans le développement de leucémies caractérisées par l'expression de protéines du SEC fusionnées à la protéine MLL (Lin et al., 2010; Monroe et al., 2011; Mueller et al., 2009). Ces protéines de fusion permettent le recrutement du SEC, dans lequel P-TEFb est incorporé, sur les gènes cibles de MLL via son domaine de liaison à l'ADN et stimulent leur expression. Une prolifération non contrôlée des progéniteurs hématopoïétiques en découle, pouvant mener à la transformation leucémique (Smith et al., 2011a). Une augmentation de l'expression de P-TEFb et de son activité est aussi associée à la transformation maligne des cellules lymphoïdes (Bellan et al., 2004). Globalement, l'activité de Cdk9 est ciblée pour le traitement de différents cancers et de nombreux inhibiteurs, comme le Flavopiridol, sont en phase d'essai clinique (Abdullah et al., 2011; Bible et al., 2012; Kretz et al., 2017; Lemke et al., 2014; Mitra et al., 2016; Mohapatra et al., 2009; Shao et al., 2013; Walsby et al., 2014).

En dehors de son implication dans la prolifération cellulaire menant au cancer, P-TEFb joue un rôle clé dans le processus d'hypertrophie cardiaque (Matrone et al., 2015; Sano and Schneider, 2004; Sano et al., 2004). Divers signaux hypertrophiques induisent en effet dans les cardiomyocytes la dissociation de la RNP 7SK/P-TEFb (voir II.iii.a. Voies de signalisation et modifications post-traductionnelles de la RNP 7SK/P-TEFb). La suractivité de P-TEFb qui en découle stimule la transcription par l'ARN Pol II et mène à une augmentation globale des niveaux d'ARNm et de protéines cellulaires, provoquant l'expansion en taille des cardiomyocytes. La destruction de la RNP 7SK/P-TEFb par déplétion de l'ARN 7SK dans des cellules cardiaques permet d'ailleurs de récapituler les effets observés en réponse aux signaux hypertrophiques. De même, la délétion du gène d'Hexim1 chez la souris (appelé CLP-1) provoque une létalité embryonnaire caractérisée par une hypertrophie cardiaque (Huang et al., 2004). Ainsi, la régulation de l'activité de P-TEFb par la RNP 7SK s'avère critique pour le développement et le fonctionnement physiologiques du cœur. L'inhibition pharmacologique de Cdk9 a d'ailleurs été proposée en traitement de l'hypertrophie cardiaque (Krystof et al., 2010).

iii. Libération de P-TEFb de la RNP 7SK/P-TEFb

La RNP 7SK/P-TEFb constitue un réservoir de P-TEFb permettant d'adapter les quantités de P-TEFb actif aux besoins transcriptionnels cellulaires. Ainsi, le blocage de la transcription par l'ARN Pol II par traitement à l'actinomycin D, qui bloque la translocation de l'ARN Pol II, au DRB ou Flavopiridol qui inhibent l'activité de P-TEFb, induit une signalisation menant à la dissociation de la RNP 7SK/P-TEFb (Nguyen et al., 2001; Yang et al., 2001). De même, certains agents inducteurs de dommages à l'ADN, commes les UV, les inhibiteurs d'histones déacétylases ou l'analogue de cytidine Aza-C, provoquent la libération de P-TEFb de la RNP 7SK (Contreras et al., 2007). Ces différents stress cellulaires entraînent des cascades de signalisation au cours desquelles l'activité de nombreuses protéines et divers facteurs changent de façon à répondre de manière coordonnée au signal. La libération de P-TEFb de la RNP 7SK, et la potentielle activation transcriptionnelle qui en découle, est une des conséquences moléculaires de l'activation de ces cascades.

a. Voies de signalisation et modifications post-traductionnelles de la RNP 7SK/P-TEFb

Le remodelage transcriptionnel est une étape obligatoire de la réponse à un stimulus. Il est ainsi naturel que différentes voies de signalisation induisent la dissociation de la RNP 7SK/P-TEFb. Lors de la transduction du signal, de nombreuses protéines à activité enzymatiques, telles que des kinases, sont activées. HEXIM1, Cdk9 et Cycline T1 sont tous les trois cibles de modifications post-traductionnelles qui influent sur leur association avec la RNP 7SK cœur (figure 18).

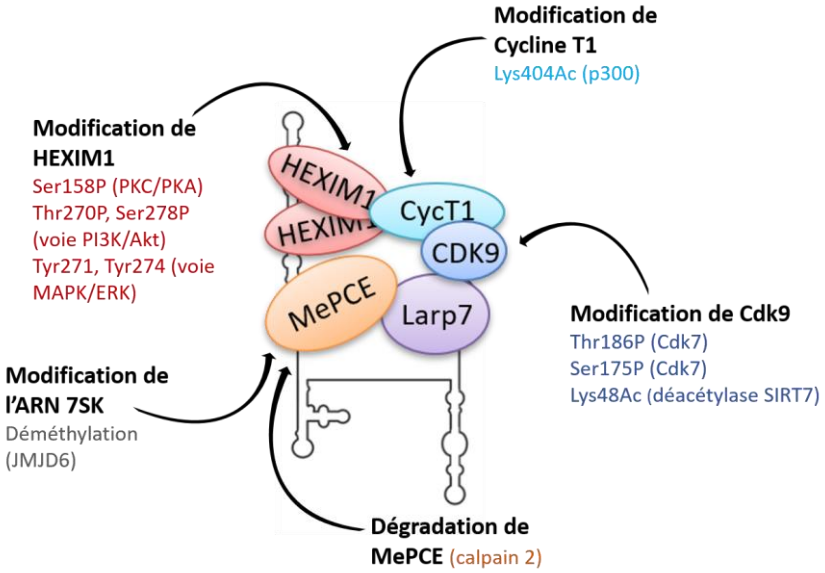


Figure 18 : Modifications post-traductionnelles des composants de la RNP 7SK/P-TEF impliqués dans la libération de P-TEFb. Les enzymes ou voies de signalisation impliquées dans ces modifications sont indiquées.

- *Phosphorylations de Cdk9 :*

Comme mentionné précédemment, la phosphorylation de la Thr186 de Cdk9 est un prérequis pour l'incorporation de P-TEFb dans la RNP 7SK. Dans les cellules HeLa, l'irradiation aux UV ou le traitement avec l'agent de différenciation cellulaire HMBA (hexamethylene bisacetamide) induisent l'activation calcium dépendante des phosphatases PP2B (calmodullin-protein phosphatase 2B) et PP1α (protein phosphatase 1α) (Ammosova et al., 2016; Chen et al., 2008). PP2B induit, par un mécanisme inconnu, un changement conformationnel de la RNP 7SK/P-TEFb, permettant à PP1α de déphosphoryler la Thr186P de Cdk9 menant à la dissociation de la RNP 7SK/P-TEFb. Le rôle de PP1α dans la déphosphorylation de Cdk9 a d'ailleurs été confirmé par l'utilisation d'un inhibiteur spécifique à cette phosphatase : une augmentation de la phosphorylation de la Thr186 de Cdk9 concomitante à une augmentation de son association avec la RNP 7SK est alors observée (Ammosova et al., 2011). De plus, la dissociation de la RNP 7SK/P-TEFb est également constatée après traitement de cellules cardiaques avec des agents inducteurs d'hypertrophie, dont PP2B (également appelée calcineurin) (Sano et al., 2002). Ainsi, la libération de P-TEFb dans le contexte hypertrophique pourrait se faire via le mécanisme PP2B/PP1α. Deux membres de la famille des phosphatases métal-dépendantes, PPM1A et PPM1G, peuvent aussi déphosphoryler spécifiquement la thr186 de Cdk9 (Budhiraja et al., 2012; Gudipaty et al., 2015; Wang et al., 2008). Tandis que l'activité de PPM1A ne semble pas impacter la formation de la RNP 7SK/P-TEFb, l'activation de PPM1G permet la libération de P-TEFb de la RNP 7SK par un mécanisme suggéré dépendant de NF-kB (Gudipaty et al., 2015). Enfin, Cdk9 peut également être phosphorylée sur la Ser175 par Cdk7 (Mbonye et al., 2018; Mbonye et al., 2013). Cette modification

est exclusivement retrouvée sur Cdk9 non associé à la RNP 7SK et pourrait aussi être liée à la dissociation de la RNP 7SK (Mbonye et al., 2013).

- *Phosphorylations d'HEXIM1 :*

La phosphorylation de HEXIM1 au niveau de son domaine d'interaction à Cycline T1 en réponse à l'activation de différentes voies de signalisation participe aussi à la dissociation de la RNP 7SK/P-TEFb (figure 17). Ainsi, le traitement au HMBA conduit à la phosphorylation des Thr270 et Ser278 via l'activation de la voie PI3K/Akt, et l'expression de mutants non-phosphorylables empêche le désassemblage de la RNP 7SK/P-TEFb (Contreras et al., 2007). La stimulation des lymphocytes T permet la phosphorylation des Tyr271 et 274, probablement via l'activation de la voie MAPK/ERK (Kim et al., 2011; Mbonye et al., 2015). La mutation phosphomimétique de ces résidus compromet la séquestration de P-TEFb dans la RNP 7SK (Mbonye et al., 2015). La région de HEXIM1 impliquée dans l'interaction avec l'ARN 7SK peut également être phosphorylée et induire la dissociation de la RNP 7SK/P-TEFb (figure 17). Notamment, la Ser158 est ciblée par la protéine kinase C θ (PKCθ) en réponse à l'activation des lymphocytes T (Fujinaga et al., 2012). Là encore, l'expression d'un mutant non-phosphorylables sur ce résidu bloque la libération de P-TEFb. L'activation de la voie de signalisation de l'adénosine monophosphate cyclique (cAMP) conduit également à la phosphorylation de la Ser 158 par la protéine kinase A (PKA) et à la libération de P-TEFb de la RNP 7SK (Sun et al., 2019). La suractivation de cette voie de signalisation est retrouvée dans les cellules rénales de patients atteints de la maladie rénale polycystique (PKD, polycystic kidney desease) (Sun et al., 2019; Yamaguchi et al., 1997). Les cellules PKD présentent des niveaux d'HEXIM1 phosphorylé sur la Ser158 anormalement élevés, associés à une diminution de l'interaction entre HEXIM1 et P-TEFb comparée à l'association observée dans des cellules saines (Sun et al., 2019). Des cellules de rein de souris récapitulant la maladie présentent d'ailleurs une suractivation de P-TEFb, tandis que l'inhibition de P-TEFb au Flavopiridol supprime l'accroissement pathologique des vésicules rénales (Sun et al., 2019).

- *Acétylations de Cdk9 et Cycline T1 :*

Enfin, l'acétylation des deux sous-unités de P-TEFb peut également impacter son association à la RNP 7SK. L'activité de Cdk9 semble pouvoir être régulée par acétylation/déactétylation de ces Lys44 et 48 (Blank et al., 2017; Fu et al., 2007; Sabo et al., 2008; Zhang et al., 2013). Notamment, l'acétylation de la Lys48 par GCN5 diminue l'activité kinase de Cdk9, probablement en interférant avec la liaison de l'ATP dans le site catalytique de Cdk9 (Blank et al., 2017; Sabo et al., 2008). La déacétylase SIRT7 apparait ainsi promouvoir l'activité de Cdk9 en catalysant le retrait du groupement Ac sur la Lys48 et en induisant de cette manière la dissociation de la RNP 7SK/P-TEFb (Blank et al., 2017). Cycline T1 apparait également modifié par l'acétyl-transférase p300 sur la Lys404, dans la région impliquée dans

sa liaison avec HEXIM1 (figure 16) (Cho et al., 2009). L'analyse par fractionnement sur gradient de glycérol montre que Cycline T1 acétylée ne co-sédimente pas avec les autres composants de la RNP 7SK/P-TEFb. De plus, traitement des cellules avec des inhibiteurs de déacétylases entraîne une diminution de l'interaction entre HEXIM1 et Cycline T1. A noter que l'acétylation de Cycline T1 ne semble pas empêcher la dissociation de la RNP 7SK induite par le traitement au DRB, à l'actinomycin D ou au HMBA, suggérant que cette modification participe à l'équilibre entre les formes actives et inactives de P-TEFb en conditions normales, ou bien en réponse à des stimuli autres.

Ainsi, la modification post-traductionnelle des composants labiles de la RNP 7SK est impliquée dans l'équilibre entre les formes actives et inactives de P-TEFb en réponse à un stimulus. Néanmoins, les mécanismes d'action et les enzymes responsables sont encore peu caractérisés. De plus, il n'est pas clair si certaines de ces modifications sont la cause ou la conséquence de la dissociation de la RNP 7SK/P-TEFb. La modification des composants de la RNP 7SK/P-TEFb participe certainement à la libération de P-TEFb via des processus plus complexes faisant intervenir d'autres acteurs et mécanismes que nous allons entrevoir dans les paragraphes suivants. Il est important de noter que ces modifications ont globalement été identifiées pour leur potentiel à induire la transcription du gène intégré du VIH, dont l'activation est très sensible aux quantités de P-TEFb actif disponible (Bartholomeeusen et al., 2013; Bartholomeeusen et al., 2012; Chiu et al., 2004). Il reste donc à identifier les gènes cellulaires impactés par la dissociation de la RNP 7SK/P-TEFb en réponse à divers stimuli.

b. Extraction de P-TEFb de la RNP 7SK/P-TEFb

Certains facteurs ont la capacité d'interagir directement avec des composants de la RNP 7SK/P-TEFb, déstabilisant ainsi les interactions déjà présentes. Ce mécanisme est d'ailleurs employé par le cofacteur du VIH Tat. Tat possède un domaine basique similaire à celui d'HEXIM, lui permettant d'interagir avec la tige-boucle en 5' de l'ARN 7SK (Yik et al., 2004), et les deux protéines partagent également un même domaine de liaison sur Cycline T1 (Schulte et al., 2005). Ainsi, Tat peut séquentiellement se lier à l'ARN 7SK et déloger les protéines HEXIM, puis s'associer à Cycline T1 par compétition avec HEXIM, l'affinité de Tat pour Cycline T1 étant supérieure à celle d'HEXIM (Barboric et al., 2007; Muniz et al., 2010). Ce processus d'extraction semble être favorisé par la protéine du SEC AFF1, qui interagit avec P-TEFb au sein de la RNP 7SK et stimule l'interaction entre Tat et Cycline T1 (Lu et al., 2014). Comme Tat, le co-facteur du virus T-lymphotropique humain (VTLH) Tax peut extraire P-TEFb de la RNP 7SK (Cho et al., 2010; Zhou et al., 2006). Le mécanisme employé par les protéines virales Tat et Tax pourrait ainsi être commun à d'autres virus. Le co-activateur cellulaire Brd4 peut aussi promouvoir la dissociation de la RNP 7SK via son domaine d'interaction à P-TEFb (PID, P-TEFb

interaction domain) (Krueger et al., 2010). L'incubation de la RNP 7SK/P-TEFb avec le PID de Brd4 permet de libérer P-TEFb *in vitro* de la même façon que la protéine virale Tat. De plus, la fraction active de P-TEFb augmente suite à la surexpression de Brd4 dans les cellules HeLa. En réponse aux UV, c'est la protéine Rbm7, activée par phosphorylation, qui, en interagissant directement avec la troisième tige-boucle de l'ARN 7SK, permet la libération de P-TEFb pour l'activation transcriptionnelle de gènes impliqués dans la réponse aux dommages à l'ADN (Bugai et al., 2019).

c. Dissociation de la RNP 7SK/P-TEFb et changement conformationnel

Comme cela a été décrit dans la partie sur l'ARN 7SK, la libération de P-TEFb de la RNP 7SK induit un changement conformationnel qui pourrait résulter de la perte de P-TEFb et HEXIM. Un remodelage actif pourrait également s'opérer et faciliter la dissociation de la RNP 7SK/P-TEFb. Les hélicases à ARN DEAD-box DDX21 et DDX6 ont été impliquées dans ce processus (Calo et al., 2015; Muck et al., 2016). DDX21 est requise pour la maturation des ARN ribosomiques (ARNr) et occupe la région transcrite de l'ADNr (Calo et al., 2015). L'identification des ARN associés à DDX21 par iCLIP (individual-nucleotide-resolution crosslinking and immunoprecipitation) a révélé qu'en plus des ARNr et de snoARN, DDX21 interagissait également avec l'ARN 7SK, au niveau de sa troisième tige-boucle. De plus, DDX21 interagit avec d'autres composants de la RNP 7SK/P-TEFb, et promeut sa dissociation *in vitro*. En revanche, d'une version catalytiquement inactive de DDX21 échoue à dissocier la RNP 7SK/P-TEFb et ses domaines d'interactions sur l'ARN 7SK changent, suggérant un changement conformationnel permis par l'activité hélicase de DDX21 et nécessaire pour le relargage de P-TEFb. Les auteurs de cette étude ont trouvé que ces événements se produisaient au niveau des promoteurs de nombreux gènes activement transcrits par l'ARN Pol II, majoritairement codant des protéines ribosomales, sur lesquels DDX21 est recruté de manière 7SK-dépendante. DDX6 semble promouvoir le transfert de P-TEFb de la RNP 7SK/P-TEFb vers le Super Elongation Complex, mais cette hypothèse n'est basée que sur des interactions détectées *in vitro* (Muck et al., 2016).

d. Remodelage de la RNP 7SK/P-TEFb

Les différents mécanismes de dissociation de la RNP 7SK/P-TEFb cités jusque-là sont réversibles. Cela implique que la RNP 7SK/P-TEFb peut donc se reformer rapidement. Cependant, une fois P-TEFb et HEXIM relarguées de la RNP 7SK, d'autres protéines peuvent interagir avec la RNP 7SK cœur pour empêcher sa réassociation précoce avec P-TEFb. Notamment, après traitement au DRB, à l'actinomycine D ou aux UV, la RNP 7SK cœur s'associe transitoirement à des protéines de liaison à l'ARN de type hnRNP (heteregeneous ribonuclear protein), les hnRNP A1, A2/B1, Q, R et l'hélicase à ARN A (RHA) (Barrandon et al., 2007; Hogg and Collins, 2007; Van Herreweghe et al., 2007). Les hnRNP

sont des protéines abondantes qui s'associent aux transcrits nouvellement formés. Leur disponibilité augmente donc en réponse au blocage transcriptionnel engendré par ces drogues, et pourrait expliquer l'association observée. Cependant, la co-déplétion des hnNRPA1 et A2 par siARN ou la déplétion de la troisième tige-boucle de l'ARN 7SK, la région liée par les hnRNPA1 et A2, entrave la libération de P-TEFb de la RNP 7SK (Barrandon et al., 2007; Van Herreweghe et al., 2007). Ces protéines pourraient donc jouer un rôle actif dans la dissociation de la RNP 7SK/P-TEFb en réponse à un blocage transcriptionnel. Comme DDX21, l'hélicase RHA pourrait provoquer un changement conformationnel de la RNP 7SK/P-TEFb nécessaire pour son désassemblage. L'association des hnRNP à la RNP 7SK étant incompatible à celle de P-TEFb, elle pourrait également empêcher la réassociation précoce de P-TEFb (Barrandon et al., 2007). De façon similaire aux hnRNP, la phosphatase PPM1G est directement impliquée dans la dissolution de la RNP 7SK/P-TEFb en réponse aux dommages à l'ADN en déphosphorylant la Thr186 de Cdk9 (Gudipaty et al., 2015). PPM1G reste liée à l'ARN 7SK et HEXIM1 après la libération de P-TEFb et pourrait ainsi empêcher sa re-séquestration le temps nécessaire à la réponse transcriptionnelle.

La dissociation de la RNP 7SK/P-TEFb peut également être régulée de manière irréversible, via la dégradation directe ou indirecte de l'ARN 7SK (Elagib et al., 2013; Liu et al., 2013). Ainsi, l'ARN 7SK pourrait être la cible de la déméthylase JMJD6 (Liu et al., 2013). JMJD6 semble ainsi pouvoir retirer la coiffe à l'extrémité 5' de l'ARN 7SK, provoquant sa dégradation et la libération de P-TEFb en conséquence. Le rôle de JMJD6 sera détaillé dans la partie suivante. Les mégacaryocytes sont issus de la différenciation de cellules hématopoiétiques et caractérisés par un accroissement hypertrophique de la masse cellulaire provoqué par une suractivation de P-TEFb (Elagib et al., 2013). Durant la mégakaryétopoièse, la protéase calpain 2 est activée et clive la protéine MePCE, provoquant la dégradation de l'ARN 7SK qui n'est plus protégé à son extrémité 5'. L'activation de P-TEFb qui en découle stimule alors l'expression de gènes codant des protéines de régulation du cytosquelette impliquées dans l'élargissement des cellules.

iv. Régulation de l'activité de P-TEFb à la chromatine

Comment P-TEFb est délivré de manière régulée sur les gènes et le rôle de la RNP/7SK P-TEFb dans ce processus restent encore mal connus. La RNP 7SK/P-TEFb a initialement été détectée dans la fraction nucléaire soluble (Biglione et al., 2007). De plus, des expériences de FISH ARN (fluorescent *in situ* hybridization) indiquent une localisation nucléoplasmique diffuse de l'ARN 7SK (Egloff et al., 2006; Prasanth et al., 2010). Ainsi, il a été proposé que la forme active de P-TEFb soit recrutée sur les gènes, provenant de sa fraction active ou bien de la fraction libérée de la RNP 7SK/P-TEFb en réponse à un stimulus. Cependant, l'analyse des ARN associés à la chromatine a révélé qu'une fraction significative

de l'ARN 7SK, estimée à ~3000 copies, y était associée (Mondal et al., 2010). Des études de fractionnement cellulaire dans notre équipe ont confirmé cet ordre de grandeur. Plusieurs études durant la dernière décennie ont d'ailleurs identifié des composants de la RNP 7SK/P-TEFb associés à l'ARN Pol II ou au niveau de promoteurs de gènes et d'enhancers (Calo et al., 2015; D'Orso and Frankel, 2010; Egloff et al., 2017b; Gudipaty et al., 2015; Ji et al., 2013; Liu et al., 2013; McNamara et al., 2013; McNamara et al., 2016b). Ainsi, la forme inactive de P-TEFb pourrait également être recrutée sur les gènes pour activer leur expression en conditions normales ou après stimulation. Alors que différentes études décrivent un recrutement de la RNP 7SK/P-TEFb pour activer la transcription, la RNP 7SK pourrait également être recrutée sur les gènes pour re-capturer P-TEFb une fois la pause transcriptionnelle levée. Une étude par microscopie à fluorescence incluant une version étiquetée de l'ARN 7SK a ainsi mis en évidence une augmentation transitoire de la concentration de l'ARN 7SK au niveau d'un gène rapporteur suite au blocage de son expression (Prasanth et al., 2010). En revanche, le recrutement de l'ARN 7SK n'a pas été détecté lors de l'activation du gène rapporteur.

a. Association de la RNP 7SK/P-TEFb à la chromatine

Le chargement de la RNP 7SK/P-TEFb à la chromatine fait intervenir différents mécanismes. Une puce chargée avec différents peptides d'histones modifiés a permis de montrer que l'ARN 7SK peut reconnaître et se lier à différentes marques d'histones, et notamment aux histones H4 di-méthylés sur l'arginine 3 (H4R3me2), une marque répressive de la chromatine (Liu et al., 2013). H4R3me2 est retrouvée au niveau d'une classe d'enhancers appelés anti-pause enhancers (A-PE), co-occupés par la RNP 7SK/P-TEFb, Brd4 et la déméthylase JMJD6 (figure 19). Les A-PE stimulent l'expression des gènes avec lesquels ils interagissent selon un mécanisme complexe : la RNP 7SK/P-TEFb est recrutée au niveau des A-PE via la marque H4K3me2, JMJD6 par Brd4, lui-même recruté via sa liaison aux histones H3 et H4 acétylés. JMJD6 efface la marque répressive (me2) sur H4R3 et retire le groupement méthyl protecteur à l'extrémité 5' de l'ARN 7SK (figure 18). L'ARN 7SK est alors dégradé, induisant la libération de P-TEFb et l'activation de la transcription. La protéine MePCE a récemment été identifiée au niveau des régions promotrices de certains gènes, recrutée via une interaction directe avec la queue des histones H4 (Shelton et al., 2018). Cependant, au lieu de représenter un point d'ancrage pour la RNP 7SK à la chromatine, MePCE semble être recrutée sur ces gènes indépendamment de la RNP 7SK mais avec Cdk9, et permettrait ainsi d'activer leur expression.

La RNP 7SK/P-TEFb peut également être recrutée au niveau de nombreuses régions promotrices par la protéine TRIM28, aussi appelée KAP1, grâce à une interaction directe entre Larp7 et TRIM28 (figure 19) (McNamara et al., 2016b). En réponse à une stimulation, P-TEFb peut alors être libéré de la RNP 7SK localement et activer rapidement l'élongation de la transcription. Un mécanisme similaire a été

proposé pour le facteur d'épissage SRSF2 (Ji et al., 2013). SRSF2, qui occupe préférentiellement les promoteurs de gènes actifs, permet d'y recruter la RNP 7SK/P-TEFb en s'associant à la troisième tige-boucle de l'ARN 7SK (figure 19). L'apparition d'un site de liaison à SRSF2 sur l'ARN naissant alors qu'il émerge de l'ARN Pol II permettrait alors le transfert de SRSF2 sur le transcrit, libérant P-TEFb de la RNP 7SK au passage pour activer l'élongation transcriptionnelle.

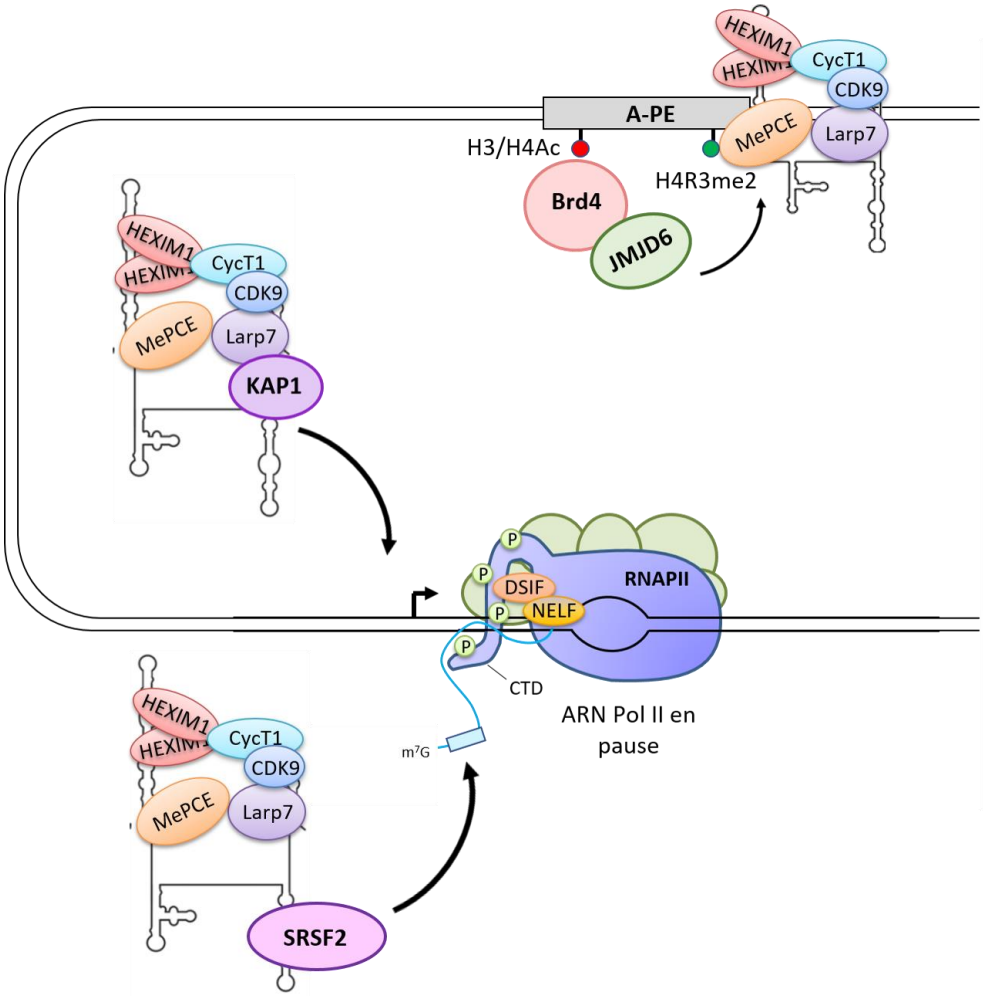


Figure 19 : Mécanismes d'association et de libération de la RNP 7SK/P-TEFb à la chromatine.

b. Dissociation de la RNP 7SK/P-TEFb à la chromatine

Les études réalisées jusqu'à présent n'ont pas mis en évidence de recrutement actif de la RNP 7SK/P-TEFb sur les gènes en réponse à un signal. La RNP 7SK/P-TEFb semble occuper constitutivement les régions promotrices d'un certain nombre de gènes, dont l'activation transcriptionnelle dépend du mécanisme par lequel la RNP 7SK se dissocie pour libérer P-TEFb. Les mécanismes de dissociation de la RNP 7SK à la chromatine restent cependant peu connus. Au niveau des gènes fortement transcrits, l'hélicase DDX21 pourrait permettre de libérer régulièrement P-TEFb de la RNP 7SK/P-TEFb selon le mécanisme qui a été décrit dans la partie précédente (Calo et al., 2015). DDX21 occupe en effet de

nombreuses régions promotrices de gènes activement transcrits par l'ARN Pol II, notamment des gènes codants des protéines ribosomales ou impliquées dans d'autres aspects de la prolifération cellulaire. De plus, son association à ces régions est dépendante de l'ARN 7SK. En ce qui concerne les gènes de réponse, la phosphatase PPM1G promouvoit la dissociation de la RNP 7SK/P-TEFb sur les gènes cibles du facteur de transcription NF-kB (Gudipaty et al., 2015). Suite à des dommages à l'ADN, NF-kB active l'expression de ses gènes cibles en stimulant le recrutement de l'ARN Pol II et en activant l'élongation transcriptionnelle via le recrutement de PPM1G, qui libère P-TEFb de la RNP 7SK. Des mécanismes similaires pourraient être envisagés pour d'autres types de gènes inductibles.

v. Les autres RNP 7SK

L'abondance de l'ARN 7SK est estimée à environ 200 000 copies par cellule. La quasi-totalité des protéines MePCE et Larp7 lui sont associées, et ensemble séquestrent environ 50 % des complexes P-TEFb nucléaires. Parmi la totalité des RNP 7SK cœurs, environ 30-50 % sont associées à P-TEFb et HEXIM, en fonction des lignées cellulaires (Barrandon et al., 2007; Haaland et al., 2003; Yik et al., 2003). Une faible fraction de la RNP 7SK cœur, environ 5 %, est associée aux protéines hnRNP A1, A2, R, Q et à la RHA en conditions normales, et augmente après traitement à l'actinomycine D ou DRB comme cela a été décrit plus précédemment. Ainsi, une grande fraction de RNP 7SK cœurs peut potentiellement interagir avec d'autres partenaires. Récemment, de nouvelles RNP 7SK, indépendantes de P-TEFb/HEXIM, ont émergé et étendent les fonctions de l'ARN 7SK au-delà du contrôle de l'activité de P-TEFb.

a. La RNP 7SK/LEC

Le LEC (little elongation complex) est un complexe régulant spécifiquement la transcription par l'ARN Pol II des gènes des petits ARN nucléaires et nucléolaires (sn et snoARN) (Smith et al., 2011b). Le LEC est composé des protéines ELL, ICE1, ICE2, EAF et ZC3H8. ICE1 semble requis pour le recrutement de l'ARN Pol II sur les gènes de sn(o)ARN et ELL pour l'élongation de la transcription de ces gènes. Il a récemment été montré au laboratoire que la RNP 7SK cœur formait un complexe avec le LEC et participait à son chargement sur les gènes des sn(o)ARN (Egloff et al., 2017b). Ainsi, la protéine Larp7 est spécifiquement retrouvée associée aux gènes des sn(o)ARN, et sa déplétion, comme celle de MePCE ou de l'ARN 7SK, affecte le recrutement de l'ARN Pol II et l'expression des sn(o)ARN. De plus, alors que la RNP 7SK/P-TEFb se dissocie après traitement à l'actinomycin D, la formation de la RNP 7SK/LEC augmente. La 7SK/hnRNP, dont la formation augmente également après blocage de la transcription, et la RNP 7SK/LEC semblent cependant former deux complexes indépendants. L'inhibition de la transcription induit donc deux mécanismes complémentaires dépendants de la RNP

7SK pour booster à la fois la production des ARNm, via la libération de P-TEFb, et celle des sn(o)ARN, via la formation de la RNP 7SK/LEC.

b. L'ARN 7SK et le complexe BAF

L'occupation génomique de l'ARN 7SK a récemment été déterminée par ChIRP-seq (Chromatin Isolation by RNA Purification and sequencing) (Flynn et al., 2016). En plus d'occuper un grand nombre de promoteurs et de régions transcrites, l'ARN 7SK a également été localisé au niveau d'enhancers et de clusters d'enhancers appelés super enhancers (SE). Au niveau de ces régions, l'ARN 7SK interagit avec le complexe de remodelage de la chromatine BAF pour limiter le recrutement de l'ARN Pol II et ainsi l'expression des eRNAs. La déplétion de l'ARN 7SK par siARN inhibe le recrutement de BAF, et le recrutement de l'ARN 7SK est lui-même inhibé par traitement des cellules au JQ1, un composé chimique qui empêche la liaison de Brd4 au niveau des histones acétylées. Ces données suggèrent donc que l'ARN 7SK pourrait être recruté au niveau des enhancers via Brd4, bien qu'aucune interaction entre les deux n'ait encore été rapportée. Cependant, seulement une faible fraction de l'ARN 7SK s'associe aux complexes BAF, il est donc possible que l'interaction entre l'ARN 7SK et Brd4 soit également faible et difficilement détectable du fait de la forte association de Brd4 à la chromatine (Biglione et al., 2007).

c. La RNP 7SK/Ars2

La protéine Ars2 est un composant du Cap Binding Complex (CBC), un complexe nucléaire qui lie la coiffe en 5' des ARN synthétisés par l'ARN Pol II. Au sein de ce complexe, Ars2 participe à la maturation en 3' de diverses classes d'ARN synthétisés par l'ARN Pol II, et notamment des micro et snARN et des ARNm des histones réplication-dépendants (Hallais et al., 2013). Les ARNm des histones réplication-dépendants ont la particularité de ne pas être poly-adénylés mais clivés au niveau d'un site particulier dans leur région 3' non-traduite par la snRNP U7, recrutée par la protéine SLBP (stem-loop binding protein) (Yang et al., 2009). Ars2 favorise ce processus via sa capacité à s'associer aux ARNm des histones. L'ARN 7SK peut aussi interagir avec Ars2, et l'empêche alors de promouvoir la maturation des ARNm des histones (Gruber et al., 2012). Ainsi, la déplétion de l'ARN 7SK favorise la production d'ARNm d'histones correctement maturés, tandis que la proportion d'ARNm d'histones poly-adénylés augmente avec la déplétion de Ars2. De plus, l'interaction entre Ars2 et l'ARN 7SK augmente après traitement des cellules à l'actinomycin D, qui bloque la transcription en se liant à l'ADN au niveau des bulles transcriptionnelles et induit l'arrêt en phase G1 du cycle cellulaire. Ainsi, la séquestration dynamique d'Ars2 par l'ARN 7SK pourrait permettre d'adapter la production d'histones aux capacités prolifératives cellulaires. Cdk9 n'est en revanche pas retrouvé dans le complexe Ars2/7SK. Le

traitement à l'actinomycin D induisant la dissociation de la RNP 7SK/P-TEFb au profit de la formation de la RNP 7SK/hnRNP, il est possible qu'Ars2 fasse également partie de ce dernier complexe, bien que cela n'ait pas été testé. Alternativement, les deux complexes, 7SK/hnRNP et 7SK/Ars2 pourraient se former après dissociation de P-TEFb et HEXIM.

III. Réponse transcriptionnelle aux dommages à l'ADN causés par les UV

L'ADN des cellules est constamment soumis à des menaces provenant de l'environnement ou émanant des processus métaboliques internes qui en altèrent sa composition. Il existe divers mécanismes de prise en charge des lésions de l'ADN dépendants de leur nature. La réponse transcriptionnelle varie également en fonction du type de lésion. Les lésions oxydatives sont les plus communes, générées par les espèces réactives de l'oxygène (ROS, reactive oxigen species) produites par la chaîne respiratoire mitochondriale, et sont généralement bien tolérées par la machinerie transcriptionnelle aidée des facteurs d'élongation tels que TFIIF et TFIIS (Charlet-Berguerand et al., 2006). Les cassures double-brins de l'ADN induisent un arrêt local de la transcription le temps de la réparation (Marnef et al., 2017). Les lésions provoquées par les UV, elles, bloquent non seulement la progression de l'ARN Pol II si elle rencontre une base altérée, mais induisent également une extinction globale de la transcription. Les bases des processus de réparation des dommages à l'ADN causés par les UV ont été posées grâce à l'étude de pathologies rares d'hypersensibilité aux UV, telles que Xeroderma Pigmentosum (XP) et le Syndrome Cockaine (CS). Récemment, des études résolutives de la transcription à l'échelle du génome ont permis d'apporter de nouveaux éléments de réponse quant au mécanisme d'extinction globale de la transcription (Geijer and Marteijn, 2018; Gregersen and Svejstrup, 2018).

i. Lésions de l'ADN provoquées par les UV et leur signalisation

Les UV occasionnent des dommages à l'ADN en convertissant deux pyrimidines adjacentes en dimères : des photoproduits 6-4 pyrimidine-pyrimidone (6-4PP) et des dimères cyclobutane-pyrimidine (CPD). Contrairement aux lésions causées par les ROS, les 6-4PP et CPD induisent des distorsions hélicoïdales qui sont spécifiquement prises en charge par la voie de réparation par excision de nucléotide (NER, nucleotide excision repair) (figure 20). Le NER peut être divisée en deux branches, selon l'endroit du génome où ont lieu les lésions. Lorsque ces lésions surviennent sur le brin matrice dans une région transcriptionnellement active, la voie de réparation couplée à la transcription (TC-NER) est déclenchée par le blocage de l'ARN Pol II sur le gène (Brueckner and Cramer, 2007; Brueckner et al., 2007; Tornaletti, 2005). Les autres lésions sont réparées via la voie de réparation générale du génome (GG-NER). Le TC-NER et le GG-NER diffèrent par le mécanisme de reconnaissance des lésions, qui convergent dans les deux cas au recrutement du facteur TFIIH pour la vérification du dommage,

avant de procéder au processus de réparation par le complexe NER cœur. Bien que les mécanismes de reconnaissance des dommages ne soient pas encore totalement élucidés, il semble que l'arrêt de la transcription par l'ARN Pol II associée à un dommage soit un signal fort pour initier la réparation par le TC-NER : les lésions ayant lieu sur le brin matrice des gènes sont réparées plus rapidement que les lésions ayant lieu dans le reste du génome (Bohr et al., 1985; Hanawalt and Spivak, 2008). Une réparation rapide étant associée à un plus faible risque de mutation, les mécanismes de réparation pourraient avoir évolué d'un mécanisme global à un mécanisme spécifique pour préserver tout particulièrement les régions d'ADN associées à la production d'ARN.

a. Détection des lésions par le TC-NER

Le TC-NER est amorcé par la détection indirecte de la lésion à travers le blocage de l'ARN Pol II en élongation. Une lésion sur le brin matrice d'un gène empêche en effet la progression de l'ARN Pol II et induit son arrêt au niveau de la lésion (Laine and Egly, 2006). La protéine Cockaine Syndrome B (CSB) est alors recrutée par liaison directe à l'ARN Pol II (figure 20, étape 1 droite) (Fousteri et al., 2006; Sin et al., 2016; van Gool et al., 1997). CSB possède une activité translocase ATP-dépendante qui permet la reprise de la transcription par l'ARN Pol II lorsque celle-ci est bloquée par un dommage mineur à l'ADN, mais qui s'avère inefficace face à des lésions plus volumineuses comme celles occasionnées par les UV (Xu et al., 2017). Ainsi, la stabilisation du complexe CSB-ARN Pol II au niveau d'une lésion causée par les UV permet le recrutement des autres protéines du TC-NER, et notamment CSA et de UVSSA (figure 20, étapes 1-3 droite) (Geijer and Marteijn, 2018). Les protéines CSA et UVSSA sont impliquées dans la régulation de la stabilité de CSB : alors que CSA, via son appartenance au complexe E3-ligase CRL4-CSA, promeut l'ubiquitinylation et la dégradation de CSB, UVSSA permet le recrutement de l'ubiquitin-specific protéase USP7, qui neutralise l'activité du complexe CRL4-CSA (Groisman et al., 2006; Schwertman et al., 2012). En présence d'UVSSA-USP7, CSA ne peut donc pas induire la dégradation de CSB. De plus, UVSSA joue un rôle crucial dans le recrutement de TFIIH via une interaction directe avec la sous-unité p62 de TFIIH (figure 20, étape 4) (Okuda et al., 2017). TFIIH et USP7 reconnaissent cependant le même domaine sur UVSSA, suggérant une compétition entre les deux. Cela interroge sur le devenir de CSB suite à la liaison de TFIIH, étant donné qu'USP7 protège CSB de la dégradation par ubiquitinylation. Une autre part d'ombre du TC-NER émane du devenir de l'ARN Pol II durant le processus de réparation. L'ARN Pol II recouvre en effet le brin transcrit contenant la lésion, sur environ 35 nt, bloquant ainsi l'accès à la machinerie de réparation (Tornaletti et al., 1999). Plusieurs scénarios ont été envisagés quant au mécanisme d'accessibilité à la lésion, incluant la dissociation de l'ARN Pol II de la matrice, sa marche arrière et sa dégradation (Geijer and Marteijn, 2018; Gregersen and Svejstrup, 2018; Marteijn et al., 2014; Wilson et al., 2013). La dégradation semble

être le dernier recours, lorsque la quantité de dommage est trop importante ou que le TC-NER est défectueux. La marche arrière est le modèle privilégié, possiblement stimulée par TFIIH, qui se lie en aval de l'ARN Pol II et pourrait la faire reculer via son activité hélicase (Xu et al., 2017). Cependant, il a récemment été suggéré que les ARN Pol II bloquées au niveau de lésion se dissociaient de la matrice après la réparation (Chiou et al., 2018).

b. Détection des lésions par le GG-NER

Pour la reconnaissance des dommages via le GG-NER, le facteur clé est XPC (XP complementation group C), stabilisé par les protéines CETN2 et RAD23B (figure 20, étape 1 gauche). Des expériences de FRAP ont suggéré que XPC sondait constamment l'ADN en s'y associant transitoirement en l'absence de dommages (Hoogstraten et al., 2008). L'affinité de XPC pour l'ADN augmente lorsqu'il rencontre une lésion induisant une configuration thermodynamique instable, comme les 6-4PP générées par les UV (Dip et al., 2004; Puumalainen et al., 2016) . La liaison de XPC au niveau des lésions, et en particulier des CPD, qui perturbent peu l'établissement des liaisons hydrogènes avec les bases du brin opposé à la lésion, est facilitée par le complexe de protéines de liaison à l'ADN endommagé, DDB1-DDB2 (figure 20, étape 2 gauche) (Kusakabe et al., 2019; Sugasawa, 2016; Wittschieben et al., 2005). XPC se lie au brin opposé à la lésion et permet le recrutement du complexe TFIIH via une interaction directe avec la sous-unité p62 de TFIIH (figure 20, étape 4) (Okuda et al., 2015).

c. Réparation des lésions par excision de nucléotide

Que le processus de réparation soit initié par le TC-NER ou le GG-NER, les deux se rejoignent avec le recrutement de TFIIH selon un mécanisme similaire (figure 20, étape 4). Cependant, l'arrivée à cette étape n'est pas une assurance sans faille de la présence d'une lésion. L'ARN Pol II est en effet fréquemment arrêtée durant la transcription, et XPC reconnait aussi l'ADN non-endommagé, notamment des structures particulières présentant une distorsion de l'ADN. Une étape de vérification de la présence d'une lésion a ainsi lieu avant l'excision de l'ADN endommagé et repose sur TFIIH et XPA (figure 20, étape 4). Le complexe TFIIH est doté de deux sous-unités à activité hélicase ATP-dépendante, XPB et XPD. Tandis que l'activité ATPase de XPB permet d'ancrer stablement le complexe au site de lésion, l'activité hélicase de XPD est requise pour le processus de réparation (Coin et al., 2007; Oksenych et al., 2009). En effet, une fois recruté via UVSSA lors du TC-NER ou via XPC lors du GG-NER, XPD permet l'ouverture de la double hélice d'ADN dans le sens 5'-3' jusqu'à atteindre la lésion. Les protéines XPA et RPA rejoignent alors la structure simple-brin ainsi formée, avec RPA liant et protégeant le brin non endommagé (figure 20, étapes 4-5) (de Laat et al., 1998). De son côté, XPA effectue, en coopération avec TFIIH, la vérification du dommage en reconnaissant les nucléotides dont

la structure est altérée et en stimulant l'activité hélicase de XPD via le relargage du domaine kinase (CAK) de TFIIH (Camenisch et al., 2006; Coin et al., 2008; Li et al., 2015). Les endonucléases XPF, en complexe avec ERCC1, et XPG sont ensuite recrutées aux extrémités de la bulle d'ADN. Elles génèrent séquentiellement les incisions en 5' et 3' de la lésion, respectivement, libérant un oligonucléotide d'environ 27 nt contenant l'ADN endommagé (figure 20, étapes 5-6) (Evans et al., 1997a; Evans et al., 1997b; Huang et al., 1992). Pour finir, les ADN Pol δ/ε comblent la brèche et l'ADN ligase I reforme la liaison phosphodiester en 3' (figure 20, étapes 7-8) (Wood, 1997).

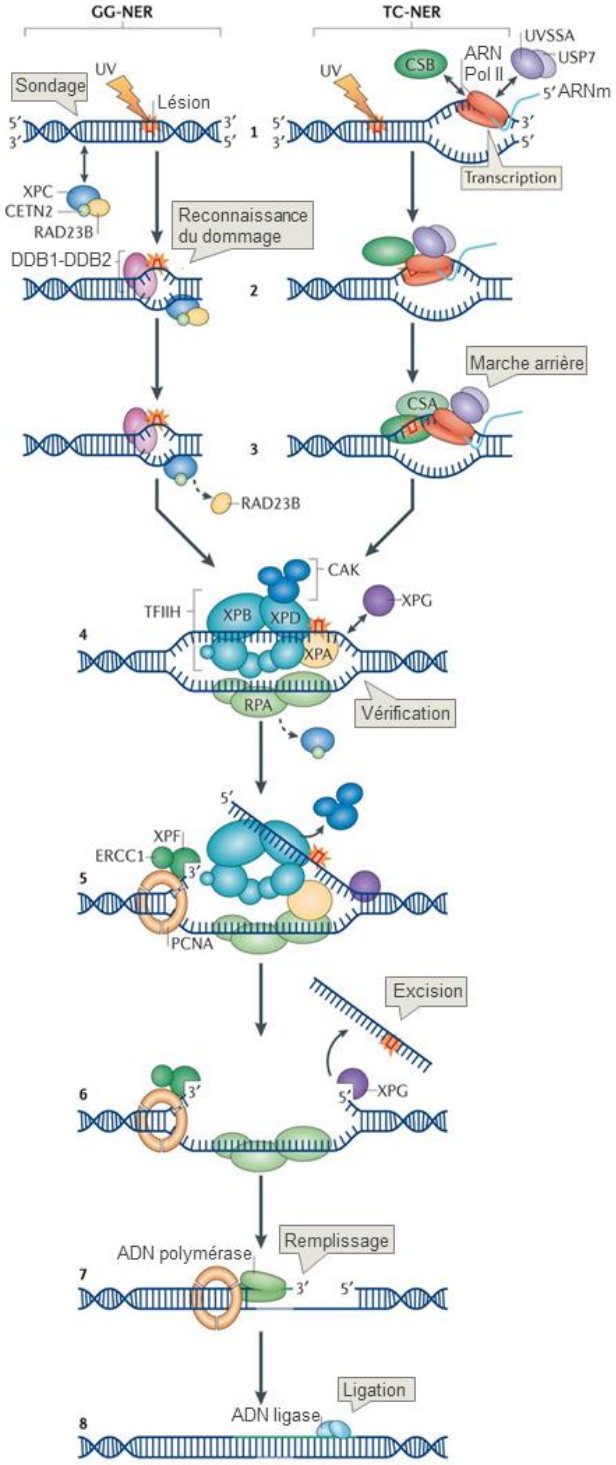


Figure 20 : Reconnaissance et réparation des lésions UV par le NER (Marteijn et al., 2014). Dans le GG-NER (étapes 1-3, gauche), la lésion est reconnue par XPC avec l'aide du complexe DDB1-DDB2. Dans le TC-NER (étapes 1-3, droite), le dommage sur le brin matrice des gènes est indirectement reconnu par le blocage de l'ARN Pol II en cours d'élongation au niveau de la lésion. CSB est recrutée au niveau de l'ARN Pol II bloquée et permet le recrutement de CSA et UVSSA-USP7, impliquée dans la stabilité de CSB. Après reconnaissance de la lésion, le processus de réparation est commun aux GG-NER et TC-NER (étapes 4-8). Le complexe TFIIH, recruté par XPC (GG-NER) ou UVSSA (TC-NER), contient deux sous-unités à activité hélicase XPB et XPD et un domaine à activité kinase (CAK). Le CAK de TFIIH est libéré suite à son recrutement, puis XPB et XPD ouvrent la double hélice d'ADN et procèdent à la vérification de la lésion avec XPA (étape 4). La protéine RPA lie et protège le brin d'ADN opposé à celui endommagé. Les endonucléases XPF-ERCC1 et XPG génèrent les incisions sur le brin endommagé en 5' et 3' de la liaison, respectivement (étapes 5-6). Le complexe PCNA est recruté suite à l'incision en 5' par XPG et recrute l'ADN Pol qui synthétise l'ADN manquant (étape 7). Le processus de réparation est complété par l'action de la ligase à ADN qui soude l'extrémité 3'OH de l'ADN néosynthétisé à l'extrémité 5' de l'ADN incisé (étape 8).

d. Cinétique de réparation

La réparation des dommages à l'ADN causées par les UV est un processus relativement rapide, avec la grande majorité des lésions éliminées 12-24h après irradiation (Baek et al., 2018). Cependant, comme évoqué plus haut, la réparation par le NER n'est pas un processus homogène. Ainsi, le mécanisme de reconnaissance de la lésion joue un rôle capital dans la cinétique de réparation : les lésions situées dans le brin matrice des gènes, prises en charge par le TC-NER, sont éliminées plus rapidement que les lésions reconnues via la GG-NER (Adar et al., 2016; Hanawalt and Spivak, 2008; Hu et al., 2015). De plus, le type de lésion impacte la prise en charge par le NER. Comme mentionné précédemment, les CPD induisent une plus faible distorsion hélicoïdale de l'ADN que les 6-4PP et sont donc moins bien reconnues. Ainsi, il faut environ 4h à la cellule pour venir à bout des 6-4PP, tandis que la résolution des CPD peut prendre jusqu'à 48h (Adar et al., 2016). Récemment, la mise au point de techniques permettant de localiser les lésions à l'échelle du génome a mis en évidence d'autres facteurs influençant la cinétique de réparation, tel que le contexte chromatinien (Adar et al., 2016). Ainsi, le taux de réparation est plus rapide dans les régions hypersensibles à la DNase et plus lent au niveau de l'hétérochromatine.

e. Voies de signalisation activées en réponse aux UV

Après irradiation aux UV, alors que la machinerie de réparation s'affaire à réparer les dommages à l'ADN générés, plusieurs voies de signalisation sont activées pour moduler d'autres processus cellulaires. La réponse aux dommages à l'ADN consiste sommairement en l'activation des processus de réparation, d'arrêt du cycle cellulaire et éventuellement d'entrée en apoptose. ATM et ATR sont deux kinases de la famille PI(3)K qui jouent des rôles pléiotropes durant la réponse aux dommages à l'ADN. Lors de la réponse aux UV, ATR est rapidement activée par les fragments d'ADN simple-brin générés par le NER (Munoz et al., 2017; Stiff et al., 2008; Vrouwe et al., 2011), tandis qu'ATM semble être activée par ATR elle-même, ou bien via les R-loop générées au niveau des ARN Pol II bloquées par des lésions sur les régions transcrites (Stiff et al., 2006; Tresini et al., 2015). ATR et ATM phosphorylent leurs cibles respectives, Chk1 et Chk2 (Checkpoint kinase 1 et 2, dès 2h +UV), qui elles-mêmes phosphorylent d'autres protéines afin d'activer les points de contrôle du cycle cellulaire (Borisova et al., 2018; Ciccia and Elledge, 2010). En réponse aux UV, ATR permet notamment d'activer rapidement le point de contrôle G2/M pour empêcher les cellules d'entrer en mitose (Stiff et al., 2008). Le facteur de transcription p53 est activé via la cascade de phosphorylation ATM-dépendante (Riley et al., 2008). p53 joue un rôle clé dans le destin cellulaire en réponse aux dommages en régulant l'expression de nombreux gènes, dont les produits permettront de retrouver l'homéostasie cellulaire ou au contraire d'activer les processus d'apoptose ou de sénescence (Ciccia and Elledge, 2010; Giono et al., 2016).

Une autre voie de signalisation clé activée lors de la réponse aux dommages à l'ADN est la voie MAPK (mitogen-activated protein kinase) p38 (Cuadrado and Nebreda, 2010). Similairement à la voie ATR/ATM, l'activation de la voie MAPK p38 induit une cascade de phosphorylation impliquée dans l'activation des gènes de réponse précoce (Bahrami and Drablos, 2016). Récemment, il a été montré que MAPK p38 était rapidement et transitoirement activée après irradiation aux UV et induisait de nombreux événements de phosphorylation impliqués dans l'activation transcriptionnelle (Borisova et al., 2018; Bugai et al., 2019). Des sites de phosphorylation p38-dépendants ont notamment été identifiés sur la sous-unité NELF-E du complexe NELF, au niveau des Ser49, 51, 115 et 251 (Borisova et al., 2018). Les phosphorylations de NELF-E après UV servent de point d'ancrage pour la protéine 14-3-3, qui permet sa dissociation de la chromatine et ainsi le relargage de la pause transcriptionnelle à proximité du promoteur. L'inhibition de P-TEFb, qui phosphoryle également NELF-E mais au niveau d'autres résidus, par traitement au DRB entraine une augmentation de cette interaction. L'activation de la voie p38 en réponse aux UV induit également la phosphorylation de la protéine Rbm7 (Blasius et al., 2014; Borisova et al., 2018). Très récemment, il a été montré que cette phosphorylation était importante pour la libération de P-TEFb de la RNP 7SK après irradiation, et l'activation subséquente de gènes de réponse p53-dépendants (Bugai et al., 2019). Comme ATR, p38 MAPK joue aussi un rôle important dans l'activation du point de contrôle G2/M après irradiation aux UV (Bulavin et al., 2001).

ii. Reprogrammation transcriptionnelle lors de la réponse aux UV

En plus des conséquences directes sur la progression des ARN Pol II bloquées au niveau des lésions à l'ADN, les UV activent une signalisation menant à l'extinction globale de la transcription par l'ARN Pol II. Ce mécanisme apparait indépendant des dommages en *cis* : des extraits nucléaires de cellules irradiées ne permettent pas la transcription *in vitro* d'une matrice non endommagée (Rockx et al., 2000). La reprise de la transcription en revanche semble suivre la cinétique de réparation, et dans les régions 3' des brins transcrits, où la réparation est plus lente, la reprise de la transcription est également retardée (Andrade-Lima et al., 2015). Cela dit, alors que la transcription est globalement réduite après l'irradiation aux UV, certains gènes échappent à cette inhibition ou profitent d'un taux d'élongation réduit pour exprimer des isoformes différentes (Williamson et al., 2017). Les différents points de contrôle au cours de la transcription par l'ARN Pol II offrent de nombreuses opportunités de régulation, particulièrement sollicitées lors de la réponse au stress. Ainsi, divers mécanismes se mettent en place afin d'éteindre transitoirement l'expression globale des gènes tout en activant la transcription de certains. Le remodelage transcriptionnel rapidement induit par les UV est nécessaire pour préserver l'intégrité du génome et éviter la mort cellulaire. Une fois les lésions réparées, la transcription reprend son cours et les cellules retrouvent progressivement une activité normale.

a. Répression de l'initiation de la transcription par l'ARN Pol II

Un des effets notoires de l'irradiation aux UV est l'inhibition de l'initiation de la transcription par l'ARN Pol II (figure 21). Le facteur d'initiation TFIID/TBP est en effet séquestré au niveau des lésions, qui sont reconnues par TBP de manière similaire à son site de liaison usuel, la TATA-box (Coin et al., 1998; Vichi et al., 1997). Le facteur de transcription ATF3, dont le niveau augmente rapidement en réponse aux UV, empêche le recrutement de l'ARN Pol II sur les promoteurs de nombreux gènes présentant les sites de liaison CRE/ATF auxquels il s'associe (Epanchintsev et al., 2017; Kristensen et al., 2013). De plus, TFIIH, en plus d'être un facteur d'initiation, est réquisitionné pour le processus de réparation des lésions, comme vu dans la section précédente (Compe and Egly, 2012). Une disparition progressive de l'ARN Pol II et de TFIIH de la plupart des promoteurs est ainsi observée après l'irradiation (Gyenis et al., 2014). De même, la forme hypophosphorylée de l'ARN Pol II, la forme recrutée sur les promoteurs, disparait au profit de la forme hyperphosphorylée, qui correspond aux ARN Pol II en élongation et augmente avec les ARN Pol II bloquées sur les régions transcrites au niveau des lésions (Lavigne et al., 2017; Rockx et al., 2000). Enfin, une fraction conséquente des ARN Pol II bloquées sur les gènes après irradiation est également dégradée par le protéasome (Anindya et al., 2007; Munoz et al., 2009; Wilson et al., 2013). Un rôle de la phosphorylation de l'ARN Pol II dans ce processus a été proposé mais reste controversé et le mécanisme est peu détaillé (Mitsui and Sharp, 1999; Munoz et al., 2009; Starita et al., 2005; Wilson et al., 2013). Ainsi, il est encore incertain si l'hyperphosphorylation de l'ARN Pol II est un mécanisme actif permettant de prévenir l'initiation, d'induire sa dégradation ou bien s'il résulte simplement du blocage de l'ARN Pol II en élongation au niveau des lésions et de son taux d'élongation réduit.

b. Libération de l'ARN Pol II en pause à proximité du promoteur

Il a été récemment montré que les UV affectaient également la pause transcriptionnelle à proximité du promoteur, étape régulée par P-TEFb comme cela a été décrit dans les chapitres précédents (Borisova et al., 2018; Lavigne et al., 2017). Un relargage massif et rapide (dès 30 min) en élongation des ARN Pol II en pause à proximité du promoteur a en effet été observé en réponse aux UV, quels que soient la taille ou le niveau d'expression du gène (figure 21) (Lavigne et al., 2017). A noter que ce mécanisme contribue, avec l'inhibition de l'initiation, à la disparition du signal ARN Pol II observée aux promoteurs (Gyenis et al., 2014). Ainsi, les formes hyposphosphorylées et phosphorylées sur les Ser5 et Ser2 de l'ARN Pol II diminuent aux promoteurs, tandis que le niveau des formes phosphorylées, en particulier de la Ser2P-ARN Pol II, augmente dans le début des gènes. La libération massive de la pause de l'ARN Pol II à proximité du promoteur est bloquée par traitement au DRB, un inhibiteur de P-TEFb, et semble donc bien dépendre de l'activité de P-TEFb. La dissociation de NELF de la chromatine après

irradiation a également été reportée (Borisova et al., 2018). En outre, ce mécanisme a été proposé jouer un rôle capital dans la signalisation des lésions : la libération globale des ARN Pol II dans les corps de gènes permettrait d'augmenter leur probabilité de rencontrer des lésions et favoriserait ainsi leur prise en charge rapide par le TC-NER (Lavigne et al., 2017). Une accumulation des Ser2P-ARN Pol II au niveau des sites potentiels de lésions (dinucléotides TT) a effectivement été observée, et ce particulièrement dans la région 5' des gènes (Lavigne et al., 2017). De plus, l'analyse à l'échelle du génome des sites de lésion par purification et séquençage des produits d'excision (XR-seq) indique un biais de réparation pour ces régions là (Hu et al., 2015; Lavigne et al., 2017). A *contrario*, les lésions générées par les UV dans les régions 3' des longs gènes semblent être réparées moins vite et dépendre plutôt du GG-NER (Andrade-Lima et al., 2015). Ainsi, alors que la réparation des lésions sur les gènes est globalement compromise dans des cellules TC-NER-déficientes (cellules déplétées de CSB), la déficience du GG-NER (cellules déplétées de XPC) affecte la vitesse de réparation des lésions uniquement dans les régions 3' des longs gènes.

c. Restriction du taux d'élongation par l'ARN Pol II

Les ARN Pol II libérées en élongation et celles déjà présentes dans les corps de gènes progressent plus lentement dans les gènes après irradiation (figure 21) (Gregersen and Svejstrup, 2018; Williamson et al., 2017). En effet, alors que le taux de transcription est d'environ 2,5 kb/min en conditions normales, il tombe rapidement en dessous des 400 pb/min, dès les premières minutes après irradiation (Williamson et al., 2017). Bien que l'apparition de lésions sur les brins transcrits et le blocage de l'ARN Pol II impactent les taux d'élongation mesurés, ils ne semblent pas en être seuls la cause. L'inhibition de facteurs d'élongation pourrait aussi participer au ralentissement de l'ARN Pol II sur les gènes. Par exemple, le facteur d'élongation FACT est la cible du complexe E3-ligase UBR5, recruté à la chromatine en réponse aux UV (Sanchez et al., 2016). Des facteurs d'élongation négatifs, tels que RECQL5, pourraient aussi jouer un rôle dans la diminution du taux d'élongation (voir I.ii.h. L'élongation productive). Une autre hypothèse pour expliquer le ralentissement de l'ARN Pol II repose sur l'hyperphosphorylation de son CTD par une kinase encore non identifiée, potentiellement Cdk9, activée par la voie ATR (Munoz et al., 2017; Munoz et al., 2009).

D'un point de vue fonctionnel, la diminution du taux d'élongation contribue à la baisse globale d'expression des gènes observée en réponse aux UV, mais impacte également les processus co-transcriptionnels, et notamment l'épissage et la polyadénylation (Ip et al., 2011; Munoz et al., 2017; Munoz et al., 2009; Tresini et al., 2015; Williamson et al., 2017). Par exemple, des événements d'épissage alternatif induits par les UV, majoritairement des inclusions d'exons, sont observés et peuvent être récapitulés par des mutants « lents » de l'ARN Pol II (Munoz et al., 2017; Munoz et al.,

2009). Certains gènes, tel que bcl-x, passent de l'expression d'une isoforme pro- à une isoforme anti-apoptotique par épissage alternatif en réponse à la diminution du taux d'élongation après irradiation ou en présence de mutants lents de l'ARN Pol II (Munoz et al., 2009). De plus, le faible taux d'élongation après UV restreint la transcription de l'ARN Pol II aux 20-25 premiers kb des gènes, ce qui favorise l'expression des gènes courts mais aussi l'utilisation de sites de polyadénylations précoces et/ou l'inclusion de derniers exons alternatifs (Williamson et al., 2017). Ainsi, en réponse aux UV, une isoforme plus courte est exprimée à partir du gène codant la protéine ASCC3 par inclusion d'un dernier exon alternatif. Etonnamment, cette isoforme fonctionne en ARN non-codant et favorise la reprise de la transcription après irradiation aux UV (Williamson et al., 2017). Cet exemple illustre le potentiel d'action des isoformes générées en réponse aux UV dont l'étude est encore à ses balbutiements.

d. Activation des gènes de réponse

Alors que la transcription de la majorité des gènes est inhibée après irradiation aux UV, celle de gènes de réponse aux dommages à l'ADN, d'apoptose et de régulation de la prolifération est activée (Gyenis et al., 2014). Différents mécanismes s'opèrent donc sur les gènes activés et ceux réprimés. Ainsi, la transcription de gènes de réponse immédiate, tels que c-jun et c-fos, est rapidement induite après irradiation aux UV (15 min) et atteint son maximum en 30-60 min (Herrlich et al., 1994). Le facteur de transcription p53, dont la durée de vie augmente après UV, stimule le recrutement de l'ARN Pol II sur ses gènes cibles (Gyenis et al., 2014; Herrlich et al., 1994). Généralement, les gènes rapidement induits en réponse aux UV sont courts, quelques kb de long. En plus de permettre une expression plus rapide que des gènes longs *per se,* le risque d'être touché par une lésion est inhéremment plus faible pour les gènes courts (McKay et al., 2004). Ainsi, il a été montré qu'il existait une relation inverse entre le nombre de gènes p53-dépendants activés (6h après) en réponse aux UV et la dose d'UV administrée, de laquelle dépend le nombre de lésions. De plus, les gènes p53 dépendants activés à haute dose d'UV sont en moyenne plus courts que ceux activés à plus faible dose, et codent préférentiellement pour des protéines pro-apoptotiques. Cependant, la dose d'UV ne semble pas influer sur la liaison de p53 à ses gènes cibles, suggérant que la balance entre survie et apoptose est en réalité un mécanisme passif lié à la probabilité d'apparition des lésions sur les gènes.

e. Reprise générale de la transcription

La transcription retrouve un niveau d'activité normal environ 24h après irradiation (figure 21), dépendant évidemment de la dose d'UV administrée et du fonctionnement des voies de réparation (Andrade-Lima et al., 2015; Gentile et al., 2003; McKay et al., 2004). En effet, la reprise de la transcription suit chronologiquement l'éviction des lésions, et reste bloquée si les mécanismes de

réparation sont défectueux, comme en l'absence de la protéine CSB par exemple (Mayne and Lehmann, 1982; Proietti-De-Santis et al., 2006). Les mécanismes permettant la reprise de la transcription après l'extinction massive induite par les UV sont encore peu connus. Il est en effet difficile de discriminer les facteurs affectant la reprise de la transcription de ceux affectant la réparation, étant donné les conséquences évidentes d'un défaut de réparation sur la reprise de la transcription. Par exemple, CSA et CSB ont des rôles bien définis dans le processus de réparation, et leur déplétion empêche le redémarrage transcriptionnel (Mayne and Lehmann, 1982). Cependant, plusieurs études ont révélé leur implication dans la reprise de la transcription indépendamment de leur fonction dans le TC-NER (Epanchintsev et al., 2017; Fousteri et al., 2006; Kristensen et al., 2013). Ainsi, CSB, le complexe E3-ligase CRL4-CSA et l'ubiquitine-ligase MDM2 stimulent la dégradation par polyubiquitylation du facteur réprimant l'initiation ATF3, permettant alors la restauration du recrutement de l'ARN Pol II (Epanchintsev et al., 2017). De plus, CSA et CSB permettent le recrutement de TFIIS pour stimuler la reprise de la transcription des ARN Pol II arrêtées par des lésions et reculées (Dutta et al., 2015; Fousteri et al., 2006; Jensen and Mullenders, 2010; Tornaletti et al., 1999). La reprise de la transcription au niveau des ARN Pol II bloquées par des lésions semble également dépendre du facteur d'élongation ELL, recruté via la kinase Cdk7 du complexe TFIIH, mais son mécanisme d'action reste à définir (Mourgues et al., 2013). Enfin, plusieurs facteurs de remodelage d'histones, tels que FACT, DOT1L et HIRA, pourraient participer à la reprise de la transcription en induisant un environnement chromatinien favorable (Adam et al., 2013; Dinant et al., 2013; Oksenych et al., 2013).

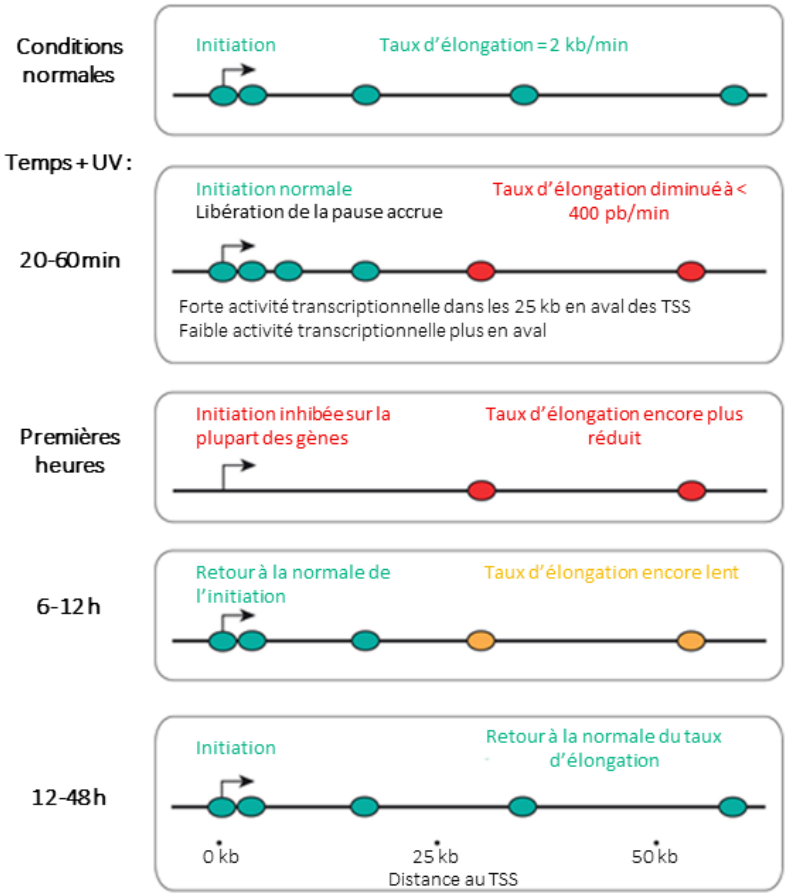


Figure 21 : Dynamique de la reprogrammation transcriptionnelle après irradiation aux UV (Gregersen and Svejstrup, 2018).TSS : transcription start site.

.Résultats

Préambule

Comme nous avons pu le voir dans l'introduction bibliographique, P-TEFb joue un rôle clé dans le processus de transcription par l'ARN Pol II. Son activité est en effet nécessaire pour le passage en élongation productive. En phosphorylant ses cibles au niveau du complexe d'élongation en pause à proximité du promoteur, P-TEFb participe à la formation d'un complexe d'élongation compétent. Le blocage de son activité inhibe globalement la transcription par l'ARN Pol II et donc la production d'ARN messagers. A l'inverse, une activité de P-TEFb accrue est également néfaste pour l'homéostasie cellulaire, soulignée par son implication dans le développement de différentes pathologies et notamment de nombreux cancers. Il est donc important de contrôler l'activité de P-TEFb en fonction des besoins cellulaires. La RNP 7SK joue un rôle majeur dans ce mécanisme en maintenant une fraction de P-TEFb dans un état inactif. Divers signaux cellulaires permettent la libération rapide de P-TEFb de la RNP 7SK, faisant de cette dernière un réservoir de P-TEFb pouvant être mobilisé par la cellule sur commande. Ces dernières années, l'identification de la RNP 7SK/P-TEFb associée à l'ARN Pol II et aux promoteurs de nombreux gènes a cependant remis en question son rôle dans la transcription par l'ARN Pol II. Ces nouvelles données tendent vers un modèle dans lequel la RNP 7SK/P-TEFb permettrait le relargage local de P-TEFb au niveau des régions promotrices afin de faciliter la libération de la pause transcriptionnelle. Mais ce modèle reste controversé au vu des divergences dans les données reportées. Par exemple, la proportion de RNP 7SK associée à la chromatine n'a pas été déterminée de manière consistante, et de grandes variations ont été rapportées, dépendant principalement de la technique employée (de quelques pourcents à plus de 50 %) (Ji et al., 2013; Mondal et al., 2010). Le nombre de gènes cibles de la RNP 7SK varie également considérablement entre les études (de ~ 2000 (Ji et al., 2013) à 12 000 (Flynn et al., 2016) dans les cellules embryonnaires de souris, de ~ 5000-6000 (Flynn et al., 2016) à plus de 15 000 (McNamara et al., 2016a) chez l'homme (cellules HeLa, H1 et HCT116 respectivement)). La contribution de la RNP 7SK/P-TEFb dans le contrôle transcriptionnel reste donc obscur. Alors que le niveau de P-TEFb séquestré au sein de la RNP 7SK dépend de l'état prolifératif de la cellule, des signaux extracellulaires peuvent moduler rapidement les quantités des P-TEFb actif par dissociation de la RNP 7SK/P-TEFb. A ce stade, de nombreuses zones d'ombre demeurent également. Les complexes P-TEFb libérés de la RNP 7SK en réponse à un stimulus participent certainement à la transcription de gènes de réponse, mais leur fonction reste globalement peu étudiée, étant indissociable de celle de la fraction active. D'ailleurs, il n'a pas été constaté si différents stimuli induisant la dissociation de la RNP 7SK induisent l'activation des mêmes gènes. En d'autres termes, P-TEFb libéré de la RNP 7SK est-il pris en charge par des facteurs de transcription ou cofacteurs stimulus-dépendant ou son rôle est-il plus global ?

Pour y voir plus clair dans la fonction de la RNP 7SK/P-TEFb dans les cellules, nous avons entrepris d'invalider l'expression de l'ARN 7SK dans des cellules humaines. Des déplétions transitoires de l'ARN 7SK, par siARN ou oligonucléotides antisens modifiés, ont été utilisés dans quelques études (Flynn et al., 2016; Ji et al., 2013; Liu et al., 2013). Mais ces modèles ne sont pas adaptés à l'étude de la fonction de la RNP 7SK/P-TEFb sur l'expression des gènes étant donné qu'ils induisent une libération massive de P-TEFb. Les effets de la perte de la RNP 7SK/P-TEFb ne peuvent donc pas être distingués des effets de la suractivation de P-TEFb. La déplétion prolongée de l'ARN 7SK déclenche cependant un mécanisme d'adaptation pour réguler l'activité de P-TEFb par diminution de l'expression de ses deux sous-unités, Cdk9 et Cycline T1 (Dai et al., 2014; Krueger et al., 2008). Nous avons donc opté pour une invalidation stable de l'ARN 7SK au moyen de la technologie d'édition du génome CRISPR/Cas9 (figure 1). Si la perte de l'ARN 7SK est viable, nous pourrions nous attendre à obtenir des cellules dans lesquelles la RNP 7SK/P-TEFb est incapable de se former sans pour autant induire la suractivation de P-TEFb via la diminution de l'expression de Cdk9 et Cycline T1. Une telle lignée cellulaire constituerait un modèle de choix pour étudier (1) le rôle de la RNP 7SK/P-TEFb sur l'expression des gènes en conditions normales (2) la fonction de la dissociation de la RNP 7SK/P-TEFb en réponse à un stimulus.

Comme modèle d'étude, nous nous sommes tournés vers les cellules HAP1, des cellules adhérentes dérivées de cellules de tumeurs myéloïdes chroniques (figure 22A, B). Ces cellules ont la particularité d'être haploïdes et sont donc particulièrement adaptées à l'utilisation du système CRISPR/Cas9 (figure 22B). De plus, les HAP1 ont la capacité de se diploïdiser spontanément après plusieurs passages successifs, générant un modèle d'étude plus physiologique (Essletzbichler et al., 2014). Nous nous sommes toutefois assurés que les mécanismes de régulation de l'activité de P-TEFb par la RNP 7SK étaient conservés dans cette lignée cellulaire (figure 22C, D). Des expériences de co-immunoprécipitation valident l'association des protéines Cycline T1, HEXIM1 et hnRNP A1 avec l'ARN 7SK (figure 22C). Les RNP 7SK/P-TEFb et 7SK/hnRNP se forment donc bien dans les cellules HAP1. Nous avons également vérifié que la dynamique de la RNP 7SK/P-TEFb en réponse à un stress était conservée en réalisant ces mêmes expériences de co-immunoprécipitation avant et après irradiation des cellules aux UV. Nous avons observé que l'interaction entre HEXIM1 et l'ARN 7SK diminue après irradiation aux UV, tandis que l'interaction entre hnRNP A1 et l'ARN 7SK augmente (figure 22D), comme cela a précédemment été montré dans d'autres lignées cellulaires (Barrandon et al., 2007; Hogg and Collins, 2007; Van Herreweghe et al., 2007). Ces données valident donc l'utilisation des cellules HAP1 pour étudier la fonction de la RNP 7SK/P-TEFb.

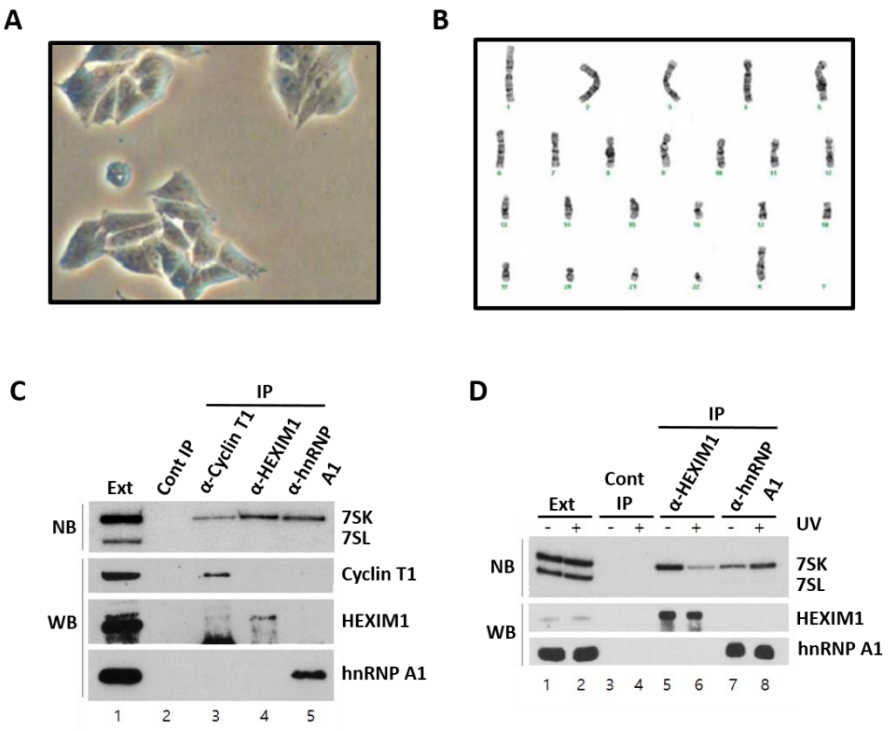


Figure 22 : Conservation du mécanisme de régulation de l'activité de P-TEFb dans les cellules humaines HAP1. (A) Cellules HAP1 observées en microscopie optique à contraste de phase et (B) caryotype de cellules HAP1 (Carette et al., 2011). (C) Immunoprécipitation (IP) des protéines Cycline T1, HEXIM1 et hnRNP A1 à partir d'extraits totaux de cellules HAP1. Les protéines et les ARN co-précipités ont été analysés par western blot (WB) et northern blot (NB), respectivement. Les IP contrôles réalisées sans anticorps sont indiquées (Cont IP). Ext, extraits totaux. (D) Les protéines HEXIM1 et hnRNP A1 ont été IP à partir d'extraits totaux de cellules HAP1 irrradiées (+) ou non irradiées (-) aux UV. Les protéines et les ARN co-précipités ont été analysés par WB et northern blot NB, respectivement. Les IP contrôles réalisées sans anticorps sont indiquées (Cont IP). Ext, extraits totaux.

Au cours de ma thèse, j'ai obtenu grâce au système d'édition du génome CRISPR/Cas9 différentes lignées cellulaires dépourvues de RNP 7SK/P-TEFb. Comme attendu, ces différentes lignées ne présentent pas d'activité accrue de P-TEFb, permettant ainsi d'étudier spécifiquement la fonction de la RNP 7SK/P-TEFb dans les cellules. J'ai ainsi pu évaluer l'effet de la perte de la RNP 7SK/P-TEFb sur la transcription par l'ARN Pol II en conditions normales et après un stress cellulaire. Mon travail de recherche s'est notamment focalisé sur la réponse transcriptionnelle en réponse aux UV. L'irradiation aux UV induit en effet la dissociation de la RNP 7SK/P-TEFb et permet l'activation des voies de réponse aux dommages à l'ADN, marquées par une reprogrammation transcriptionnelle dans laquelle P-TEFb a été impliqué (Bugai et al., 2019; Lavigne et al., 2017). Dans ce chapitre, vous retrouverez l'article scientifique résumant l'ensemble des travaux de recherche que j'ai effectué durant ma thèse. Ensuite, je reviendrai sur les résultats majeurs de cette étude et y apporterai des résultats et discussions complémentaires. En annexe, vous trouverez également la revue que j'ai rédigée au cours de ma thèse.

The 7SK/P-TEFb snRNP controls UV-induced transcriptional reprogramming.

Cécilia Studniarek[1], Michael Tellier[2], Pascal Martin[3], Shona Murphy[2], Tamás Kiss[1,4*] and Sylvain Egloff[1*]

[1]Laboratoire de Biologie Moléculaire Eucaryote du CNRS, UMR5099, Centre de Biologie Intégrative, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 9, France
[2]Sir William Dunn School of Pathology, University of Oxford, UK
[3]INRA UMR1332, ToxAlim, University of Toulouse, Toulouse, France
[4]Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary

*Corresponding authors: Tamás Kiss e-mail: tamas@ibcg.biotoul.fr
phone: (33)561335907
fax: (33)561335886
Sylvain Egloff e-mail: sylvain.egloff@ibcg.biotoul.fr
phone: (33)561335991

Running title: function of 7SK/P-TEFb snRNP during UV-induced stress response

Abstract

RNA polymerase II (RNAPII) pausing is a widely spread regulatory step of mRNA synthesis particularly important for the synchronized expression of stimulus-induced and developmental genes. The positive transcription elongation factor b (P-TEFb), comprising the Cdk9 kinase and a cyclin T, is required to convert paused RNAPII into an elongation-competent form. In human cells, a significant fraction of P-TEFb is kept in a catalytically inactive state, bound to the 7SK snRNA and the MePCE, LARP7 and HEXIM1 cellular proteins. Following stress, the 7SK-inactivated fraction of P-TEFb is rapidly released, yet the functional significance of increased Cdk9 activity during stress response remains unclear. Here, we generated human cell lines that lack 7SK snRNP and are therefore unable to sequester P-TEFb in an inactive state. Using mammalian nascent elongating transcript sequencing (mNET-seq), we show that lack of 7SK RNP impairs RNAPII pause release and transcriptional reprogramming that normally occur shortly after UV exposure. In particular, lack of 7SK snRNP lessens transcriptional activation of a set of key early-responsive genes and provokes cellular hypersensitivity to UV light, thus reflecting a weakened DNA-damage response. Mechanistically, we demonstrate that P-TEFb is specifically extracted from nucleoplasmic 7SK RNP after UV exposure and relocated onto chromatin, where it enhances escape of RNAPII from pause sites and recruitment of key RNAPII elongation factors on responsive genes. Altogether, our results show that P-TEFb activity that is stored within the 7SK snRNP is specifically exploited under stressful conditions for establishment of an effective stress response.

Keywords: P-TEFb; 7SK snRNA; transcriptional reprogramming; RNA polymerase II pausing; stress response

Introduction

Promoter-proximal pausing is a key rate-limiting regulatory step of the expression of most RNA polymerase II (RNAPII)-transcribed genes (Core and Adelman, 2019; Jonkers and Lis, 2015; Zhou et al., 2012). RNAPII pausing is particularly prominent on developmentally-regulated, stimulus-induced and signaling pathway genes (Adelman and Lis, 2012; Williams et al., 2015). After transcription initiation, the negative elongation factors DSIF (DRB-Sensitivity Inducing Factor) and NELF (Negative Elongation Factor) bind to RNAPII to block its progression along the gene. To release paused RNAPII and to support transcription progression through the entire gene body, the positive transcription elongation factor b (P-TEFb), a cyclin-dependent protein kinase composed of Cdk9 and cyclin T1 or T2 (Cyc T1 and T2) (Gressel et al., 2017; Jonkers et al., 2014; Peterlin and Price, 2006; Price, 2000), phosphorylates NELF, the Spt5 subunit of DSIF and the heptapeptide repeat ($Y_1S_2P_3T_4S_5P_6S_7$) of the C-terminal domain (CTD) of RNAPII at Ser2. Consequently, the phosphorylated NELF-P dissociates from the transcription machinery, DSIF-P functions as an elongation stimulating factor, and Ser2-P promotes downstream co-transcriptional events such as splicing and 3' end processing of the nascent pre-mRNA (Liu et al., 2015). Among these Cdk9-mediated phosphorylation events, Spt5 phosphorylation emerges as the decisive step for RNAPII pause release as it triggers NELF dissociation and resumption of transcription elongation (Cheng and Price, 2007; Lu et al., 2016; Yamada et al., 2006). Releasing paused RNAPII is now recognized as a central determinant of gene expression (AJ et al., 2016; Jonkers and Lis, 2015; Smith and Shilatifard, 2013; Zhou et al., 2012).

The activity of P-TEFb is controlled mainly by the 7SK small nuclear RNP (snRNP) that is composed of the 7SK small nuclear (sn)RNA, and two 7SK snRNP core proteins, Larp7 (La-related protein 7) and MePCE (methylphosphate capping enzyme) (Jeronimo et al., 2007; Krueger et al., 2008). The 7SK snRNP, together with HEXIM1/2, sequesters P-TEFb into large kinase-inactive 7SK/HEXIM/P-TEFb snRNP, hereafter called 7SK/P-TEFb snRNP (Nguyen et al., 2001; Yang et al., 2001). While MePCE and Larp7 stabilize 7SK snRNA (He et al., 2008; Jeronimo et al., 2007; Krueger et al., 2008; Markert et al., 2008), HEXIM1/2 binds to and inhibits the kinase activity of P-TEFb in a 7SK-dependent manner (Kobbi et al., 2016;

Michels et al., 2004). A number of environmental conditions and signaling pathways which globally impact cell growth or induce stress response trigger dynamic and reversible dissociation of the 7SK/P-TEFb snRNP, presumably to increase the pool of active P-TEFb available for transcription stimulation (AJ et al., 2016). Thus, the inactive 7SK/P-TEFb snRNP represents an important P-TEFb reservoir from which active P-TEFb can rapidly be mobilized in response to various transcriptional stimuli (Zhou et al., 2012).

Recruitment of active P-TEFb to transcriptionally active genes is directed by diverse transcription factors and RNA binding proteins, including the BET bromodomain protein Brd4 (AJ et al., 2016; Jang et al., 2005; Yang et al., 2005). Another fraction of P-TEFb associates with the super elongation complexes (SECs) which contain additional elongation factors, such as ELL, and therefore, stimulate RNAPII processivity very efficiently (He et al., 2010; Lin et al., 2010; Luo et al., 2012; Sobhian et al., 2010; Yokoyama et al., 2010; Zhou et al., 2012). Moreover, recent studies showed that the inactive 7SK/P-TEFb snRNP can also be loaded onto various gene promoters, where active P-TEFb is locally extracted and used for RNAPII pause release (D'Orso and Frankel, 2010; McNamara et al., 2013; McNamara et al., 2016). Genome-wide analysis detected enrichments of 7SK snRNP components at the promoter regions of most human protein-coding genes, arguing for widespread 7SK/P-TEFb snRNP anchoring to chromatin (Calo et al., 2015; Ji et al., 2013; McNamara et al., 2016). Coordinated 7SK/P-TEFb loading on selected promoters may stimulate rapid and synchronous transcriptional activation in response to environmental cues (Bugai et al., 2019; McNamara et al., 2013). Apparently, P-TEFb mobilization from the 7SK snRNP is fundamental to gene activation, but it is difficult to appraise the impacts of nucleoplasmic and chromatin-anchored 7SK/P-TEFb disassembly played in transcriptional activation of different genes (AJ et al., 2016).

Exposure to ultraviolet (UV) light provokes rapid P-TEFb release from the 7SK/P-TEFb inhibitory snRNP (Zhou et al., 2012; Zhou and Yik, 2006) and also induces P-TEFb-dependent transcriptional activation of key DNA damage response (DDR) genes (Bugai et al., 2019). In addition, UV-induced DNA damage induces a global release of paused RNAPII from promoter-proximal sites (Lavigne et al., 2017). This wave of elongating polymerases is thought to allow rapid and efficient lesion

detection by the transcription machinery and to facilitate repair by the transcription-coupled nucleotide excision repair (TC-NER) pathway. UV-released polymerases display lower elongation rates and their transcriptional range is restricted to the first 25 kilobases (Williamson et al., 2017). However, it is unknown whether UV-induced 7SK/P-TEFb dissociation, either in the nucleoplasm or on the chromatin, is connected to coordinated P-TEFb recruitment and global RNAPII pause release at transcribed genes, which may participate in DDR and maintenance of genome integrity.

To further investigate the transcriptional regulatory function of 7SK/P-TEFb snRNP, we have generated a human 7SK knock-out (KO) cell line that accumulates only the active form of P-TEFb. Thus, 7SK KO cells lack the ability (i) to load 7SK-associated inactive P-TEFb onto the chromatin and (ii) to rapidly increase P-TEFb activity under stress conditions. Unexpectedly, under normal growth conditions the 7SK KO cells show no significant deficiency in cell proliferation and they display only subtle changes in RNA steady-state levels. This suggests that 7SK/P-TEFb loading onto gene promoters may be a fine-tuning regulatory mechanism that is not critical for basal level of transcription. Mammalian native elongating transcript sequencing (mNET-seq) revealed that some key early-responsive genes which are readily induced in control cells by UV irradiation show impaired transcriptional activation in 7SK KO cells, indicating that proper activation of these 7SK-dependent genes relies on intact 7SK/P-TEFb snRNP. Mechanistically, P-TEFb extraction from 7SK/P-TEFb and its loading on selected genes as part of SECs stimulates RNAPII entry into productive elongation and subsequent recruitment of RNAPII elongation factors. In addition, the lack of 7SK snRNA impairs UV-induced release of ‘lesion-scanning’ polymerases into gene bodies, which may account for reduced viability of 7SK KO cells after UV irradiation (Lavigne et al., 2017). Altogether, our results demonstrate that the 7SK/P-TEFb snRNP is dispensable for cell growth, but it plays a key role in orchestrating transcriptional response to UV-induced stress.

Results

CRISPR/Cas9-mediated inactivation of human 7SK snRNA gene

The 7SK snRNA is an abundant and evolutionarily conserved non-coding RNA (Diribarne and Bensaude, 2009). The RNAPIII-mediated transcription of human 7SK gene is governed exclusively by upstream promoter elements, the TATA-box and the proximal and distal sequence elements (PSE and DSE) (Fig S1A). While the DSE only stimulates RNAPIII transcription, the TATA-box and PSE are crucial for RNAPIII transcription since they recruit essential transcription factors, the TATA box-binding protein (TBP) and the snRNA activating protein complex (SNAPc), respectively (Dergai and Hernandez, 2019). Besides one active 7SK gene, the human genome contains a plethora of 7SK pseudogenes which lack functional promoter elements. To avoid DNA breakage at all pseudogenes, human HAP1 cells (derived from chronic myelogenous leukemia) were transfected with a single CRISPR targeting the 7SK PSE element, disruption of which is expected to abolish SNAPc recruitment and RNAPIII transcription initiation (Fig S1A). As demonstrated by northern blot analysis and fluorescence *in situ* hybridization (FISH), interruption of the PSE regulatory sequence fully abrogated 7SK snRNA accumulation (Fig 1A and S1A). To exclude potential off-target effects of the CRISPR-Cas9 reaction, 7SK expression was rescued in the 7SK KO cells through stable re-insertion of the full-length 7SK gene, restoring 7SK accumulation at about 70% of the original level observed in HAP1 cells (Fig 1A, 7SK RST). To further corroborate the correctness and biological significance of our 7SK KO cell line utilized in this work, we also generated two alternative control 7SK KO cell lines with excision of the entire coding region of the HAP1 7SK gene (Fig S1B) or with disruption of LARP7 expression that reduced 7SK accumulation to about 3% of the wild-type level (Fig S1C).

To characterize P-TEFb-containing complexes in HAP1, 7SK KO and 7SK RST cells, total cell lysates were size-fractionated by velocity sedimentation on 10-40% glycerol gradients (Fig 1B). The gradients were fractionated into 11 fractions and distribution of P-TEFb (Cdk9 and Cyc T1), 7SK snRNP proteins (MePCE and Larp7) and HEXIM1 was monitored by western blot analyses. In the control HAP1 cells, about half of Cdk9 and Cyc T1 and a small proportion (~5-10%) of HEXIM1 were found in fractions 6–7 that had been reported to contain the large 7SK/P-TEFb snRNP containing HEXIM1 (large complex, LC) (Yang et al., 2001). The other half of Cdk9 and Cyc T1, corresponding to active free P-TEFb (small complex, SC) was detected in fractions 3-4. As expected, most of MePCE and a significant proportion of

Larp7 co-sedimented with P-TEFb in the large 7SK/P-TEFb complex (Krueger et al., 2008). The 7SK snRNA provides a central scaffold for 7SK/P-TEFb assembly. Consistent with this, the Cdk9, Cyc T1, HEXIM1, MePCE and Larp7 7SK-associated proteins showed altered sedimentation profiles in 7SK KO cells, with apparent shifts from the LC fractions (6-7) towards lighter fractions containing smaller particles. Importantly, the wild-type sedimentation properties of all 7SK proteins (Fig 1B) and the 7SK snRNA (Fig S1D) were restored in 7SK RST cells. The sedimentation profile of the highly abundant hnRNP A1 protein that interacts with a small fraction of 7SK snRNA was not detectably affected by 7SK loss (Barrandon et al., 2007; Van Herreweghe et al., 2007).

Co-immunoprecipitation experiments confirmed that the protein-protein interactions supporting 7SK/HEXIM/P-TEFb assembly in HAP1 cells are lost in 7SK KO and are reestablished in 7SK RST cells (Fig 1C). In the absence of 7SK, the interaction of HEXIM1 with P-TEFb, MePCE and Larp7, are fully abolished, confirming the 7SK-dependency of HEXIM1-mediated P-TEFb inhibition. Co-precipitation of Cyc T1 and Cdk9 confirms that 7SK KO cells express only active free P-TEFb, but as expected, they lack its 7SK/HEXIM1-associated inactive form.

Increased P-TEFb activity can lead to uncontrolled growth and tumorigenesis (He et al., 2008). Disruption of 7SK expression was expected to augment P-TEFb activity and to promote cell proliferation. However, cell counting (Fig 1D) and MTS cell proliferation assay (Fig S1E) failed to detect growth anomalies for 7SK KO cells, arguing for a compensatory mechanism that prevents P-TEFb over-activation upon 7SK depletion. Indeed, when the protein levels of cell extracts were adjusted to the GAPDH and RPL5 control proteins, we noticed that accumulation of both CycT1 and Cdk9 was significantly lessened in 7SK KO (Fig 1A and Fig S1B) and Larp7 KO cells (Fig S1C), while Larp7 and hnRNP A1 accumulation remained unaffected (Fig 1A). Importantly, the wild-type P-TEFb level was reestablished by restoration of 7SK expression in 7SK RST cells, demonstrating that down-regulation of P-TEFb subunits is directly linked to the intracellular level of 7SK snRNA. In addition to Cdk9 and Cyc T1 down-regulation, 7SK KO cells showed increased HEXIM1 accumulation, concurring with the observation that HEXIM1 transcription is upregulated upon 7SK depletion (Castelo-Branco et al., 2013) or after compound-induced disruption of the

7SK/P-TEFb snRNP (Liu et al., 2014). Thus, depletion of the 7SK/P-TEFb snRNP leads to the activation of HEXIM1 transcription through a yet unknown mechanism.

Human P-TEFb phosphorylates the C-terminal repeat 1 (CTR1) of Spt5 at Thr806 to potentiate the elongation stimulatory activity of DSIF (Larochelle et al., 2006; Sanso et al., 2016; Yamada et al., 2006). To compare Cdk9 activities in HAP1 and 7SK KO cells, we performed *in vitro* kinase assays by using recombinant Spt5 CTR1 (aa 720-830) as a substrate (Fig 1E) (Larochelle et al., 2006). Total extracts from HAP1 and 7SK KO cells were incubated with GST-fused CTR1 in the presence of radiolabeled ATP and Cdk9-dependent phosphorylation was measured by autoradiography following purification of GST-CTR1 on glutathione resin. Despite the reduced levels of Cdk9 and Cyc T1 observed in 7SK KO cells, the two extracts showed the same protein kinase activity targeting GST-CTR1, indicating that the levels of active P-TEFb in control and 7SK KO cells are highly comparable. Importantly, CTR1 phosphorylation was largely inhibited by pre-treatment of the cells with Cdk9 inhibitors such as DRB (Fig 1E, compare lanes 4 and 5, 9 and 10) or FP (Fig S1F), demonstrating that CTR1 phosphorylation was P-TEFb-dependent. Together, our data suggest that in the absence of 7SK snRNP, the 7SK KO cells reduce Cdk9 and CycT1 expression to prevent P-TEFb hyper-activation and uncontrolled cell proliferation.

Impact of 7SK snRNA loss on RNAPII transcription

Next, we assessed the effect of 7SK depletion on steady-state transcript levels using DNA microarray (MA) (Fig 2A). The lack of 7SK had a modest impact on the transcriptome, with only 485 misregulated genes representing about 2 % of the analyzed transcription units (FC≥1.5, pval<0.05). The relative accumulation levels of six selected mRNAs were confirmed by RT-qPCR (Fig 2B). Hierarchical clustering identified two major clusters corresponding to genes with increased (group 1) or decreased (group 2) expression in 7SK KO vs HAP1 cells (Fig S2A). Gene Ontology (GO) enrichment analysis showed that the 305 upregulated genes are enriched for genes involved in nervous system development (Fig S2B). Enrichment of functional categories among the 180 downregulated genes was consistent with a function in neuronal cells. Interestingly, recent studies highlighted a function for 7SK snRNA in

neural/neuronal differentiation (Bazi et al., 2018; Briese et al., 2016; Skreka et al., 2012) and axon elongation (Briese et al., 2018), which could potentially be linked to misregulation of these genes. Nevertheless, besides a relatively small number of affected genes, the RNA steady-state levels of most human genes remain unchanged in the absence 7SK snRNA, making it unlikely that the 7SK/P-TEFb snRNP significantly contributes to the output of a specific gene regulatory program in HAP1 cells under normal growth conditions.

Since P-TEFb regulates productive elongation, we next assessed the impact of 7SK loss directly on RNAPII transcription instead of measuring RNA steady-state levels. Mammalian native elongation transcript sequencing (mNET-seq) was conducted to determine the landscape of nascent transcription in HAP1 and 7SK KO cells at single-nucleotide-resolution. The assay is based on sequencing nascent RNAs associated with the catalytic center of transcriptionally engaged RNAPII. Regions harboring higher read counts reflect greater polymerase occupancies, making mNET-seq a method of choice to study transcription dynamics and RNAPII pausing genome-wide (Nojima et al., 2016; Nojima et al., 2015). Previously published RNA-seq data from HAP1 cells were used to divide protein-coding genes into 4 quartiles (Q1 to Q4) with increasing expression levels (Rodriguez-Castaneda et al., 2018)(GSE111272). mNET-seq signals around the Transcription Start Sites (TSSs) reproduced the characteristic promoter-proximal peaks with divergent transcription on the opposite strand (Fig 2C). A good correlation between RNA-seq and mNET-seq signals was observed, with highly expressed genes displaying higher mNET-seq signals downstream of TSSs. Thus, quantification of reads downstream of TSSs was used to compare transcription levels in HAP1 and 7SK KO cells (Fig 2D, left panel). This approach identified 1512 upregulated and 852 down-regulated genes (FC≥1.5) in 7SK KO cells (see Fig S2C for GO enrichment analysis). The higher number of affected genes revealed by mNET-seq (~15 % of analyzed genes), as compared to MA, likely reflects a better sensitivity of mNET-seq and a deeper analysis of HAP1-expressed genes by high-throughput sequencing. Importantly, a significant fraction of genes displaying altered mRNA steady-state level in MA analysis was found to be affected at the transcriptional level (144 out of 279 upregulated genes; 65 out of 179 downregulated genes) (Fig 2D, right panel). Accordingly, mNET-seq metaprofile analysis on genes which were up- and downregulated in MA confirmed the increased

and decreased level of RNAPII nascent transcription, respectively (Fig 2E). Misregulated genes show uniform changes in RNAPII occupancies at both TSS and gene body (GB), suggesting that 7SK snRNA modulates the level of RNAPII transcription rather than promoter-proximal pausing. This observation holds true for all misregulated genes in mNET-seq, as shown by metaprofiles and read quantification (Fig 2F and S2D). However, pausing index (PI), which measures the enrichment of RNAPII in the promoter-proximal region (TSS to +250 bp) relative to the gene body (Fig 2D), was significantly reduced in 7SK KO cells for the up-regulated genes (Fig S2E), suggesting that promoter-proximal pausing may also be perturbed. Thus, lack of 7SK impacts RNAPII transcription of a subset of HAP1 genes, through an unknown mechanism that may be indirect and primarily occurs at the initiation step.

7SK KO leads to a decrease in promoter-proximal pausing

mNET-seq revealed that transcription of the majority of HAP1 genes remains unaffected by 7SK loss. However, when meta-profiles for all RNAPII and heatmaps over the TSS were generated, we noticed that RNAPII occupancy is globally reduced at promoter-proximal regions in the 7SK KO cells in two biological replicates (Fig 3A, S3A and S3B). In contrast, RNAPII densities in the gene body (GB) and beyond the Transcription End Site (TES) are highly comparable in 7SK KO and HAP1 cells (Nojima, 2015) (Fig S3C). These results point to a specific function of 7SK snRNP in RNAPII promoter-proximal pausing. Representative HAP1 and 7SK KO mNET-seq profiles with sharp mNET-seq peaks detected over TSS are shown for the *Ago4, Rtf2 and Wdr25* genes (Fig 3B and S3D). We performed chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR) to assess the interaction of RNAPII and the NELF pausing factor with the promoters of *STX7* and *HELQ* genes for which mNET-seq analysis detected reduced RNAPII occupancy at the TSS (Fig 3C, top panel), but not in the GB (Fig S3E). Both RNAPII and NELF-E occupancies at TSS were significantly reduced in 7SK KO cells as compared to HAP1 cells (Fig 3C, bottom panel), demonstrating that establishment and/or maintenance of RNAPII pausing requires 7SK snRNA. Pausing is a general feature of RNAPII transcription. Synthesis of long non-coding RNAs (lncRNAs), promoter upstream transcripts (PROMPTs) and transcribed enhancer RNAs (eRNAs) also experiences RNAPII

pausing (Bunch et al., 2016; Core et al., 2014; Henriques et al., 2018). Examination of these classes of non-coding genes in our mNET-seq data sets demonstrated that they all exhibit reduced RNAPII accumulation at TSS in 7SK KO cells compared to WT cells (Fig 3D and Fig S3F). This suggests that the mechanisms supporting RNAPII pause establishment/release are identical for protein-coding and non-coding genes and both are dependent on the 7SK/P-TEFb snRNP. Interestingly, we noticed that the highly paused genes are the most efficiently expressed ones (Fig 2C) and they are highly affected by 7SK loss (see Fig 3B, 3C and S3D). Genes displaying more reads at TSS showed the largest reduction in promoter-proximal pausing in the absence of 7SK (Fig 3E and S3G). It seems unlikely that reduced RNAPII pausing in 7SK KO cells is due to uncontrolled release of poised RNAPII into the GB, as no concomitant increase in RNAPII density was observed on the GBs of highly affected genes (Fig 3E and S3G). Thus, we conclude that 7SK snRNA helps to maintain stable pausing across the genome regardless of the gene types, with the effect being strongest for highly paused and highly expressed genes. Since HAP1 and 7SK KO cells possess the same level of P-TEFb activity (Fig 1E), we propose that reduced RNAPII pausing observed in 7SK KO cells is due to the lack of the 7SK/P-TEFb negative transcriptional regulatory snRNP.

The 7SK/P-TEFb snRNP promotes pause release following UV irradiation

The 7SK RNP keeps a significant fraction of P-TEFb in an inactive form that is prone to be rapidly mobilized under stress conditions. Exposure of human cells to UV light has been shown to release P-TEFb from the 7SK RNP within an hour (Nguyen et al., 2001). Accordingly, *in vitro* kinase assays detected a 1.6-fold increase in CDK9 activity one hour after UV exposure (40J/m2), which was fully abrogated in 7SK KO cells (Fig 4A). Thus, the 7SK/P-TEFb RNP provides active P-TEFb under stress conditions and the 7SK KO cells are useful to study its function during stress response. We treated HAP1 and 7SK KO cells with UV and monitored changes in nascent transcription by mNET-seq during the early recovery phase (0-4 h after UV irradiation). In HAP1 cells, we observed a pronounced depletion of actively transcribing RNAPII at both TSS and GB at 2 and 4 hours post-UV (Fig S4A, upper panel), likely reflecting a global shut-down of transcription at the initiation stage

(Rockx et al., 2000). Blocking transcription after exposure to UV light presumably limits production of aberrant transcripts and facilitates lesion repair (Giono et al., 2016). Transcription initiation is still timely inhibited in the 7SK KO cells (Fig S4A, bottom panel), demonstrating that the 7SK/P-TEFb complex is not required for transcriptional repression after UV treatment. However, we noticed a profound impact of 7SK loss one hour after irradiation, when active P-TEFb is normally released from 7SK/P-TEFb, as highlighted by PI ratio calculation from two independent mNET-seq experiments (Fig 4B and S4B). In HAP1 cells, UV exposure released RNAPII from promoter-proximal sites, leading to a consequential increase of elongating polymerases in the proximal region of GB (Lavigne et al., 2017) (Fig 4C). To visualize transition from pausing to elongation, we analyzed RNAPII pausing based on PI calculation on individual genes (Fig 4D). As expected, UV treatment caused a marked decreased in PI (RNAPII depletion at promoters relative to GB), which was strongly attenuated in the 7SK KO cells. Indeed, release of poised RNAPII was substantially reduced in 7SK KO cells, as exemplified by metaprofiles (Fig 4C and S4C), heatmaps over TSS (Fig 4E), and the mNET-seq profile on the HELQ gene (Fig S4D). Importantly, the PI changes in 7SK KO cells resulted from both higher mNET-seq signals at TSS and reduced signals within the GB (Fig 4F and S4A). We conclude that the 7SK/P-TEFb RNP is necessary to promptly release paused RNAPII into productive elongation after UV exposure. This UV-induced transcriptional ‘wave’ has previously been shown to enhance detection and repair of lesions within transcribed regions (Lavigne et al., 2017). Since the propagation of transcription wave is reduced in 7SK KO cells (Fig 4C), they may display decreased DNA repair efficiency and higher mutational rates.

The 7SK/P-TEFb snRNP is required for UV-induced transcriptional reprogramming

UV exposure induces dramatic transcriptional alterations, the mechanisms of which remain poorly defined. While transcription is globally repressed (Fig S4A), some key responsive genes are quickly induced to enable a pro-survival transcriptional response that is crucial for cell fate (Christmann and Kaina, 2013). mNET-seq identified 2998 genes up-regulated at least 2 folds one hour after UV exposure.

Repression of transcription occurs later, spreading to around half of the 14000 analyzed protein-coding genes after 4 hours (Fig 5A). Again, 7SK loss is mostly materialized one hour after UV irradiation, with a significant fraction of early-responsive genes showing impaired transcriptional activation in 7SK KO cells, regardless of the activation level (Fig 5B). We identified at least 498 genes with highly compromised induction in absence of 7SK. Importantly, these genes encode key factors implicated in apoptosis, cell proliferation, signaling pathways and cellular response to stimuli, as shown by GO enrichment analysis (Fig S5A). As an example, the mNET-seq profile of the defective Gadd45b DNA-damage response gene is shown in Fig 5C. The transcriptional activation defect on this gene was confirmed by RNAPII ChIP experiment in different types of 7SK KO cells (Fig 5D and S5B). Analysis of selected mRNAs and their pre-mRNAs by qRT-PCR shows reduced inductions in 7SK and Larp7 KO cells (Fig 5E, S5C and S5D). Metaprofile analysis demonstrates that the transcriptional defect observed for the 498 non-induced genes arises, at least in part, from default RNAPII pause release (Fig 5F and S5E). This suggests that pause release, besides allowing RNAPII to scan DNA lesions, is also critical for expression of early-responsive genes. Interestingly, the transcriptional wave emanating from RNAPII pause release has a reduced elongation speed (400 bp/min instead of 2 kb/min) and therefore, it favors transcription of shorter genes (Williamson et al., 2017). We found that genes which are readily induced in response to UV appear to be shorter as they are strongly induced (Fig 5G). The short average length of early responsive genes may allow their expression by slowly elongating RNAPII and it can reduce the probability of RNAPII encountering bulky photo-products that interfere with its progression. Theoretically, defective detection of DNA lesions and reduced expression of early-responsive genes after UV could have deleterious effects on DNA damage response efficiency and cell fate. We thus tested the ability of HAP1 and 7SK KO cells to recover from lower dose of UV light (10 J/m2). As expected, cells lacking 7SK RNA failed to resume growth 3 days after irradiation whereas HAP1 and 7SK RST cells properly recovered from UV-induced stress (Fig 5H). Thus, we conclude that lack of 7SK RNA provokes cellular hypersensitivity to UV light, which is likely due to compromised P-TEFb-dependent transcriptional reprogramming.

The 7SK/P-TEFb snRNP stimulates UV-dependent recruitment of key elongation factors

We monitored the occupancy of key RNAPII-associated factors on the Gadd45b early-responsive gene by ChIP-qPCR, starting with RNAPII and the DSIF/NELF pausing factors. As mentioned before, we detected a lower level of NELF-E near TSS in 7SK KO cells under normal growth conditions (see Fig 3C, 6A and S6A), and an efficient promoter-proximal clearing of NELF-E after UV irradiation in both cell lines (Fig 6A and S6A). The same pattern and the same dissociation ratio were observed for two other early-responsive genes (Fig S6A). This result demonstrates that dissociation of NELF from chromatin still occurs in the absence of the 7SK/P-TEFb snRNP. Profiles of RNAPII and the Spt5 subunit of DSIF along the Gadd45b transcribed region were highly similar (Fig 6B). In HAP1 cells, the level of RNAPII and DSIF decreases at TSS and increases in the gene body after UV irradiation (Fig S6B), reflecting pause release and gene activation. This increase in GB was largely attenuated in 7SK KO cells for both RNAPII and DSIF (Fig 6B and S6B). Accordingly, the Spt5/RNAPII ratio is highly comparable in HAP1 and 7SK KO cells before and after UV treatment alongside the entire gene (Fig S6C), showing that association of DSIF with the transcriptional machinery is not affected by the lack of 7SK. We next attempted to analyze the recruitment of P-TEFb to Gadd45b gene, but ChIP with different anti-CDK9 were not successful in our hands. Since a vast majority of active P-TEFb is found within the super elongation complex (Luo et al., 2012) that delivers P-TEFb to active genes (Lin et al., 2010), we performed ChIP with an antibody against the AFF4 subunit of SEC as a proxy for P-TEFb analysis. After UV irradiation of HAP1 cells, we detected an increased AFF4 association with the promoter region, what was fully abolished in 7SK KO cells (Fig 6C and S6D). These results indicate that UV-induced SEC loading on early-response genes depends on the 7SK snRNP, and strongly suggest that defective P-TEFb recruitment is responsible for the weaker induction of these genes in 7SK KO cells.

We also followed the recruitment of two other key RNAPII-associated elongation factors, the PAF1 complex (PAF1C) and SPT6. Mammalian PAF1C consists in five subunits (Cdc73, Ctr9, Leo1, Paf1, and Ski8) and its role in regulation of promoter-proximal pausing remains controversial, since PAF1C has been reported to function both in enforcement of RNAPII pausing (Chen et al., 2015; Chen et al.,

2017) and in promotion of pause release through a P-TEFb-dependent recruitment (Vos et al., 2018; Yu et al., 2015). Interestingly, promoter-proximal occupancy of the Paf1 subunit on Gadd45b gene dramatically increased following UV irradiation of HAP1 cells, but this Paf1 recruitment was significantly reduced in 7SK KO cells (Fig 6D and S6E). Importantly, reduced PAF1C recruitment is not due to reduced RNAPII pausing, since paused RNAPII levels are approximately the same in HAP1 and 7SK KO cells after UV irradiation (Fig 6B). Therefore, on this class of RNAPII-transcribed genes, our data support a function of PAF1C in pause release rather than pause reinforcement. The same effect was observed for SPT6 (Fig 6D and S6E), recruitment of which is required both for escape of RNAPII from promoter pausing (Vos et al., 2018) and subsequent elongation (Endoh et al., 2004). Since the SEC, Paf1C and Spt6 positively regulate elongation and RNAPII progression beyond the pause release step (Hou et al., 2019), defective recruitment of these factors on early-responsive genes in 7SK KO cells could lead to reduced elongation rates and reduced synthesis of full-length transcripts (or prematurely terminated transcripts). Together, our results suggest that the 7SK/P-TEFb RNP stimulates expression of stress-response genes through providing additional P-TEFb, the lack of which impairs pause release and subsequent recruitment of elongation factors.

P-TEFb is extracted from nucleoplasmic 7SK/P-TEFb RNP during UV response

Defective activation of early responsive genes in 7SK KO cells arises from insufficient recruitment of P-TEFb to promoters (Fig 6C). The inactive 7SK/P-TEFb RNP has been proposed to be loaded onto gene promoters, where it is dissociated by specific transcription factors to trigger activation of various sets of genes (Gudipaty et al., 2015; McNamara et al., 2013). This model would allow synchronous gene activation in response to environmental cues such as UV irradiation. However, our attempts to localize the 7SK/P-TEFb RNP on early-responsive genes were unsuccessful. First, we failed to detect association of MePCE with the TSS of Gadd45b gene by ChIP (Fig 7A) while its occupancy at the TSS of 7SK gene was readily visualized, as previously shown (Xue et al., 2010). Association of MePCE with the 7SK snRNA gene was reduced after UV irradiation and it was fully abolished in 7SK KO cells (Fig S7A), likely mirroring the level of RNAPIII-mediated 7SK transcription. Surprisingly,

all tested protein-coding genes were negative for MePCE association at TSS, including hnRNP H1 where loading of the 7SK/P-TEFb RNP was previously reported in mouse (Ji et al., 2013) (Fig S7B). ChIP against two other proteins of the RNP, Larp7 and Hexim1, also came up negative (data not shown), suggesting that association of the 7SK/P-TEFb RNP with TSS of protein-coding genes cannot be visualized by ChIP in HAP1 cells. Therefore, we analyzed the association of 7SK RNA with chromatin before and after UV irradiation by performing chromatin isolation by RNA purification followed by sequencing (ChIRP-seq). After crosslinking and shearing of DNA, antisense biotinylated oligonucleotides covering the 7SK snRNA sequence retrieved around 75% of total 7SK RNA from HAP1 cells, as quantified by Northern blot (Fig S7C). The 7SK KO cell line was used as negative control for this experiment, with ChIRP-seq signals in this cell line considered as non-specific. Again, we failed to detect significant 7SK RNA association with protein-coding gene promoters genome-wide before or after UV exposure (Fig 7B and S7D), with the signal in HAP1 cells being even lower than in 7SK KO cells. The 498 genes which are not induced in the 7SK KO cells do not display specific association with 7SK RNA either (Fig 7C and S7E). In contrast, previously reported association of 7SK snRNP with snRNA genes (Egloff, 2017) and specific protein-coding genes such as EEF1A1 (Flynn 2016; Mumbach 2019) were readily detected in our ChIRP-seq analysis (Fig 7D). We conclude that the 7SK snRNA, similarly to the 7SK RNP proteins, MePCE and Larp7 (Egloff et al., 2017), does not associate with most protein-coding genes in HAP1 cells.

We next followed the recruitment/dissociation of P-TEFb and Hexim1 to chromatin after UV irradiation through cellular fractionation (Fig 7E). We noticed that both Cdk9 and CyclinT1 become more associated with the chromatin fraction one hour after UV exposure, demonstrating that some nucleoplasmic P-TEFb is recruited to genes. Importantly, UV-induced P-TEFb recruitment fully depends on 7SK RNA as it is fully abolished in the 7SK KO cells. In contrast to P-TEFb, the chromatin-associated level of Hexim1 remains unchanged after UV irradiation and it is not affected by 7SK loss, demonstrating that P-TEFb is not recruited in its inactive 7SK-associated form. This result implicates that active P-TEFb derived from the nucleoplasmic 7SK/P-TEFb RNP is specifically recruited to chromatin in response to stress. We then artificially destroyed the 7SK snRNP in HAP1 cells by knocking down

Larp7 and tested the induction of early-response genes after UV irradiation (Fig 7F). Surprisingly, cells treated with Larp7 siRNAs did not sustain full activation of the Gadd45b and Rnd1 early response genes, even though the levels of Cdk9 and Cyc T1 were not affected (Fig S7D). A previous study showed that Larp7 KD strongly increases CDK9 kinase activity in human cells (Ji et al., 2014). Therefore, our data suggest that an increase in P-TEFb activity is not sufficient to stimulate gene induction and demonstrate that the integrity of the 7SK/P-TEFb RNP is critical for efficient stress response.

DISCUSSION

Pausing during early elongation is a universally-experienced step of the RNAPII transcription cycle. Escape from promoter-proximal pausing is strongly dependent on CDK9 activity of P-TEFb that is primarily controlled by the 7SK snRNP. Together with HEXIM1, the 7SK snRNP sequesters active P-TEFb into transcriptionally inactive 7SK/P-TEFb RNP. Many cellular stresses liberate P-TEFb from 7SK/P-TEFb to activate CDK9 activity but, paradoxically, stress conditions often halt transcription. Therefore, the biological significance of the 7SK/P-TEFb complex remained difficult to ascertain. Here, we generated a 7SK KO human cell line that accumulates the active form of P-TEFb at normal levels (Fig 1E), but lacks its inactive 7SK-associated P-TEFb reservoir (Fig 1). We investigated the impact of the lack of the inactive fraction of P-TEFb on RNAPII transcription under normal and stress conditions.

Under normal growth conditions, the 7SK KO cells show no significant deficiency in cell proliferation and they display only subtle changes in RNA steady-state levels (Fig 2A). It suggests that 7SK/P-TEFb snRNP delivery onto promoter-proximal regions is not critical for basal transcription and may only fine-tune gene expression. Nevertheless, 7SK depletion reduced transcriptionally engaged RNAPII from the pause sites at most genes, with the strongest effects on highly expressed and paused genes. How the lack of 7SK snRNA impacts RNAPII pausing remains unknown, but it likely happens through regulation of P-TEFb activity. In theory, reduced pausing can result from a decreased rate of RNAPII initiation (Core and Adelman, 2019; Gressel et al., 2017). However, it is more probable that the lack of 7SK/P-TEFb influences pause establishment and/or release. Supporting this idea, at

TSS we detected lower levels of the NELF pausing factor, occupancy of which is directly regulated by P-TEFb activity (Core and Adelman, 2019). The Cdk9 activities are highly comparable in HAP1 and 7SK KO cells (Fig 1E), and the reduced mNET-seq reads at pause sites are not accompanied by increased reads in gene bodies (Fig 3E), arguing against increased pause release by P-TEFb. One possible explanation would be that recruitment of the inactive 7SK/P-TEFb RNP to pause sites would lengthen average pause duration on genes. In support to this hypothesis, published ChIP (Ji et al., 2013; McNamara et al., 2016) and 7SK ChIRP-seq (Flynn et al., 2016) detected low levels of 7SK RNP at most protein-coding gene promoters. However, our attempts to localize 7SK RNP components on protein-coding genes failed (Fig 7) (Egloff et al., 2017). Our ChIRP-seq data suggest that stable loading of 7SK snRNA onto genes is restricted to snRNA genes and few particular protein-coding genes (Fig 7D) (Egloff et al., 2017; Flynn et al., 2016; Mumbach et al., 2019). It is important to note that ChIRP-seq conducted in HAP1, but also in the 7SK KO cells, engender high background signal at TSS (Fig 7B and C), suggesting that open chromatin regions might occasion non-specific signals in crosslinking-based experiments. Thus, special care should be taken in analysis of ChIP/ChIRP data and the use of KO cell lines as negative controls is highly recommended for these experiments. In conclusion, while our results show that the 7SK/P-TEFb snRNP plays an important role in maintaining RNAPII pause at promoter-proximal sites, they argue against solid 7SK/P-TEFb loading onto protein-coding genes.

Our mNET-seq data highlighted a two-step mechanism during the early recovery phase from UV exposure, both of which depends on 7SK snRNP. Within the first hour, paused RNAPII is released from promoter-proximal sites and this is followed by a progressive transcriptional shutdown (Fig 4 and 5A). This timing certainly provides a window for transcribing early stress-response genes before transcriptional repression is fully established. As 7SK/P-TEFb disassembly occurs within the first hour post-UV, it is expected to influence early events of the UV transcriptional response. Indeed, we found that the 7SK/P-TEFb snRNP is required for rapid and synchronous escape of paused RNAPII from promoter-proximal regions following UV-exposure. This generates an elongation wave that was reported to promote uniform DNA-damage sensing by RNAPII on virtually all transcribed genes (Lavigne et al., 2017). RNAPII stalled at bulky DNA lesions can serve as a sensor of

damaged DNA that could initiate repair by the TC-NER system (Lindsey-Boltz and Sancar, 2007). The P-TEFb-induced RNAPII wave is restricted to a region of 25 kb post-TSS (Williamson et al., 2017) that primarily allows expression of short genes such as early-response genes (Fig 5G), activation of which is critical for efficient DNA damage response (Christmann and Kaina, 2013). The UV-induced elongation boost is largely attenuated in the 7SK KO cells, where additional P-TEFb cannot be provided to trigger transition into productive elongation, as shown by kinase assays (Fig 4A). As a result, some key early-response genes show impaired transcriptional activation in 7SK KO cells that certainly arises from lower release of elongating RNAPII (Fig 5F), but also from defective recruitment of key elongation factors such as the SEC, Paf1C and Spt6, as shown on the Gadd45b gene (Fig 6). The SEC is one of the most active P-TEFb-containing complexes that encompasses other elongation factors and that is required for rapid transcriptional gene induction (Luo et al., 2012). Paf1C was shown to help the release of paused RNAPII and also to positively regulate RNAPII elongation and progression beyond the pause release step (Hou et al., 2019). Finally, besides chaperoning histones during transcription, Spt6 stimulates transcription elongation *in vivo* (Ardehali et al., 2009). Structural studies have demonstrated that P-TEFb facilitates PAF1 binding that in turn, could help RNAPII release from pausing (Vos et al., 2018). This could explain why pausing is affected even though NELF is efficiently released from chromatin (Fig 6A). In addition, P-TEFb-dependent phosphorylation of the RNAPII CTD linker promotes SPT6 binding to the elongation complex (Vos et al., 2018). Thus, defective recruitment of these factors following UV irradiation could have synergic effects on RNAPII pause release and elongation rate. Since the elongation rate governs co-transcriptional alternative splicing (Munoz et al., 2009) and expression of alternative isoforms (Williamson et al., 2017), the impact of 7SK loss on transcription could even be reinforced by RNA processing defects.

Fractionation experiments showed that P-TEFb released from the 7SK/P-TEFb snRNP by UV exposure relocates to the chromatin (Fig 7E). Importantly, disassembly of the 7SK/P-TEFb snRNP in WT cells prior to UV irradiation also reduces transcriptional activation of responsive genes, demonstrating that this process necessitates an intact 7SK snRNP and that increased CDK9 activity is not sufficient to support gene activation (Fig 7F). Thus, specific transcription factors depending on

the actual stress stimulus extract active P-TEFb from the nucleoplasmic 7SK/P-TEFb snRNP and target it to selected genes. Both BRD4 (BET bromodomain protein 4) and RBM7 (RNA-binding motif protein 7) have been shown to release P-TEFb from the 7SK/P-TEFb RNP under stress conditions (Bugai et al., 2019; Yang et al., 2005), and are therefore attractive candidates for triggering UV-induced recruitment of Cdk9 to chromatin.

In summary, our data implicate relocation of P-TEFb from the nucleoplasmic 7SK/P-TEFb snRNP to the chromatin as a major regulatory step of UV-induced stress response, that is necessary to release genome-wide RNAPII pausing and to induce transcription of key responsive genes. Therefore, the 7SK/P-TEFb snRNP likely exerts genome-protective effects through promoting efficient DNA damage response that could be critical for cell fate, as suggested by the inability of the 7SK KO cells to recover normal cell growth after low dose of UV irradiation.

MATERIALS AND METHODS

General procedures. DNA and RNA manipulations were performed according to standard laboratory procedures. Oligodeoxynucleotides were purchased from Eurofins MWG Operon (sequences are available on request).

Cell culture. Human HAP1 cells were purchased from Haplogen Genomics and were grown in IMDM media supplemented with 10% (v/v) fetal bovine serum and 100 µg/ml penicillin and streptomycin (GIBCO, Invitrogen). Unless specified in the text, cells were UV-irradiated at 40 J/m2 with a stratalinker (Fischer Bioblock Scientific) using 254 nm wavelength lamp.

Genetic Inactivation of 7SK and LARP7 genes by CRISPR/Cas9. Overlapping oligonucleotides harboring the appropriate sgRNA target sequences were annealed for 5 min at 95° before slowly cooling down to room temperature. Annealed oligos were subsequently ligated (T4 DNA ligase, NEB) overnight at room temperature into BbsI-digested pX459 (Addgene plasmid #62988). Guide RNA/Cas9 expression vectors were transfected into HAP1 with Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. 24 h after transfection puromycin was added to the

cells at 2 µg/ml concentration to select for plasmid-expressing cells. After 48 h of puromycin selection, the medium was changed for non-selective conditions. Cells were left to recover for 48 h and individualized by the limiting dilution method. Monoclonal cells were screened by Northern Blot (for 7SK) or Western Blot (for Larp7), and the nature of the mutation was verified by sequencing.

Glycerol gradient analysis. HAP1, 7SK KO and 7SK RST cells were scraped, centrifuged at 2,000 rpm and lysed on ice in 10 mM HEPES, 150 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 0.1% NP40, complemented with PMSF and EDTA-free complete protease inhibitor cocktail (Roche). Cell were disrupted with a Bioruptor Plus Sonicator (Diagenode) by sonication five times for 30 s with 30 s interval at high setting and lysates were clarified by centrifugation for 10 min at 14,000 rpm. Extracts were loaded onto 10-40% glycerol gradients (10 mM HEPES, 150 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, PMSF and protease inhibitor cocktail) and gradients were run at 45,000 rpm for 16h in a Beckman SW41-Ti rotor before being fractionated.

siRNA-mediated knockdown. Control siRNA (D-001210-01) and siRNAs targeting Larp7 (M-020996-01) mRNA were purchased from Dharmacon and transfected into HAP1 cells by electroporation. HAP1 cells were transfected twice (at 0 and 48 h) with siRNAs and incubated for additional 24 h.

Microarray analysis. mRNA gene expression analysis was performed at the GeT-Trix facility (Toulouse) using Agilent SurePrint G3 Human GE microarrays V3 (8x60K) according to the manufacturer's instructions.

Chromatin fractionation. Approximately 8×10^6 cells were washed with ice-cold PBS, lysed with ice-cold 4 ml of HLB/NP40 buffer (10 mM Tris-Hcl, pH 7.5, 10 mM NaCl, 0.5% NP40, and 2.5 mM $MgCl_2$) and incubated on ice for 5 min. After incubation, 1 ml of ice-cold HLB/NP40/Sucrose buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.5% NP40, 2.5 mM $MgCl_2$, and 10% Sucrose) was under-laid and then the nuclei were collected under 1,400 rpm centrifuge at 4°C for 5 min. Isolated nuclei were resuspended in 125 µl of NUN1 solution (20 mM Tri-HCl, pH 8.0, 75mM NaCl, 0.5 mM EDTA, 50% Glycerol and proteinase inhibitor 1xComplete (Roche) followed by 1.2 ml NUN2 buffer (20 mM HEPES-KOH pH 7.6, 7.5 mM $MgCl_2$, 0.2 mM EDTA,

300 mM NaCl, 1 M Urea, 1% NP40, proteinase inhibitor 1xComplete and phosphatase inhibitor 1xPhosStop (Roche)). Fifteen minute incubation was carried out on ice with mixing by max speed vortex for 5 s every ~3 min and then chromatin pellets were precipitated under 13,000 rpm centrifuge at 4°C for 10 min. Chromatin pellet was resuspended in 500 µl NET-2 (50 mM Tris–HCl pH 7.5, 200 mM NaCl and 0.1% NP40) and solubilized by sonication with Bioruptor Plus (Diagenode) sonifier (high power setting, 10 cycles for 30 s on and 30 s off). The chromatin extract was clarified by centrifugation prior to SDS-PAGE analysis.

Chromatin immunoprecipitation. Formaldehyde cross-linking and IP of chromatin was performed essentially as described previously (Boyd et al., 1998). Briefly, HAP1 cells were cross-linked with 1% formaldehyde for 10 min at RT and quenched for 5 min by adding glycine to 125 mM final concentration. Cells were resuspended in Cell Lysis Buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP40 supplemented with protease inhibitor cocktail, Roche) and incubated on ice for 10 min. After collection by centrifugation at 2,300 *g* for 5 min, nuclei were lysed in Nuclei Lysis Buffer (50 mM Tris–HCl pH 8.1, 10 mM EDTA, 1% SDS) for 20 min. Chromatin was disrupted with Bioruptor Plus (Diagenode) sonifier (high power setting, 20 cycles for 30 s on and 30 s off) on ice. The extract was cleared by centrifugation and incubated with 2 µg of appropriate antibody overnight at 4°C. Immunocomplexes were immobilized by reacting with 10 µl of BSA-saturated Pansorbin cells (Calbiochem) for 15 min at 4°C. After extensive and stringent washing and elution, protein-DNA cross-links were reversed by incubation at 67°C for 5 h and DNA was deproteinized with proteinase K treatment at 50°C for 30 min. DNA was purified by the Nucleospin PCR cleanup kit of Macherey-Nagel. ChIP samples were directly analyzed by qPCR.

Mammalian Native Elongating Transcript sequencing (mNET-seq). mNET-seq was essentially carried out as previously described (Nojima et al., 2016). In brief, the chromatin fraction was isolated from $1x10^7$ HAP1 cells. Chromatin was digested in 100 µL of MNase (40 units/µL) reaction buffer for ~2 min at 37°C in a thermomixer (1,400 rpm). After addition of 10 µL EGTA (25mM) to inactivate MNase, soluble digested chromatin was collected by 13,000 rpm centrifuge for 5 min. The supernatant was diluted with 400 µL of NET-2 buffer and RNAPII antibody-conjugated beads were added. 10 µg of Pol II antibody was used per mNET-seq

experiments. Immunoprecipitation was performed at 4°C for 1 hr. The beads were washed with 1 mL of NET-2 buffer six times with 100 µL of 1xPNKT (1xPNK buffer and 0.05% Triton X-100) buffer once in cold room. Washed beads were incubated in 50 µL PNK reaction mix (1xPNKT, 1 mM ATP and 0.05 U/ml T4 PNK 3′phosphatase minus (NEB) in Thermomixer (1,400 rpm) at 37°C for 6 min. Beads were washed with 1 mL of NET-2 buffer once and RNA was extracted with Trizol reagent. RNA was suspended in urea Dye (7M Urea, 1xTBE, 0.1% BPB and 0.1% XC) and resolved on 6% TBU gel (Invitrogen) at 200 V for 5 min. In order to size select 30-160 nt RNAs, a gel fragment was cut between BPB and XC dye markers. 0.5 mL tube was prepared with 3-4 small holes made with 25G needle and placed in a 1.5 mL tube. Gel fragments were placed in the layered tube and broken down by centrifugation at 12,000 rpm for 1 min. The small RNAs were eluted from gel using RNA elution buffer (1 M NaOAc and 1 mM EDTA) at 25°C for 1 h in Thermomixer (900 rpm). Eluted RNA was purified with SpinX column (Coster) with 2 glass filters (Millipore) and the flow-through RNA was ethanol precipitated. mNET-seq libraries were prepared using TruSeq small RNA library preparation kit (Illumina, cat. no. RS-200-0012), and T4 RNA ligase, deletion mutant 2 (Epicenter, cat. no. LR2D1132K), according to Illumina instructions. 14 cycles of PCR were used to amplify the library. Before sequencing, the libraries were size-selected on a 6% TBE gel selecting only the 150-230 bp PCR products to exclude primer-primer ligated DNA. Gel elution was performed as described above. Deep sequencing (Hiseq4000, Illumina) was conducted at the Wellcome Trust Centre for Human Genetics (WTCHG), Oxford.

Chromatin isolation by RNA purification (ChIRP) and deep sequencing. Chromatin isolation by RNA purification experiments were performed essentially as described earlier (Chu et al., 2012). Briefly, HAP1 cells (~100 million) were harvested and cross-linked in PBS containing 1% of glutaraldehyde for 10 min at RT with continuous gentle agitation. The reaction was stopped by adding glycine to 125 mM final concentration. Cells were collected, washed with PBS, resuspended in Lysis Buffer (50 mM Tris-Cl pH 7.0, 10 mM EDTA, 1% SDS, protease inhibitors), and sonicated for 2 h in a 4°C water bath (Bioruptor Plus, Diagenode, 30 s pulses with 30 s intervals at high setting). Hybridization of 7SK- and control biotinylated oligonucleotides (~100 pmol/ml of chromatin) to cellular RNAs was performed at 37°C for 4 h in 50 mM Tris-Cl pH 7.0, 750 mM NaCl, 1% SDS, 1 mM EDTA, 15%

formamide supplemented with protease inhibitors. Biotinylated oligonucleotides were affinity-selected by Streptavidin magnetic beads (Dynabeads MyOne Streptavidin C1, 65002, Invitrogen) for 30 min at RT. Beads were extensively washed in 2× SSC, 0.5% SDS supplemented with protease inhibitors, and the bound chromatin was eluted in DNA Elution Buffer (50 mM $NaHCO_3$, 1% SDS). DNA was deproteinized with proteinase K treatment and purified by the Nucleospin PCR cleanup kit of Macherey-Nagel. ChIRP samples were analyzed by deep sequencing using Illumina HiSeq 4000 (Wellcome Trust Centre for Human Genetics, University of Oxford).

Immunoprecipitation and protein analysis. HAP1 were rinsed with PBS, scraped, and collected by centrifugation. Cells were placed on ice, resuspended in 0.5 ml of cold NET-2 buffer (50 mM Tris–HCl pH 7.5, 200 mm NaCl and 0.1% NP40), and disrupted with a Bioruptor Plus Sonicator (Diagenode) by sonication five times for 30 s with 30 s interval at high setting. Cell debris were removed by sedimentation at 16,000 *g* for 10 min, and the clarified cell extracts were incubated with 2 µg of antibodies coupled to 20 µl of packed protein A agarose beads (Sigma). After immunoprecipitation, beads were washed six times in NET-2 buffer and the associated proteins were eluted by 1× SDS–PAGE loading buffer. For IP, Western blot analysis and ChIP, the following antibodies were used: anti-RNAPII (A304-405A, Bethyl), anti-CDK9 (#2316S, CST), anti-Cyclin T1 (ab2098, Abcam and sc-271348, Santa Cruz), anti-AFF4 (A302-539A, Bethyl), anti-Spt5 (sc-133217, Santa Cruz), anti-PAF1 (ab137519, Abcam), anti-Spt6 (NB100-2582, Novus Biologicals), anti-H3 (ab1791, Abcam), anti-NELFe (HPA007187, Sigma), anti-HEXIM1 (ab25388, Abcam), anti-Larp7 (ab134746, Abcam), anti-hnRNP A1 (sc-32301, Santa Cruz), anti-GST (G7781, Sigma), anti-GAPDH (MAB374, Millipore) and anti-MePCE (ab185991, Abcam). Unless indicated differently by the manufacturers, antibodies were diluted to 1:1,000 for Western blot analyses. As secondary antibodies, horseradish peroxidase-conjugated anti-rabbit and anti-mouse IgGs (W401B and W402B, Promega) were used.

RNA analysis. RNAs from HAP1, cell extracts and protein A agarose beads were purified by the guanidine thiocyanate/phenol–chloroform RNA extraction procedure. For Northern blot analysis, RNAs were size-fractionated on a 6% sequencing gel, electroblotted onto a Hybond-N nylon membrane and probed with labeled

oligonucleotides. Real-time quantitative RT-PCR was performed on 2 µg of total RNA. DNA first strands were synthesized by GoScript reverse transcriptase (Promega), and qPCR reactions were performed on a BioRad C1000 Touch thermal cycler using the iQ SYBR Green Supermix of BioRad. The delta-delta Ct method was used for relative quantification of RNA accumulation. Sequences of the oligonucleotides used for Northern blot analysis and RT-qPCR are available on request.

In vitro kinase assay. HAP1 and 7SK KO cells were collected, resuspended in cold NET-2 buffer (50 mM Tris–HCl pH 7.5, 200 mm NaCl and 0.1% NP40) and disrupted with a Bioruptor Plus Sonicator (Diagenode) by sonication five times for 30 s with 30 s interval at high setting. Cell debris were removed by sedimentation at 16,000 *g* for 10 min, and protein concentrations in clarified cell extracts were quantitated by the Bradford method. Extract, in indicated amounts, were incubated with 1 µg of GST-fused Spt5 CTR1 substrate (Sanso et al., 2016) (provided by Dr R.P. Fisher) in 10 mM HEPES pH 7.9, 150mM NaCl, 1 mM DTT, 10 mM MgCl2, 100 µM unlabeled ATP and γ-32P-ATP (10 µCi). Reactions were incubated 15 min at 30°C, stopped by addition of 20 mM EDTA and incubated 1 h with 20 µl of Glutathione sepharose beads (GE Healthcare). Beads were washed five times in NET-2 buffer and purified GST-CTR1 was eluted by addition of 1× SDS–PAGE loading buffer. After size fractionation by SDS-PAGE and transfer onto a Hybond-C nylon membrane, phosphorylated GST-CTR1 was visualized by autoradiography and quantified by phosphorimager.

Fluorescence *in situ* hybridization. Fluorescence *in situ* hybridization, image acquisition, and processing have been described (Darzacq et al., 2002). The following oligonucleotide probe was used to detect HAP1 7SK snRNA: (5'-G*TGTCTGGAGTCTTGGAAGC*-3'). The aminoallyl-nucleotides labeled with FluoroLink Cy3 monofunctional reactive dye (GE Healthcare) are indicated by asterisks. Nuclear DNA was visualized with 1 µg/ml DAPI.

BIOINFORMATICS ANALYSIS

mNET-seq data processing. mNET-seq data were processed as previously described (Nojima et al., 2018): adapters were trimmed with Cutadapt in paired-end mode with the following parameters: -q 15, 10 **--**minimum-length 10 -a TGGAATTCTCGGGTGCCAAGG -A GATCGTCGGACTGTAGAACTCTGAAC. Trimmed reads were mapped to the human hg38 reference sequence with STAR (Dobin et al., 2013) and the parameters:

--runThreadN 16 --genomeLoad LoadAndKeep --readFilesCommand gunzip -c -k --limitBAMsortRAM 20000000000 --outSAMtype BAM SortedByCoordinate.

SAMtools (Li et al., 2009) was used to retain only properly paired and mapped reads (-f 3). A custom python script (Nojima et al., 2015) was used to obtain the 3' nucleotide of the second read and the strandedness of the first read. Strand-specific bam files were generated with SAMtools (-f 97 and -f 81). FPKM normalized bigwig files were generated for each bam files with deepTools2 (Ramirez et al., 2016) bamCoverage tool (-bs 1 –p max –normalizeUsingRPKM).

ChIRP-seq data processing. Adapters were trimmed with Cutadapt in paired-end mode with the following parameters: -q 15, 10 **--**minimum-length 10 -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC. Trimmed reads were mapped to the human hg38 reference genome with STAR and the parameters:

--runThreadN 16 --genomeLoad LoadAndKeep --readFilesCommand gunzip -c -k --alignIntronMax 1 --limitBAMsortRAM 20000000000 --outSAMtype BAM SortedByCoordinate.

Properly paired and mapped reads were filtered with SAMtools (-f 3). PCR duplicates were removed with Picard MarkDuplicates tool. FPKM normalized bigwig files were generated for each bam files with deepTools2 bamCoverage tool (-bs 10 –p max –normalizeUsingRPKM –e).

RNA-seq data processing. HAP1 poly(A)+ RNA-seq were retrieved from (Soneson et al., 2019), ArrayExpress number E-MTAB-7757. Adapters were trimmed with

Cutadapt in paired-end mode with the following parameters: --minimum-length 10 -q 15,10 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT. Trimmed reads were mapped to the human hg38 reference sequence with STAR and the parameters:

--runThreadN 16 --genomeLoad LoadAndKeep --readFilesCommand gunzip -c -k --limitBAMsortRAM 20000000000 --quantMode TranscriptomeSAM GeneCounts --outSAMtype BAM SortedByCoordinate.

SAMtools was used to retain only properly paired and mapped reads (-f 3). Quantification of the different mRNA isoforms from the RNA-seq data was performed with salmon version 0.13.1 (Patro et al., 2017) and the salmon quant command.

ChIRP-seq peak calling. Peak calling was performed with the MACS2 software version 2.1.1.20160309 (Zhang et al., 2008). The first step of peak calling was achieved by comparing each IP to their respective Input with the following command: macs2 callpeak -B -f BAMPE --nomodel --extsize 200 -g 2.9e9. To obtain the list of 7SK ChIRP-seq peaks, the peaks in HAP1 were compared to the peaks in 7SK-KO with the macs2 bdgdiff command and the parameters -g 60 -l 200.

Transcription unit (TU) definition. Gencode V29 annotation, based on the hg38 version of the human genome, was used to extract TUs. For each protein-coding gene, we defined the start and end sites as the TSS and the poly(A) site from the most expressed mRNA isoform defined in the HAP1 RNA-seq data. For meta-analyses, only non-overlapping protein-coding genes (with a 2.5 kb-free window before TSS and after poly(A) site) with a TU>2 kb and RNA-seq signal>0 were analyzed. PROMPTs were taken from (Chen et al., 2016). The eRNAs were taken from the PrESSTo database, which is part of the FANTOM5 project.

Reads quantification. Total read base count for mNET-seq data were computed with samtools bedcov tool using strand-specific bam files and normalized to 100 million paired-end reads and to the region's length. For ChIRP-seq, total read base

count were computed with samtools bedcov, normalized to 100 million paired-end reads, then the Input of the 7SK-KO signal were subtracted to the HAP1 IP signal and normalized to the region's length. Only the regions with a positive signal in UV- or UV+ were kept. For the samples having a signal ≤ 0 on the remaining regions, their values were put to the minimal value divided by two.

Metagene profiles. Metagene profiles were generated with deepTools2 computeMatrix tool with a bin size of 10 bp and the plotting data obtained with plotProfile –outFileNameData tool. Graphs representing the ChIRP-seq signal or the mNET-seq signal were then created with GraphPad Prism 8.1.1.

FIGURE LEGENDS

Figure 1. Characterization of the 7SK knock-out (KO) cell line. **(A)** RNAs and whole cell extracts isolated from HAP1, 7SK KO and 7SK-restored (RST) cells were analyzed by Northern Blot (NB) or Western Blot (WB), respectively. **(B)** Glycerol gradient (10-40 %) sedimentation analysis of HAP1, 7SK KO and 7SK RST cells. Distribution of Cdk9, cyclin T1, HEXIM1, MePCE, Larp7 and hnRNP A1 in the gradient fractions was monitored by western blot using antibodies to the proteins noted at the right. Ext, total extract; SC, small complexes; LC, large complexes. **(C)** Endogenous HEXIM1, Cyc T1 and CDK9 were immunoprecipitated in HAP1, 7SK KO and 7SK RST cells. Co-purification of 7SK-associated proteins (HEXIM1, Cyc T1, Cdk9, Larp7 and MePCE) was tested by western blotting. **(D)** Growth curves of HAP1, 7SK KO and 7SK RST cells. **(E)** Radioactive *in vitro* kinase assay using increasing amount of HAP1 or 7SK KO total extracts and recombinant Spt5-CTR1 fused to GST (GST-CTR1) as substrate. Cdk9-dependent phosphorylation was measured by autoradiography following purification of GST-CTR1 on glutathione resin. As a control, the kinase reaction was also performed with cells treated with DRB, a Cdk9 inhibitor. Levels of Cdk9, Cyc T1 and GAPDH in total extracts, as well as level of purified GST-CTR1, were analyzed by Western Blot. Radioactive signals were quantitated by phosphorimager from at least three independent experiments. $^{***}p < 0.001$, determined by Student's t-test. ns, not significant.

Figure 2. Impact of 7SK loss on gene expression. **(A)** Volcano plot of the microarray analysis. The vertical lines correspond to 1.5-fold up and down, and the horizontal cut-off delineates a p-value of 0.05. Red (up) and green (down) dots represents differentially expressed RNAs with statistical significance. **(B)** Relative expression of selected protein-coding genes in HAP1 and 7SK KO cells was determined by RT-qPCR. The signal were normalized to b-Actin mRNA levels that showed no significant change in microarray analysis. Bars indicate average values of three independent experiments with error bars representing standard deviations. $^{**}p < 0.01$; $^{***}p < 0.001$, determined by Student's t-test. **(C)** Metagene analysis showing transcriptionally engaged RNAPII occupancy on protein-coding genes measured by mNET-seq on sense and antisense strands in HAP1 cells. Genes were divided in four groups (Q1-Q4) based on their expression level in RNA-seq data. Read density (RPKM) of mNET-seq data was plotted across a 6 Kb window surrounding the Transcription Start Site (TSS). Horizontal dotted lines represent the average level of RNAPII promoter-proximal pausing for each gene set. **(D)** Left panel: depiction of pausing index (PI) calculation by ratio of reads density at TSS (0 to +250) over density in gene body (+250 to TES). Right panel: Venn diagrams comparing up- and down-regulated genes in 7SK KO cells in microarray and mNET-seq analyses, from a list of genes that are in common for both analyses. **(E)** Meta-analysis of mNET-seq signal over genes found to be 1.5-fold up- (left) and down-regulated (right) in 7SK KO cells in microarray analysis. FPKM, fragments per kilobase per million mapped reads. **(F)** Meta-analysis of mNET-seq over genes found to be up- (left) and down-regulated (right) after quantification of mNET-seq reads downstream of TSSs.

Figure 3. 7SK KO leads to a decrease in promoter-proximal pausing. **(A)** meta-analysis of mNET-seq over protein-coding genes in HAP1 and 7SK KO cells. Horizontal dotted lines represent the average level of RNAPII promoter-proximal pausing. **(B)** The mNET-seq positive strand reads are shown on the AGO4 gene. Peaks corresponding to promoter-proximal pausing are highlighted by shadowed box. **(C)** The mNET-seq negative strand reads are shown on the STX7 and HELQ gene (upper panel). Peaks corresponding to promoter-proximal pausing are highlighted by shadowed box. Quantitative ChIP analyses of endogenous RNAPII and NELF-E occupancy at TSS of HELQ and STX7 genes in HAP1 and 7SK KO cells is shown (bottom panel). $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$, determined by Student's

t-test. **(D)** meta-analysis of mNET-seq over lncRNA genes and PROMPTS (negative strand) in HAP1 and 7SK KO cells. Horizontal dotted lines represent the average level of RNAPII promoter-proximal pausing. **(E)** Box plots showing quantification of reads at TSS (left panel) and in gene bodies (right panel) in HAP1 and 7SK KO cells, measured by mNET-seq. Genes were sorted in three equal groups based on reads number at TSS in HAP1 cells (low, moderate and high). Genes displaying more reads at TSS showed the largest reduction in promoter-proximal pausing in the absence of 7SK. $^{*}p < 0.05$; $^{***}p < 0.001$, determined by Wilcoxon test. ns, not significant.

Figure 4. The 7SK/P-TEFb snRNP promotes pause release following UV irradiation. **(A)** Radioactive *in vitro* kinase assay using HAP1 or 7SK KO total extracts and recombinant Spt5-CTR1 fused to GST (GST-CTR1) as substrate. Cdk9-dependent phosphorylation was measured before and after UV irradiation by autoradiography following purification of GST-CTR1 on glutathione resin. Levels of Cdk9, Cyc T1 and GAPDH in total extracts, as well as level of purified GST-CTR1, were analyzed by Western Blot. Radioactive signals were quantitated by phosphorimager from at least three independent experiments. $^{**}p < 0.01$, determined by Student's t-test. ns, not significant. **(B)** boxplot analysis comparing PI ratio (+/- UV) between HAP1 (blue) and 7SK KO (red) cells 1, 2 and 4 hour after UV irradiation. $^{*}p < 0.05$; $^{***}p < 0.001$, determined by Wilcoxon test. **(C)** Metagene analysis showing transcriptionally engaged RNAPII occupancy on protein-coding gene before and after UV in HAP1 (left) and 7SK KO cells (right) measured by mNET-seq. **(D)** Cumulative distribution function analysis comparing PI in HAP1 (left) and 7SK KO (right) before and after UV exposure. **(E)** Heatmaps showing ratio of mNET-seq signals around TSS regions of protein-coding genes. Genes were sorted by decreasing read density at TSS. Color-scale depicts increased (in blue) or decreased mNET-seq signals (in red). **(F)** Violin plots depicting reads quantification in HAP1 (blue) and 7SK KO cells (red) at TSS and in gene body (GB). Thick dotted line represent the median, while the thin dotted lines represent the interquartile range (25-75%).

Figure 5. The 7SK/P-TEFb snRNP is required for UV-induced transcriptional reprogramming. **(A)** Bar charts showing the number of 2-folds induced and repressed genes 1, 2 and 4 hours after UV exposure in HAP1 cells, as assessed by mNET-seq. **(B)** Left panel: bar charts showing the number of 2-folds induced genes 1, 2 and 4

hours after UV exposure in HAP1 and 7SK KO cells. Right panel: bar charts showing the number of induced transcripts in HAP1 and 7SK KO cells 1h after UV irradiation. The degrees of induction are presented according to the legend. **(C)** Schematic diagram of the human Gadd45b gene. The corresponding mNET-seq positive strand reads are shown below. pA: polyA site. **(D)** Quantitative ChIP analyses of endogenous RNAPII occupancy on Gadd45b gene. Signals are represented as fold-change enrichment after UV (ratio +/-UV) for HAP1 and 7SK KO cells. Positions of the PCR-amplified (1)–(5) fragments are indicated below the schematic diagram of the human Gadd45b gene. $^{*}p < 0.05$; $^{**}p < 0.01$, determined by Student's t test. ns, not significant. **(E)** RT-qPCR of the indicated UV-induced mRNA (left) and their corresponding pre-mRNA transcripts (right). Results were normalized to the non-irradiated control and error bars represent standard deviations from at least three independent experiments. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$, determined by Student's t test. **(F)** Metagene analysis showing transcriptionally engaged RNAPII occupancy on 498 genes that are not properly induced in the 7SK KO cells after UV exposure, measured by mNET-seq. **(G)** Box plots showing that strongly UV-induced genes appears shorter. The fold-change induction (FC) and the number of analyzed genes are indicated below. $^{***}p < 0.001$, determined by Wilcoxon test. **(H)** Growth curves of HAP1, 7SK KO and 7SK RST cells after UV irradiation (10J/m2) at Day 0. Error bars represent standard deviations from three independent experiments.

Figure 6. The 7SK/P-TEFb snRNP stimulates UV-dependent recruitment of key elongation factors. **(A)** Quantitative ChIP analyses of endogenous NELF-E occupancy at TSS of Gadd45b gene in HAP1 and 7SK KO cells before and 1 hour after UV exposure. Signals are represented as % of input. Position of the PCR-amplified (2) fragment is indicated below the schematic diagram of the human Gadd45b gene in Fig 6B. In all panels: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$, determined by Student's t test. ns, not significant. **(B)** Quantitative ChIP analyses of endogenous RNAPII and Spt5 occupancies over the human Gadd45b gene in HAP1 and 7SK KO cells before and 1 hour after UV exposure. Positions of the PCR-amplified fragments are indicated below the schematic diagram of Gadd45b gene. **(C)** Quantitative ChIP analyses of endogenous AFF4 occupancy at TSS of Gadd45b gene in HAP1 and 7SK KO cells before and 1 hour after UV exposure. Other details are identical to panel (A). **(D)** Quantitative ChIP analyses of endogenous PAF1 and SPT6

occupancies at TSS of Gadd45b gene in HAP1 and 7SK KO cells before and 1 hour after UV exposure.

Figure 7. P-TEFb is extracted from nucleoplasmic 7SK/P-TEFb RNP during UV response. **(A)** Quantitative ChIP analyses of endogenous MePCE occupancy at TSS of 7SK and Gadd45b gene in HAP1 cells before and 1 hour after UV exposure. Signals are represented as % of input. $^{**}p < 0.01$, determined by Student's t test. ns : not significant. **(B)** Metagene analysis showing 7SK snRNA occupancy on protein-coding genes in HAP1 cells, as measured by ChIRP-seq. Signals detected in 7SK KO cells are considered as non-specific. FPKM, fragments per kilobase per million mapped reads; TSS, Transcription Start Site. **(C)** Metagene analysis showing 7SK snRNA occupancy in HAP1 cells on the 498 genes that are not properly induced after UV exposure in the absence of 7SK, measured by ChIRP-seq. **(D)** Upper panel: metagene analysis showing 7SK snRNA occupancy on snRNA genes before and after UV, as measured by ChIRP-seq. Bottom panel: ChIRP-seq reads in HAP1 and 7SK KO cells are shown for the Eef1a1 gene. **(E)** Chromatin fractionation in HAP1 and 7SK KO cells. Chromatin was purified before and after UV treatment (40 J/m2) and the abundance of the indicated proteins were tested by Western blotting. **(F)** RT-qPCR of Gadd45b and SGK1 pre-mRNA transcripts in control- and Larp7-depleted HAP1 cells, before and 1 hour after UV irradiation. Signals were normalized to b-Actin mRNA levels. Bars indicate average values of three independent experiments with error bars representing standard deviations. Levels of Larp7, Cdk9, Cyc T1 and GAPDH were verified by Western Blot (WB), while co-depletion of 7SK snRNA was assessed by Northern Blot (NB). $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$, determined by Student's t test.

Acknowledgments

We are indebted to Dr. R.P Fischer (Icahn School of Medecine at Mount Sinai, New York, USA) for providing us with the GST-CTR1 construct. Our work was supported by grants from INSERM Plan Cancer and Agence Nationale de la Recherche (to T.K.).

Author contribitions

SE, SM, and TK conceived and designed the study. CS and SE carried out all experiments. CS, MT and PM analyzed the mNET-seq and ChIRP-seq data. SE and TK wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature Reviews Genetics *13*, 720-731.

AJ, C.Q., Bugai, A., and Barboric, M. (2016). Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. Nucleic Acids Res *44*, 7527-7539.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J *28*, 1067-1077.

Barrandon, C., Bonnet, F., Nguyen, V.T., Labas, V., and Bensaude, O. (2007). The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. Mol Cell Biol *27*, 6996-7006.

Bazi, Z., Bertacchi, M., Abasi, M., Mohammadi-Yeganeh, S., Soleimani, M., Wagner, N., and Ghanbarian, H. (2018). Rn7SK small nuclear RNA is involved in neuronal differentiation. J Cell Biochem *119*, 3174-3182.

Boyd, K.E., Wells, J., Gutman, J., Bartley, S.M., and Farnham, P.J. (1998). c-Myc target gene specificity is determined by a post-DNAbinding mechanism. Proc Natl Acad Sci U S A *95*, 13887-13892.

Briese, M., Saal-Bauernschubert, L., Ji, C., Moradi, M., Ghanawi, H., Uhl, M., Appenzeller, S., Backofen, R., and Sendtner, M. (2018). hnRNP R and its main interactor, the noncoding RNA 7SK, coregulate the axonal transcriptome of motoneurons. Proc Natl Acad Sci U S A *115*, E2859-E2868.

Briese, M., Saal, L., Appenzeller, S., Moradi, M., Baluapuri, A., and Sendtner, M. (2016). Whole transcriptome profiling reveals the RNA content of motor axons. Nucleic Acids Res *44*, e33.

Bugai, A., Quaresma, A.J.C., Friedel, C.C., Lenasi, T., Duster, R., Sibley, C.R., Fujinaga, K., Kukanja, P., Hennig, T., Blasius, M*., et al.* (2019). P-TEFb Activation by RBM7 Shapes a Pro-survival Transcriptional Response to Genotoxic Stress. Mol Cell *74*, 254-267 e210.

Bunch, H., Lawney, B.P., Burkholder, A., Ma, D., Zheng, X., Motola, S., Fargo, D.C., Levine, S.S., Wang, Y.E., and Hu, G. (2016). RNA polymerase II promoter-proximal pausing in mammalian long non-coding genes. Genomics *108*, 64-77.

Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature *518*, 249-253.

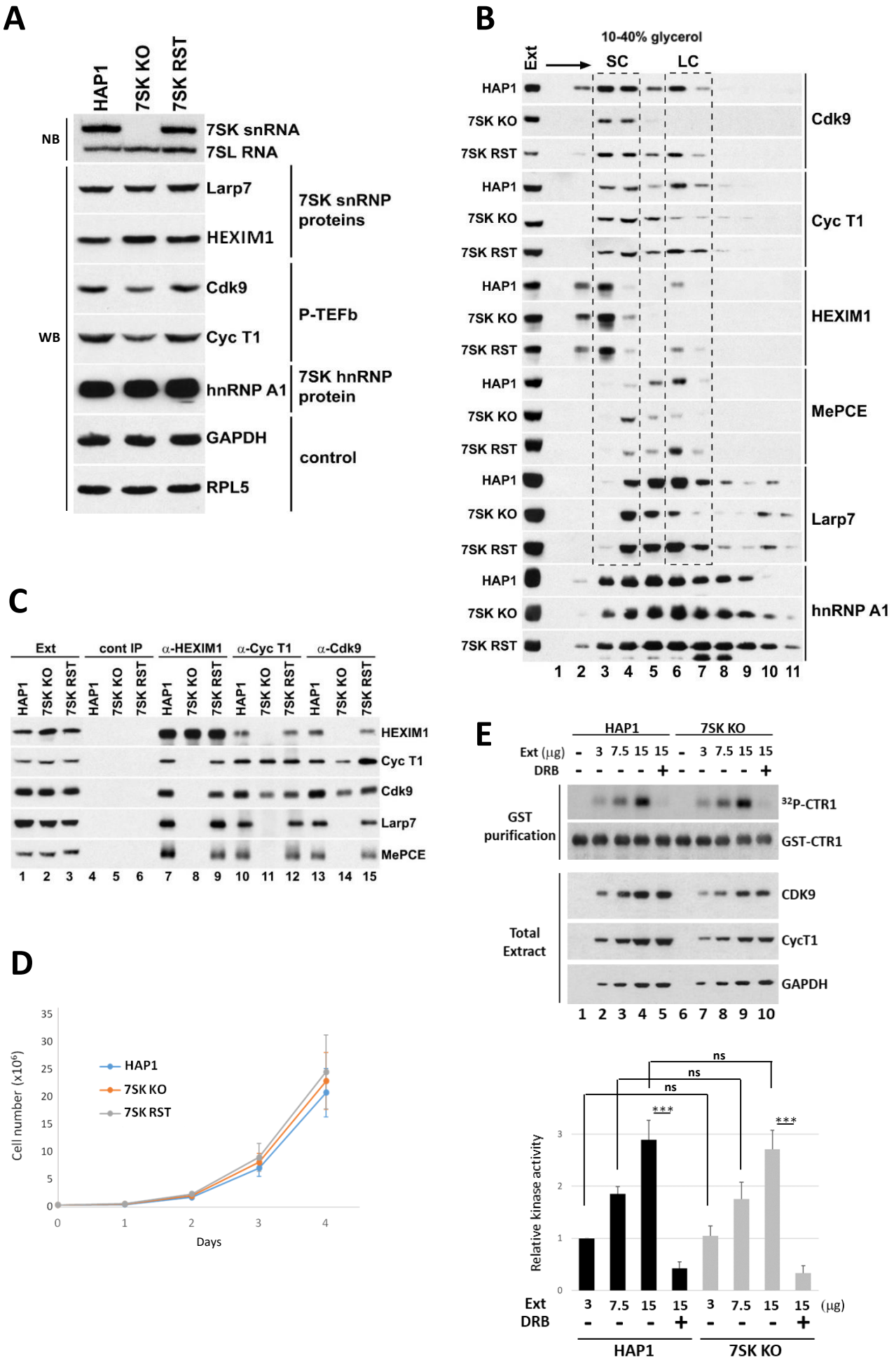
Castelo-Branco, G., Amaral, P.P., Engstrom, P.G., Robson, S.C., Marques, S.C., Bertone, P., and Kouzarides, T. (2013). The non-coding snRNA 7SK controls transcriptional termination, poising, and bidirectionality in embryonic stem cells. Genome Biol *14*, R98.

Chen, F.X., Woodfin, A.R., Gardini, A., Rickels, R.A., Marshall, S.A., Smith, E.R., Shiekhattar, R., and Shilatifard, A. (2015). PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. Cell *162*, 1003-1015.
Chen, F.X., Xie, P., Collings, C.K., Cao, K., Aoi, Y., Marshall, S.A., Rendleman, E.J., Ugarenko, M., Ozark, P.A., Zhang, A.*, et al.* (2017). PAF1 regulation of promoter-proximal pause release via enhancer activation. Science *357*, 1294-1298.
Chen, Y., Pai, A.A., Herudek, J., Lubas, M., Meola, N., Jarvelin, A.I., Andersson, R., Pelechano, V., Steinmetz, L.M., Jensen, T.H.*, et al.* (2016). Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. Nat Genet *48*, 984-994.
Cheng, B., and Price, D.H. (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.
Christmann, M., and Kaina, B. (2013). Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation. Nucleic Acids Res *41*, 8403-8420.
Chu, C., Quinn, J., and Chang, H.Y. (2012). Chromatin isolation by RNA purification (ChIRP). J Vis Exp.
Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev.
Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311-1320.
D'Orso, I., and Frankel, A.D. (2010). RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat Struct Mol Biol *17*, 815-821.
Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E., and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. EMBO J *21*, 2746-2756.
Dergai, O., and Hernandez, N. (2019). How to Recruit the Correct RNA Polymerase? Lessons from snRNA Genes. Trends Genet *35*, 457-469.
Diribarne, G., and Bensaude, O. (2009). 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. RNA Biol *6*, 122-128.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
Egloff, S., Vitali, P., Tellier, M., Raffel, R., Murphy, S., and Kiss, T. (2017). The 7SK snRNP associates with the little elongation complex to promote snRNA gene expression. EMBO J *36*, 934-948.
Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A.*, et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol Cell Biol *24*, 3324-3336.
Flynn, R.A., Do, B.T., Rubin, A.J., Calo, E., Lee, B., Kuchelmeister, H., Rale, M., Chu, C., Kool, E.T., Wysocka, J.*, et al.* (2016). 7SK-BAF axis controls pervasive transcription at enhancers. Nat Struct Mol Biol *23*, 231-238.
Giono, L.E., Nieto Moreno, N., Cambindo Botto, A.E., Dujardin, G., Munoz, M.J., and Kornblihtt, A.R. (2016). The RNA Response to DNA Damage. J Mol Biol *428*, 2636-2651.
Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. Elife *6*.
Gudipaty, S.A., McNamara, R.P., Morton, E.L., and D'Orso, I. (2015). PPM1G Binds 7SK RNA and Hexim1 To Block P-TEFb Assembly into the 7SK snRNP and Sustain Transcription Elongation. Mol Cell Biol *35*, 3810-3828.
He, N., Jahchan, N.S., Hong, E., Li, Q., Bayfield, M.A., Maraia, R.J., Luo, K., and Zhou, Q. (2008). A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. Mol Cell *29*, 588-599.

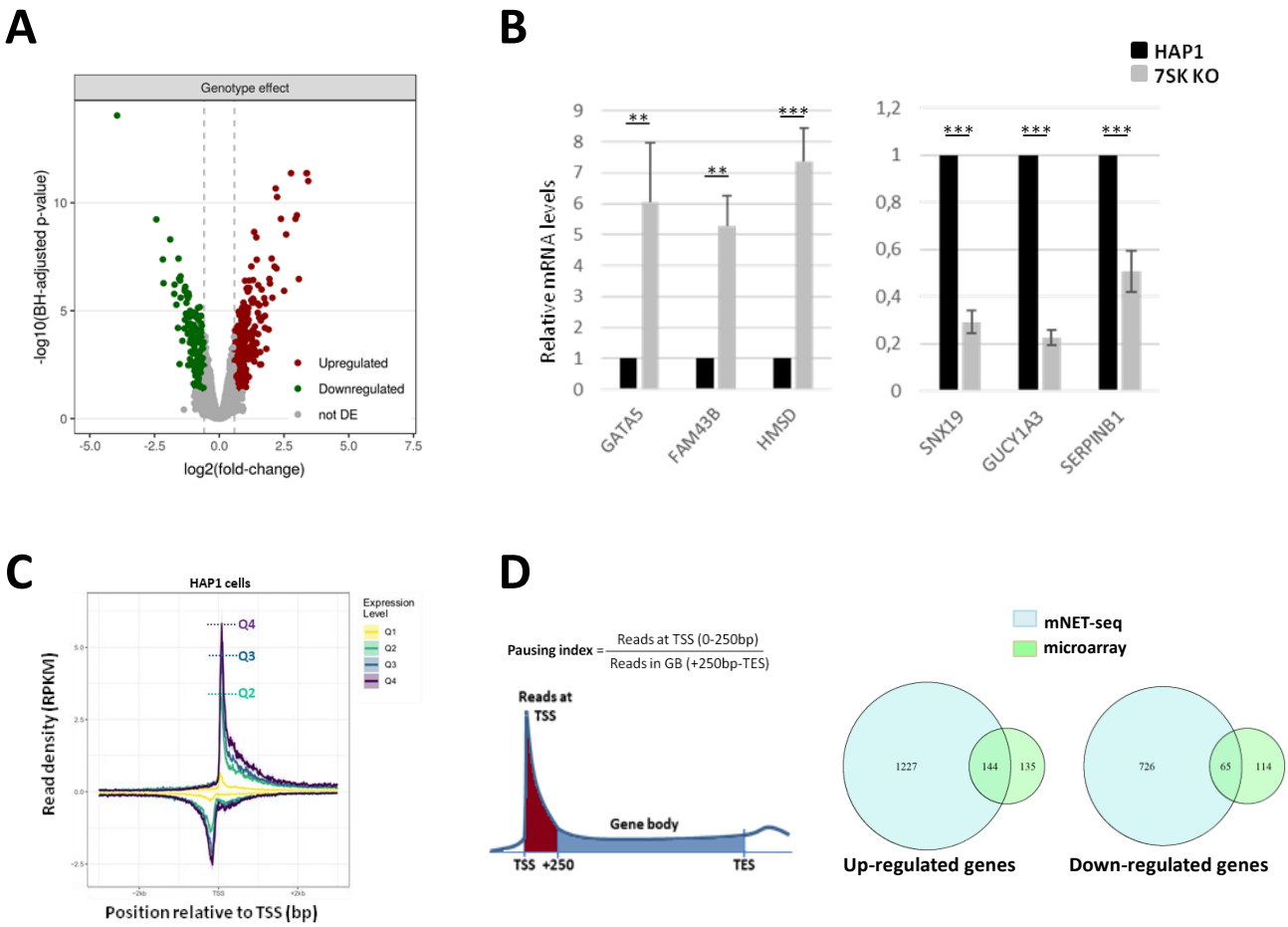
He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N.J., Alber, T., and Zhou, Q. (2010). HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. Mol Cell *38*, 428-438.
Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev *32*, 26-41.
Hou, L., Wang, Y., Liu, Y., Zhang, N., Shamovsky, I., Nudler, E., Tian, B., and Dynlacht, B.D. (2019). Paf1C regulates RNA polymerase II progression by modulating elongation rate. Proc Natl Acad Sci U S A *116*, 14583-14592.
Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell *19*, 523-534.
Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J.*, et al.* (2007). Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol Cell *27*, 262-274.
Ji, X., Lu, H., Zhou, Q., and Luo, K. (2014). LARP7 suppresses P-TEFb activity to inhibit breast cancer progression and metastasis. Elife *3*, e02907.
Ji, X., Zhou, Y., Pandit, S., Huang, J.E., Li, H.R., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.D. (2013). SR Proteins Collaborate with 7SK and Promoter-Associated Nascent RNA to Release Paused Polymerase. Cell *153*, 855-868.
Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407.
Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 167-177.
Kobbi, L., Demey-Thomas, E., Braye, F., Proux, F., Kolesnikova, O., Vinh, J., Poterszman, A., and Bensaude, O. (2016). An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb. Proc Natl Acad Sci U S A.
Krueger, B.J., Jeronimo, C., Roy, B.B., Bouchard, A., Barrandon, C., Byers, S.A., Searcey, C.E., Cooper, J.J., Bensaude, O., Cohen, E.A.*, et al.* (2008). LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic Acids Res.
Larochelle, S., Batliner, J., Gamble, M.J., Barboza, N.M., Kraybill, B.C., Blethrow, J.D., Shokat, K.M., and Fisher, R.P. (2006). Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics. Nat Struct Mol Biol *13*, 55-62.
Lavigne, M.D., Konstantopoulos, D., Ntakou-Zamplara, K.Z., Liakos, A., and Fousteri, M. (2017). Global unleashing of transcription elongation waves in response to genotoxic stress restricts somatic mutation rate. Nat Commun *8*, 2076.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
Lin, C., Smith, E.R., Takahashi, H., Lai, K.C., Martin-Brown, S., Florens, L., Washburn, M.P., Conaway, J.W., Conaway, R.C., and 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.
Lindsey-Boltz, L.A., and Sancar, A. (2007). RNA polymerase: the most specific damage recognition protein in cellular responses to DNA damage? Proc Natl Acad Sci U S A *104*, 13213-13214.
Liu, P., Xiang, Y., Fujinaga, K., Bartholomeeusen, K., Nilson, K.A., Price, D.H., and Peterlin, B.M. (2014). Release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein (snRNP) activates hexamethylene bisacetamide-inducible protein (HEXIM1) transcription. J Biol Chem *289*, 9918-9925.
Liu, X., Kraus, W.L., and Bai, X. (2015). Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. Trends Biochem Sci *40*, 516-525.

Lu, X., Zhu, X., Li, Y., Liu, M., Yu, B., Wang, Y., Rao, M., Yang, H., Zhou, K., Wang, Y.*, et al.* (2016). Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. Nucleic Acids Res *44*, 6853-6867.
Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. Nat Rev Mol Cell Biol *13*, 543-547.
Markert, A., Grimm, M., Martinez, J., Wiesner, J., Meyerhans, A., Meyuhas, O., Sickmann, A., and Fischer, U. (2008). The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. EMBO Rep *9*, 569-575.
McNamara, R.P., McCann, J.L., Gudipaty, S.A., and D'Orso, I. (2013). Transcription factors mediate the enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for transcription elongation. Cell Rep *5*, 1256-1268.
McNamara, R.P., Reeder, J.E., McMillan, E.A., Bacon, C.W., McCann, J.L., and D'Orso, I. (2016). KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. Mol Cell *61*, 39-53.
Michels, A.A., Fraldi, A., Li, Q., Adamson, T.E., Bonnet, F., Nguyen, V.T., Sedore, S.C., Price, J.P., Price, D.H., Lania, L.*, et al.* (2004). Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. Embo J *23*, 2608-2619.
Mumbach, M.R., Granja, J.M., Flynn, R.A., Roake, C.M., Satpathy, A.T., Rubin, A.J., Qi, Y., Jiang, Z., Shams, S., Louie, B.H.*, et al.* (2019). HiChIRP reveals RNA-associated chromosome conformation. Nat Methods *16*, 489-492.
Munoz, M.J., Perez Santangelo, M.S., Paronetto, M.P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J.J.*, et al.* (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. Cell *137*, 708-720.
Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature *414*, 322-325.
Nojima, T., Gomes, T., Carmo-Fonseca, M., and Proudfoot, N.J. (2016). Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. Nat Protoc *11*, 413-428.
Nojima, T., Gomes, T., Grosso, A.R., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526-540.
Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S.M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., and Proudfoot, N.J. (2018). Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. Mol Cell *72*, 970-984 e977.
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods *14*, 417-419.
Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol Cell *23*, 297-305.
Price, D.H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. Mol Cell Biol *20*, 2629-2634.
Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.
Rockx, D.A., Mason, R., van Hoffen, A., Barton, M.C., Citterio, E., Bregman, D.B., van Zeeland, A.A., Vrieling, H., and Mullenders, L.H. (2000). UV-induced inhibition of transcription involves repression of transcription initiation and phosphorylation of RNA polymerase II. Proc Natl Acad Sci U S A *97*, 10503-10508.
Rodriguez-Castaneda, F., Lemma, R.B., Cuervo, I., Bengtsen, M., Moen, L.M., Ledsaak, M., Eskeland, R., and Gabrielsen, O.S. (2018). The SUMO protease SENP1 and the chromatin remodeler CHD3 interact and jointly affect chromatin accessibility and gene expression. J Biol Chem *293*, 15439-15454.

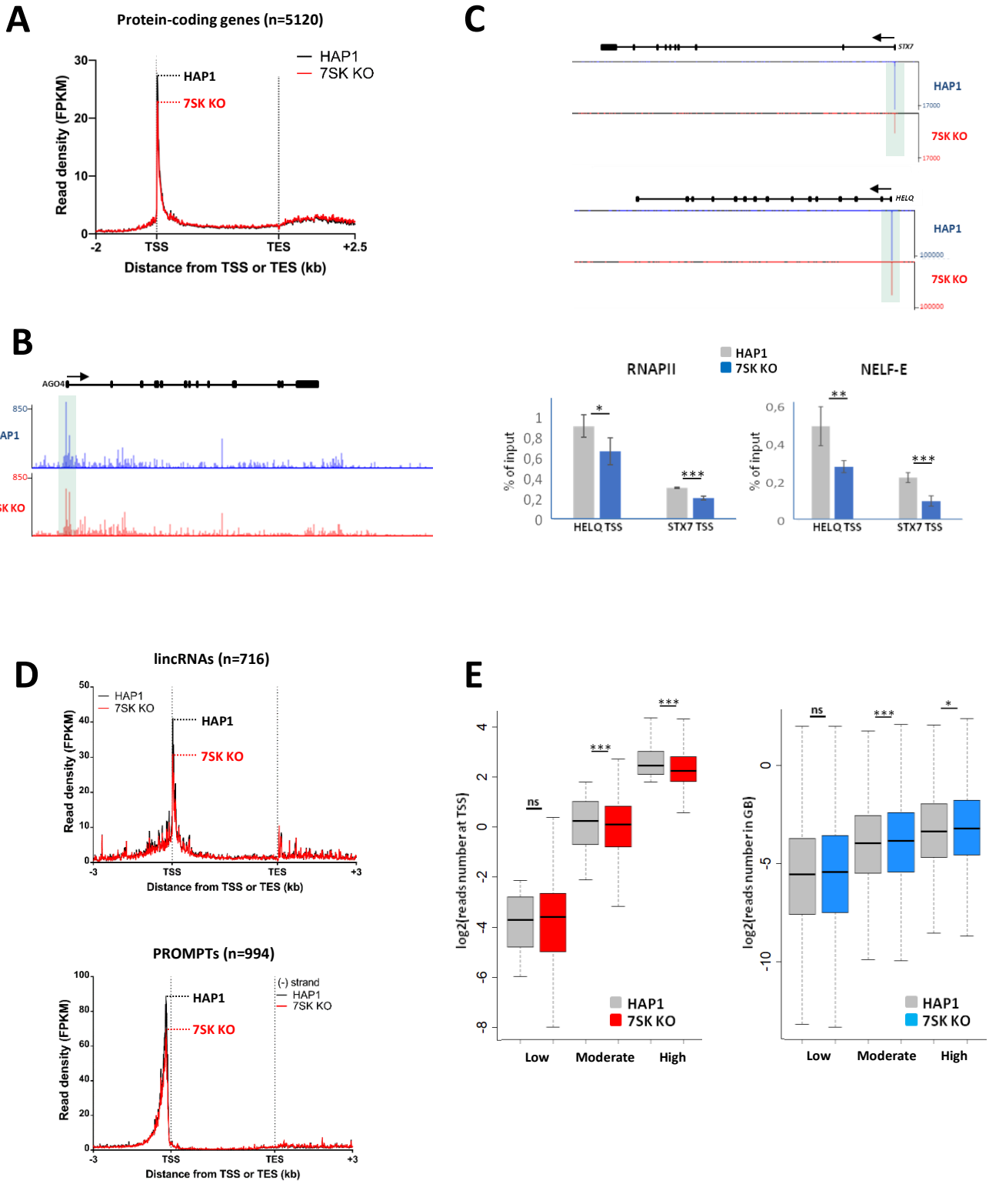
Sanso, M., Levin, R.S., Lipp, J.J., Wang, V.Y., Greifenberg, A.K., Quezada, E.M., Ali, A., Ghosh, A., Larochelle, S., Rana, T.M.*, et al.* (2016). P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. Genes Dev *30*, 117-131.
Skreka, K., Schafferer, S., Nat, I.R., Zywicki, M., Salti, A., Apostolova, G., Griehl, M., Rederstorff, M., Dechant, G., and Huttenhofer, A. (2012). Identification of differentially expressed non-coding RNAs in embryonic stem cell neural differentiation. Nucleic Acids Res *40*, 6001-6015.
Smith, E., and Shilatifard, A. (2013). Transcriptional elongation checkpoint control in development and disease. Genes Dev *27*, 1079-1088.
Sobhian, B., Laguette, N., Yatim, A., Nakamura, M., Levy, Y., Kiernan, R., and Benkirane, M. (2010). HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. Mol Cell *38*, 439-451.
Soneson, C., Yao, Y., Bratus-Neuenschwander, A., Patrignani, A., Robinson, M.D., and Hussain, S. (2019). A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. bioRxiv.
Van Herreweghe, E., Egloff, S., Goiffon, I., Jady, B.E., Froment, C., Monsarrat, B., and Kiss, T. (2007). Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. Embo J *26*, 3570-3580.
Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature *560*, 607-612.
Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., and Adelman, K. (2015). Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol Cell *58*, 311-322.
Williamson, L., Saponaro, M., Boeing, S., East, P., Mitter, R., Kantidakis, T., Kelly, G.P., Lobley, A., Walker, J., Spencer-Dene, B.*, et al.* (2017). UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. Cell *168*, 843-855 e813.
Xue, Y., Yang, Z., Chen, R., and Zhou, Q. (2010). A capping-independent function of MePCE in stabilizing 7SK snRNA and facilitating the assembly of 7SK snRNP. Nucleic Acids Res *38*, 360-369.
Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell *21*, 227-237.
Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol Cell *19*, 535-545.
Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature *414*, 317-322.
Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., and Cleary, M.L. (2010). A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. Cancer Cell *17*, 198-212.
Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., and Roeder, R.G. (2015). RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science *350*, 1383-1386.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.
Zhou, Q., Li, T., and Price, D.H. (2012). RNA Polymerase II Elongation Control. Annu Rev Biochem *81*, 119-143.
Zhou, Q., and Yik, J.H. (2006). The Yin and Yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. Microbiol Mol Biol Rev *70*, 646-659.



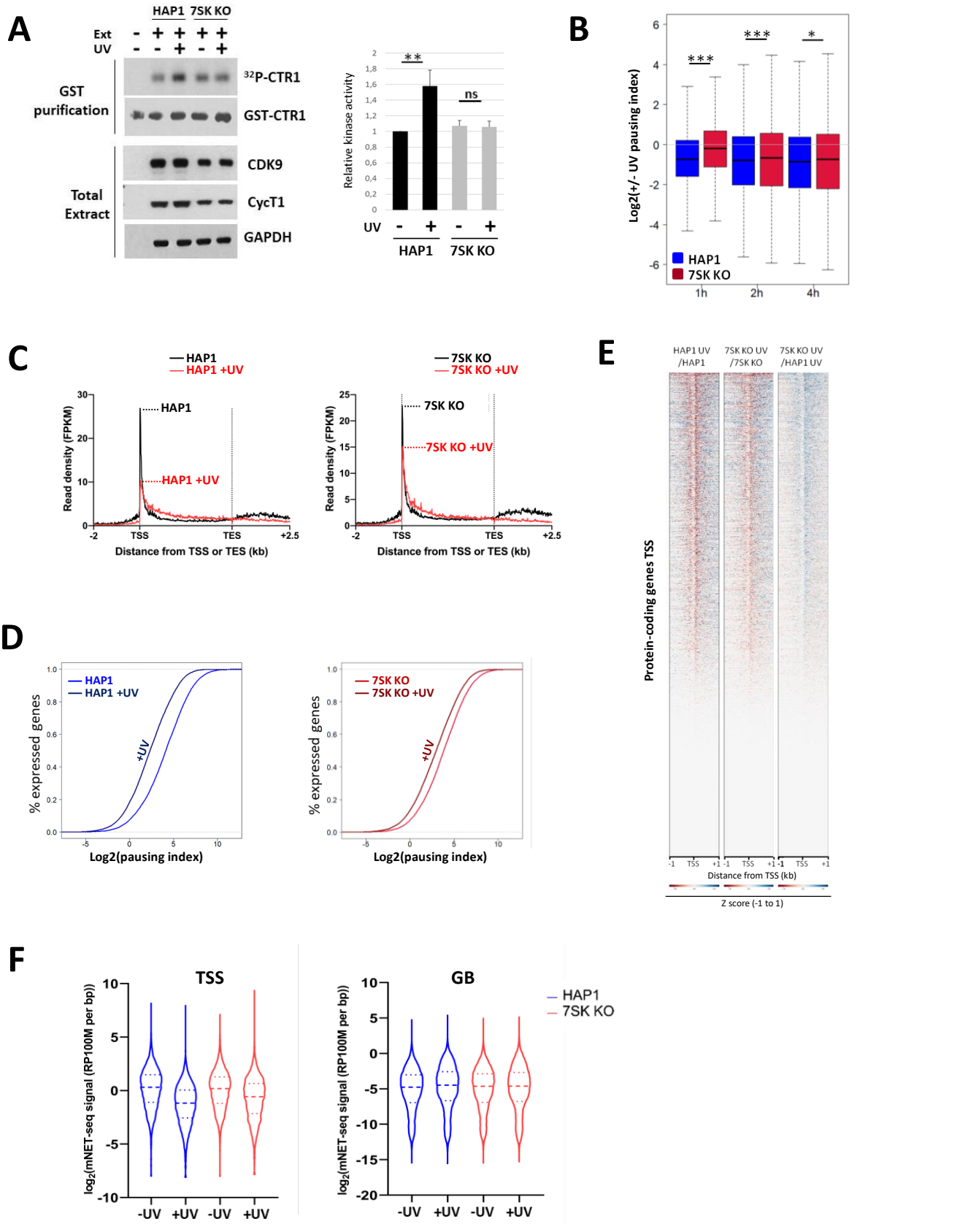
Studniarek et al., Figure 1



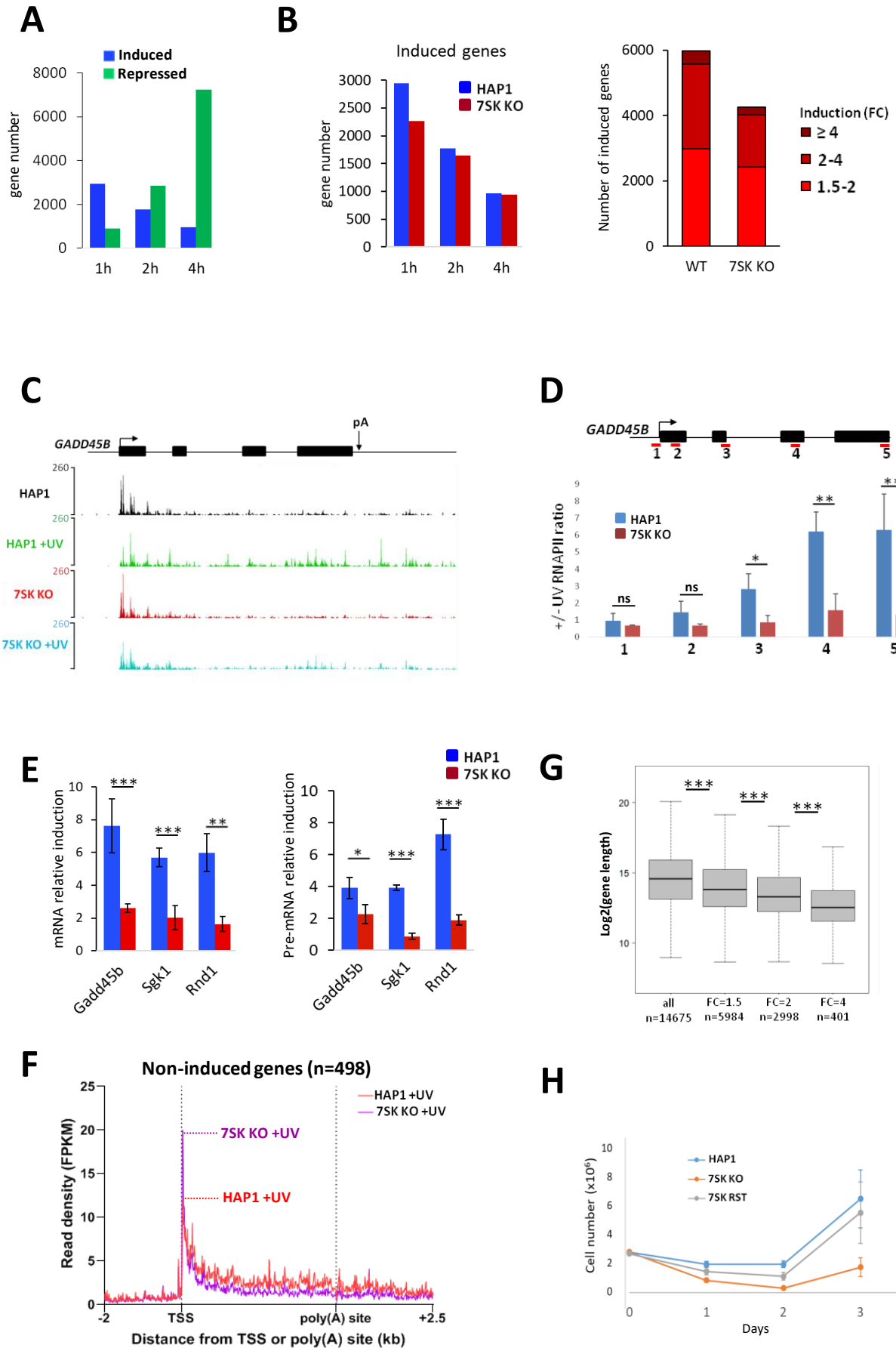
Studniarek et al., Figure 2



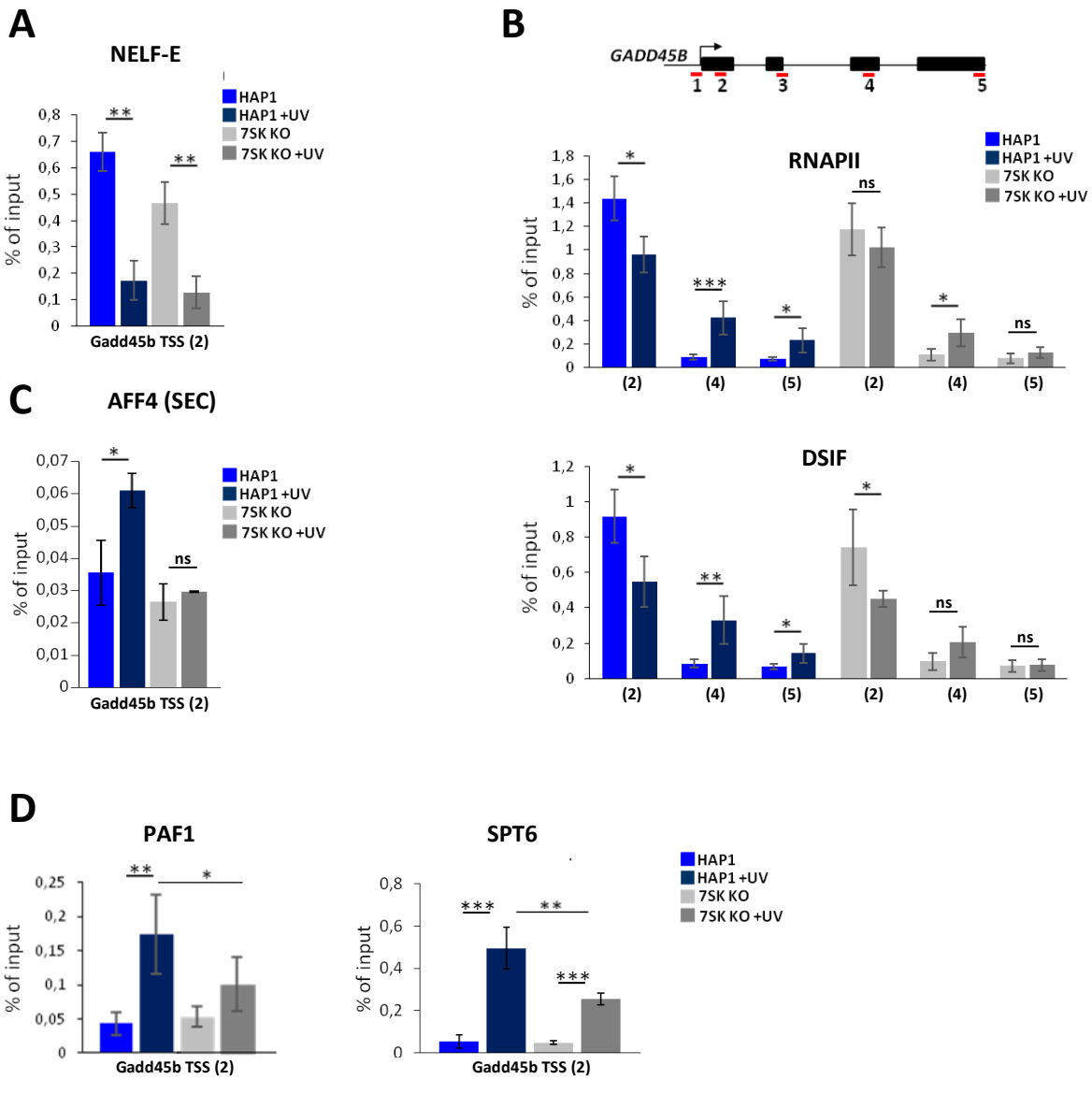
Studniarek et al., Figure 3



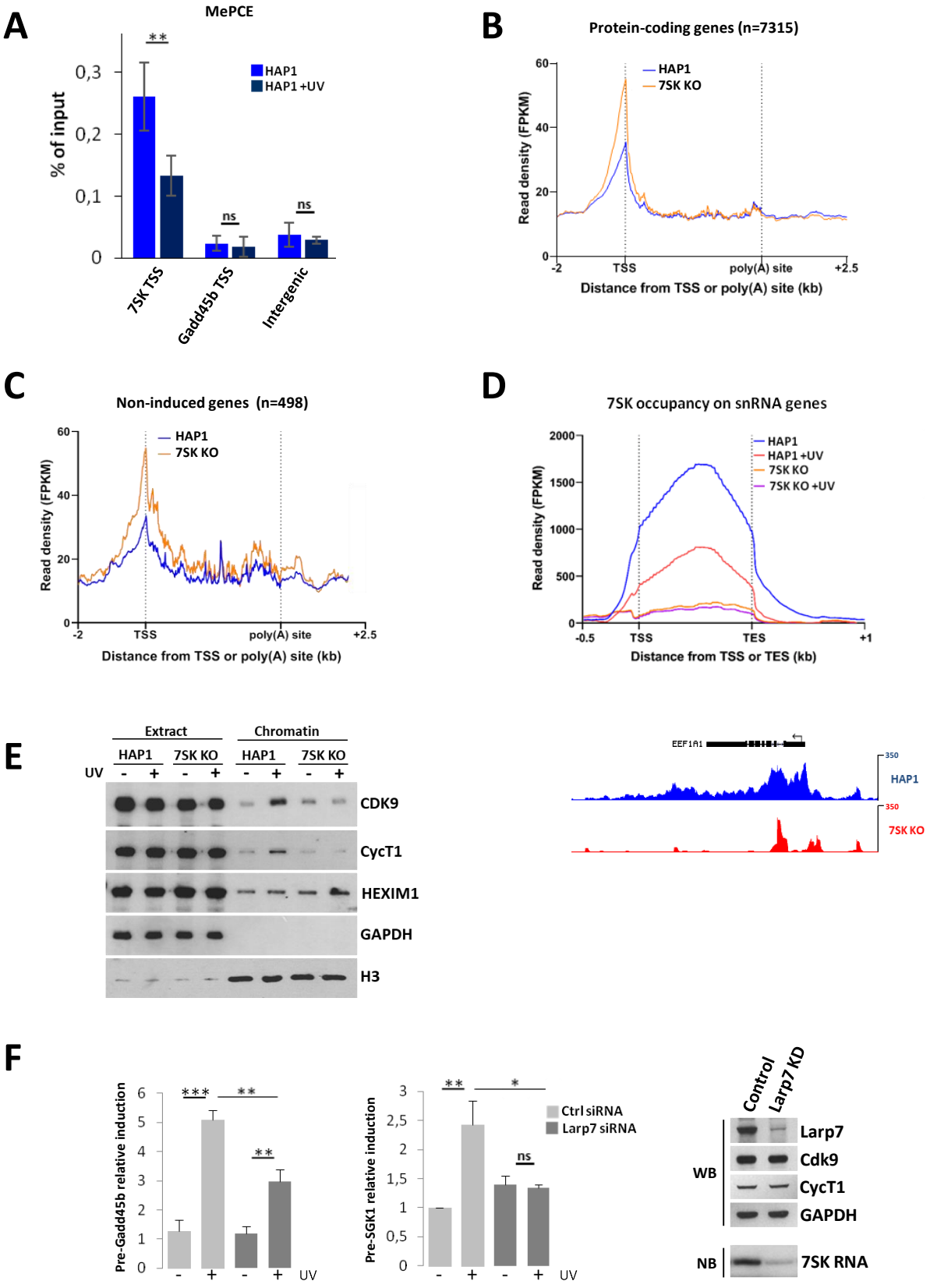
Studniarek et al., Figure 4



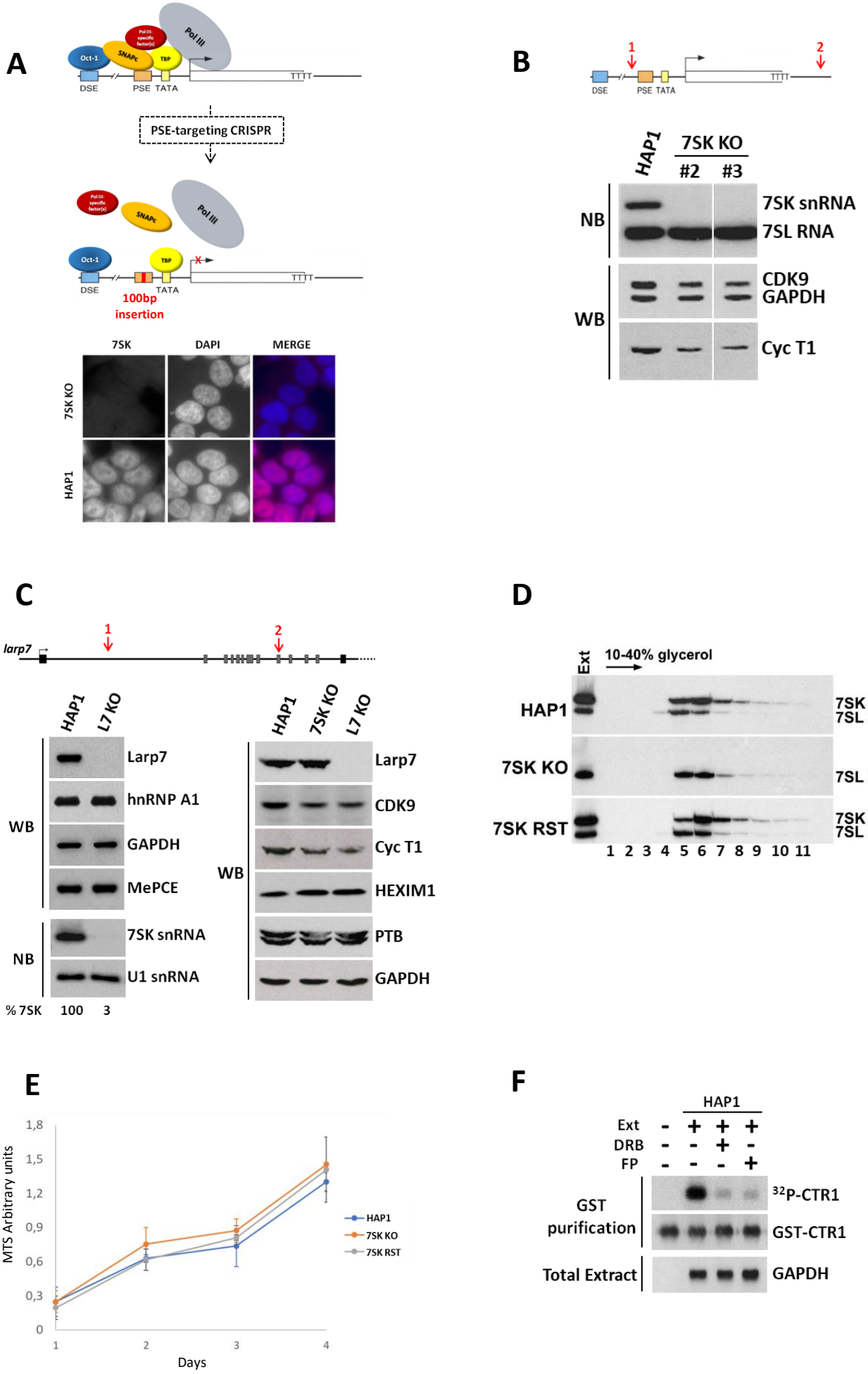
Studniarek et al., Figure 5



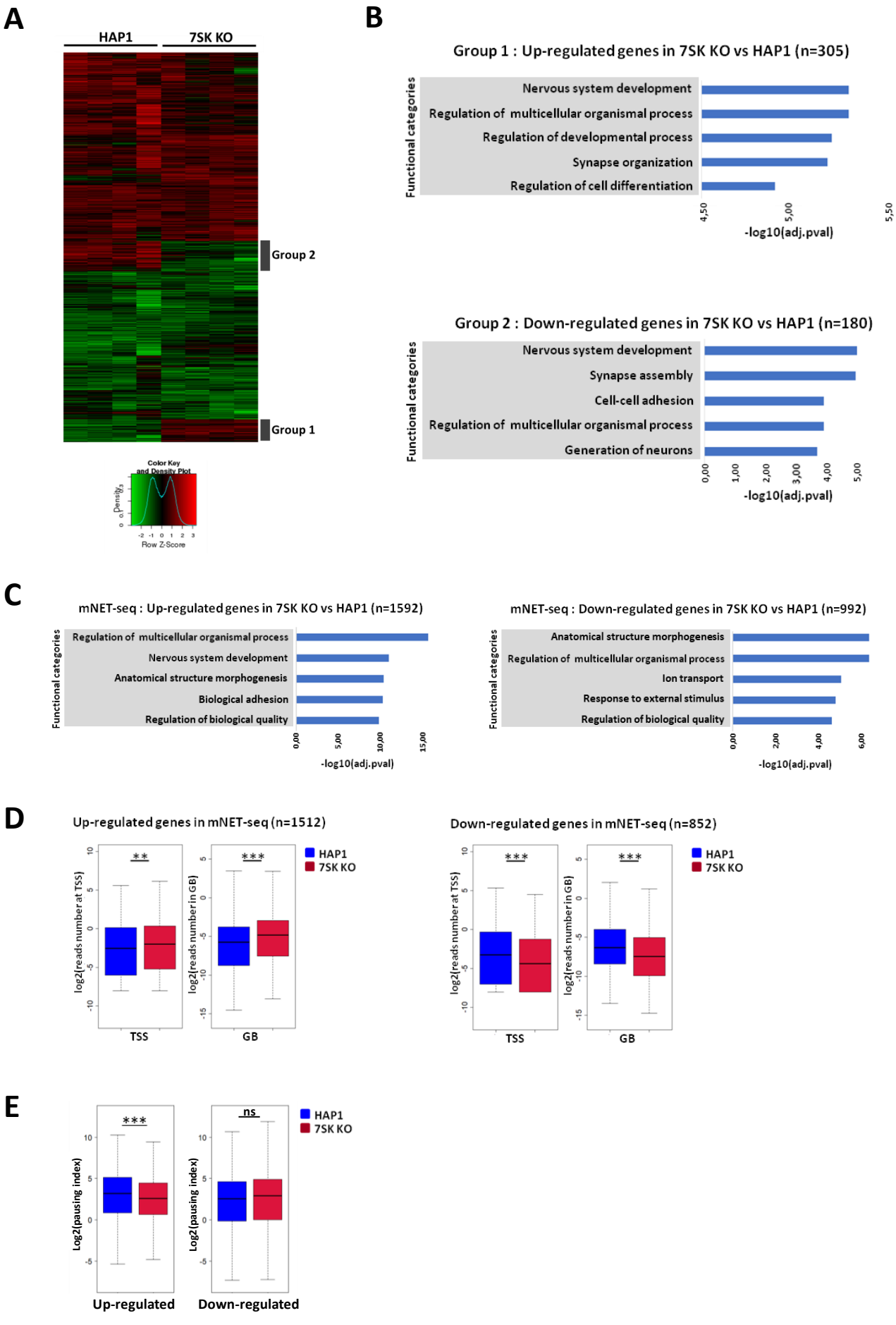
Studniarek et al., Figure 6



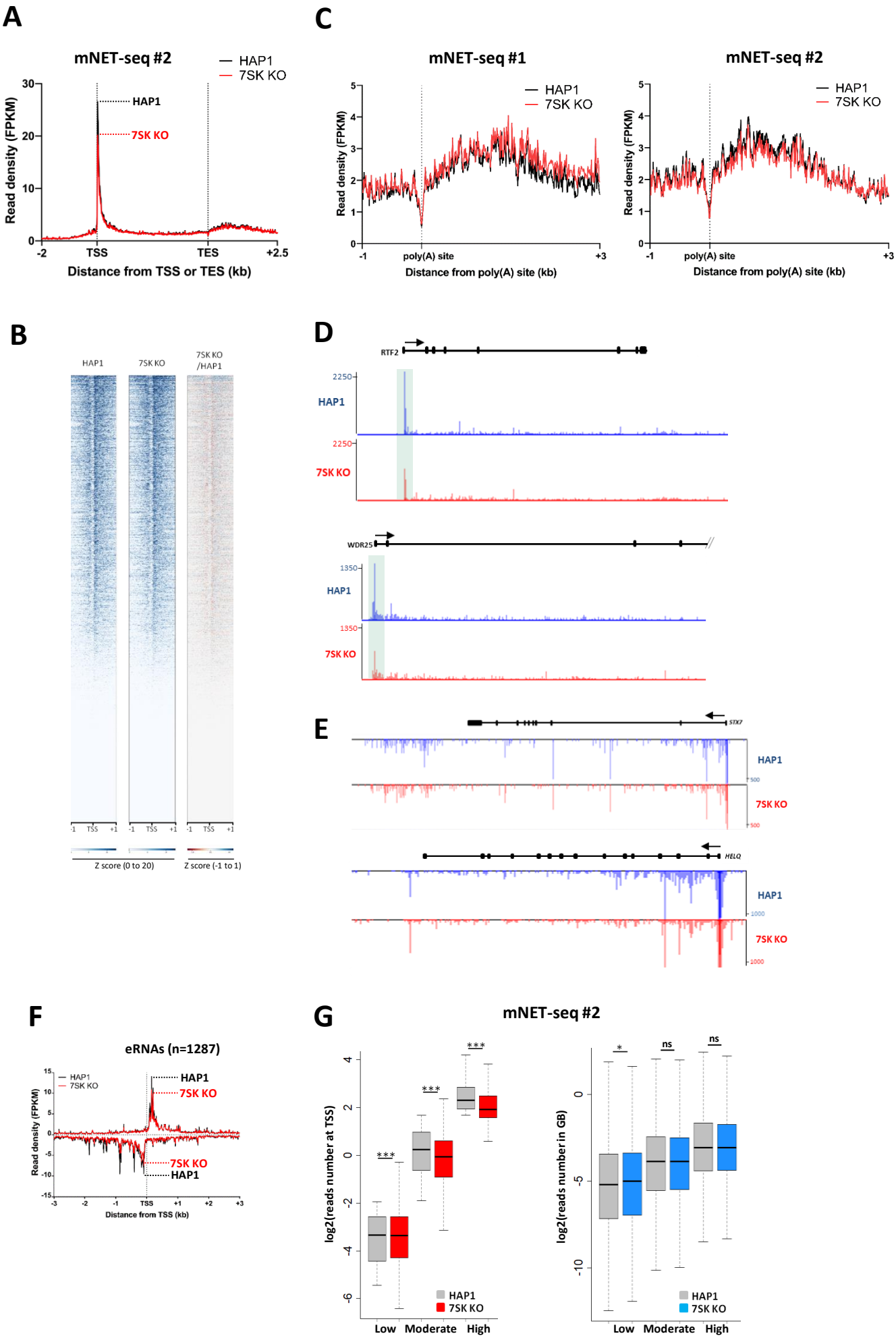
Studniarek et al., Figure 7



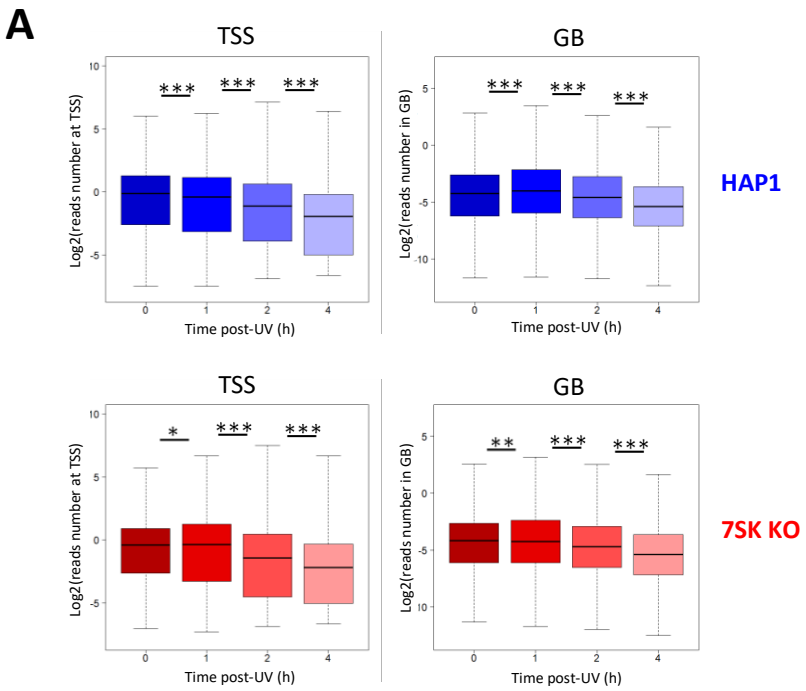
Studniarek et al., Figure S1

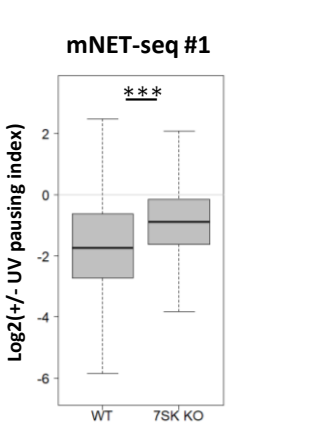


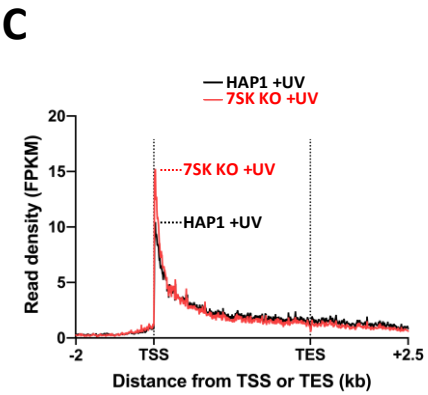
Studniarek et al., Figure S2

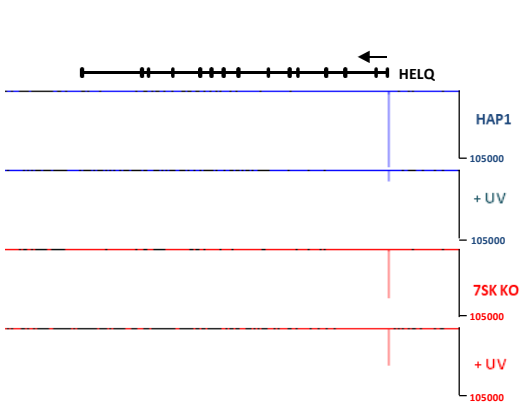


Studniarek et al., Figure S3

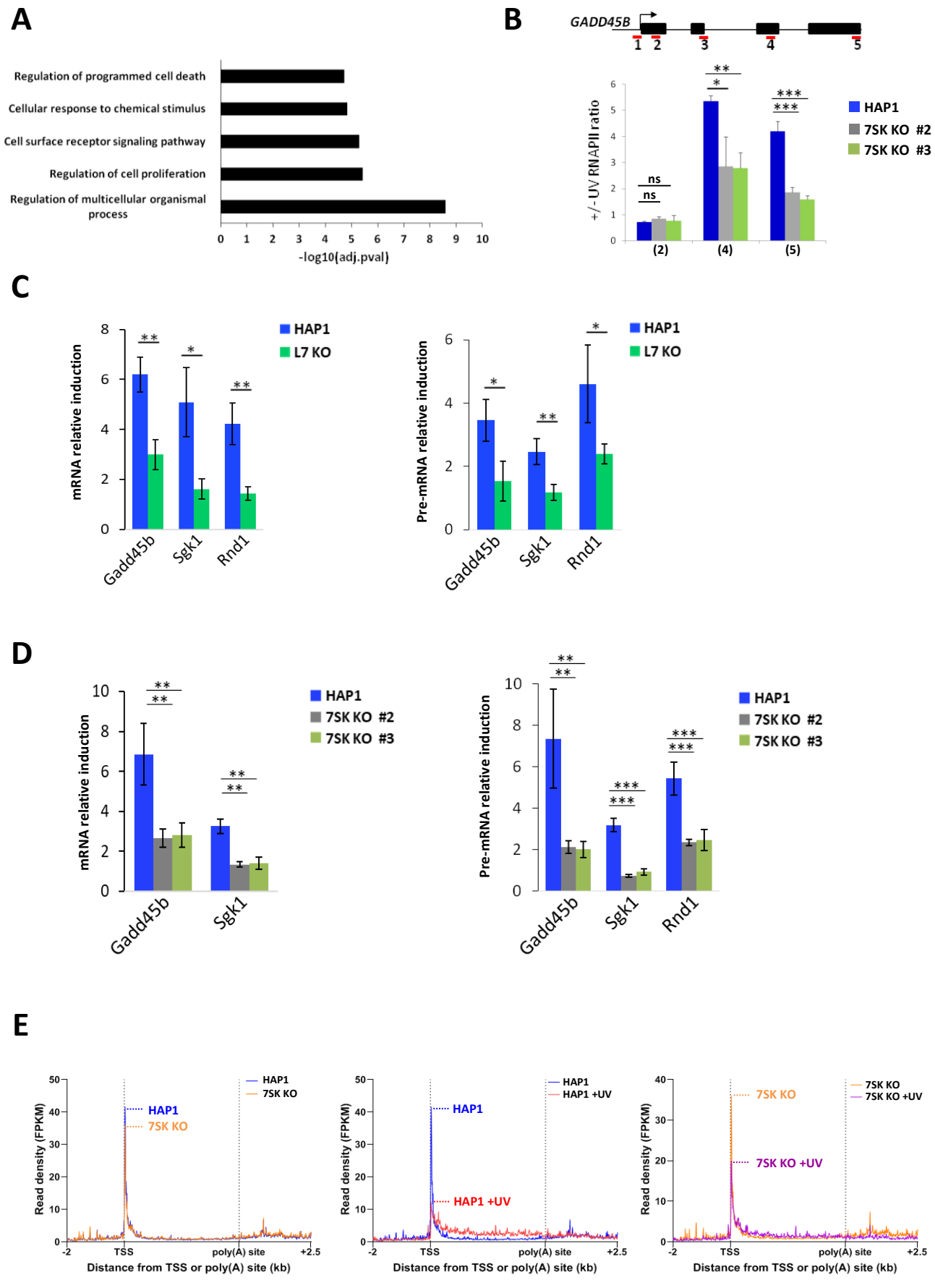




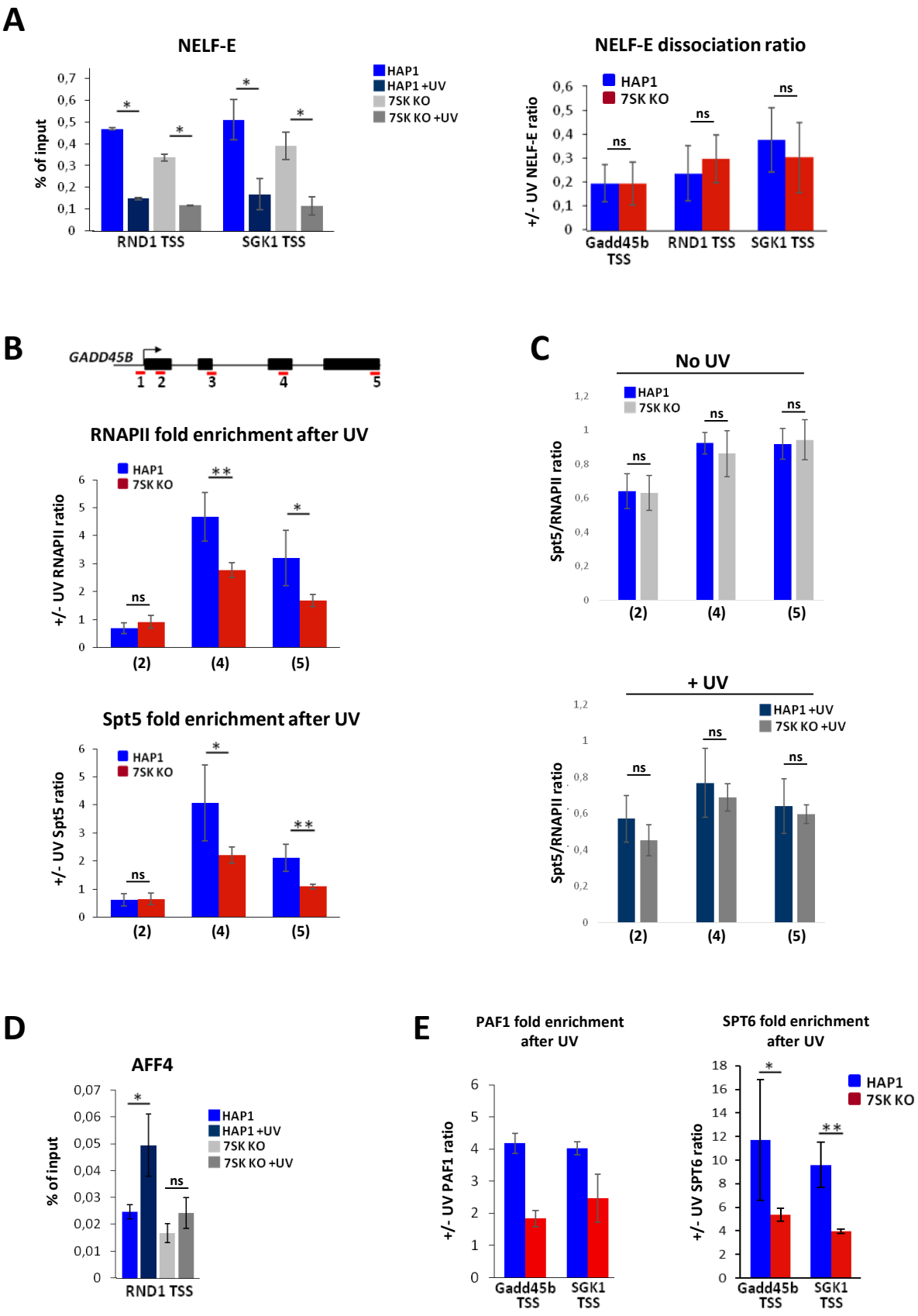




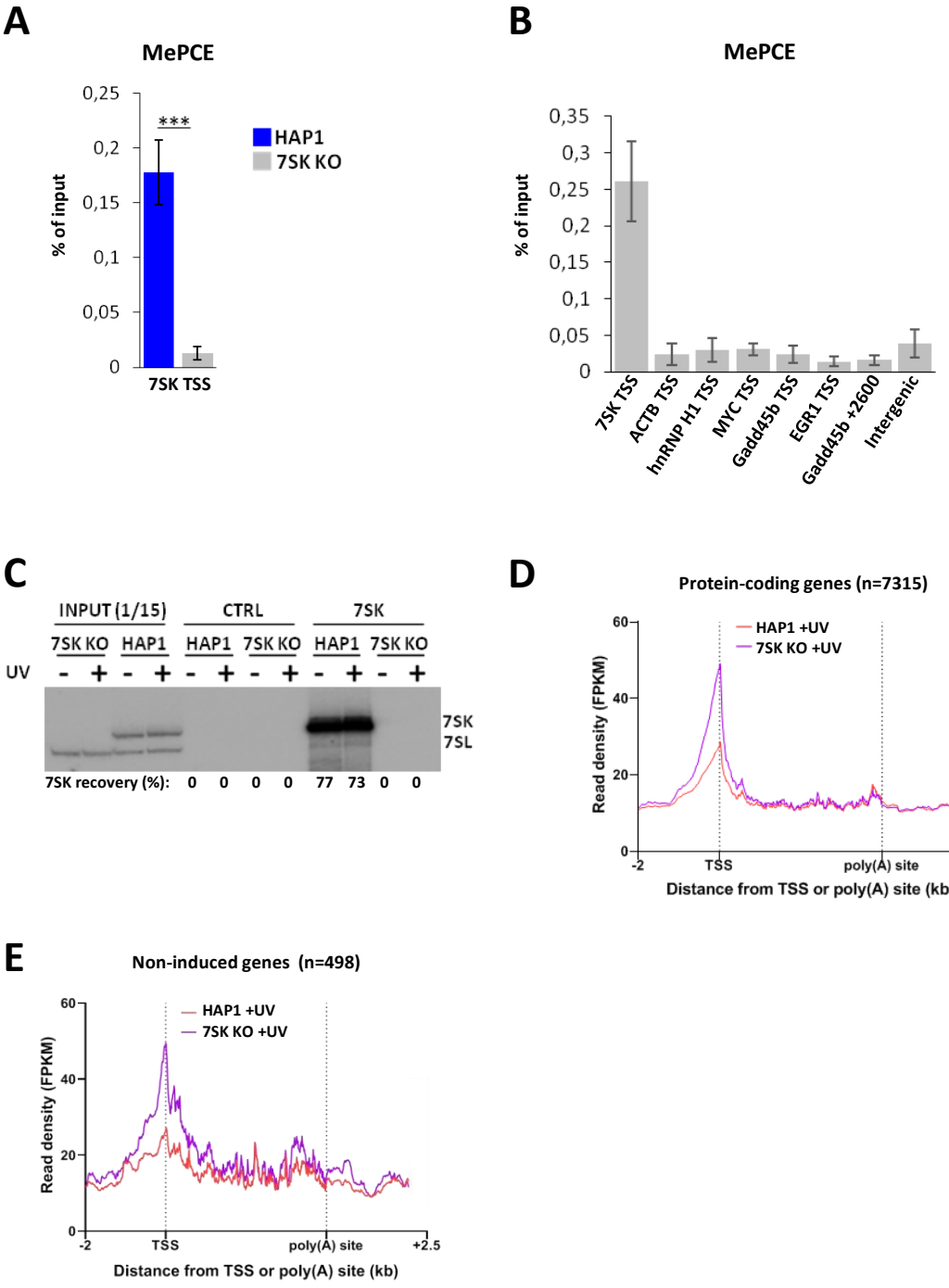
Studniarek et al., Figure S4



Studniarek et al., Figure S5



Studniarek et al., Figure S6



Studniarek et al., Figure S7

Supplementary Figure legends

Figure S1. Characterization of 7SK and Larp7 KO cell lines. **(A)** Upper panel: Strategy used for generating the 7SK KO cell line. Schematic diagram of the human 7SK gene with the promoter regulatory elements and their respective binding partners. CRISPR-induced disruption of the PSE is expected to abrogate RNA polymerase III (Pol III)-mediated transcription of 7SK RNA gene. DSE, distal sequence element; PSE, proximal sequence element, TATA, TATA box. Bottom panel: human HAP1 and 7SK KO cells were probed with a fluorescently labeled oligonucleotide complementary to 7SK snRNA. **(B)** CRISPR-mediated deletion of the entire 7SK snRNA gene. Schematic diagram of the human 7SK gene indicating the position of the predicted CRISPR target sites. RNAs and whole cell extract from HAP1, 7SK KO #2 and 7SK KO #3 cells were analyzed by Northern Blot (NB) or Western Blot (WB), respectively. **(C)** CRISPR-mediated inactivation of the Larp7 gene. Schematic diagram of the human Larp7 gene indicating the position of the predicted CRISPR target sites. RNAs and whole cell extract from HAP1 and Larp7 KO (L7 KO) were analyzed by Northern Blot (NB) or Western Blot (WB), respectively. The accumulation of 7SK snRNA in L7 KO cells, quantified by phosphorimager, is indicated. **(D)** Glycerol gradient (10-40 %) sedimentation analysis of HAP1, 7SK KO and 7SK RST cells. Northern blotting was performed with antisense oligonucleotides directed against 7SL RNA and 7SK snRNA. Ext: total extract. **(E)** MTS cell proliferation assay of HAP1, 7SK KO and 7SK RST cells. **(F)** Radioactive *in vitro* kinase assay using 7.5 µg of HAP1 total extract and recombinant Spt5-CTR1 fused to GST (GST-CTR1) as substrate. Cdk9-dependent phosphorylation was measured by autoradiography following purification of GST-CTR1 on glutathione resin. As a control, the kinase reaction was performed with cells treated with DRB and Flavopiridol (FP). Levels of GAPDH in total extracts, as well as level of purified GST-CTR1, were analyzed by Western Blot.

Figure S2. Impact of 7SK loss on gene expression. **(A)** Hierarchical clustering of the microarray data identified two major clusters corresponding to genes with increased (Group 1) or decreased (Group 2) expression in 7SK KO vs HAP1 cells. The color scale depicts the fold difference from the mean (bottom), red indicating increased transcript expression and green indicating decreased transcript expression. **(B)**

Functional categories of genes significantly up- (Group 1) and down-regulated (Group 2) in 7SK KO cells, based on microarray data. All GO groups demonstrated enhanced statistical representation. **(C)** Functional categories of genes significantly up- and down-regulated in 7SK KO cells, based on mNET-seq data. All GO groups demonstrated enhanced statistical representation. **(D)** boxplot analysis comparing reads density at TSS and in GB between HAP1 (blue) and 7SK KO (red) cells for the up- and down-regulated genes in mNETseq. $^{**}p < 0.01$; $^{***}p < 0.001$, determined by Wilcoxon test. **(E)** Boxplot analysis comparing PI between HAP1 and 7SK KO cells for the up- and down-regulated genes in mNET-seq. $^{***}p < 0.001$, determined by Wilcoxon test. ns, not significant.

Figure S3. 7SK loss leads to decreased RNAPII promoter-proximal pausing. **(A)** Metagene analysis of a biological replicate of mNET-seq, showing transcriptionally engaged RNAPII occupancy over protein-coding genes in HAP1 and 7SK KO cells. Horizontal dotted lines represent the average level of RNAPII promoter-proximal pausing. **(B)** Heatmaps showing mNET-seq signals around TSS regions of protein-coding genes. Genes were sorted by decreasing read density at TSS. For the ratio heatmap, color-scale depicts increased (in blue) or decreased mNET-seq signals (in red). **(C)** Metagene analysis of transcriptionally engaged RNAPII occupancy over termination regions (up to 3 kb 3' to polyA site) in HAP1 and 7SK KO cells, measured for two biological replicates of mNET-seq. **(D)** The mNET-seq positive strand reads are shown on the RTF2 and WDR25 genes. Peaks corresponding to promoter-proximal pausing are highlighted by shadowed box. **(E)** mNET-seq reads in gene bodies of STX7 and HELQ genes are shown. **(F)** Metagene analysis of mNET-seq over eRNA genes in HAP1 and 7SK KO cells. Horizontal dotted lines represent the average level of RNAPII promoter-proximal pausing. **(G)** Box plots showing quantification of reads at TSS (left panel) and in gene bodies (right panel) for a biological replicate of mNET-seq. Genes were sorted in three equal groups based on reads number at TSS in HAP1 cells (low, moderate and high). $^{*}p < 0.05$; $^{***}p < 0.001$, determined by Wilcoxon test. ns, not significant.

Figure S4. The 7SK/P-TEFb snRNP promotes pause release following UV irradiation. **(A)** Box plots showing quantification of reads at TSS (left panel) and in gene bodies (right panel) of protein-coding genes in HAP1 (in blue) and 7SK KO cells

(in red) following UV irradiation. *p < 0.05; **p < 0.01; ***p < 0.001, determined by Wilcoxon test. ns, not significant. **(B)** Boxplot analysis comparing PI ratio (+/- UV) between HAP1 and 7SK KO cells 1 hour after UV irradiation, measured by a biological replicate of mNET-seq. ***p < 0.001, determined by Wilcoxon test. **(C)** Metagene analysis comparing transcriptionally engaged RNAPII occupancy on protein-coding gene one hour after UV in HAP1 and 7SK KO cells, measured by mNET-seq. **(D)** The mNET-seq negative strand reads are shown on the HELQ gene.

Figure S5. The 7SK/P-TEFb snRNP is required for UV-induced transcriptional reprogramming. **(A)** Functional categories enrichment for the 498 genes which are not properly induced after UV exposure in 7SK KO cells, based on mNET-seq data. All GO groups demonstrated enhanced statistical representation. **(B)** Quantitative ChIP analyses of endogenous RNAPII occupancy on Gadd45b gene in HAP1, 7SK KO #2 and 7SK KO #3 cells. Signals are represented as fold-change enrichment after UV (ratio +/-UV). Positions of the PCR-amplified fragments are indicated below the schematic diagram of the human Gadd45b gene. In all panels: *p < 0.05; **p < 0.01; ***p < 0.001, determined by Student's t test. ns, not significant. **(C)(D)** RT-qPCR of the indicated UV-induced mRNA (left) and their corresponding pre-mRNA transcripts (right) in the Larp7 KO (panel C) and the 7SK KO #2 and 7SK KO #3 cell lines (panel D). Results were normalized to the non-irradiated control and error bars represent standard deviations from at least three independent experiments. *p < 0.05; **p < 0.01, ***p < 0.001, determined by Student's t test. **(E)** Metagene analyses comparing transcriptionally engaged RNAPII occupancy on the 498 genes that are not properly induced in the 7SK KO cells after UV exposure, measured by mNET-seq.

Figure S6. The 7SK/P-TEFb snRNP stimulates UV-dependent recruitment of key elongation factors. **(A)** Quantitative ChIP analyses of endogenous NELF-E occupancy at TSS of RND1 and SGK1 genes in HAP1 and 7SK KO cells, before and 1 hour after UV exposure. Signals are represented as % of input. In all panels: *p < 0.05; **p < 0.01; ***p < 0.001, determined by Student's t test. ns, not significant. The NELF dissociation ratio (+/- UV) on Gadd45b, RND1 and SGK1 genes is shown on the right. **(B)** Enrichment over the human Gadd45b gene of RNAPII and SPT5 (DSIF) after UV exposure (ratio +/- UV), measured by quantitative ChIP analyses performed

in HAP1 and 7SK KO cells. Positions of the PCR-amplified fragments are indicated below the schematic diagram of Gadd45b gene. **(C)** Spt5/RNAPII ratio over the human Gadd45b gene in HAP1 and 7SK KO cells, before (top) and after (bottom) UV irradiation. **(D)** Quantitative ChIP analyses of endogenous AFF4 occupancy at TSS of RND1 gene in HAP1 and 7SK KO cells, before and after UV exposure. Signals are represented as % of input. **(E)** Enrichment of PAF1 and SPT6 after UV exposure (ratio +/- UV) at Gadd45b and SGK1 TSS, measured by quantitative ChIP analyses performed in HAP1 and 7SK KO cells.

Figure S7. P-TEFb is extracted from nucleoplasmic 7SK/P-TEFb RNP during UV response. **(A)** Quantitative ChIP analyses of endogenous MePCE occupancy at TSS of 7SK snRNA gene in HAP1 and 7SK KO cells. Signals are represented as % of input. $^{***}p < 0.001$, determined by Student's t test. **(B)** Quantitative ChIP analyses of endogenous MePCE occupancy at TSS of the indicated gene in HAP1 cells. Signals are represented as % of input. **(C)** Northern blot showing retrieval of 7SK snRNA in ChIRP experiments. 7SK purification was performed in HAP1 and 7SK KO cells, with or without UV treatment, using either non-specific (CTRL) or 7SK-specific (7SK) mixture of antisense biotinylated oligonculeotides. Phosphorimager quantification is indicated below. **(D)** Metagene analysis showing 7SK snRNA occupancy after UV irradiation in HAP1 cells on protein-coding genes, measured by ChIRP-seq. Signals detected in 7SK KO cells are considered as non-specific. **(E)** Metagene analysis showing 7SK snRNA occupancy after UV irradiation in HAP1 cells on the 498 genes that are not properly induced after UV exposure in the absence of 7SK, measured by ChIRP-seq. Signals detected in 7SK KO cells were considered as non-specific.

Résultats complémentaires

I. Rôle de la RNP 7SK/P-TEFb dans la régulation de la transcription par l'ARN Pol II

Résumé des résultats présentés dans l'article :

La RNP 7SK/P-TEFb a été détectée au niveau des régions promotrices de nombreux gènes. L'ARN 7SK a également été retrouvé au niveau des enhancers et super-enhancers, indépendamment de P-TEFb, où il pourrait réguler leur activité en coopération avec le complexe BAF, ainsi qu'au niveau des gènes des snARN dont il stimule la production via le LEC (Egloff et al., 2017a; Flynn et al., 2016; McNamara et al., 2016a). Ces données révèlent de potentielles fonctions directes de l'ARN 7SK dans la régulation de la transcription par l'ARN Pol II. Pour étudier son rôle dans ce processus, nous avons généré à l'aide du système CRISPR-Cas9 différentes lignées cellulaires humaines dans lesquelles l'expression de l'ARN 7SK est abolie, empêchant la formation de la RNP 7SK/P-TEFb mais dans lesquelles l'activité de P-TEFb n'est pas altérée (Fig.1, S1, figure 23).

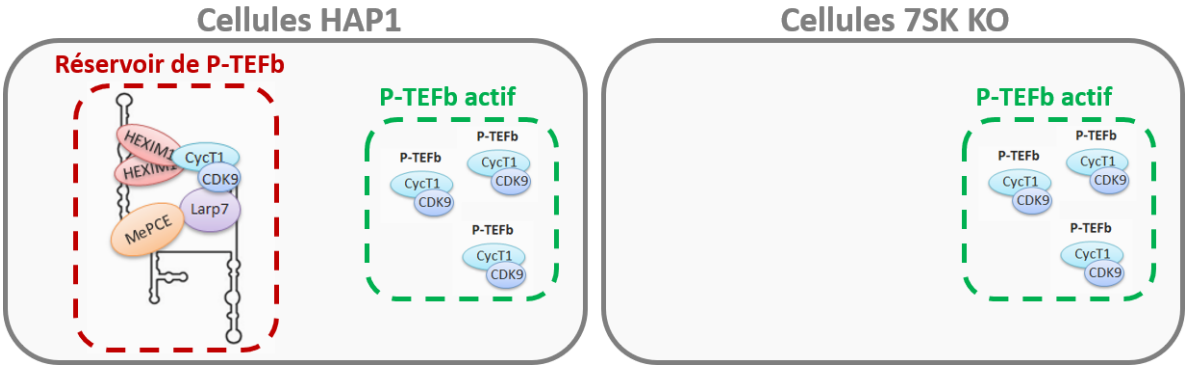


Figure 23 : Représentation schématique de l'activité de P-TEFb dans les cellules HAP1 et 7SK KO.

L'analyse du transcriptome (par puce à ADN) et de la transcription naissante (par mNET-seq) a permis de mettre en évidence un effet discret de la perte de l'ARN 7SK sur l'expression génique, principalement transcriptionnel. La perte de l'ARN 7SK affecte la transcription de ~15 % des gènes actifs transcrits par l'ARN Pol II, mais seulement 2 % du transcriptome (Fig. 2, S2). Au niveau transcriptionnel, l'ARN 7SK joue également un rôle dans le maintien de la pause de l'ARN Pol II à proximité du promoteur (Fig. 3, S3). En l'absence de l'ARN 7SK, le niveau d'ARN Pol II en pause est en effet globalement diminué alors que la transcription dans les corps de gènes est peu affectée (Fig. 3E, S3A).

i. Rôle spécifique de l'ARN 7SK dans la régulation des gènes

La différence entre l'impact sur le transcriptome et sur la transcription naissante reflète la sensibilité de la technique de mNET-seq à détecter des changements transcriptionnels subtils. De plus, des variations transcriptionnelles pourraient ne pas systématiquement se répercuter sur l'accumulation des ARNm, étant donné que d'autres facteurs entrent en compte, comme la stabilité des ARNm. L'analyse approfondie des résultats de mNET-seq confirme un effet discret de la perte de l'ARN 7SK sur la transcription par l'ARN Pol II, illustré par le faible impact sur le transcriptome. Nous avons comparé les résultats de deux réplicats indépendants de mNET-seq générés sur les cellules HAP1 et 7SK KO. Bien que les données des deux réplicats corrèlent fortement (figure 24A), nous avons observé une différence significative dans le nombre de gènes dont l'expression est impactée négativement (gènes down) et positivement (gènes up) par la perte de l'ARN 7SK (d'un facteur ≥ 1.5, figure 24B). Ceci est en partie dû au seuil fixé pour considérer les gènes comme différentiellement exprimés (FC ≥ 1.5). En effet, en comparant l'induction des gènes down et up dans un réplicat avec leur induction dans le second réplicat, nous observons que la tendance de répression/activation transcriptionnelle est globalement conservée dans les deux réplicats. La figure 24C montre la comparaison de l'expression des gènes down d'un facteur ≥ 1.5 identifiés dans le premier réplicat (Rep1) à leur expression dans le second réplicat (Rep2). La majorité des gènes down du Rep1 sont en effet en-dessous de 0 dans le Rep 2 (83%, figure 24C), 0 indiquant un ratio d'expression 7SK KO/HAP1 inchangé. De même, la majorité des gènes up du Rep1 sont au-dessus de 0 dans le Rep 2 (74%, figure 24D). Nous pouvons également observer que la plupart des gènes sont faiblement réprimés/induits en l'absence de l'ARN 7SK, le ratio d'expression 7SK KO/HAP1 étant proche du seuil fixé (figure 24C, D). Nos données suggèrent donc que la perte de l'ARN 7SK affecte subtilement la transcription par l'ARN Pol II, l'expression de la majorité des gènes n'étant pas ou peu altérée. Pour la suite des analyses, nous avons décidé de ne considérer que les gènes les plus différentiellement exprimés (FC ≥ 1.5) communs aux deux réplicats de mNET-seq et donc robustement dérégulés par la perte de l'ARN 7SK (figure 24B, intersection).

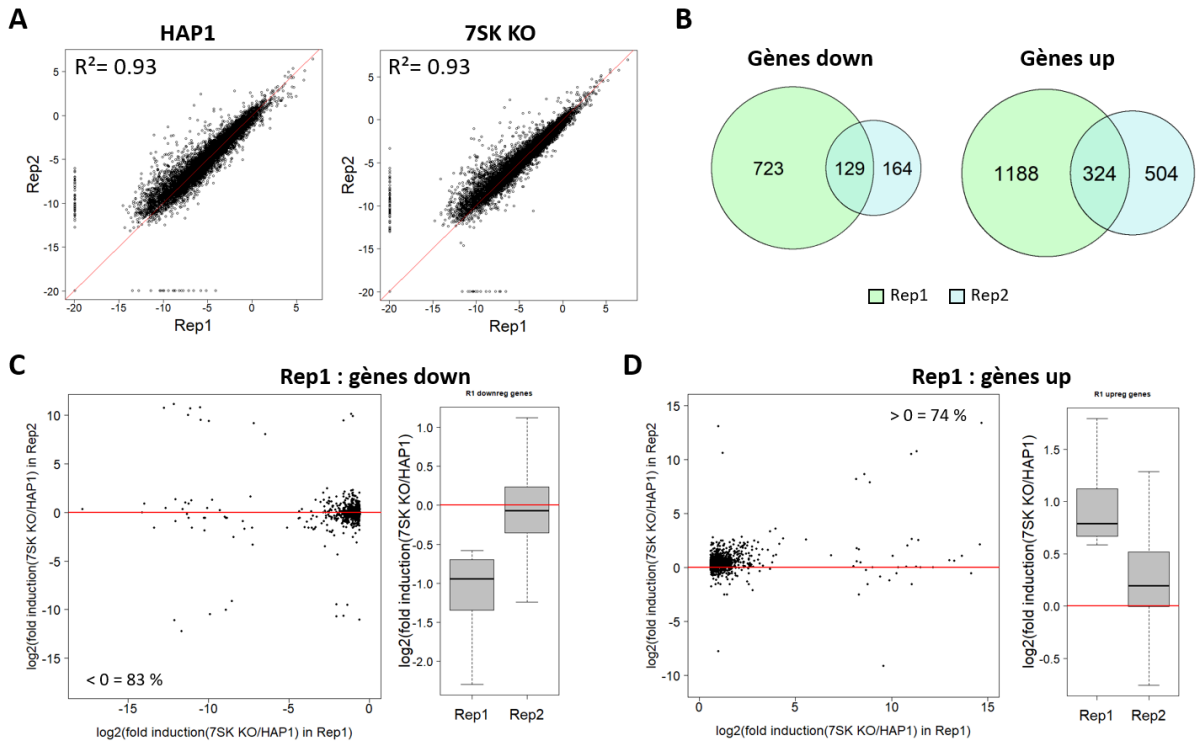


Figure 24 : La perte de l'ARN 7SK a un effet discret sur la transcription par l'ARN Pol II. (A) Comparaison de l'expression des gènes de cellules HAP1 et 7SK KO dans deux réplicats biologiques de mNET-seq (Rep1 et Rep2). Les coefficients de corrélation sont indiqués (R^2). (B) Diagramme de Venn comparant les gènes négativement (gènes down) et positivement (gènes up) régulés par la perte de l'ARN dans deux réplicats biologiques de mNET-seq. (C) Représentation du log2 du ratio d'expression 7SK KO/HAP1 pour les gènes down dans le Rep1 comparé au log2 de leur ratio d'expression (7SK KO/HAP1) dans le Rep2. Le pourcentage de gènes pour lesquels le log2 du ratio d'expression dans le Rep2 est inférieur à 0 (ratio d'expression < 1) est indiqué. (D) Comme en (C) pour les gènes up dans le Rep1. Le pourcentage de gènes pour lesquels le log2 du ratio d'expression dans le Rep2 est supérieur à 0 (ratio d'expression > 1) est indiqué.

Le mécanisme d'action de l'ARN 7SK dans la régulation de l'expression de ces gènes est encore en cours d'étude. Un point important de cette analyse repose sur la détermination des gènes ciblés directement par la RNP 7SK/P-TEFb. Pour cela, nous avons réalisé des ChIRP-seq afin de déterminer l'occupation de l'ARN 7SK à l'échelle du génome. Ces expériences consistent à purifier l'ARN 7SK au moyen d'oligonucléotides antisens biotinylés et de séquencer l'ADN co-purifié (figure 25A). Le ChIRP-seq réalisé dans les cellules 7SK KO constitue un contrôle optimal pour déterminer l'ADN co-purifié de manière non-spécifique. Les résultats préliminaires de cette analyse n'indiquent cependant pas la présence de l'ARN 7SK au niveau des gènes dérégulés par la perte de l'ARN 7SK, alors qu'elle est bien détectée au niveau des gènes des snARN (figure 25B, Fig. 7D). Ces résultats sont donc en défaveur d'un contrôle de la transcription de ces gènes par le chargement de la RNP 7SK/P-TEFb au niveau de leurs régions promotrices. Il a aussi été rapporté que l'ARN 7SK occupait de nombreux enhancers et super enhancers (Flynn et al., 2016; Liu et al., 2013). Il sera donc intéressant de rechercher la présence de l'ARN 7SK au niveau de ces régions régulatrices.

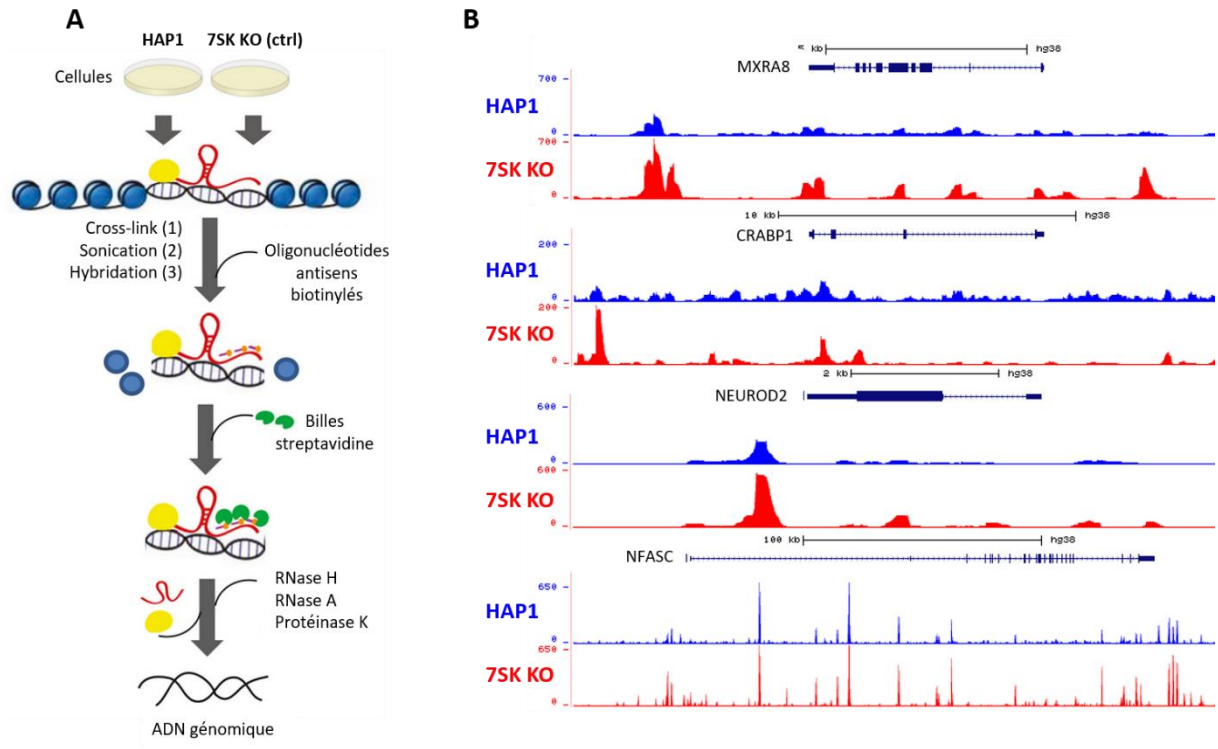


Figure 25 : L'ARN 7SK n'est pas détecté au niveau des régions promotrices des gènes dérégulés en son absence.(A) Schéma du protocol de ChIRP-seq (Samudyata et al., 2019). (B) Signaux de ChIRP-seq 7SK pour les gènes MXRA8, CRABP1, NEUROD2 et NFASC visualisés sur le genome browser de (UCSC). Le signal dans les cellules HAP1 est indiqué en bleu, celui dans les cellules 7SK KO en rouge. L'axe y montre la densité du signal ChiRP-seq. L'axe x indique la position génomique.

Pour essayer de comprendre le mécanisme de régulation des gènes dont l'expression est altérée par l'absence de l'ARN 7SK, nous avons recherché s'ils pouvaient être la cible de facteurs de transcription spécifiques. L'analyse d'enrichissement a été réalisée avec EnrichR (http://amp.pharm.mssm.edu/Enrichr/) sur la liste des gènes dérégulés communs aux deux réplicats de mNET-seq (figure 24B). Les cibles de trois facteurs de transcription, REST, SUZ12 et EZH2, sont significativement enrichies parmi ces gènes (tableau 2). REST est notamment connu pour réprimer la transcription des gènes neuronaux (Bruce et al., 2004; Bruce et al., 2009; Schoenherr and Anderson, 1995), en accord avec les résultats d'enrichissement obtenus pour les gènes dérégulés en l'absence de l'ARN 7SK par GeneOntology (Fig. S1B, C). Suz12 et EZH2 font tous deux partie du complexe répresseur PRC2 (Polycomb repressor complex 2). Ces données soulèvent l'intrigante possibilité que l'ARN 7SK pourrait participer à la répression des gènes en favorisant le recrutement de ces facteurs au niveau de leurs gènes cibles. Alternativement, un(des) facteur(s) dérégulé(s) par la perte de l'ARN 7SK pourrai(en)t être impliqué(s) dans ce processus. Suz12 semble aussi impliqué dans la régulation des gènes négativement régulés par la perte de l'ARN 7SK (tableau 2), suggérant plutôt une activité indirecte de l'ARN 7SK dans le contrôle de la transcription par PRC2. De plus, l'analyse ciblée de nos données de ChIRP-seq sur des gènes régulés par REST, SUZ12 et EZH2 n'argumente pas en faveur de

la présence de l'ARN 7SK au niveau de leurs régions promotrices. L'identification des protéines associées à l'ARN 7SK par ChIRP suivi de spectrométrie de masse (ChIRP-MS) n'indique pas d'interaction entre l'ARN 7SK et ces facteurs de transcription. Ainsi, aucun de nos résultats n'indiquent de lien direct entre l'ARN 7SK et la transcription des gènes transcriptionnellement dérégulés en son absence.

Gènes dont la transcription est augmentée par la perte de l'ARN 7SK (n=324, facteur ≥ 1.5)				
Facteur de transcription	**Total**	**Liste**	**P-val. ajustée**	**Description**
SUZ12	1684	68	1.36E-10	Membre du complexe répresseur Polycombe PRC2/EED-EZH2. Réprime la transcription en catalysant la méthylation de H3K27
EZH2	237	20	1.06E-07	Membre du complexe répresseur Polycombe PRC2/EED-EZH2. Réprime la transcription en catalysant la méthylation de H3K27
REST	1280	45	2.68E-05	Répresseur de la transcription des gènes neuronaux. Se lie au niveau d'éléments NRS (neuron-restrictive silencer) dans les promoteurs et recrute le complexe BHC composé de HDAC et de HMT.
Gènes dont la transcription est diminuée par la perte de l'ARN 7SK (n=129, facteur ≥ 1.5)				
Facteur de transcription	**Total**	**Liste**	**P-val. ajustée**	**Description**
SUZ12	1684	39	7.43E-11	Membre du complexe répresseur Polycombe PRC2/EED-EZH2. Réprime la transcription en catalysant la méthylation de H3K27

Tableau 2 : Facteurs de transcription pour lesquels les gènes cibles sont significativement enrichis parmi les gènes positivement et négativement régulés en l'absence de l'ARN 7SK.

ii. Rôle spécifique de la protéine Larp7 dans la régulation de l'expression des gènes

Afin de vérifier la spécificité des effets observés par la perte de l'ARN 7SK, nous avons également généré des lignées Larp7 KO (Fig. S1C). La protéine Larp7 étant essentielle à la stabilité de l'ARN 7SK, l'accumulation de l'ARN 7SK est profondément réduite dans ces cellules. En conséquence, la RNP 7SK/P-TEFb est également incapable de se former dans les cellules Larp7 KO, démontré par des expériences de co-immunoprécipitation (Figure 26A). Les mêmes mécanismes d'adaptation à la perte de l'ARN 7SK semblent se mettre en place en l'absence de la protéine Larp7, avec la diminution des niveaux protéiques des deux sous-unités de P-TEFb (Fig. S1C). En accord avec le maintien d'une activité physiologique de P-TEFb, la croissance des cellules 7SK KO et Larp7 KO est fortement similaire à celle des cellules HAP1 (Figure 26B). Ainsi, comme attendu, la perte de la protéine Larp7 reflète fortement

la perte de l'ARN 7SK. Les cellules Larp7 KO constituent donc de bons contrôles pour corroborer les effets dépendants de la RNP 7SK détectés dans les cellules 7SK KO.

Nous avons vérifié que la dérégulation de l'expression génique observée en l'absence de l'ARN 7SK était conservée dans les cellules Larp7 KO. Nous avons pour cela mesuré par RT-qPCR le niveau d'expression de certains gènes identifiés dérégulés par la perte de l'ARN 7SK dans les lignées HAP1, 7SK KO et Larp7 KO (Figure 26C). L'expression des gènes PTGER3 et GUCY1A3 est effectivement réduite dans les lignées 7SK KO et Larp7 KO comparé au niveau d'expression dans les cellules HAP1, tandis que celle de STRA6 et CRABP1 est augmentée. Curieusement, nous avons identifié un gène, Bmp7, pour lequel l'expression est affectée de manière opposée par la perte de l'ARN 7SK KO et par celle de la protéine Larp7. Ce gène est en effet activé dans les cellules 7SK KO et réprimé dans les cellules Larp7 KO (Figure 26C). L'ARN 7SK étant fonctionnellement dépendant de la protéine Larp7, cette régulation différentielle pourrait provenir de fonctions de Larp7 indépendantes de l'ARN 7SK. La quasi-totalité des protéines Larp7 est normalement associée à l'ARN 7SK (Muniz et al., 2013). Les niveaux de Larp7 n'étant pas affectés par la perte de l'ARN 7SK, la fraction existant en dehors de la RNP 7SK est donc en large excès dans les cellules 7SK KO. Les effets d'une fonction spécifique de Larp7 pourraient donc être amplifiés en l'absence de l'ARN 7SK, et pourraient être à l'origine de la régulation différentielle observée pour l'expression du gène Bmp7. De précédentes études ont rapporté que la protéine Larp7 stabilisait l'ARN messager de Lin28 en stimulant son interaction avec la poly(A) polymérase Star-PAP (Dai et al., 2014). Dans le cas de Bmp7, la régulation semble cependant être transcriptionnelle, étant donné que nous retrouvons son expression augmentée dans nos données de mNET-seq. Larp7 pourrait donc être impliqué dans la régulation de l'activation des gènes au niveau transcriptionnel.

En ce sens, nous avons observé la présence d'un complexe Larp7 de haut poids moléculaire indépendant de l'ARN 7SK. En effet, la séparation des complexes cellulaires sur gradient de glycérol montre qu'une fraction de Larp7 à fort coefficient de sédimentation n'est pas impactée par la perte de l'ARN 7SK (Fig. 1B, figure 26D, fractions 10-11). Au contraire, l'abondance de ces complexes semble augmenter dans les cellules 7SK KO. L'intégrité des complexes Larp7 présents dans ces fractions est cependant dépendante d'ARN, étant donné que le traitement des extraits cellulaires à la RNase avant leur fractionnement sur gradient de glycérol induit la disparition des complexes Larp7 à fort coefficient de sédimentation (figure 26D). Cette fraction de Larp7 pourrait correspondre à celle associée à des ARNm. L'immunoprécipitation de Larp7 réalisée dans des extraits de cellules HAP1 suivie d'un marquage radioactif des ARN associés, montre qu'en plus de l'ARN 7SK, Larp7 interagit également avec un petit ARN dont la taille correspondant à celle de l'ARN U6 (figure 26E). Cette interaction semble augmenter en l'absence de l'ARN 7SK, rappelant le comportement de la fraction de Larp7 à fort coefficient de sédimentation. L'interaction entre Larp7 et l'ARN U6 a d'ailleurs été précédemment

rapportée, mais suggérée se former via la protéine MePCE, qui s'associe effectivement à l'ARN U6 et catalyse l'ajout de la coiffe mono-méthyle à son extrémité 5' (Jeronimo et al., 2007; Krueger et al., 2008; Muniz et al., 2013; Yang et al., 2019). Nous détectons également une interaction de Larp7 avec un petit ARN dont la taille correspond à celle de l'ARN U4, bien qu'à un niveau moins élevé comparé à celui de l'interaction Larp7-7SK et Larp7-U6. L'ARN U4 pourrait donc s'associer indirectement à Larp7 via l'ARN U6 au sein de la RNP U4/U6, ou être moins réactif au marquage radioactif réalisé dans cette expérience. Sans pour autant révéler l'identité des complexes Larp7 présents dans les fractions à fort coefficient de sédimentation, nous observons la co-sédimentation de Larp7 et des ARN U4 et U6 dans les fractions 10 et 11 (figure 26D). L'ensembles de ces données suggèrent que Larp7 pourrait former un complexe avec la RNP U4/U6, et jouer un rôle dans l'assemblage de cette RNP ou dans l'épissage des pré-ARNm.

Ainsi, en plus de constituer des contrôles importants pour distinguer les effets de la perte de l'ARN 7SK de potentiels effets secondaires, les cellules Larp7 KO permettront de mettre en lumière le rôle joué par la protéine Larp7 indépendamment de l'ARN 7SK.

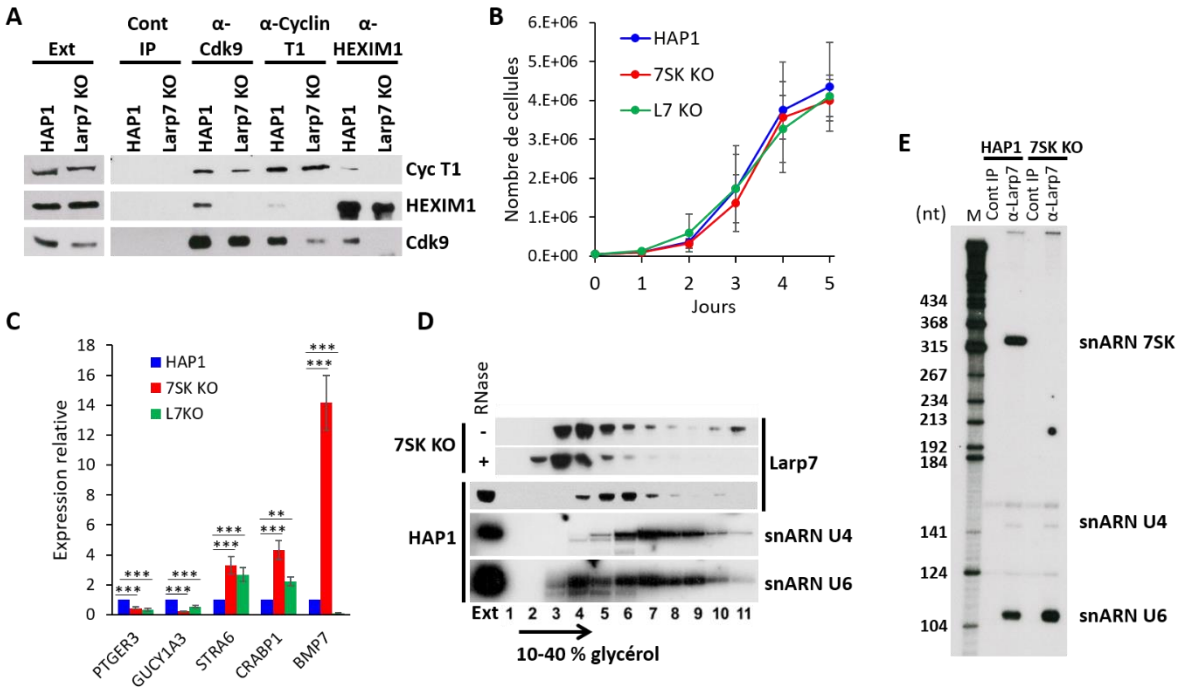


*Figure 26 : Mise en évidence de potentielles fonctions de Larp7 indépendantes de l'ARN 7SK. (A) IP des protéines Cdk9, Cycline T1 et HEXIM1 à partir d'extraits totaux de cellules HAP1 et Larp7 KO. Les protéines co-précipitées ont été analysées par WB. Les IP contrôles réalisées sans anticorps sont indiquées (Cont IP). Ext, extraits totaux. (B) Prolifération des cellules HAP1, 7SK KO et Larp7 KO (L7 KO). (C) Expression relative des ARNm des gènes indiqués dans les cellules HAP1, 7SK KO et L7 KO déterminée par RT-qPCR. Les valeurs moyennes ± écart-type sont représentées (n=3). **p < 0.01; ***p < 0.001 (D) Fractionnement d'extraits cellulaires totaux de cellules 7SK KO et HAP1 sur gradient de glycérol (10-40 %). Les extraits cellulaires de cellules 7SK KO ont été traités (+) ou non (-) à la RNase avant fractionnement. La distribution des protéines et des ARN dans les fractions du gradient a été*

analysée par WB et NB. Ext, extraits totaux. (E) Détection des ARN marqués radioactivement après IP de la protéine Larp7 dans les cellules HAP1 et 7SK KO. Les IP contrôles réalisées sans anticorps sont indiquées (Cont IP). Piste M, marqueurs de taille. nt, nucléotides.

I. Rôle de la RNP 7SK/P-TEFb dans la réponse transcriptionnelle aux UV

Résumé des résultats présentés dans l'article :

L'irradiation des cellules aux UV induit une reprogrammation transcriptionnelle massive (Gregersen and Svejstrup, 2018). Nous avons mis à profit notre lignée 7SK KO afin d'étudier le rôle joué par la RNP 7SK/P-TEFb dans ce processus. Le réservoir de P-TEFb séquestré au sein de la RNP 7SK est en effet libéré en réponse aux UV. Nous avons montré que cette libération de P-TEFb était associée à une augmentation de son activité (Fig. 4A). La formation de la RNP 7SK/P-TEFb étant abolie dans les cellules 7SK KO, elles ne disposent pas de réservoir de P-TEFb pouvant être mobilisé après irradiation aux UV (Figure 27). En conséquence, l'augmentation de l'activité de P-TEFb est perdue dans les cellules 7SK KO (Fig. 4A), nous permettant d'étudier comment les cellules mettent à profit cette activité accrue de P-TEFb dans la réponse au stress. Nous avons, notamment, interrogé la fonction du réservoir de P-TEFb dans deux aspects de la réponse transcriptionnelle induite par les UV : (1) la régulation de la pause de l'ARN Pol II à proximité du promoteur et (2) l'induction des gènes de réponse au stress. Nos résultats indiquent que la RNP 7SK/P-TEFb participe aux deux. En analysant la transcription naissante par mNET-seq, nous avons mis en évidence un impact prépondérant de l'absence de l'ARN 7SK 1h après exposition aux UV, temps auquel la RNP 7SK/P-TEFb se dissocie (Fig. 4, 5). L'absence de l'ARN 7SK affecte la libération globale des ARN Pol II en pause à proximité du promoteur, potentiellement affectant la détection des lésions dans les corps de gènes (Fig. 4, S4). L'absence de l'ARN 7SK affecte également l'induction de certains gènes de réponse au stress (Fig. 5, S5). D'un point de vu mécanistique, nous avons montré par des expériences de ChIP-qPCR que l'induction de ces gènes était compromise en l'absence de l'ARN 7SK via un défaut de libération de la pause à proximité du promoteur et de recrutement de facteurs d'élongation, et notamment du SEC (Fig. 6, S6). Nos résultats indiquent également que l'induction de ces gènes après irradiation dépend du réservoir nucléoplasmique de P-TEFb (Fig. 7, S7). En effet, ni l'évaluation de la localisation génomique de l'ARN 7SK par ChIRP-seq, ni celle de MePCE par ChIP-qPCR n'indiquent la présence de la RNP 7SK dans les régions promotrices de ces gènes avant ou après l'exposition aux UV (Fig. 7, S7). De plus, des expériences de fractionnement cellulaire suggèrent que P-TEFb est recruté à la chromatine sous sa forme active après irradiation, de manière dépendante de l'ARN 7SK (Fig. 7E). L'intégrité de la RNP 7SK/P-TEFb est d'ailleurs nécessaire à ce processus d'induction transcriptionnelle, étant donné que la déplétion de Larp7 par siARN dans les cellules HAP1, bien qu'augmentant la fraction active de P-TEFb, ne restaure pas l'activation des gènes en réponse aux UV (Fig. 7F).

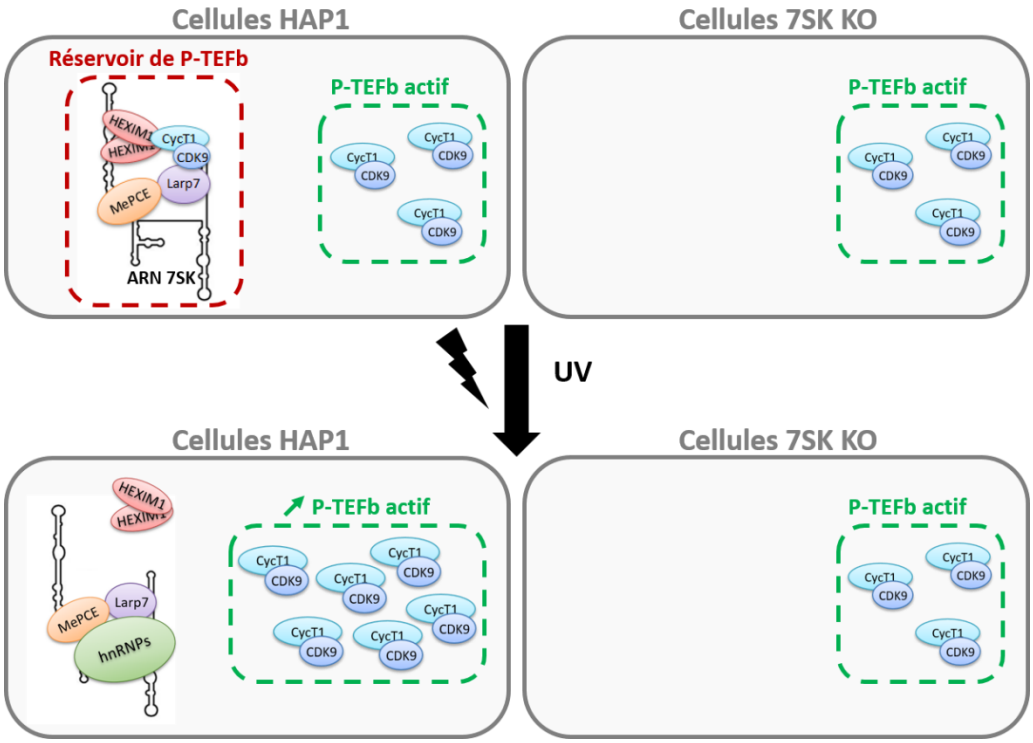


Figure 27 : Représentation schématique de l'activité de P-TEFb en conditions normales et après stress UV dans les cellules HAP1 et 7SK KO.

Résultats complémentaires :

P-TEFb se dissocie rapidement de la RNP 7SK après irradiation aux UV, et reste dissocié pendant plusieurs heures. Nous avons effectivement observé par des expériences de co-immunoprécipitation que l'association de P-TEFb à l'ARN 7SK ne commençait à se rétablir qu'à partir de ~ 8h après irradiation (figure 28A), générant une large fenêtre temporelle durant laquelle l'activité de P-TEFb reste élevée. L'activation transcriptionnelle en réponse à un stimulus est un processus dynamique. Différents groupes de gènes sont en effet induits à différents temps après un stimulus. Notamment, les gènes activés lors de la première vague d'induction (IEG, immediate early genes) permettent l'expression de facteurs impliqués dans l'activation de gènes lors d'une seconde vague d'induction (SRG, secondary response genes) (Bahrami and Drablos, 2016). L'effet de 7SK sur l'expression des IEG pourrait donc se répercuter sur les SRG. Afin d'identifier les effets directs de la libération de P-TEFb de la RNP 7SK, nous avons choisi d'étudier l'induction des gènes à des temps relativement courts après irradiation. Les conditions optimales ont été déterminées en analysant l'activation transcriptionnelle de plusieurs gènes de réponse à différents temps suivant l'irradiation (figure 28B). Pour l'ensemble des gènes testés, l'induction transcriptionnelle déterminée par RT-qPCR est maximale 1h après irradiation. Ce timing correspond également au temps auquel a été observé la libération globale des ARN Pol II en pause à proximité du promoteur. Nous avons donc choisi de nous concentrer sur les effets

transcriptionnels apparaissant 1h après exposition aux UV, temps auquel les défauts transcriptionnels majeurs directement liés à la perte de la RNP 7SK/P-TEFb devraient être observés.

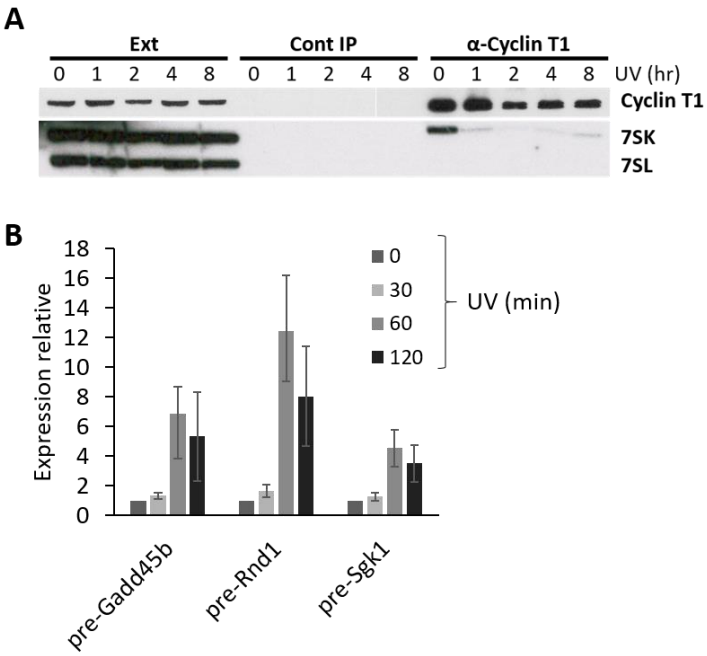


Figure 28 : Dissociation de la RNP 7SK/P-TEFb et induction transcriptionnelle après irradiation aux UV. (A) IP de Cycline T1 à partir d'extraits totaux de cellules HAP1 récoltées à différents temps après irradiation aux UV (40 J/m²). Les IP ont été analysées par WB et les ARN co-précipités par NB. Les IP contrôles réalisées sans anticorps sont indiquées (Cont IP). Ext, extraits totaux. (B) Expression relative des ARN pré-messagers des gènes indiqués dans les cellules HAP1 déterminée par RT-qPCR. Les niveaux d'expression ont été mesurés à différents temps après irradiation aux UV comme indiqué. min, minutes. Les résultats ont été normalisés aux valeurs avant irradiation aux UV (t=0). Les valeurs moyennes ± ecart-type sont représentées (n=3).

i. Libération de l'ARN Pol II en pause par le réservoir inactif de P-TEFb

Les ARN Pol II en pause à proximité du promoteur sont rapidement libérées en réponse aux UV (lavigne 2017). P-TEFb a été impliqué dans ce mécanisme étant donné son rôle majeur dans la libération de la pause de l'ARN Pol II. La libération des ARN Pol II après irradiation est effectivement inhibée si les cellules sont traitées au DRB, un inhibiteur de P-TEFb (Lavigne et al., 2017). En mesurant la transcription naissante par mNET-seq dans les cellules HAP1 et 7SK KO avant et après irradiation, nous avons montré que cette libération massive des ARN Pol II en pause était dépendante du réservoir de P-TEFb, libéré de la RNP 7SK après exposition aux UV (Fig. 4, S4). En effet, nous observons dans les HAP1 une diminution globale du niveau de pause accompagnée d'une augmentation de la transcription dans les corps de gènes, événements caractéristiques de l'entrée en élongation productive (Fig. 4C, E, F). En revanche, la libération de la pause de l'ARN Pol II est significativement diminuée en l'absence de la RNP 7SK/P-TEFb (Fig. 4D, F).

La libération de la pause de l'ARN Pol II en réponse aux UV n'est cependant pas associée à une activation transcriptionnelle globale. Le taux d'élongation est en effet rapidement et fortement diminué après irradiation aux UV (Williamson et al., 2017). Nous avons d'ailleurs montré que les gènes activés en réponse aux UV dans les HAP1 sont globalement plus courts que la moyenne (Fig. 5G), en accord avec de précédentes études (McKay et al., 2004; Williamson et al., 2017). En analysant la transcription naissante selon la taille des gènes, nous pouvons approximer la distance parcourue par les ARN Pol II libérées de la pause dans l'heure suivant l'irradiation (Figure 29A, flèches). Pour les gènes de taille inférieure à 12 kb, la vague de polymérase libérée après UV parvient à la fin du gène (TES, transcription end site). Pour les gènes plus longs en revanche, peu de polymérases parviennent au TES, matérialisé par une diminution de la transcription naissante moyenne en fin de gène. En effet, l'avancée de l'ARN Pol II dans les corps de gènes diminue proportionnellement à l'augmentation de la taille des gènes, avec une distance moyenne parcourue en 1h estimée aux alentours des 25 kb. Ces données sont donc cohérentes avec une réduction du taux d'élongation après irradiation, car la vitesse moyenne de l'ARN Pol II est de 2 Kb/min, soit 120 kb en 1h (Jonkers et al., 2014; Williamson et al., 2017). En l'absence de l'ARN 7SK, la progression de la vague d'ARN Pol II est d'autant plus réduite (Figure 29A, droite). La distance approximative parcourue par l'ARN Pol II dans les corps de gènes est en effet systématiquement inférieure dans les cellules 7SK KO à celle parcourue en moyenne par l'ARN Pol II dans les cellules HAP1. Ces données confirment à nouveau la perturbation de la libération des ARN Pol II en pause observée en l'absence de la RNP 7SK/P-TEFb : moins d'ARN Pol II entrant dans les corps de gènes, le signal moyen de transcription naissante diminue d'autant plus dans les régions distales.

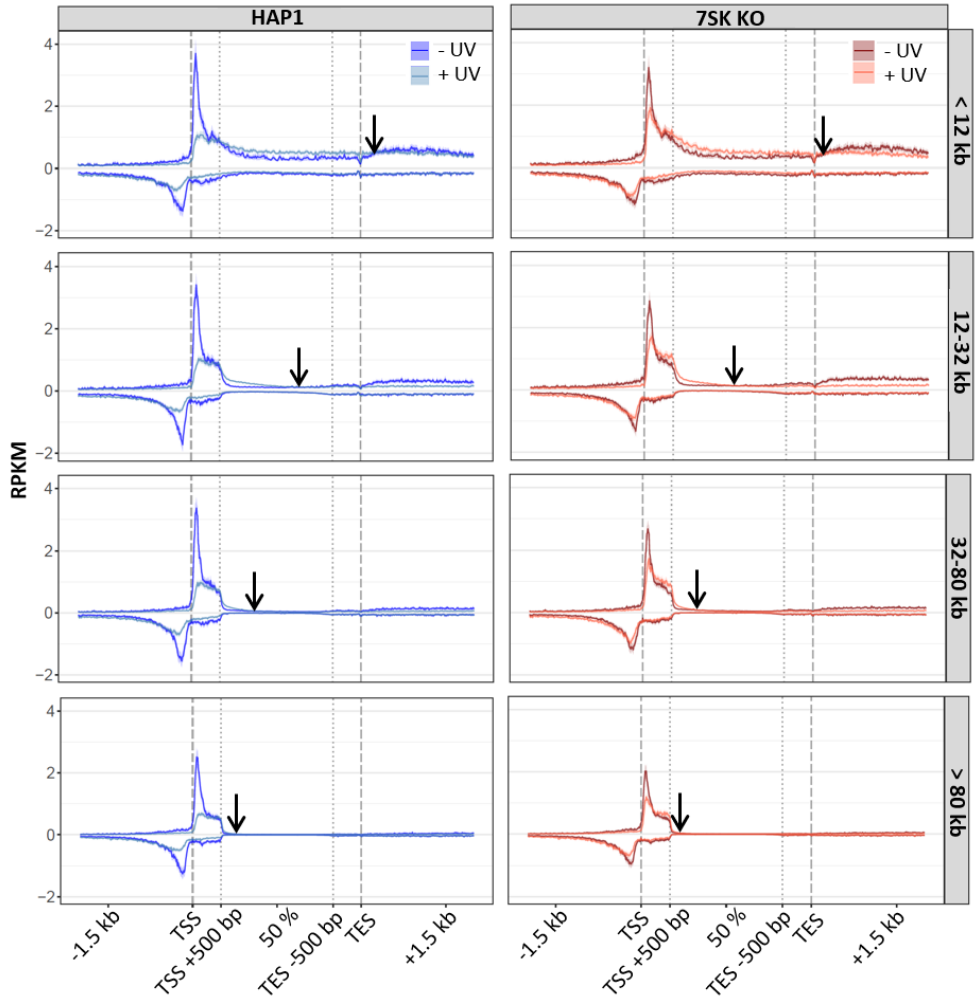


Figure 29 : Propagation de la vague transcriptionnelle en réponse aux UV. Les méta-profiles du signal de mNET-seq sur les gènes dans les cellules HAP1 (gauche) et 7SK KO (droite) ont été réalisés pour différents groupes de gènes établis en fonction de leur taille. Les flèches indiquent l'avancée moyenne approximative de l'ARN Pol II dans les corps de gènes. Les intervalles de taille considérées sont indiqués à droite. L'axe y représente l'intensité du signal (RPKM). L'axe x indique la position relative dans le gène. TSS, transcription start site. TES, transcription end site.

ii. Activation transcriptionnelle dépendante de l'ARN 7SK en réponse aux UV

A l'instar de la comparaison des gènes dérégulés par l'absence de l'ARN 7SK en conditions normales (section I.a., figure 24), la comparaison des gènes transcriptionnellement induits en réponse aux UV dans les deux réplicats de mNET-seq révèle l'induction commune (d'un facteur ≥ 2) de seulement la moitié des gènes (figure 30A). Le seuil fixé est, comme indiqué dans la partie I.a., en partie responsable de la différence d'induction observée. Cependant, il a récemment été montré que les appareils d'irradiation (UV-crosslinker) généraient une grande variabilité dans la dose d'UV délivrée (Tufegdzic Vidakovic et al., 2019). Le nombre de gènes activés et leur niveau d'induction étant sensibles à la dose d'UV (Gentile et al., 2003), la variation induite par l'appareil d'irradiation participe sans doute également à la variabilité d'induction observée entre les deux réplicats de mNET-seq.

Comme nous venons de le voir, l'irradiation aux UV induit une libération globale de la pause à proximité du promoteur. Ainsi, il est possible que certains gènes identifiés comme activés, en particulier les gènes courts étant donné le taux d'élongation moyen réduit après irradiation, ne donnent pas lieu à des transcrits matures. Dans l'idéal, la comparaison entre nos données de mNET-seq et des données de RNA-seq réalisés dans les mêmes conditions permettrait de supprimer ces faux-positifs. En l'absence de telles données, nous estimons que le seuil d'induction choisi (≥ 2) permet de limiter l'apparition de ces faux-positifs.

Parmi les gènes induits en réponse aux UV, nous avons montré que l'activation de certains d'entre eux était fortement réduite voir abolie en l'absence de la RNP 7SK/P-TEFb, et notamment de nombreux gènes impliqués dans les processus d'apoptose, de prolifération et de signalisation cellulaire (Fig. 5B, S5A). Plusieurs modèles peuvent être envisagés pour expliquer la dépendance de l'ARN 7SK dans ce processus d'induction (figure 30B). Premièrement, la RNP 7SK/P-TEFb pourrait être présente au niveau des régions promotrices des gènes inductibles aux UV avant l'irradiation, ou y être recrutée en réponse aux UV, afin de délivrer P-TEFb de manière ciblée (figure 30B, (1)). Ce modèle a été proposé pour l'induction des gènes dépendant du facteur de transcription NF-kB en réponse au traitement à l'etoposide, un agent inducteur de cassure doubles-brins, ou au TNF-α (Gudipaty et al., 2015; McNamara et al., 2016a). Toutefois, nos résultats ne sont pas en faveur de ce modèle, étant donné que nous n'avons pas détecté la RNP 7SK au niveau des régions promotrices des gènes induits, que ce soit avant ou après irradiation (Fig 7, S7).

Deuxièmement, l'activation des gènes pourrait nécessiter de forts niveaux de P-TEFb actif, permis par la libération du réservoir de P-TEFb en réponse aux UV (figure 30B, (2)). Ce modèle correspond au mécanisme d'activation du gène intégré du VIH, particulièrement sensible aux quantités de P-TEFb actif (Bartholomeeusen et al., 2013; Bartholomeeusen et al., 2012; Chiu et al., 2004). Pour tester ce deuxième modèle, nous avons procédé à la déplétion de Larp7 par siARN dans les cellules HAP1. La perte transitoire de Larp7 entraîne la co-déplétion de l'ARN 7SK et empêche ainsi la formation de la RNP 7SK/P-TEFb, mais sans impacter les niveaux protéiques des deux composants de P-TEFb (Fig. 7F, S7D). Les niveaux de P-TEFb actifs induits par la déplétion de Larp7 sont donc théoriquement similaires, voir supérieurs, à ceux retrouvés après dissociation de la RNP 7SK/P-TEFb en réponse aux UV. Cependant, l'induction en réponse aux UV des gènes que nous avons testés n'est pas rétablie après déplétion de Larp7 (Fig. 7F). Pour corroborer ces résultats, nous avons mesuré par RT-qPCR l'activation transcriptionnelle de gènes de réponse à différents temps après irradiation (figure 30C). En effet, l'activité transcriptionnelle de l'ARN Pol II étant progressivement réduite après irradiation aux UV (Fig. 4, S4), l'activité de P-TEFb devient théoriquement moins limitante. Ainsi, nous pourrions nous attendre à un retard de l'induction des gènes de réponse dans les cellules 7SK KO, qui ne possèdent pas de

réservoir de P-TEFb. Dans les HAP1, le profil d'induction des trois gènes testés est sensiblement similaire et montre une induction rapide, maximale à 1h et 2h après l'irradiation (figure 30C). Dans les cellules 7SK KO, le profil d'induction est plus variable, l'expression des gènes Gadd45b et Rnd1 bénéficiant d'une légère activation, maximale à 3 et 2h, respectivement. L'expression de Sgk1 n'est induite à aucun temps analysé. Les niveaux d'induction de Gadd45b et Rnd1 restent cependant plus bas dans les cellules 7SK KO par rapport à ceux dans les cellules HAP1, suggérant un recrutement inefficace de P-TEFb. Ceci est également observé de manière générale sur les gènes différentiellement induits 1h après irradiation aux UV dans les cellules HAP1 et 7SK KO (figure 30D). L'induction est en effet globalement plus basse en l'absence de l'ARN 7SK à 1h, 2H et 4h après irradiation. Ces données, couplées à celles obtenues avec la déplétion de Larp7, sont donc en défaveur du deuxième modèle.

Troisièmement, un facteur de transcription activé par les UV pourrait spécifiquement cibler la RNP 7SK/P-TEFb pour en extraire P-TEFb et le cibler sur les gènes de réponse (figure 30B, (3)). Ce modèle a récemment été proposé pour le facteur Rbm7, activé par p38 en réponse aux UV (Bugai et al., 2019). Ayant démontré la nécessité de l'intégrité de la RNP 7SK mais pas sa localisation chromatinienne, nos résultats convergent vers ce troisième modèle. Cependant, malgré les efforts entrepris, nous n'avons pas identifié le facteur permettant d'extraire P-TEFb de la RNP 7SK/P-TEFb ou de le cibler sur les gènes. Nos résultats préliminaires n'indiquent par exemple pas le recrutement de Brd4, à même de recruter la forme active de P-TEFb sur les gènes, ou de KAP1, susceptible de recruter la RNP 7SK/P-TEFb, au niveau des régions promotrices de gènes candidats (données non montrées). Nos conditions expérimentales n'ont pas non plus permis d'observer l'interaction entre les différents composants de la RNP 7SK/P-TEFb et Rbm7, un candidat pourtant intéressant au vu de données récentes (Bugai et al., 2019).

Enfin, nous avons aussi considéré la possibilité que l'ARN 7SK puisse promouvoir l'induction des gènes après exposition aux UV par un mécanisme indépendant de P-TEFb, comme précédemment montré pour l'expression des snARN du spliceosome et la répression des eRNA. Pour tester cette hypothèse, nous avons généré des lignées exprimant stablement une version mutante de l'ARN 7SK, incapable de lier HEXIM1 et donc de séquestrer P-TEFb (figure 30E) (Egloff et al., 2006). Il est important de noter que cette mutation, localisée dans la région distale de la tige-boucle en 5', n'affecte pas la structure de l'ARN 7SK. La liaison des autres partenaires identifiés de l'ARN 7SK, qui dépend principalement de l'intégrité de la troisième tige-boucle, ne sera donc pas affectée. L'induction transcritpionnelle en réponse aux UV de deux gènes modèles, Gadd45b et Sgk1, n'est cependant pas restaurée dans les cellules exprimant la version mutante de l'ARN 7SK, tandis qu'elle l'est partiellement dans les cellules exprimant la version sauvage de l'ARN 7SK (figure 30F). Comme l'expression de l'ARN 7SK n'est que partiellement restaurée dans ces cellules (Fig. 1A), l'induction intermédiaire était attendue. Ces

résultats indiquent que l'induction des gènes est bien dépendante de la formation de la RNP 7SK/P-TEFb.

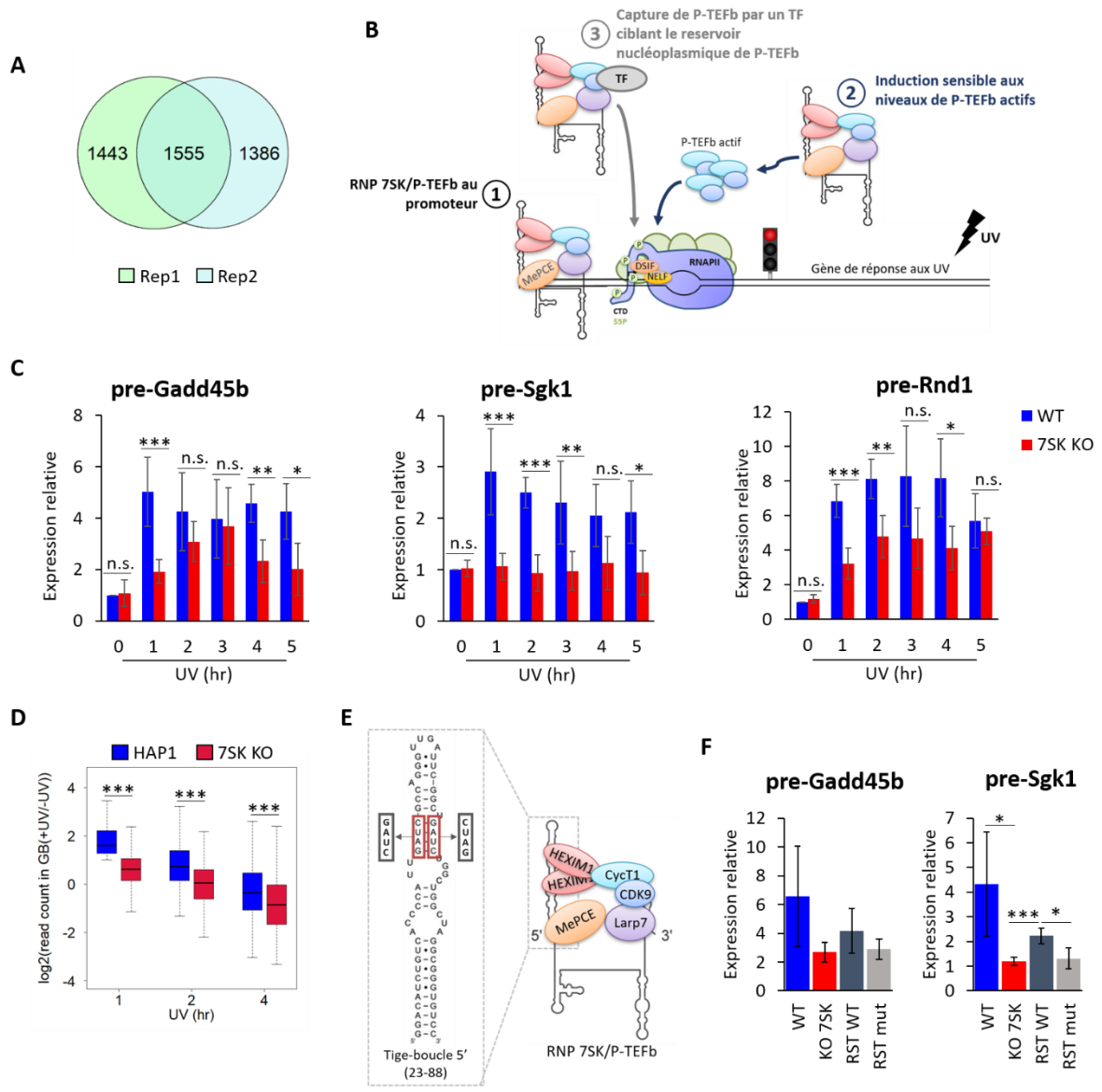


Figure 30 : Induction des gènes dépendante de la RNP 7SK/P-TEFb. (A) Diagramme de Venn comparant les gènes induits après irradiation aux UV dans les HAP1 (facteur ≥ 2) dans deux réplicats biologiques de mNET-seq. (B) Modèles envisagés pour l'induction des gènes dépendante de la RNP 7SK/P-TEFb. (C) Expression relative des ARN pré-messagers des gènes indiqués dans les cellules HAP1 et 7SK KO déterminée par RT-qPCR. Les niveaux d'expression ont été mesurés à différents temps après irradiation aux UV comme indiqué. hr, heures. Les résultats ont été normalisés aux valeurs avant irradiation aux UV (t=0). Les valeurs moyennes ± ecart-type sont représentées (n=3) (D) Quantification par boxplot du niveau d'induction transcriptionnelle des gènes dépendants de la RNP 7SK/P-TEFb à différents temps après irradiation aux UV (n=235). L'induction a été calculée par normalisation du signal de mNET-seq à 1, 2 et 4h après UV au signal avant UV pour chaque gène. (E) Schéma de la mutation dans la tige-boucle 5' de l'ARN 7SK. (F) Expression relative des ARN pré-messagers des gènes indiqués dans les cellules HAP1 (WT), 7SK KO ou restaurées avec une version sauvage (RST WT) ou mutée (RST mut) déterminée par RT-qPCR. Les résultats ont été normalisés aux valeurs avant irradiation aux UV (t=0). Les valeurs moyennes ± ecart-type sont représentées (n=3).

.Discussion

I. Fonction de la RNP 7SK/P-TEFb dans la régulation de l'expression génique en conditions normales

Les travaux que je présente ici permettent de mettre à jour la fonction de l'ARN 7SK dans la régulation de la transcription par l'ARN Pol II. Plusieurs études ont en effet rapporté la présence de l'ARN 7SK au niveau de nombreuses régions génomiques régulatrices, promoteurs et enhancers, ainsi que sur les gènes des snARN (Calo et al., 2015; Egloff et al., 2017a; Flynn et al., 2016; Ji et al., 2013; McNamara et al., 2016a). Bien que l'analyse de la localisation de l'ARN 7SK à l'échelle du génome soit encore en cours d'analyse, les cellules 7SK KO constituent un contrôle idéal pour identifier les régions ciblées spécifiquement par cet ARN. Les techniques permettant de cartographier à l'échelle du génome la liaison d'un facteur d'intérêt à l'ADN souffrent en effet de différents biais pouvant résulter en une amplification non-spécifique de certaines régions du génome. Nos données préliminaires de ChIRP-seq 7SK indiquent par exemple un enrichissement du signal au niveau du TSS des gènes équivalent dans les cellules HAP1 et 7SK KO (Fig. 7B, S7D), pouvant provenir d'une forte accessibilité de ces régions (Auerbach et al., 2009; Ramachandran et al., 2015; Teytelman et al., 2009). Etant donné que la RNP 7SK a principalement été détectée dans ces régions-là, nos données de ChIRP-seq indiquent que certains de ces enrichissements pourraient en réalité être des faux-positifs et nécessiteraient une réévaluation avec un contrôle adéquat. Des ChIP-seq contre une version étiquetée de la protéine Larp7 dans des cellules HeLa n'avait déjà pas permis de visualiser un enrichissement de cette protéine au niveau des régions promotrices des gènes codants les protéines (Egloff et al., 2017b). Les différentes expériences réalisées dans le laboratoire vont donc à l'encontre d'une association stable de la RNP 7SK avec la plupart des régions promotrices des gènes codants les protéines.

Malgré les données rapportées dans la littérature suggérant un rôle de l'ARN 7SK dans la régulation globale de la transcription par l'ARN Pol II, l'étude transcriptomique effectuée dans les cellules HAP1 et 7SK KO a révélé un impact discret de la perte de l'ARN 7SK sur le niveau d'expression des gènes. Cependant, le nombre de gènes dérégulés en l'absence de l'ARN 7SK pourrait être sous-estimé du fait de l'apparition de mécanisme compensatoires, propres aux lignées cellulaires KO. Une précédente étude a par exemple mis en évidence un rôle spécifique de la RNP 7SK/P-TEFb dans l'expression des gènes codants les protéines ribosomiques et snoARN, au niveau desquels son recrutement permettrait de faciliter l'entrée en élongation productive (Calo et al., 2015). Or, ni l'analyse de l'accumulation des ARN cellulaires ni celle des ARN en cours de transcription n'indiquent d'effets de la perte de l'ARN 7SK sur cette catégorie de gènes. Le même constat s'applique pour les gènes des snARN, dont la transcription est pourtant stimulée par la RNP 7SK/LEC au sein de laquelle l'ARN 7SK joue un rôle

structural majeur (Egloff et al., 2017b). L'ARN 7SK est pourtant bien détecté dans les HAP1 sur cette catégorie de gènes (Fig. 7D), confortant l'hypothèse d'un mécanisme compensatoire. Etant donné que les gènes des snARN sont présents en copies multiples dans le génome (Lund and Dahlberg, 1984; Van Arsdell and Weiner, 1984), la perte de l'ARN 7SK pourrait être compensée par l'activation d'un plus grand nombre de copies. Un tel mécanisme de compensation est par exemple employé pour réguler l'expression des gènes des ARN ribosomiques durant l'embryogénèse chez la souris (Zheng et al., 2012). Enfin, nous avons également mis en évidence la diminution de l'expression des deux composants de P-TEFb, Cdk9 et Cyclin T1, probablement via la diminution de la stabilité de ces deux protéines. La transcription des gènes codants Cdk9 et Cycline T1 l'accumulation des ARNm correspondants ne sont en effet pas impactés dans les cellules 7SK KO. Il est concevable qu'un excès de P-TEFb dans les cellules, en dehors d'un contexte particulier retrouvé par exemple dans certains cancers (voir Introduction bibliographique II.ii.c.), ne soit pas stable et donc rapidement dégradé par le protéasome. Il serait alors intéressant de comparer la dérégulation de l'expression des gènes dans les lignées 7SK KO et dans des cellules transitoirement déplétées de l'ARN 7SK au moyen de siARN, dans lesquelles les mécanismes compensatoires potentiellement affectant des gènes cibles de l'ARN 7SK ne seraient pas activés.

A l'inverse, il est également possible que les gènes dérégulés en l'absence de l'ARN 7SK le soient par un mécanisme indirect. En effet, l'analyse préliminaire des ChIRP-seq de l'ARN 7SK ne semble pas indiquer sa présence au niveau des gènes dont la transcription est positivement ou négativement impactée par sa perte. Etant donné que l'ARN 7SK a été impliqué dans la régulation de l'activité des enhancers, nous considérerons tout particulièrement la présence de l'ARN 7SK au niveau de ces éléments régulateurs au voisinage des gènes différentiellement exprimés. Cependant, comme je l'ai brièvement mentionné dans la partie résultats I.a., la perte de l'ARN 7SK peut également affecter l'expression de facteurs de transcription ou cofacteurs qui à leur tour peuvent affecter l'expression d'autres gènes. La liste des gènes dérégulés en l'absence de l'ARN 7SK serait alors composée de cibles réelles et de gènes indirectement régulés, compliquant l'identification d'un mécanisme 7SK-dépendant. De plus, comme cela a été discuté pour la protéine Larp7 (résultats complémentaires I.b.), pour laquelle les niveaux protéiques ne sont pas affectés et qui est donc en excès en l'absence de l'ARN 7SK, les niveaux protéiques de MePCE restent également constants en l'absence de l'ARN 7SK tandis que ceux d'HEXIM augmentent légèrement. Nous ne pouvons donc exclure un effet de l'excès de ces protéines sur l'expression des gènes dans les cellules 7SK KO. MePCE semble en effet pouvoir interagir indépendamment de l'ARN 7SK avec Cdk9 et le cibler au niveau des régions promotrices des gènes (Shelton et al., 2018). HEXIM peut quant à lui interagir avec le récepteur nucléaire aux glucocorticoïdes et inhiber ses gènes cibles (Shimizu et al., 2005; Yoshikawa et al., 2008). La comparaison des gènes

cibles de chacun de ces facteurs par ChIP-seq/ChIRP-seq, ainsi que des gènes différentiellement exprimés dans leurs lignées KO respectives, pourrait aider à discerner les effets individuels et collectifs des composants de la RNP 7SK/P-TEFb. De plus, des données de ChIRP-MS ont récemment été générées dans l'équipe afin d'identifier les protéines associées à l'ARN 7SK (données non montrées). Leur analyse, et leur recoupement avec les protéines associées à la protéine Larp7 (Gudipaty et al., 2015; Krueger et al., 2008)(données non publiées), permettra également de mieux comprendre les mécanismes par lesquels la RNP 7SK régule l'expression des gènes.

En dépit des potentiels effets indirects énoncés ci-dessus, l'analyse de l'enrichissement pour des catégories biologiques parmi les gènes dérégulés en l'absence de l'ARN 7SK indique clairement un enrichissement pour des gènes impliqués dans le développement neuronal, aussi bien pour les gènes positivement que négativement régulés (Fig. S2B, S2C). Un enrichissement pour des gènes impliqués dans les processus neuronaux a également été rapporté pour les gènes dérégulés après déplétion de l'ARN 7SK au moyen d'oligonucléotides antisens dans des cellules souches embryonnaires de souris (Castelo-Branco et al., 2013). De plus, des variations d'expression de l'ARN 7SK ont été observées au cours du développement neuronal murin et suite à la différenciation de cellules souches en cellules de neurones (Bazi et al., 2018; Skreka et al., 2012). L'ensemble de ces données pointent vers un rôle spécifique de l'ARN 7SK dans le contrôle de l'expression des gènes neuronaux. L'identification d'un enrichissement pour les gènes cibles du facteur de transcription REST, lui-même impliqué dans la régulation transcriptionnelle des gènes neuronaux, parmi les gènes dérégulés en l'absence de l'ARN 7SK laisse supposer un lien étroit entre les deux. L'étude d'une coopération potentielle entre l'ARN 7SK et REST au cours de la différentiation neuronale pourrait apporter de nouveaux éléments quant à leur fonction au cours de ce processus.

II. Contrôle de la pause de l'ARN Pol II à proximité du promoteur par la RNP 7SK/P-TEFb

La fraction active de P-TEFb contrôle le niveau d'ARN Pol II en pause à proximité du promoteur en régulant l'entrée en élongation productive. Nos travaux suggèrent que la fraction inactive de P-TEFb joue également un rôle dans ce contrôle, bien que le mécanisme sous-jacent soit encore élusif et ne semble pas avoir un impact majeur sur l'expression des gènes associés. En l'absence de l'ARN 7SK, le niveau d'ARN Pol II en pause à proximité du promoteur est en effet globalement diminué de manière conséquente (Fig. 3, S3). En théorie, une diminution de la pause peut être induite soit par une diminution de l'initiation de la transcription, soit par une augmentation de l'entrée en élongation productive, soit par une augmentation de la terminaison précoce (figure 31) (Brannan et al., 2012; Gressel et al., 2017; Wagschal et al., 2012). Les deux premières hypothèses impliquent clairement des

conséquences sur le niveau de transcription dans les corps de gènes. L'augmentation de l'entrée en élongation résulterait effectivement en une augmentation de la transcription dans les corps de gènes, alors que la diminution du taux d'initiation aurait l'effet opposé. Or, nos données indiquent que la diminution du niveau de pause ne s'accompagne pas d'une variation globale du signal de transcription dans les corps de gènes. Les gènes dont l'expression est augmentée en l'absence de l'ARN 7SK ont d'ailleurs tendance à avoir un niveau d'ARN Pol II en pause plus important (Fig. 2E, F). De plus, l'activité de P-TEFb n'est pas significativement altérée en l'absence de l'ARN 7SK, défavorisant l'hypothèse d'une entrée en élongation accrue due à un excès de P-TEFb. La RNP 7SK pourrait donc être impliquée dans la stabilité des ARN Pol II résidant aux sites de pause. Des résultats similaires à ce que nous observons dans les cellules 7SK KO ont récemment été identifiés après déplétion du complexe FACT (Tettey et al., 2019). La perte de la sous-unité SSRP1 du complexe FACT induit en effet la diminution globale du niveau d'ARN Pol II en pause qui semble résulter d'un turnover plus rapide de l'ARN Pol II aux promoteurs. Il sera intéressant d'évaluer si la durée de la pause est effectivement diminuée en l'absence de l'ARN 7SK, ce qui pourrait signifier une coopération entre l'ARN 7SK et le complexe FACT. En ce sens, nous avons identifié les deux sous-unités de ce complexe, SSRP1 et Spt16, dans la liste des protéines associées à l'ARN 7SK obtenue par ChIRP-MS. Bien que l'ARN 7SK ne semble pas être stablement présent au niveau des régions promotrices des gènes, il pourrait interagir avec FACT et favoriser son recrutement afin de stabiliser les ARN Pol II en pause. Après validation des interactions suggérées par les données de spectrométrie de masse, la localisation génomique de FACT dans les cellules HAP1 et 7SK KO permettra de déterminer si effectivement une coopération entre l'ARN 7SK et FACT existe. Plus généralement, identifier l'impact de la perte de l'ARN 7SK sur le recrutement d'autres facteurs impliqués dans la mise en place de la pause ou dans sa libération, tels que Gdown1, PAF1C et TFIIS, ainsi que ceux impliqués dans la terminaison précoce, TTF2 et Xrn2, permettra de mieux comprendre le mécanisme sous-jacent.

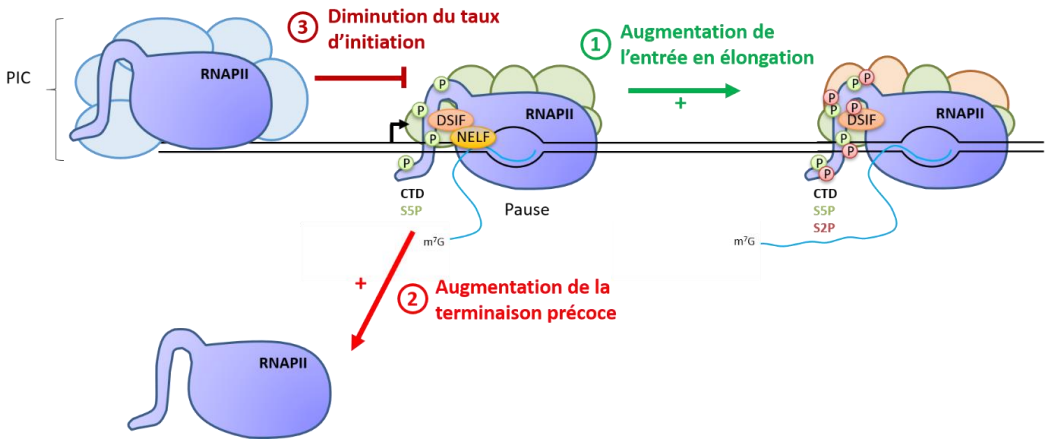


Figure 31 : Mécanismes théoriques induisant la diminution du niveau de pause de l'ARN Pol II.

III. Rôle déterminant du réservoir inactif de P-TEFb dans la réponse aux UV

L'irradiation aux UV induit la dissociation rapide de la RNP 7SK/P-TEFb, conduisant à une augmentation de la fraction de P-TEFb actif. Son rôle dans la réponse au stress reste cependant peu étudié. Nos résultats montrent que l'ARN 7SK joue un rôle majeur dans ce processus, illustré par une sensibilité aux UV accrue en son absence. L'absence de l'ARN 7SK cause en effet un défaut de libération global des ARN Pol II en pause et d'induction de gènes de réponse aux UV, tous deux pouvant participer à la faible survie cellulaire observée après irradiation.

i. Libération massive des ARN Pol II en pause par le réservoir inactif de P-TEFb

La mesure de la transcription naissante par mNET-seq avant et après irradiation aux UV a permis de visualiser la libération globale des ARN Pol II en pause induite par les UV (Fig. 4, S4, figure 28), en accord avec de précédentes études (Borisova et al., 2018; Lavigne et al., 2017; Williamson et al., 2017). Cette libération globale a précédemment été montrée dépendre de P-TEFb (Lavigne et al., 2017). Nos travaux ont permis de démontrer que le réservoir de P-TEFb, libéré de la RNP 7SK en réponse aux UV, jouait un rôle critique dans ce processus. Il est en effet facilement concevable que cette libération massive de la pause nécessite une forte activité de P-TEFb. En absence du réservoir de P-TEFb, nous avons observé un ralentissement de la propagation de l'ARN Pol II dans les corps de gènes, matérialisé par une persistance de la transcription naissante dans les régions à proximité des promoteurs, et ce particulièrement aux sites de pause. Ces données indiquent un processus de libération effectivement ralenti. En conséquence, l'ARN Pol II progresse à une plus faible distance dans les corps de gènes à un temps donné. La perte de l'ARN 7SK pourrait donc avoir de sérieuses conséquences sur la vitesse de réparation des dommages à l'ADN localisés sur les brins matrices des gènes. En effet, l'initiation du processus de réparation par le TC-NER repose sur la reconnaissance par CSB de l'ARN Pol II bloquée au niveau d'une lésion (Gregersen and Svejstrup, 2018). Il a été proposé que la libération globale des ARN Pol II en pause après l'irradiation permette de favoriser la rencontre des lésions par l'ARN Pol II, et donc de promouvoir leur réparation (Lavigne et al., 2017). Ainsi, la réparation des dommages à l'ADN dans les régions transcrites par le TC-NER pourrait être significativement retardée dans les cellules 7SK KO et compromettre leur taux de survie. Des expériences supplémentaires pour comparer la cinétique de réparation dans les cellules HAP1 et 7SK KO, par exemple par analyse de la disparition des produits d'excision générés par le NER après irradiation par XR-seq, pourraient permettre d'évaluer l'importance de l'ARN 7SK dans la récupération d'une activité cellulaire normale après irradiation.

ii. Rôle de la RNP 7SK/P-TEFb dans l'activation des gènes de réponse aux UV

a. Libération spécifique de la pause au niveau des gènes activés

Nous avons montré qu'au moins 500 gènes, dont des acteurs majeurs de la réponse au stress, présentaient un défaut d'induction en l'absence de l'ARN 7SK ressemblant à un défaut d'activité de P-TEFb (Fig. 5, S5). Cependant, nos tentatives pour évaluer l'impact de la perte de l'ARN 7SK sur le recrutement de P-TEFb au niveau des régions promotrices des gènes ont été endiguées par une absence d'anticorps permettant sa détection par ChIP. Les anticorps utilisés dans toutes les études présentant des ChIP-Cdk9 ne sont en effet plus disponibles (Santa Cruz), et les différents anticorps que nous avons testés se sont révélés inefficaces à l'immunoprécipitation de la fraction chromatinienne de Cdk9. Des difficultés ont d'ailleurs été rapportées par d'autres laboratoires (Shelton et al., 2018). Nous testerons prochainement l'utilisation d'un mix d'anticorps anti-Cdk9/anti-Cycline T1, qui permet d'augmenter la qualité du signal de ChIP et s'est avéré efficace pour la détection chromatinienne de P-TEFb ((Hertweck et al., 2016), R. Jenner, communication personnelle). En dépit de cette déconvenue, nos résultats argumentent fortement pour un défaut de recrutement de P-TEFb en l'absence de la RNP 7SK : nous observons effectivement un défaut de recrutement de la protéine du SEC AFF4 au niveau des promoteurs de gènes induits, ainsi que celui des facteurs d'élongation PAF1 et Spt6 (Fig. 6C, D, S6D, E), dont le recrutement dépend de l'activité de P-TEFb (Vos et al., 2018a). De plus, le niveau d'ARN Pol II et du facteur bivalent Spt5 (pause/élongation) dans les corps des gènes induits en réponse aux UV est négativement affecté par la perte de l'ARN 7SK (Fig. 6B, S6B, C). Ces données suggèrent donc un défaut de libération de la pause sur ces gènes de réponse, comme cela est observé de manière globale après irradiation en l'absence de l'ARN 7SK. Cependant, des données préliminaires suggèrent que les mécanismes permettant l'activation des gènes de réponse et la libération globale de la pause après l'irradiation sont différents. En effet, les facteurs d'élongation PAF1 et Spt6 apparaissent spécifiquement recrutés sur les gènes induits. Ces données seront plus amplement discutées lors de la présentation de notre modèle général.

Dans le but de confirmer le défaut de libération de la pause de l'ARN Pol II sur les gènes dont l'induction en réponse aux UV est défectueuse dans les cellules 7SK KO, nous avons analysé l'occupation des régions promotrices par NELF avant et après irradiation. NELF joue en effet un rôle majeur dans la mise en place et le maintien de la pause par l'ARN Pol II et sa dissociation est communément considérée dépendante de P-TEFb (Zhou et al., 2012). De façon inattendue, nous avons observé le relargage de NELF-E des régions promotrices en réponse aux UV dans les cellules HAP1 comme dans les cellules 7SK KO (6A, S6A). Nous nous attendions en effet à sa persistance au niveau de la pause en raison du défaut d'activité de P-TEFb dans les cellules 7SK KO. Plusieurs éléments permettent cependant d'expliquer ce

résultat. Premièrement, il a récemment été montré que NELF était aussi ciblé par une autre kinase, MK2, activée par p38 en réponse aux UV (Borisova et al., 2018). L'inhibition de p38 compromet la dissociation de NELF-E de la chromatine après irradiation, mais la dissociation des autres sous-unités de NELF semble moins impactée. Deuxièmement, P-TEFb phosphoryle plusieurs sous-unités de NELF, NELF-A et NELF-E (Lu et al., 2016b; Vos et al., 2018a). Tandis que NELF-E est nécessaire pour le recrutement de NELF aux complexes d'élongation, l'effet inhibiteur de NELF sur l'ARN Pol II semble principalement dépendre de NELF-A et NELF-C/D (Vos et al., 2018b). Il est ainsi possible que la phosphorylation de NELF-E par MK2 soit nécessaire et suffisante pour sa dissociation de la chromatine, mais que la dissociation efficace du complexe dépende de la phosphorylation de NELF-E et NELF-A par Cdk9. NELF-E n'est en effet pas nécessaire à la formation du complexe NELF, contrairement à NELF-A et NELF-C/D (Yamaguchi et al., 2002). La dissociation de NELF-E n'implique donc pas nécessairement la dissociation de l'intégralité du complexe. L'analyse de la dissociation d'autres sous-unités de NELF des complexes en pause après irradiation, notamment NELF-A et NELF-C/D, permettra de tester cette hypothèse pour le moins intéressante. Ces données permettraient en effet une révision du mode de dissociation de NELF, qui pourrait donc, au moins pour l'induction des gènes de réponse aux UV, résulter de la coopération entre les kinases MK2 et Cdk9.

b. Activation transcriptionnelle dépendante du réservoir inactif de P-TEFb

Seule une fraction des gènes de réponse aux UV est impactée par la perte du réservoir inactif de P-TEFb, indiquant que la RNP 7SK/P-TEFb est directement impliquée dans leur induction. La fraction active de P-TEFb semble être, elle, suffisante pour permettre l'activation transcriptionnelle des autres gènes de réponse.

La raison pour laquelle l'induction de certains gènes en réponse aux UV est plus sensible à la perte de la RNP 7SK/P-TEFb que celle d'autres gènes est encore en cours d'investigation. Le chargement de l'ARN 7SK au niveau des régions promotrices des gènes avant l'irradiation ou son recrutement en réponse aux UV est un modèle intéressant pour expliquer sa dépendance dans ce mécanisme d'induction. Il a effectivement été montré que la RNP 7SK/P-TEFb occupait les régions promotrices de nombreux gènes via la protéine KAP1, et se dissociait localement en réponse à la stimulation au TNF-α pour permettre l'activation de gènes de réponse (McNamara et al., 2016a; McNamara et al., 2013). Comme précédemment vu, nous n'avons cependant pas identifié la présence de l'ARN 7SK sur les régions promotrices des gènes ni avant ni 1h après l'irradiation. Modifiant notre angle d'approche, des résultats préliminaires indiquent que la déplétion de KAP1 par siARN dans les cellules HAP1 n'a pas d'incidence sur l'activation de gènes de réponse (données non montrées). Il semble donc que le recrutement de la RNP 7SK/P-TEFb sur les gènes de réponse ne soit pas le mécanisme prévalent pour

l'induction des gènes dépendant du réservoir inactif de P-TEFb dans les cellules HAP1. Nous privilégions donc un mécanisme d'extraction de P-TEFb de la RNP 7SK suivi de son ciblage spécifique sur certains gènes de réponse aux UV. Un rôle de la protéine Rbm7 dans ce processus a récemment été proposé (Bugai et al., 2019). Rbm7 semble en effet interagir transitoirement avec la RNP 7SK/P-TEFb après irradiation aux UV et être nécessaire à sa dissociation. L'action de Rbm7 permet alors l'interaction concomitante de P-TEFb avec l'ARN Pol II et la stimulation de l'expression de gènes impliqués dans la réponse aux dommages à l'ADN. Il sera donc intéressant d'étendre le rôle proposé pour Rbm7 dans la réponse aux UV à notre modèle d'étude. Ayant commencé à examiner si Rbm7 était effectivement impliqué dans l'activation après irradiation des gènes dépendants de la RNP 7SK/P-TEFb dans les cellules HAP1, nous n'avons cependant pas validé d'interactions entre P-TEFb et Rbm7, et une faible interaction entre l'ARN 7SK et Rbm7 a été identifiée par ChIRP-MS (données non montrées). Il est possible que le complexe formé soit de faible abondance et transitoire, nécessitant des conditions expérimentales particulières pour être identifié. Evaluer l'interaction entre la RNP 7SK/P-TEFb à différents temps après l'irradiation pourrait ainsi s'avérer plus fructueux. Il sera également important d'examiner la dissociation de la RNP 7SK/P-TEFb et l'induction des gènes en réponse aux UV après déplétion de Rbm7 dans les cellules HAP1. Nous essaierons également de détecter le recrutement de Rbm7 sur ces gènes en réponse aux UV.

c. Conclusion et modèle

La réponse aux UV est caractérisée par une intense reprogrammation transcriptionnelle (Gregersen and Svejstrup, 2018). Rapidement après exposition aux UV, les ARN Pol II en pause à proximité du promoteur sont globalement libérées pour faciliter la détection des lésions dans les corps de gènes. Alors que la transcription générale est ensuite progressivement diminuée par l'effet combiné d'une diminution de l'initiation et d'un taux d'élongation réduit, les gènes de réponse rapide bénéficient d'une forte activation transcriptionnelle. Ces étapes sont nécessaires pour la mise en place de la réponse cellulaire aux dommages à l'ADN et pour permettre la réparation des lésions. Une fois la majorité des dommages à l'ADN éliminée, l'activité transcriptionnelle est peu à peu rétablie. Mon travail de thèse a permis de mieux comprendre le rôle clé joué par P-TEFb dans les étapes précoces de la réponse transcriptionnelle aux UV (figure 32). Nous avons en effet montré que l'augmentation de l'activité de P-TEFb générée par la libération de son réservoir inactif était cruciale pour la libération globale des ARN Pol II en pause. De plus, tandis que la fraction de P-TEFb actif déjà présente avant l'irradiation permet l'activation transcriptionnelle d'une majorité de gènes de réponse, l'activation de certains d'entre eux repose sur la RNP 7SK/P-TEFb. Ces données impliquent le recrutement différentiel de P-TEFb sur les gènes en fonction de l'activité transcriptionnelle escomptée. Il est peu probable que

la libération globale de la pause nécessite un recrutement ciblé de P-TEFb, et pourrait simplement résulter de la diffusion de grandes quantités de P-TEFb actif. A l'inverse, l'activation des gènes de réponse dépend plus certainement d'un recrutement actif de P-TEFb, dont le mécanisme exact est encore à l'étude. Nous avons montré qu'une partie nécessite cependant la RNP 7SK/P-TEFb, et s'accompagne du recrutement de facteurs d'élongation importants. Nos résultats suggèrent également que P-TEFb est recruté sur ces gènes au sein du SEC, composé de plusieurs facteurs d'élongation. Curieusement, une interaction entre l'ARN 7SK et AFF1, un composant du SEC, a précédemment été rapportée (Lu et al., 2014). Nos données de ChIRP-MS suggèrent également une interaction entre l'ARN 7SK et AFF4, un autre composant du SEC, perdue après l'exposition aux UV. Un comportement similaire, suggérant une dissociation de l'ARN 7SK, est également observé pour les protéines du complexe PAF1. Bien que ces données préliminaires requièrent une validation rigoureuse, elles suggèrent qu'une fraction de RNP 7SK/P-TEFb pourrait être associée à des facteurs d'élongation et être sollicitée lors de la réponse aux UV pour activer spécifiquement l'expression de gènes de réponse. Ainsi, une population hétérogène de RNP 7SK/P-TEFb existerait dans le noyau, et permettrait de répondre à diverse sollicitations transcriptionnelles.

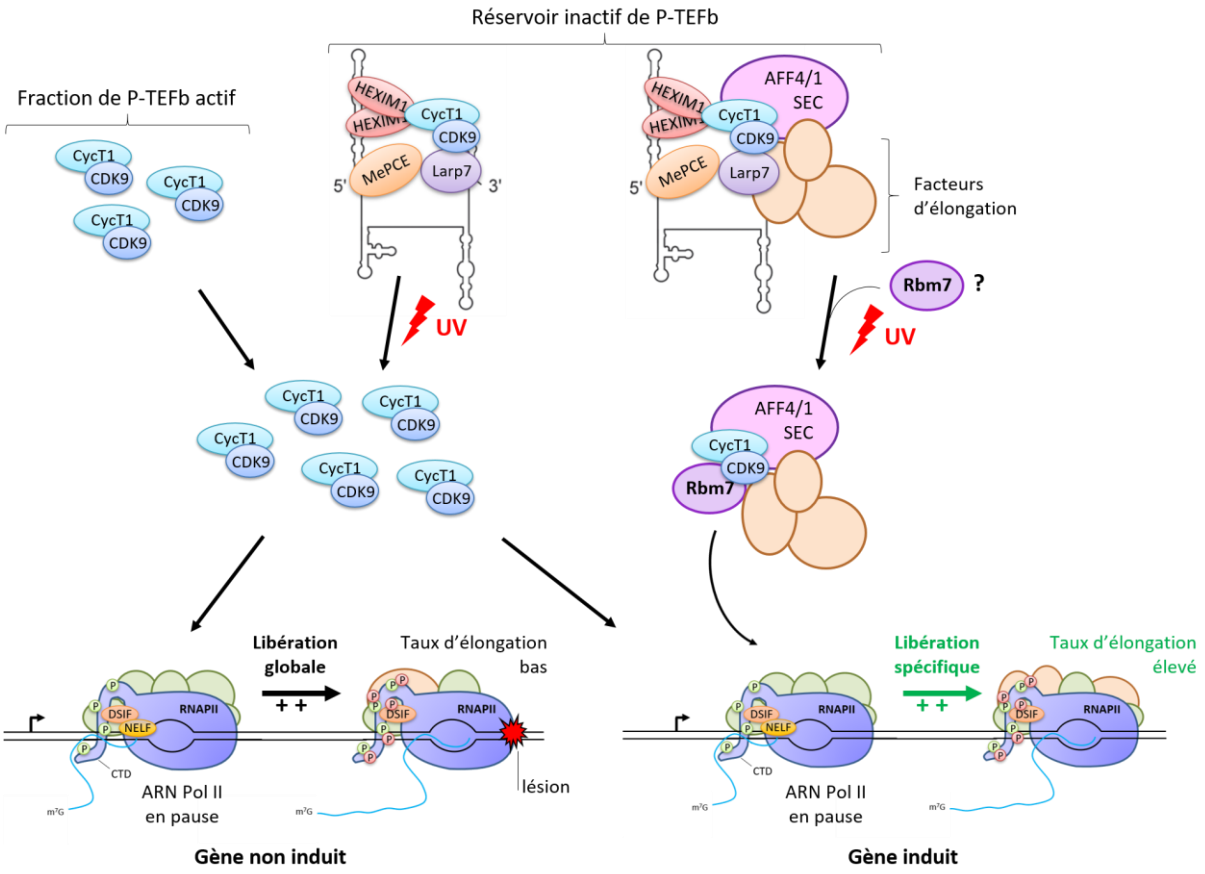


Figure 32 : Modèle d'action de la RNP 7SK/P-TEFb dans la réponse aux UV. Une fraction du réservoir inactif de P-TEFb est libérée pour augmenter l'activité de Cdk9 et permettre la libération globale des ARN Pol II en pause. La vague d'ARN Pol II générée favorise la détection et la réparation des lésions. Des RNP 7SK/P-TEFb spécialisées permettent de cibler des facteurs d'élongation sur des gènes de réponse pour promouvoir leur induction.

En plus d'étoffer notre compréhension des mécanismes de régulation de l'expression génique en réponse aux UV, mon travail de thèse ouvre sur de nouvelles perspectives concernant la régulation transcriptionnelle initiée en réponse à divers stimuli. Par exemple, le stress oxydatif génère une stabilisation rapide des ARN Pol II en pause à proximité du promoteur (Nilson et al., 2017). En réponse au choc thermique, l'expression de la majorité des gènes est réprimée par le blocage de l'entrée en élongation productive, tandis que la pause transcriptionnelle est levée pour les gènes de réponse (Mahat et al., 2016). La privation en sérum induit la dissociation de la RNP 7SK/P-TEFb et l'activation des gènes MEF2-dépendants (Nojima et al., 2008). La modulation de l'activité de P-TEFb par la RNP 7SK joue certainement un rôle majeur dans ces processus. Plus globalement, tout stress ayant un impact sur la pause de l'ARN Pol II fera potentiellement intervenir la RNP 7SK. Ainsi, les lignées 7SK KO constituent des outils considérables pour discerner les effets dépendants des fractions actives et inactives de P-TEFb.

Références bibliographiques

Abasi, M., Bazi, Z., Mohammadi-Yeganeh, S., Soleimani, M., Haghpanah, V., Zargami, N., and Ghanbarian, H. (2016). 7SK small nuclear RNA transcription level down-regulates in human tumors and stem cells. Med Oncol *33*, 128.

Abdullah, C., Wang, X., and Becker, D. (2011). Expression analysis and molecular targeting of cyclin-dependent kinases in advanced melanoma. Cell Cycle *10*, 977-988.

Adam, S., Polo, S.E., and Almouzni, G. (2013). Transcription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA. Cell *155*, 94-106.

Adar, S., Hu, J., Lieb, J.D., and Sancar, A. (2016). Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. Proc Natl Acad Sci U S A *113*, E2124-2133.

Adelman, K., Kennedy, M.A., Nechaev, S., Gilchrist, D.A., Muse, G.W., Chinenov, Y., and Rogatsky, I. (2009). Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. Proc Natl Acad Sci U S A *106*, 18207-18212.

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature Reviews Genetics *13*, 720-731.

Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. Mol Cell *17*, 103-112.

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Mol Cell *13*, 67-76.

Aiyar, S.E., Sun, J.L., Blair, A.L., Moskaluk, C.A., Lu, Y.Z., Ye, Q.N., Yamaguchi, Y., Mukherjee, A., Ren, D.M., Handa, H.*, et al.* (2004). Attenuation of estrogen receptor alpha-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor. Genes Dev *18*, 2134-2146.

AJ, C.Q., Bugai, A., and Barboric, M. (2016). Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. Nucleic Acids Res *44*, 7527-7539.

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Mol Cell *34*, 387-393.

Alazami, A.M., Al-Owain, M., Alzahrani, F., Shuaib, T., Al-Shamrani, H., Al-Falki, Y.H., Al-Qahtani, S.M., Alsheddi, T., Colak, D., and Alkuraya, F.S. (2012). Loss of function mutation in LARP7, chaperone of 7SK ncRNA, causes a syndrome of facial dysmorphism, intellectual disability, and primordial dwarfism. Hum Mutat *33*, 1429-1434.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol *16*, 155-166.

Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360-363.

Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B.*, et al.* (2012). Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. Cell Rep *1*, 543-556.

Ammosova, T., Obukhov, Y., Kotelkin, A., Breuer, D., Beullens, M., Gordeuk, V.R., Bollen, M., and Nekhai, S. (2011). Protein phosphatase-1 activates CDK9 by dephosphorylating Ser175. PLoS One *6*, e18985.

Ammosova, T., Washington, K., Rotimi, J., Kumari, N., Smith, K.A., Niu, X., Jerebtsova, M., and Nekhai, S. (2016). Protein Phosphatase-1 Regulates Expression of Neuregulin-1. Biology (Basel) *5*.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T.*, et al.* (2014a). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014b). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. Nat Commun *5*, 5336.

Andrade-Lima, L.C., Veloso, A., Paulsen, M.T., Menck, C.F., and Ljungman, M. (2015). DNA repair and recovery of RNA synthesis following exposure to ultraviolet light are delayed in long genes. Nucleic Acids Res *43*, 2744-2756.

Andrulis, E.D., Guzman, E., Doring, P., Werner, J., and Lis, J.T. (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., and Svejstrup, J.Q. (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.

Ansari, S.A., Paul, E., Sommer, S., Lieleg, C., He, Q., Daly, A.Z., Rode, K.A., Barber, W.T., Ellis, L.C., LaPorta, E.*, et al.* (2014). Mediator, TATA-binding protein, and RNA polymerase II contribute to low histone occupancy at active gene promoters in yeast. J Biol Chem *289*, 14981-14995.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J *28*, 1067-1077.

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

Arter, J., and Wegner, M. (2015). Transcription factors Sox10 and Sox2 functionally interact with positive transcription elongation factor b in Schwann cells. J Neurochem *132*, 384-393.

Auerbach, R.K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrancois, P., Struhl, K., Gerstein, M., and Snyder, M. (2009). Mapping accessible chromatin regions using Sono-Seq. Proc Natl Acad Sci U S A *106*, 14926-14931.

Aygun, O., Svejstrup, J., and Liu, Y. (2008). A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. Proc Natl Acad Sci U S A *105*, 8580-8584.

Aygun, O., and Svejstrup, J.Q. (2010). RECQL5 helicase: connections to DNA recombination and RNA polymerase II transcription. DNA Repair (Amst) *9*, 345-353.

Aygun, O., Xu, X., Liu, Y., Takahashi, H., Kong, S.E., Conaway, R.C., Conaway, J.W., and Svejstrup, J.Q. (2009). Direct inhibition of RNA polymerase II transcription by RECQL5. J Biol Chem *284*, 23197-23203.

Baek, S., Han, S., Kang, D., Kemp, M.G., and Choi, J.H. (2018). Simultaneous detection of nucleotide excision repair events and apoptosis-induced DNA fragmentation in genotoxin-treated cells. Sci Rep *8*, 2265.

Bahrami, S., and Drablos, F. (2016). Gene regulation in the immediate-early response process. Adv Biol Regul *62*, 37-49.

Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and 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.

Barboric, M., Kohoutek, J., Price, J.P., Blazek, D., Price, D.H., and Peterlin, B.M. (2005). Interplay between 7SK snRNA and oppositely charged regions in HEXIM1 direct the inhibition of P-TEFb. EMBO J *24*, 4291-4303.

Barboric, M., Lenasi, T., Chen, H., Johansen, E.B., Guo, S., and Peterlin, B.M. (2009). 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. Proc Natl Acad Sci U S A *106*, 7798-7803.

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

Barboric, M., Yik, J.H., Czudnochowski, N., Yang, Z., Chen, R., Contreras, X., Geyer, M., Matija Peterlin, B., and Zhou, Q. (2007). Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. Nucleic Acids Res *35*, 2003-2012.

Barrandon, C., Bonnet, F., Nguyen, V.T., Labas, V., and Bensaude, O. (2007). The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. Mol Cell Biol *27*, 6996-7006.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bartholomeeusen, K., Fujinaga, K., Xiang, Y., and Peterlin, B.M. (2013). Histone deacetylase inhibitors (HDACis) that release the positive transcription elongation factor b (P-TEFb) from its inhibitory complex also activate HIV transcription. J Biol Chem *288*, 14400-14407.

Bartholomeeusen, K., Xiang, Y., Fujinaga, K., and Peterlin, B.M. (2012). Bromodomain and extra-terminal (BET) bromodomain inhibition activate transcription via transient release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein. J Biol Chem *287*, 36609-36616.

Bartkowiak, B., Liu, P., Phatnani, H.P., Fuda, N.J., Cooper, J.J., Price, D.H., Adelman, K., Lis, J.T., and Greenleaf, A.L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev *24*, 2303-2316.

Baskaran, R., Chiang, G.G., Mysliwiec, T., Kruh, G.D., and Wang, J.Y. (1997). Tyrosine phosphorylation of RNA polymerase II carboxyl-terminal domain by the Abl-related gene product. J Biol Chem *272*, 18905-18909.

Baskaran, R., Dahmus, M.E., and Wang, J.Y. (1993). Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. Proc Natl Acad Sci U S A *90*, 11167-11171.

Batta, K., Zhang, Z., Yen, K., Goffman, D.B., and Pugh, B.F. (2011). Genome-wide function of H2B ubiquitylation in promoter and genic regions. Genes Dev *25*, 2254-2265.

Baumli, S., Lolli, G., Lowe, E.D., Troiani, S., Rusconi, L., Bullock, A.N., Debreczeni, J.E., Knapp, S., and Johnson, L.N. (2008). The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. EMBO J *27*, 1907-1918.

Bayfield, M.A., Yang, R., and Maraia, R.J. (2010). Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). Biochim Biophys Acta *1799*, 365-378.

Bazi, Z., Bertacchi, M., Abasi, M., Mohammadi-Yeganeh, S., Soleimani, M., Wagner, N., and Ghanbarian, H. (2018). Rn7SK small nuclear RNA is involved in neuronal differentiation. J Cell Biochem *119*, 3174-3182.

Bear, D.G., Fomproix, N., Soop, T., Bjorkroth, B., Masich, S., and Daneholt, B. (2003). Nuclear poly(A)-binding protein PABPN1 is associated with RNA polymerase II during transcription and accompanies the released transcript to the nuclear pore. Exp Cell Res *286*, 332-344.

Beck, C.F., and Warren, R.A. (1988). Divergent promoters, a common form of gene organization. Microbiol Rev *52*, 318-326.

Behera, V., Stonestrom, A.J., Hamagami, N., Hsiung, C.C., Keller, C.A., Giardine, B., Sidoli, S., Yuan, Z.F., Bhanu, N.V., Werner, M.T.*, et al.* (2019). Interrogating Histone Acetylation and BRD4 as Mitotic Bookmarks of Transcription. Cell Rep *27*, 400-415 e405.

Belanger, F., Baigude, H., and Rana, T.M. (2009). U30 of 7SK RNA forms a specific photo-cross-link with Hexim1 in the context of both a minimal RNA-binding site and a fully reconstituted 7SK/Hexim1/P-TEFb ribonucleoprotein complex. J Mol Biol *386*, 1094-1107.

Bellan, C., De Falco, G., Lazzi, S., Micheli, P., Vicidomini, S., Schurfeld, K., Amato, T., Palumbo, A., Bagella, L., Sabattini, E.*, et al.* (2004). CDK9/CYCLIN T1 expression during normal lymphoid differentiation and malignant transformation. J Pathol *203*, 946-952.

Bentley, D. (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. Curr Opin Cell Biol *11*, 347-351.

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol *17*, 251-256.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat Rev Genet *15*, 163-175.

Bentley, D.L., and Groudine, M. (1986). A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature *321*, 702-706.

Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature *529*, 551-554.

Bernecky, C., Plitzko, J.M., and Cramer, P. (2017). Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. Nat Struct Mol Biol *24*, 809-815.

Bible, K.C., Peethambaram, P.P., Oberg, A.L., Maples, W., Groteluschen, D.L., Boente, M., Burton, J.K., Gomez Dahl, L.C., Tibodeau, J.D., Isham, C.R.*, et al.* (2012). A phase 2 trial of flavopiridol (Alvocidib)

and cisplatin in platin-resistant ovarian and primary peritoneal carcinoma: MC0261. Gynecol Oncol *127*, 55-62.

Biewenga, P., Buist, M.R., Moerland, P.D., Ver Loren van Themaat, E., van Kampen, A.H., ten Kate, F.J., and Baas, F. (2008). Gene expression in early stage cervical cancer. Gynecol Oncol *108*, 520-526.

Biglione, S., Byers, S.A., Price, J.P., Nguyen, V.T., Bensaude, O., Price, D.H., and Maury, W. (2007). Inhibition of HIV-1 replication by P-TEFb inhibitors DRB, seliciclib and flavopiridol correlates with release of free P-TEFb from the large, inactive form of the complex. Retrovirology *4*, 47.

Bisgrove, D.A., Mahmoudi, T., Henklein, P., and Verdin, E. (2007). Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. Proc Natl Acad Sci U S A *104*, 13690-13695.

Blank, M.F., Chen, S., Poetz, F., Schnolzer, M., Voit, R., and Grummt, I. (2017). SIRT7-dependent deacetylation of CDK9 activates RNA polymerase II transcription. Nucleic Acids Res *45*, 2675-2686.

Blasius, M., Wagner, S.A., Choudhary, C., Bartek, J., and Jackson, S.P. (2014). A quantitative 14-3-3 interaction screen connects the nuclear exosome targeting complex to the DNA damage response. Genes Dev *28*, 1977-1982.

Blazek, D., Barboric, M., Kohoutek, J., Oven, I., and Peterlin, B.M. (2005). Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. Nucleic Acids Res *33*, 7000-7010.

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., Cimermancic, P., Ule, J., and Peterlin, B.M. (2011). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes Dev *25*, 2158-2172.

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet *13*, 770-780.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P.*, et al.* (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat Struct Mol Biol *25*, 833-840.

Bohr, V.A., Smith, C.A., Okumoto, D.S., and Hanawalt, P.C. (1985). DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. Cell *40*, 359-369.

Borisova, M.E., Voigt, A., Tollenaere, M.A.X., Sahu, S.K., Juretschke, T., Kreim, N., Mailand, N., Choudhary, C., Bekker-Jensen, S., Akutsu, M.*, et al.* (2018). p38-MK2 signaling axis regulates RNA metabolism after UV-light-induced DNA damage. Nat Commun *9*, 1017.

Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J.*, et al.* (2012). mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. Mol Cell *46*, 311-324.

Bresson, S.M., and Conrad, N.K. (2013). The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. PLoS Genet *9*, e1003893.

Brogie, J.E., and Price, D.H. (2017). Reconstitution of a functional 7SK snRNP. Nucleic Acids Res *45*, 6864-6880.

Bruce, A.W., Donaldson, I.J., Wood, I.C., Yerbury, S.A., Sadowski, M.I., Chapman, M., Gottgens, B., and Buckley, N.J. (2004). Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. Proc Natl Acad Sci U S A *101*, 10458-10463.

Bruce, A.W., Lopez-Contreras, A.J., Flicek, P., Down, T.A., Dhami, P., Dillon, S.C., Koch, C.M., Langford, C.F., Dunham, I., Andrews, R.M.*, et al.* (2009). Functional diversity for REST (NRSF) is defined by in vivo binding affinity hierarchies at the DNA sequence level. Genome Res *19*, 994-1005.

Brueckner, F., and Cramer, P. (2007). DNA photodamage recognition by RNA polymerase II. FEBS Lett *581*, 2757-2760.

Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD damage recognition by transcribing RNA polymerase II. Science *315*, 859-862.

Budhiraja, S., Ramakrishnan, R., and Rice, A.P. (2012). Phosphatase PPM1A negatively regulates P-TEFb function in resting CD4(+) T cells and inhibits HIV-1 gene expression. Retrovirology *9*, 52.

Bugai, A., Quaresma, A.J.C., Friedel, C.C., Lenasi, T., Duster, R., Sibley, C.R., Fujinaga, K., Kukanja, P., Hennig, T., Blasius, M.*, et al.* (2019). P-TEFb Activation by RBM7 Shapes a Pro-survival Transcriptional Response to Genotoxic Stress. Mol Cell *74*, 254-267 e210.

Bulavin, D.V., Higashimoto, Y., Popoff, I.J., Gaarde, W.A., Basrur, V., Potapova, O., Appella, E., and Fornace, A.J., Jr. (2001). Initiation of a G2/M checkpoint after ultraviolet radiation requires p38 kinase. Nature *411*, 102-107.

Bunch, H., Zheng, X., Burkholder, A., Dillon, S.T., Motola, S., Birrane, G., Ebmeier, C.C., Levine, S., Fargo, D., Hu, G.*, et al.* (2014). TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. Nat Struct Mol Biol *21*, 876-883.

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

Burger, K., Schlackow, M., and Gullerova, M. (2019). Tyrosine kinase c-Abl couples RNA polymerase II transcription to DNA double-strand breaks. Nucleic Acids Res *47*, 3467-3484.

Byers, S.A., Price, J.P., Cooper, J.J., Li, Q., and Price, D.H. (2005). HEXIM2, a HEXIM1-related protein, regulates positive transcription elongation factor b through association with 7SK. J Biol Chem *280*, 16360-16367.

Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature *518*, 249-253.

Camenisch, U., Dip, R., Schumacher, S.B., Schuler, B., and Naegeli, H. (2006). Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair. Nat Struct Mol Biol *13*, 278-284.

Carette, J.E., Raaben, M., Wong, A.C., Herbert, A.S., Obernosterer, G., Mulherkar, N., Kuehne, A.I., Kranzusch, P.J., Griffin, A.M., Ruthel, G.*, et al.* (2011). Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. Nature *477*, 340-343.

Caron, P., Pankotai, T., Wiegant, W.W., Tollenaere, M.A.X., Furst, A., Bonhomme, C., Helfricht, A., de Groot, A., Pastink, A., Vertegaal, A.C.O.*, et al.* (2019). WWP2 ubiquitylates RNA polymerase II for DNA-PK-dependent transcription arrest and repair at DNA breaks. Genes Dev *33*, 684-704.

Castelo-Branco, G., Amaral, P.P., Engstrom, P.G., Robson, S.C., Marques, S.C., Bertone, P., and Kouzarides, T. (2013). The non-coding snRNA 7SK controls transcriptional termination, poising, and bidirectionality in embryonic stem cells. Genome Biol *14*, R98.

Chambers, R.S., and Dahmus, M.E. (1994). Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II. J Biol Chem *269*, 26243-26248.

Chang, H., Lim, J., Ha, M., and Kim, V.N. (2014). TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. Mol Cell *53*, 1044-1052.

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

Charlet-Berguerand, N., Feuerhahn, S., Kong, S.E., Ziserman, H., Conaway, J.W., Conaway, R., and Egly, J.M. (2006). RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. EMBO J *25*, 5481-5491.

Chen, C.Y., Chang, C.C., Yen, C.F., Chiu, M.T., and Chang, W.H. (2009). Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. Proc Natl Acad Sci U S A *106*, 127-132.

Chen, F., Gao, X., and Shilatifard, A. (2015a). Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. Genes Dev *29*, 39-47.

Chen, F.X., Smith, E.R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *19*, 464-478.

Chen, F.X., Woodfin, A.R., Gardini, A., Rickels, R.A., Marshall, S.A., Smith, E.R., Shiekhattar, R., and Shilatifard, A. (2015b). PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. Cell *162*, 1003-1015.

Chen, F.X., Xie, P., Collings, C.K., Cao, K., Aoi, Y., Marshall, S.A., Rendleman, E.J., Ugarenko, M., Ozark, P.A., Zhang, A.*, et al.* (2017). PAF1 regulation of promoter-proximal pause release via enhancer activation. Science *357*, 1294-1298.

Chen, R., Liu, M., Li, H., Xue, Y., Ramey, W.N., He, N., Ai, N., Luo, H., Zhu, Y., Zhou, N.*, et al.* (2008). PP2B and PP1alpha cooperatively disrupt 7SK snRNP to release P-TEFb for transcription in response to Ca2+ signaling. Genes Dev *22*, 1356-1368.

Chen, R., Yang, Z., and Zhou, Q. (2004). Phosphorylated positive transcription elongation factor b (P-TEFb) is tagged for inhibition through association with 7SK snRNA. J Biol Chem *279*, 4153-4160.

Cheng, B., Li, T., Rahl, P.B., Adamson, T.E., Loudas, N.B., Guo, J., Varzavand, K., Cooper, J.J., Hu, X., Gnatt, A.*, et al.* (2012a). Functional association of Gdown1 with RNA polymerase II poised on human genes. Mol Cell *45*, 38-50.

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

Cheng, S.W., Kuzyk, M.A., Moradian, A., Ichu, T.A., Chang, V.C., Tien, J.F., Vollett, S.E., Griffith, M., Marra, M.A., and Morin, G.B. (2012b). Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II. Mol Cell Biol *32*, 4691-4704.

Cheng, Y., Jin, Z., Agarwal, R., Ma, K., Yang, J., Ibrahim, S., Olaru, A.V., David, S., Ashktorab, H., Smoot, D.T.*, et al.* (2012c). LARP7 is a potential tumor suppressor gene in gastric cancer. Lab Invest *92*, 1013-1019.

Chinenov, Y., Gupte, R., Dobrovolna, J., Flammer, J.R., Liu, B., Michelassi, F.E., and Rogatsky, I. (2012). Role of transcriptional coregulator GRIP1 in the anti-inflammatory actions of glucocorticoids. Proc Natl Acad Sci U S A *109*, 11776-11781.

Chiou, Y.Y., Hu, J., Sancar, A., and Selby, C.P. (2018). RNA polymerase II is released from the DNA template during transcription-coupled repair in mammalian cells. J Biol Chem *293*, 2476-2486.

Chiu, Y.L., Cao, H., Jacque, J.M., Stevenson, M., and Rana, T.M. (2004). Inhibition of human immunodeficiency virus type 1 replication by RNA interference directed against human transcription elongation factor P-TEFb (CDK9/CyclinT1). J Virol *78*, 2517-2529.

Cho, S., Schroeder, S., Kaehlcke, K., Kwon, H.S., Pedal, A., Herker, E., Schnoelzer, M., and Ott, M. (2009). Acetylation of cyclin T1 regulates the equilibrium between active and inactive P-TEFb in cells. Embo J *28*, 1407-1417.

Cho, W.K., Jang, M.K., Huang, K., Pise-Masison, C.A., and Brady, J.N. (2010). Human T-lymphotropic virus type 1 Tax protein complexes with P-TEFb and competes for Brd4 and 7SK snRNP/HEXIM1 binding. J Virol *84*, 12801-12809.

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

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Mol Cell *40*, 179-204.

Coin, F., Frit, P., Viollet, B., Salles, B., and Egly, J.M. (1998). TATA binding protein discriminates between different lesions on DNA, resulting in a transcription decrease. Mol Cell Biol *18*, 3907-3914.

Coin, F., Oksenych, V., and Egly, J.M. (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. Mol Cell *26*, 245-256.

Coin, F., Oksenych, V., Mocquet, V., Groh, S., Blattner, C., and Egly, J.M. (2008). Nucleotide excision repair driven by the dissociation of CAK from TFIIH. Mol Cell *31*, 9-20.

Collart, M.A., Tourkine, N., Belin, D., Vassalli, P., Jeanteur, P., and Blanchard, J.M. (1991). c-fos gene transcription in murine macrophages is modulated by a calcium-dependent block to elongation in intron 1. Mol Cell Biol *11*, 2826-2831.

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

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

Compe, E., Genes, C.M., Braun, C., Coin, F., and Egly, J.M. (2019). TFIIE orchestrates the recruitment of the TFIIH kinase module at promoter before release during transcription. Nat Commun *10*, 2084.

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

Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T.*, et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Contreras, X., Barboric, M., Lenasi, T., and Peterlin, B.M. (2007). HMBA releases P-TEFb from HEXIM1 and 7SK snRNA via PI3K/Akt and activates HIV transcription. PLoS Pathog *3*, 1459-1469.

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the status of RNA polymerase at promoters. Cell Rep *2*, 1025-1035.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J.*, et al.* (2008). Structure of eukaryotic RNA polymerases. Annu Rev Biophys *37*, 337-352.

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

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.

Cuadrado, A., and Nebreda, A.R. (2010). Mechanisms and functions of p38 MAPK signalling. Biochem J *429*, 403-417.

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

Czudnochowski, N., Vollmuth, F., Baumann, S., Vogel-Bachmayr, K., and Geyer, M. (2010). Specificity of Hexim1 and Hexim2 complex formation with cyclin T1/T2, importin alpha and 7SK snRNA. J Mol Biol *395*, 28-41.

D'Orso, I., and Frankel, A.D. (2010). RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat Struct Mol Biol *17*, 815-821.

Dai, Q., Luan, G., Deng, L., Lei, T., Kang, H., Song, X., Zhang, Y., Xiao, Z.X., and Li, Q. (2014). Primordial dwarfism gene maintains Lin28 expression to safeguard embryonic stem cells from premature differentiation. Cell Rep *7*, 735-746.

Dames, S.A., Schonichen, A., Schulte, A., Barboric, M., Peterlin, B.M., Grzesiek, S., and Geyer, M. (2007). Structure of the Cyclin T binding domain of Hexim1 and molecular basis for its recognition of P-TEFb. Proc Natl Acad Sci U S A *104*, 14312-14317.

Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L., Lis, J.T., Siepel, A., and Kraus, W.L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol Cell *50*, 212-222.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. Nat Struct Mol Biol *14*, 796-806.

David, C.J., Boyne, A.R., Millhouse, S.R., and Manley, J.L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. Genes Dev *25*, 972-983.

Davidson, L., Muniz, L., and West, S. (2014). 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. Genes Dev *28*, 342-356.

Day, D.S., Zhang, B., Stevens, S.M., Ferrari, F., Larschan, E.N., Park, P.J., and Pu, W.T. (2016). Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types. Genome Biol *17*, 120.

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

de Laat, W.L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N.G., and Hoeijmakers, J.H. (1998). DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. Genes Dev *12*, 2598-2609.

De Luca, A., De Falco, M., Baldi, A., and Paggi, M.G. (2003). Cyclin T: three forms for different roles in physiological and pathological functions. J Cell Physiol *194*, 101-107.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

DeLaney, E., and Luse, D.S. (2016). Gdown1 Associates Efficiently with RNA Polymerase II after Promoter Clearance and Displaces TFIIF during Transcript Elongation. PLoS One *11*, e0163649.

Descostes, N., Heidemann, M., Spinelli, L., Schuller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I.*, et al.* (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. Elife *3*, e02105.

Devaiah, B.N., Lewis, B.A., Cherman, N., Hewitt, M.C., Albrecht, B.K., Robey, P.G., Ozato, K., Sims, R.J., 3rd, and Singer, D.S. (2012). BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. Proc Natl Acad Sci U S A *109*, 6927-6932.

Dey, A., Chitsaz, F., Abbasi, A., Misteli, T., and Ozato, K. (2003). The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. Proc Natl Acad Sci U S A *100*, 8758-8763.

Dey, A., Nishiyama, A., Karpova, T., McNally, J., and Ozato, K. (2009). Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. Mol Biol Cell *20*, 4899-4909.

Di Cerbo, V., Mohn, F., Ryan, D.P., Montellier, E., Kacem, S., Tropberger, P., Kallis, E., Holzner, M., Hoerner, L., Feldmann, A.*, et al.* (2014). Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. Elife *3*, e01632.

Dinant, C., Ampatziadis-Michailidis, G., Lans, H., Tresini, M., Lagarou, A., Grosbart, M., Theil, A.F., van Cappellen, W.A., Kimura, H., Bartek, J.*, et al.* (2013). Enhanced chromatin dynamics by FACT promotes transcriptional restart after UV-induced DNA damage. Mol Cell *51*, 469-479.

Dip, R., Camenisch, U., and Naegeli, H. (2004). Mechanisms of DNA damage recognition and strand discrimination in human nucleotide excision repair. DNA Repair (Amst) *3*, 1409-1423.

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

Dutta, A., Babbarwal, V., Fu, J., Brunke-Reese, D., Libert, D.M., Willis, J., and Reese, J.C. (2015). Ccr4-Not and TFIIS Function Cooperatively To Rescue Arrested RNA Polymerase II. Mol Cell Biol *35*, 1915-1925.

Duttke, S.H.C., Lacadie, S.A., Ibrahim, M.M., Glass, C.K., Corcoran, D.L., Benner, C., Heinz, S., Kadonaga, J.T., and Ohler, U. (2015). Human promoters are intrinsically directional. Mol Cell *57*, 674-684.

Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev *32*, 127-139.

Eaton, J.D., and West, S. (2018). An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. Transcription *9*, 321-326.

Ebmeier, C.C., Erickson, B., Allen, B.L., Allen, M.A., Kim, H., Fong, N., Jacobsen, J.R., Liang, K., Shilatifard, A., Dowell, R.D.*, et al.* (2017). Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. Cell Rep *20*, 1173-1186.

Eddy, J., Vallur, A.C., Varma, S., Liu, H., Reinhold, W.C., Pommier, Y., and Maizels, N. (2011). G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. Nucleic Acids Res *39*, 4975-4983.

Egloff, S., Al-Rawaf, H., O'Reilly, D., and Murphy, S. (2009). Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. Mol Cell Biol *29*, 4002-4013.

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

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

Egloff, S., Studniarek, C., and Kiss, T. (2017a). 7SK small nuclear RNA, a multifunctional transcriptional regulatory RNA with gene-specific features. Transcription, 1-7.

Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., and 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.

Egloff, S., Van Herreweghe, E., and Kiss, T. (2006). Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. Mol Cell Biol *26*, 630-642.

Egloff, S., Vitali, P., Tellier, M., Raffel, R., Murphy, S., and Kiss, T. (2017b). The 7SK snRNP associates with the little elongation complex to promote snRNA gene expression. EMBO J *36*, 934-948.

Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., and Murphy, S. (2012). Ser7 Phosphorylation of the CTD Recruits the RPAP2 Ser5 Phosphatase to snRNA Genes. Mol Cell.

Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., and Sekine, S.I. (2017). Structure of the complete elongation complex of RNA polymerase II with basal factors. Science *357*, 921-924.

Ehrensberger, A.H., Kelly, G.P., and Svejstrup, J.Q. (2013). Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. Cell *154*, 713-715.

Eichhorn, C.D., Yang, Y., Repeta, L., and Feigon, J. (2018). Structural basis for recognition of human 7SK long noncoding RNA by the La-related protein Larp7. Proc Natl Acad Sci U S A *115*, E6457-E6466.

Elagib, K.E., Rubinstein, J.D., Delehanty, L.L., Ngoh, V.S., Greer, P.A., Li, S., Lee, J.K., Li, Z., Orkin, S.H., Mihaylov, I.S*., et al.* (2013). Calpain 2 activation of P-TEFb drives megakaryocyte morphogenesis and is disrupted by leukemogenic GATA1 mutation. Dev Cell *27*, 607-620.

Epanchintsev, A., Costanzo, F., Rauschendorf, M.A., Caputo, M., Ye, T., Donnio, L.M., Proietti-de-Santis, L., Coin, F., Laugel, V., and Egly, J.M. (2017). Cockayne's Syndrome A and B Proteins Regulate Transcription Arrest after Genotoxic Stress by Promoting ATF3 Degradation. Mol Cell *68*, 1054-1066 e1056.

Erickson, B., Sheridan, R.M., Cortazar, M., and Bentley, D.L. (2018). Dynamic turnover of paused Pol II complexes at human promoters. Genes Dev *32*, 1215-1225.

Escoubet-Lozach, L., Benner, C., Kaikkonen, M.U., Lozach, J., Heinz, S., Spann, N.J., Crotti, A., Stender, J., Ghisletti, S., Reichart, D*., et al.* (2011). Mechanisms establishing TLR4-responsive activation states of inflammatory response genes. PLoS Genet *7*, e1002401.

Essletzbichler, P., Konopka, T., Santoro, F., Chen, D., Gapp, B.V., Kralovics, R., Brummelkamp, T.R., Nijman, S.M., and Burckstummer, T. (2014). Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. Genome Res *24*, 2059-2065.

Evans, E., Fellows, J., Coffer, A., and Wood, R.D. (1997a). Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. EMBO J *16*, 625-638.

Evans, E., Moggs, J.G., Hwang, J.R., Egly, J.M., and Wood, R.D. (1997b). Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. EMBO J *16*, 6559-6573.

Eychenne, T., Novikova, E., Barrault, M.B., Alibert, O., Boschiero, C., Peixeiro, N., Cornu, D., Redeker, V., Kuras, L., Nicolas, P*., et al.* (2016). Functional interplay between Mediator and TFIIB in preinitiation complex assembly in relation to promoter architecture. Genes Dev *30*, 2119-2132.

Eychenne, T., Werner, M., and Soutourina, J. (2017). Toward understanding of the mechanisms of Mediator function in vivo: Focus on the preinitiation complex assembly. Transcription *8*, 328-342.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol Cell *11*, 1549-1561.

Fishburn, J., Tomko, E., Galburt, E., and Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. Proc Natl Acad Sci U S A *112*, 3961-3966.

Fitz, J., Neumann, T., and Pavri, R. (2018). Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. EMBO J *37*.

Flores, O., Lee, G., Kessler, J., Miller, M., Schlief, W., Tomassini, J., and Hazuda, D. (1999). Host-cell positive transcription elongation factor b kinase activity is essential and limiting for HIV type 1 replication. Proc Natl Acad Sci U S A *96*, 7208-7213.

Flynn, R.A., Almada, A.E., Zamudio, J.R., and Sharp, P.A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proc Natl Acad Sci U S A *108*, 10460-10465.

Flynn, R.A., Do, B.T., Rubin, A.J., Calo, E., Lee, B., Kuchelmeister, H., Rale, M., Chu, C., Kool, E.T., Wysocka, J.*, et al.* (2016). 7SK-BAF axis controls pervasive transcription at enhancers. Nat Struct Mol Biol *23*, 231-238.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. Mol Cell *60*, 256-267.

Fousteri, M., Vermeulen, W., van Zeeland, A.A., and Mullenders, L.H. (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.

Franco, L.C., Morales, F., Boffo, S., and Giordano, A. (2018). CDK9: A key player in cancer and other diseases. J Cell Biochem *119*, 1273-1284.

Fu, J., Yoon, H.G., Qin, J., and Wong, J. (2007). Regulation of P-TEFb elongation complex activity by CDK9 acetylation. Mol Cell Biol *27*, 4641-4651.

Fuchs, G., Hollander, D., Voichek, Y., Ast, G., and Oren, M. (2014). Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. Genome Res *24*, 1572-1583.

Fuda, N.J., Guertin, M.J., Sharma, S., Danko, C.G., Martins, A.L., Siepel, A., and Lis, J.T. (2015). GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. PLoS Genet *11*, e1005108.

Fujinaga, K., Barboric, M., Li, Q., Luo, Z., Price, D.H., and Peterlin, B.M. (2012). PKC phosphorylates HEXIM1 and regulates P-TEFb activity. Nucleic Acids Res *40*, 9160-9170.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Mol Cell Biol *24*, 787-795.

Gaillard, H., Garcia-Muse, T., and Aguilera, A. (2015). Replication stress and cancer. Nat Rev Cancer *15*, 276-289.

Galbraith, M.D., Allen, M.A., Bensard, C.L., Wang, X., Schwinn, M.K., Qin, B., Long, H.W., Daniels, D.L., Hahn, W.C., Dowell, R.D*., et al.* (2013). HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell *153*, 1327-1339.

Ganguli, D., Chereji, R.V., Iben, J.R., Cole, H.A., and Clark, D.J. (2014). RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. Genome Res *24*, 1637-1649.

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. Mol Cell *56*, 128-139.

Gargano, B., Amente, S., Majello, B., and Lania, L. (2007). P-TEFb is a crucial co-factor for Myc transactivation. Cell Cycle *6*, 2031-2037.

Garriga, J., Peng, J., Parreno, M., Price, D.H., Henderson, E.E., and Grana, X. (1998). Upregulation of cyclin T1/CDK9 complexes during T cell activation. Oncogene *17*, 3093-3102.

Geijer, M.E., and Marteijn, J.A. (2018). What happens at the lesion does not stay at the lesion: Transcription-coupled nucleotide excision repair and the effects of DNA damage on transcription in cis and trans. DNA Repair (Amst) *71*, 56-68.

Gentile, M., Latonen, L., and Laiho, M. (2003). Cell cycle arrest and apoptosis provoked by UV radiation-induced DNA damage are transcriptionally highly divergent responses. Nucleic Acids Res *31*, 4779-4790.

Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M., and Ponticelli, A.S. (2004). Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. Mol Cell Biol *24*, 10975-10985.

Ghosh, A., Shuman, S., and Lima, C.D. (2011). Structural insights to how mammalian capping enzyme reads the CTD code. Mol Cell *43*, 299-310.

Giacinti, C., Bagella, L., Puri, P.L., Giordano, A., and Simone, C. (2006). MyoD recruits the cdk9/cyclin T2 complex on myogenic-genes regulatory regions. J Cell Physiol *206*, 807-813.

Giardina, C., and Lis, J.T. (1993). DNA melting on yeast RNA polymerase II promoters. Science *261*, 759-762.

Gibbs, E.B., Lu, F., Portz, B., Fisher, M.J., Medellin, B.P., Laremore, T.N., Zhang, Y.J., Gilmour, D.S., and Showalter, S.A. (2017). Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. Nat Commun *8*, 15233.

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

Gilchrist, D.A., Fromm, G., dos Santos, G., Pham, L.N., McDaniel, I.E., Burkholder, A., Fargo, D.C., and Adelman, K. (2012). Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. Genes Dev *26*, 933-944.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev *22*, 1921-1933.

Gilmour, D.S., and Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol Cell Biol *6*, 3984-3989.

Giono, L.E., Nieto Moreno, N., Cambindo Botto, A.E., Dujardin, G., Munoz, M.J., and Kornblihtt, A.R. (2016). The RNA Response to DNA Damage. J Mol Biol *428*, 2636-2651.

Giraud, S., Hurlstone, A., Avril, S., and Coqueret, O. (2004). Implication of BRG1 and cdk9 in the STAT3-mediated activation of the p21waf1 gene. Oncogene *23*, 7391-7398.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (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.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science *292*, 1876-1882.

Gregersen, L.H., and Svejstrup, J.Q. (2018). The Cellular Response to Transcription-Blocking DNA Damage. Trends Biochem Sci *43*, 327-341.

Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. Elife *6*.

Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R.H., Michaelis, J., and Werner, F. (2011). The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. Mol Cell *43*, 263-274.

Groisman, R., Kuraoka, I., Chevallier, O., Gaye, N., Magnaldo, T., Tanaka, K., Kisselev, A.F., Harel-Bellan, A., and 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, A.R., de Almeida, S.F., Braga, J., and Carmo-Fonseca, M. (2012). Dynamic transitions in RNA polymerase II density profiles during transcription termination. Genome Res *22*, 1447-1456.

Gruber, J.J., Olejniczak, S.H., Yong, J., La Rocca, G., Dreyfuss, G., and Thompson, C.B. (2012). Ars2 promotes proper replication-dependent histone mRNA 3' end formation. Mol Cell *45*, 87-98.

Grunberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. Nat Struct Mol Biol *19*, 788-796.

Gu, B., Eick, D., and Bensaude, O. (2013). CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. Nucleic Acids Res *41*, 1591-1603.

Gudipaty, S.A., McNamara, R.P., Morton, E.L., and D'Orso, I. (2015). PPM1G Binds 7SK RNA and Hexim1 To Block P-TEFb Assembly into the 7SK snRNP and Sustain Transcription Elongation. Mol Cell Biol *35*, 3810-3828.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Guo, J., and Price, D.H. (2013). RNA polymerase II transcription elongation control. Chem Rev *113*, 8583-8603.

Guo, J., Turek, M.E., and Price, D.H. (2014). Regulation of RNA polymerase II termination by phosphorylation of Gdown1. J Biol Chem *289*, 12657-12665.

Gupta, S., Busch, R.K., Singh, R., and Reddy, R. (1990). Characterization of U6 small nuclear RNA cap-specific antibodies. Identification of gamma-monomethyl-GTP cap structure in 7SK and several other human small RNAs. J Biol Chem *265*, 19137-19142.

Gupte, R., Muse, G.W., Chinenov, Y., Adelman, K., and Rogatsky, I. (2013). Glucocorticoid receptor represses proinflammatory genes at distinct steps of the transcription cycle. Proc Natl Acad Sci U S A *110*, 14616-14621.

Gurumurthy, M., Tan, C.H., Ng, R., Zeiger, L., Lau, J., Lee, J., Dey, A., Philp, R., Li, Q., Lim, T.M.*, et al.* (2008). Nucleophosmin interacts with HEXIM1 and regulates RNA polymerase II transcription. J Mol Biol *378*, 302-317.

Guzman, E., and Lis, J.T. (1999). Transcription factor TFIIH is required for promoter melting in vivo. Mol Cell Biol *19*, 5652-5658.

Gyenis, A., Umlauf, D., Ujfaludi, Z., Boros, I., Ye, T., and Tora, L. (2014). UVB induces a genome-wide acting negative regulatory mechanism that operates at the level of transcription initiation in human cells. PLoS Genet *10*, e1004483.

Haaland, R.E., Herrmann, C.H., and Rice, A.P. (2003). Increased association of 7SK snRNA with Tat cofactor P-TEFb following activation of peripheral blood lymphocytes. AIDS *17*, 2429-2436.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol *19*, 621-637.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622-634.

Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche Nel, H., Gostan, T., Vandermoere, F., Robert, M.C., Cusack, S.*, et al.* (2013). CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. Nat Struct Mol Biol *20*, 1358-1366.

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

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

Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat Rev Mol Cell Biol *18*, 263-273.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep *15*, 2147-2158.

Harreman, M., Taschner, M., Sigurdsson, S., Anindya, R., Reid, J., Somesh, B., Kong, S.E., Banks, C.A., Conaway, R.C., Conaway, J.W.*, et al.* (2009). Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. Proc Natl Acad Sci U S A *106*, 20705-20710.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445-458.

Hawkes, N.A., and Roberts, S.G. (1999). The role of human TFIIB in transcription start site selection in vitro and in vivo. J Biol Chem *274*, 14337-14343.

He, N., Chan, C.K., Sobhian, B., Chou, S., Xue, Y., Liu, M., Alber, T., Benkirane, M., and Zhou, Q. (2011). Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. Proc Natl Acad Sci U S A *108*, E636-645.

He, N., Jahchan, N.S., Hong, E., Li, Q., Bayfield, M.A., Maraia, R.J., Luo, K., and Zhou, Q. (2008). A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. Mol Cell *29*, 588-599.

He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N.J., Alber, T., and Zhou, Q. (2010). HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. Mol Cell *38*, 428-438.

He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. Nature *495*, 481-486.

He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., and Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. Nature *533*, 359-365.

Heidemann, M., Hintermair, C., Voss, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim Biophys Acta *1829*, 55-62.

Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S., and Levine, M.S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc Natl Acad Sci U S A *105*, 7762-7767.

Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2013). Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. Mol Cell *52*, 517-528.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev *32*, 26-41.

Herrlich, P., Sachsenmaier, C., Radler-Pohl, A., Gebel, S., Blattner, C., and Rahmsdorf, H.J. (1994). The mammalian UV response: mechanism of DNA damage induced gene expression. Adv Enzyme Regul *34*, 381-395.

Hertweck, A., Evans, C.M., Eskandarpour, M., Lau, J.C., Oleinika, K., Jackson, I., Kelly, A., Ambrose, J., Adamson, P., Cousins, D.J.*, et al.* (2016). T-bet Activates Th1 Genes through Mediator and the Super Elongation Complex. Cell Rep *15*, 2756-2770.

Hieb, A.R., Halsey, W.A., Betterton, M.D., Perkins, T.T., Kugel, J.F., and Goodrich, J.A. (2007). TFIIA changes the conformation of the DNA in TBP/TATA complexes and increases their kinetic stability. J Mol Biol *372*, 619-632.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E.*, et al.* (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. Embo J.

Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. Mol Cell *48*, 365-374.

Hogg, J.R., and Collins, K. (2007). RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. Rna *13*, 868-880.

Hollink, I.H., Alfadhel, M., Al-Wakeel, A.S., Ababneh, F., Pfundt, R., de Man, S.A., Jamra, R.A., Rolfs, A., Bertoli-Avella, A.M., and van de Laar, I.M. (2016). Broadening the phenotypic spectrum of pathogenic LARP7 variants: two cases with intellectual disability, variable growth retardation and distinct facial features. J Hum Genet *61*, 229-233.

Holohan, B., Kim, W., Lai, T.P., Hoshiyama, H., Zhang, N., Alazami, A.M., Wright, W.E., Meyn, M.S., Alkuraya, F.S., and Shay, J.W. (2016). Impaired telomere maintenance in Alazami syndrome patients with LARP7 deficiency. BMC Genomics *17*, 749.

Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E.T., Nijmeijer, B., Kotthoff, C., Rybin, V., Amlacher, S.*, et al.* (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. Nature *499*, 111-114.

Hoogstraten, D., Bergink, S., Ng, J.M., Verbiest, V.H., Luijsterburg, M.S., Geverts, B., Raams, A., Dinant, C., Hoeijmakers, J.H., Vermeulen, W.*, et al.* (2008). Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC. J Cell Sci *121*, 2850-2859.

Hsieh, F.K., Kulaeva, O.I., Patel, S.S., Dyer, P.N., Luger, K., Reinberg, D., and Studitsky, V.M. (2013). Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. Proc Natl Acad Sci U S A *110*, 7654-7659.

Hsin, J.P., Li, W., Hoque, M., Tian, B., and Manley, J.L. (2014). RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. Elife *3*, e02112.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev *26*, 2119-2137.

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

Hu, J., Adar, S., Selby, C.P., Lieb, J.D., and Sancar, A. (2015). Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. Genes Dev *29*, 948-960.

Huang, B., Yang, X.D., Zhou, M.M., Ozato, K., and Chen, L.F. (2009). Brd4 coactivates transcriptional activation of NF-kappaB via specific binding to acetylated RelA. Mol Cell Biol *29*, 1375-1387.

Huang, F., Shao, W., Fujinaga, K., and Peterlin, B.M. (2018). Bromodomain-containing protein 4-independent transcriptional activation by autoimmune regulator (AIRE) and NF-kappaB. J Biol Chem *293*, 4993-5004.

Huang, F., Wagner, M., and Siddiqui, M.A. (2004). Ablation of the CLP-1 gene leads to down-regulation of the HAND1 gene and abnormality of the left ventricle of the heart and fetal death. Mech Dev *121*, 559-572.

Huang, J.C., Svoboda, D.L., Reardon, J.T., and Sancar, A. (1992). Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5' and the 6th phosphodiester bond 3' to the photodimer. Proc Natl Acad Sci U S A *89*, 3664-3668.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet *46*, 205-212.

Hussain, R.H., Zawawi, M., and Bayfield, M.A. (2013a). Conservation of RNA chaperone activity of the human La-related proteins 4, 6 and 7. Nucleic Acids Res *41*, 8715-8725.

Hussain, S., Sajini, A.A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., Paramor, M., Gleeson, J.G., Odom, D.T., Ule, J.*, et al.* (2013b). NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep *4*, 255-261.

Ibrahim, M.M., Karabacak, A., Glahs, A., Kolundzic, E., Hirsekorn, A., Carda, A., Tursun, B., Zinzen, R.P., Lacadie, S.A., and Ohler, U. (2018). Determinants of promoter and enhancer transcription directionality in metazoans. Nat Commun *9*, 4472.

Imbert-Bouteille, M., Mau Them, F.T., Thevenon, J., Guignard, T., Gatinois, V., Riviere, J.B., Boland, A., Meyer, V., Deleuze, J.F., Sanchez, E.*, et al.* (2019). LARP7 variants and further delineation of the Alazami syndrome phenotypic spectrum among primordial dwarfisms: 2 sisters. Eur J Med Genet *62*, 161-166.

Ip, J.Y., Schmidt, D., Pan, Q., Ramani, A.K., Fraser, A.G., Odom, D.T., and Blencowe, B.J. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. Genome Res *21*, 390-401.

Ishibashi, T., Dangkulwanich, M., Coello, Y., Lionberger, T.A., Lubkowska, L., Ponticelli, A.S., Kashlev, M., and Bustamante, C. (2014). Transcription factors IIS and IIF enhance transcription efficiency by differentially modifying RNA polymerase pausing dynamics. Proc Natl Acad Sci U S A *111*, 3419-3424.

Ito, M., Yuan, C.X., Okano, H.J., Darnell, R.B., and Roeder, R.G. (2000). Involvement of the TRAP220 component of the TRAP/SMCC coactivator complex in embryonic development and thyroid hormone action. Mol Cell *5*, 683-693.

Itzen, F., Greifenberg, A.K., Bosken, C.A., and Geyer, M. (2014). Brd4 activates P-TEFb for RNA polymerase II CTD phosphorylation. Nucleic Acids Res *42*, 7577-7590.

Ivanovska, I., Jacques, P.E., Rando, O.J., Robert, F., and Winston, F. (2011). Control of chromatin structure by spt6: different consequences in coding and regulatory regions. Mol Cell Biol *31*, 531-541.

Izban, M.G., and Luse, D.S. (1992). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. Genes Dev *6*, 1342-1356.

James, K., Gamba, P., Cockell, S.J., and Zenkin, N. (2017). Misincorporation by RNA polymerase is a major source of transcription pausing in vivo. Nucleic Acids Res *45*, 1105-1113.

Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell *19*, 523-534.

Jawdekar, G.W., and Henry, R.W. (2008). Transcriptional regulation of human small nuclear RNA genes. Biochim Biophys Acta *1779*, 295-305.

Jensen, A., and Mullenders, L.H. (2010). Transcription factor IIS impacts UV-inhibited transcription. DNA Repair (Amst) *9*, 1142-1150.

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J.*, et al.* (2007). Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol Cell *27*, 262-274.

Jeronimo, C., and Robert, F. (2017). The Mediator Complex: At the Nexus of RNA Polymerase II Transcription. Trends Cell Biol *27*, 765-783.

Ji, X., Lu, H., Zhou, Q., and Luo, K. (2014). LARP7 suppresses P-TEFb activity to inhibit breast cancer progression and metastasis. Elife *3*, e02907.

Ji, X., Zhou, Y., Pandit, S., Huang, J.E., Li, H.R., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.D. (2013). SR Proteins Collaborate with 7SK and Promoter-Associated Nascent RNA to Release Paused Polymerase. Cell *153*, 855-868.

Jimeno-Gonzalez, S., Ceballos-Chavez, M., and Reyes, J.C. (2015). A positioned +1 nucleosome enhances promoter-proximal pausing. Nucleic Acids Res *43*, 3068-3078.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., Chait, B.T., Gnatt, A., Ren, B., and Roeder, R.G. (2012). Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. Mol Cell *45*, 51-63.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 167-177.

Kanazawa, S., Okamoto, T., and Peterlin, B.M. (2000). Tat competes with CIITA for the binding to P-TEFb and blocks the expression of MHC class II genes in HIV infection. Immunity *12*, 61-70.

Kanazawa, S., Soucek, L., Evan, G., Okamoto, T., and Peterlin, B.M. (2003). c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. Oncogene *22*, 5707-5711.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.

Kaplan, C.D., Morris, J.R., Wu, C., and 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.

Kellner, W.A., Bell, J.S., and Vertino, P.M. (2015). GC skew defines distinct RNA polymerase pause sites in CpG island promoters. Genome Res *25*, 1600-1609.

Kemble, D.J., McCullough, L.L., Whitby, F.G., Formosa, T., and Hill, C.P. (2015). FACT Disrupts Nucleosome Structure by Binding H2A-H2B with Conserved Peptide Motifs. Mol Cell *60*, 294-306.

Keramati, F., Seyedjafari, E., Fallah, P., Soleimani, M., and Ghanbarian, H. (2015). 7SK small nuclear RNA inhibits cancer cell proliferation through apoptosis induction. Tumour Biol *36*, 2809-2814.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010a). Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat Struct Mol Biol *17*, 1279-1286.

Kim, J., Guermah, M., McGinty, R.K., Lee, J.S., Tang, Z., Milne, T.A., Shilatifard, A., Muir, T.W., and Roeder, R.G. (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell *137*, 459-471.

Kim, J., Guermah, M., and Roeder, R.G. (2010b). The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. Cell *140*, 491-503.

Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520-527.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010c). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, Y.K., Mbonye, U., Hokello, J., and Karn, J. (2011). T-cell receptor signaling enhances transcriptional elongation from latent HIV proviruses by activating P-TEFb through an ERK-dependent pathway. J Mol Biol *410*, 896-916.

Kininis, M., Chen, B.S., Diehl, A.G., Isaacs, G.D., Zhang, T., Siepel, A.C., Clark, A.G., and Kraus, W.L. (2007). Genomic analyses of transcription factor binding, histone acetylation, and gene expression reveal mechanistically distinct classes of estrogen-regulated promoters. Mol Cell Biol *27*, 5090-5104.

Kininis, M., Isaacs, G.D., Core, L.J., Hah, N., and Kraus, W.L. (2009). Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes. Mol Cell Biol *29*, 1123-1133.

Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V.M. (2002). Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. Mol Cell *9*, 541-552.

Kobbi, L., Demey-Thomas, E., Braye, F., Proux, F., Kolesnikova, O., Vinh, J., Poterszman, A., and Bensaude, O. (2016). An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb. Proc Natl Acad Sci U S A.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C.*, et al.* (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat Struct Mol Biol *18*, 956-963.

Kohoutek, J., Li, Q., Blazek, D., Luo, Z., Jiang, H., and Peterlin, B.M. (2009). Cyclin T2 is essential for mouse embryogenesis. Mol Cell Biol *29*, 3280-3285.

Komissarova, N., and Kashlev, M. (1997). RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. J Biol Chem *272*, 15329-15338.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krause, M.W., Love, D.C., Ghosh, S.K., Wang, P., Yun, S., Fukushige, T., and Hanover, J.A. (2018). Nutrient-Driven O-GlcNAcylation at Promoters Impacts Genome-Wide RNA Pol II Distribution. Front Endocrinol (Lausanne) *9*, 521.

Kremer, S.B., Kim, S., Jeon, J.O., Moustafa, Y.W., Chen, A., Zhao, J., and Gross, D.S. (2012). Role of Mediator in regulating Pol II elongation and nucleosome displacement in Saccharomyces cerevisiae. Genetics *191*, 95-106.

Kretz, A.L., Schaum, M., Richter, J., Kitzig, E.F., Engler, C.C., Leithauser, F., Henne-Bruns, D., Knippschild, U., and Lemke, J. (2017). CDK9 is a prognostic marker and therapeutic target in pancreatic cancer. Tumour Biol *39*, 1010428317694304.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 Is an RNA polymerase II CTD phosphatase. Mol Cell *14*, 387-394.

Kristensen, U., Epanchintsev, A., Rauschendorf, M.A., Laugel, V., Stevnsner, T., Bohr, V.A., Coin, F., and Egly, J.M. (2013). Regulatory interplay of Cockayne syndrome B ATPase and stress-response gene ATF3 following genotoxic stress. Proc Natl Acad Sci U S A *110*, E2261-2270.

Krueger, B.J., Jeronimo, C., Roy, B.B., Bouchard, A., Barrandon, C., Byers, S.A., Searcey, C.E., Cooper, J.J., Bensaude, O., Cohen, E.A.*, et al.* (2008). LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic Acids Res.

Krueger, B.J., Varzavand, K., Cooper, J.J., and Price, D.H. (2010). The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK. PLoS One *5*, e12335.

Krumm, A., Hickey, L.B., and Groudine, M. (1995). Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. Genes Dev *9*, 559-572.

Krystof, V., Chamrad, I., Jorda, R., and Kohoutek, J. (2010). Pharmacological targeting of CDK9 in cardiac hypertrophy. Med Res Rev *30*, 646-666.

Kuhn, U., Gundel, M., Knoth, A., Kerwitz, Y., Rudel, S., and Wahle, E. (2009). Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. J Biol Chem *284*, 22803-22814.

Kusakabe, M., Onishi, Y., Tada, H., Kurihara, F., Kusao, K., Furukawa, M., Iwai, S., Yokoi, M., Sakai, W., and Sugasawa, K. (2019). Mechanism and regulation of DNA damage recognition in nucleotide excision repair. Genes Environ *41*, 2.

Kuzmina, A., Verstraete, N., Galker, S., Maatook, M., Bensaude, O., and Taube, R. (2014). A single point mutation in cyclin T1 eliminates binding to Hexim1, Cdk9 and RNA but not to AFF4 and enforces repression of HIV transcription. Retrovirology *11*, 51.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950-953.

Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D.S., Zeitlinger, J*., et al.* (2013). Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. Cell *153*, 976-987.

Laine, J.P., and Egly, J.M. (2006). Initiation of DNA repair mediated by a stalled RNA polymerase IIO. EMBO J *25*, 387-397.

Laribee, R.N., Krogan, N.J., Xiao, T., Shibata, Y., Hughes, T.R., Greenblatt, J.F., and Strahl, B.D. (2005). BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex. Curr Biol *15*, 1487-1493.

Larochelle, S., Amat, R., Glover-Cutter, K., Sanso, M., Zhang, C., Allen, J.J., Shokat, K.M., Bentley, D.L., and Fisher, R.P. (2012). Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. Nat Struct Mol Biol *19*, 1108-1115.

Lau, J., Lew, Q.J., Diribarne, G., Michels, A.A., Dey, A., Bensaude, O., Lane, D.P., and Chao, S.H. (2009). Ubiquitination of HEXIM1 by HDM2. Cell Cycle *8*, 2247-2254.

Lavigne, M.D., Konstantopoulos, D., Ntakou-Zamplara, K.Z., Liakos, A., and Fousteri, M. (2017). Global unleashing of transcription elongation waves in response to genotoxic stress restricts somatic mutation rate. Nat Commun *8*, 2076.

Lebars, I., Martinez-Zapien, D., Durand, A., Coutant, J., Kieffer, B., and Dock-Bregeon, A.C. (2010). HEXIM1 targets a repeated GAUC motif in the riboregulator of transcription 7SK and promotes base pair rearrangements. Nucleic Acids Res *38*, 7749-7763.

Lee, D.K., Duan, H.O., and Chang, C. (2001). Androgen receptor interacts with the positive elongation factor P-TEFb and enhances the efficiency of transcriptional elongation. J Biol Chem *276*, 9978-9984.

Lee, H., Kraus, K.W., Wolfner, M.F., and Lis, J.T. (1992). DNA sequence requirements for generating paused polymerase at the start of hsp70. Genes Dev *6*, 284-295.

Lee, J.H., and Skalnik, D.G. (2008). Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. Mol Cell Biol *28*, 609-618.

Lee, J.S., Garrett, A.S., Yen, K., Takahashi, Y.H., Hu, D., Jackson, J., Seidel, C., Pugh, B.F., and Shilatifard, A. (2012). Codependency of H2B monoubiquitination and nucleosome reassembly on Chd1. Genes Dev *26*, 914-919.

Legendre, M., and Gautheret, D. (2003). Sequence determinants in human polyadenylation site selection. BMC Genomics *4*, 7.

Lemke, J., von Karstedt, S., Abd El Hay, M., Conti, A., Arce, F., Montinaro, A., Papenfuss, K., El-Bahrawy, M.A., and Walczak, H. (2014). Selective CDK9 inhibition overcomes TRAIL resistance by concomitant suppression of cFlip and Mcl-1. Cell Death Differ *21*, 491-502.

Lenasi, T., and Barboric, M. (2010). P-TEFb stimulates transcription elongation and pre-mRNA splicing through multilateral mechanisms. RNA Biol *7*, 145-150.

Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. Cell *145*, 502-511.

Lew, Q.J., Chia, Y.L., Chu, K.L., Lam, Y.T., Gurumurthy, M., Xu, S., Lam, K.P., Cheong, N., and Chao, S.H. (2012). Identification of HEXIM1 as a positive regulator of p53. J Biol Chem *287*, 36443-36454.

Lewis, B.A., Burlingame, A.L., and Myers, S.A. (2016). Human RNA Polymerase II Promoter Recruitment in Vitro Is Regulated by O-Linked N-Acetylglucosaminyltransferase (OGT). J Biol Chem *291*, 14056-14061.

Li, C.L., Golebiowski, F.M., Onishi, Y., Samara, N.L., Sugasawa, K., and Yang, W. (2015). Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair. Mol Cell *59*, 1025-1034.

Li, J., and Gilmour, D.S. (2011). Promoter proximal pausing and the control of gene expression. Curr Opin Genet Dev.

Li, J., Liu, Y., Rhee, H.S., Ghosh, S.K., Bai, L., Pugh, B.F., and Gilmour, D.S. (2013a). Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. Mol Cell *50*, 711-722.

Li, Q., Price, J.P., Byers, S.A., Cheng, D., Peng, J., and Price, D.H. (2005). Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. J Biol Chem *280*, 28819-28826.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., *et al.* (2013b). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516-520.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet *17*, 207-223.

Lian, Z., Karpikov, A., Lian, J., Mahajan, M.C., Hartman, S., Gerstein, M., Snyder, M., and Weissman, S.M. (2008). A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation. Genome Res *18*, 1224-1237.

Lidschreiber, M., Leike, K., and Cramer, P. (2013). Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. Mol Cell Biol *33*, 3805-3816.

Lin, C., Garrett, A.S., De Kumar, B., Smith, E.R., Gogol, M., Seidel, C., Krumlauf, R., and Shilatifard, A. (2011). Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). Genes Dev *25*, 1486-1498.

Lin, C., Smith, E.R., Takahashi, H., Lai, K.C., Martin-Brown, S., Florens, L., Washburn, M.P., Conaway, J.W., Conaway, R.C., and 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.

Ling, T.T., and Sorrentino, S. (2016). Compound heterozygous variants in the LARP7 gene as a cause of Alazami syndrome in a Caucasian female with significant failure to thrive, short stature, and developmental disability. Am J Med Genet A *170A*, 217-219.

Liu, H., and Herrmann, C.H. (2005). Differential localization and expression of the Cdk9 42k and 55k isoforms. J Cell Physiol *203*, 251-260.

Liu, W., Ma, Q., Wong, K., Li, W.B., Ohgi, K., Zhang, J., Aggarwal, A.K., and Rosenfeld, M.G. (2013). Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. Cell *155*, 1581-1595.

Liu, X., Kraus, W.L., and Bai, X. (2015). Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. Trends Biochem Sci *40*, 516-525.

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., and Hahn, S. (2009). Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. Mol Cell Biol *29*, 4852-4863.

Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell *49*, 203-210.

Louder, R.K., He, Y., Lopez-Blanco, J.R., Fang, J., Chacon, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature *531*, 604-609.

Lu, H., Flores, O., Weinmann, R., and Reinberg, D. (1991). The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. Proc Natl Acad Sci U S A *88*, 10004-10008.

Lu, H., Li, Z., Xue, Y., Schulze-Gahmen, U., Johnson, J.R., Krogan, N.J., Alber, T., and Zhou, Q. (2014). AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of SECs for HIV transactivation. Proc Natl Acad Sci U S A *111*, E15-24.

Lu, L., Fan, D., Hu, C.W., Worth, M., Ma, Z.X., and Jiang, J. (2016a). Distributive O-GlcNAcylation on the Highly Repetitive C-Terminal Domain of RNA Polymerase II. Biochemistry *55*, 1149-1158.

Lu, X., Zhu, X., Li, Y., Liu, M., Yu, B., Wang, Y., Rao, M., Yang, H., Zhou, K., Wang, Y.*, et al.* (2016b). Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. Nucleic Acids Res *44*, 6853-6867.

Luebben, W.R., Sharma, N., and Nyborg, J.K. (2010). Nucleosome eviction and activated transcription require p300 acetylation of histone H3 lysine 14. Proc Natl Acad Sci U S A *107*, 19254-19259.

Lund, E., and Dahlberg, J.E. (1984). True genes for human U1 small nuclear RNA. Copy number, polymorphism, and methylation. J Biol Chem *259*, 2013-2021.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat Struct Mol Biol *17*, 1195-1201.

Luo, X., Chae, M., Krishnakumar, R., Danko, C.G., and Kraus, W.L. (2014). Dynamic reorganization of the AC16 cardiomyocyte transcriptome in response to TNFalpha signaling revealed by integrated genomic analyses. BMC Genomics *15*, 155.

Luo, Z., Lin, C., Guest, E., Garrett, A.S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M.P., and Shilatifard, A. (2012a). The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. Mol Cell Biol *32*, 2608-2617.

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

Luse, D.S. (2012). Rethinking the role of TFIIF in transcript initiation by RNA polymerase II. Transcription *3*, 156-159.

Luse, D.S. (2013). Promoter clearance by RNA polymerase II. Biochim Biophys Acta *1829*, 63-68.

Luse, D.S., Spangler, L.C., and Ujvari, A. (2011). Efficient and rapid nucleosome traversal by RNA polymerase II depends on a combination of transcript elongation factors. J Biol Chem *286*, 6040-6048.

Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E., and Caceres, J.F. (2012). DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. Nat Struct Mol Biol *19*, 760-766.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Mol Cell *62*, 63-78.

Majello, B., and Napolitano, G. (2001). Control of RNA polymerase II activity by dedicated CTD kinases and phosphatases. Front Biosci *6*, D1358-1368.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci *65*, 1099-1122.

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

Maraia, R.J., and Bayfield, M.A. (2006). The La protein-RNA complex surfaces. Mol Cell *21*, 149-152.

Markert, A., Grimm, M., Martinez, J., Wiesner, J., Meyerhans, A., Meyuhas, O., Sickmann, A., and Fischer, U. (2008). The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. EMBO Rep *9*, 569-575.

Marnef, A., Cohen, S., and Legube, G. (2017). Transcription-Coupled DNA Double-Strand Break Repair: Active Genes Need Special Care. J Mol Biol *429*, 1277-1288.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J Biol Chem *271*, 27176-27183.

Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol Cell Biol *12*, 2078-2090.

Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem *270*, 12335-12338.

Marteijn, J.A., Lans, H., Vermeulen, W., and Hoeijmakers, J.H. (2014). Understanding nucleotide excision repair and its roles in cancer and ageing. Nat Rev Mol Cell Biol *15*, 465-481.

Martinez-Zapien, D., Legrand, P., McEwen, A.G., Proux, F., Cragnolini, T., Pasquali, S., and Dock-Bregeon, A.C. (2017). The crystal structure of the 5 functional domain of the transcription riboregulator 7SK. Nucleic Acids Res *45*, 3568-3579.

Martinez-Zapien, D., Ruiz, F.X., Poirson, J., Mitschler, A., Ramirez, J., Forster, A., Cousido-Siah, A., Masson, M., Vande Pol, S., Podjarny, A.*, et al.* (2016). Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. Nature *529*, 541-545.

Martinez-Zapien, D., Saliou, J.M., Han, X., Atmanene, C., Proux, F., Cianferani, S., and Dock-Bregeon, A.C. (2015). Intermolecular recognition of the non-coding RNA 7SK and HEXIM protein in perspective. Biochimie *117*, 63-71.

Marz, M., Donath, A., Verstraete, N., Nguyen, V.T., Stadler, P.F., and Bensaude, O. (2009). Evolution of 7SK RNA and its protein partners in metazoa. Mol Biol Evol *26*, 2821-2830.

Mason, P.B., and 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.

Matrone, G., Wilson, K.S., Maqsood, S., Mullins, J.J., Tucker, C.S., and Denvir, M.A. (2015). CDK9 and its repressor LARP7 modulate cardiomyocyte proliferation and response to injury in the zebrafish heart. J Cell Sci *128*, 4560-4571.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C.*, et al.* (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

Maxon, M.E., Goodrich, J.A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev *8*, 515-524.

Mayer, A., Landry, H.M., and Churchman, L.S. (2017). Pause & go: from the discovery of RNA polymerase pausing to its functional implications. Curr Opin Cell Biol *46*, 72-80.

Mayne, L.V., and Lehmann, A.R. (1982). Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. Cancer Res *42*, 1473-1478.

Mbogning, J., Nagy, S., Page, V., Schwer, B., Shuman, S., Fisher, R.P., and Tanny, J.C. (2013). The PAF complex and Prf1/Rtf1 delineate distinct Cdk9-dependent pathways regulating transcription elongation in fission yeast. PLoS Genet *9*, e1004029.

Mbonye, U., Wang, B., Gokulrangan, G., Shi, W., Yang, S., and Karn, J. (2018). Cyclin-dependent kinase 7 (CDK7)-mediated phosphorylation of the CDK9 activation loop promotes P-TEFb assembly with Tat and proviral HIV reactivation. J Biol Chem *293*, 10009-10025.

Mbonye, U.R., Gokulrangan, G., Datt, M., Dobrowolski, C., Cooper, M., Chance, M.R., and Karn, J. (2013). Phosphorylation of CDK9 at Ser175 enhances HIV transcription and is a marker of activated P-TEFb in CD4(+) T lymphocytes. PLoS Pathog *9*, e1003338.

Mbonye, U.R., Wang, B., Gokulrangan, G., Chance, M.R., and Karn, J. (2015). Phosphorylation of HEXIM1 at Tyr271 and Tyr274 Promotes Release of P-TEFb from the 7SK snRNP Complex and Enhances Proviral HIV Gene Expression. Proteomics *15*, 2078-2086.

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

McKay, B.C., Stubbert, L.J., Fowler, C.C., Smith, J.M., Cardamore, R.A., and Spronck, J.C. (2004). Regulation of ultraviolet light-induced gene expression by gene size. Proc Natl Acad Sci U S A *101*, 6582-6586.

McNamara, R.P., Bacon, C.W., and D'Orso, I. (2016a). Transcription elongation control by the 7SK snRNP complex: Releasing the pause. Cell Cycle *15*, 2115-2123.

McNamara, R.P., McCann, J.L., Gudipaty, S.A., and D'Orso, I. (2013). Transcription factors mediate the enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for transcription elongation. Cell Rep *5*, 1256-1268.

McNamara, R.P., Reeder, J.E., McMillan, E.A., Bacon, C.W., McCann, J.L., and D'Orso, I. (2016b). KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. Mol Cell *61*, 39-53.

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

Michel, M., and Cramer, P. (2013). Transitions for regulating early transcription. Cell *153*, 943-944.

Michels, A.A., Nguyen, V.T., Fraldi, A., Labas, V., Edwards, M., Bonnet, F., Lania, L., and Bensaude, O. (2003). MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. Mol Cell Biol *23*, 4859-4869.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev *25*, 742-754.

Missra, A., and Gilmour, D.S. (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.

Mitra, P., Yang, R.M., Sutton, J., Ramsay, R.G., and Gonda, T.J. (2016). CDK9 inhibitors selectively target estrogen receptor-positive breast cancer cells through combined inhibition of MYB and MCL-1 expression. Oncotarget *7*, 9069-9083.

Mitsui, A., and Sharp, P.A. (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.

Mochizuki, K., Nishiyama, A., Jang, M.K., Dey, A., Ghosh, A., Tamura, T., Natsume, H., Yao, H., and Ozato, K. (2008). The bromodomain protein Brd4 stimulates G1 gene transcription and promotes progression to S phase. J Biol Chem *283*, 9040-9048.

Mohapatra, S., Chu, B., Zhao, X., Djeu, J., Cheng, J.Q., and Pledger, W.J. (2009). Apoptosis of metastatic prostate cancer cells by a combination of cyclin-dependent kinase and AKT inhibitors. Int J Biochem Cell Biol *41*, 595-602.

Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A., and Kanduri, C. (2010). Characterization of the RNA content of chromatin. Genome Res *20*, 899-907.

Monroe, S.C., Jo, S.Y., Sanders, D.S., Basrur, V., Elenitoba-Johnson, K.S., Slany, R.K., and Hess, J.L. (2011). MLL-AF9 and MLL-ENL alter the dynamic association of transcriptional regulators with genes critical for leukemia. Exp Hematol *39*, 77-86 e71-75.

Morchikh, M., Cribier, A., Raffel, R., Amraoui, S., Cau, J., Severac, D., Dubois, E., Schwartz, O., Bennasser, Y., and Benkirane, M. (2017). HEXIM1 and NEAT1 Long Non-coding RNA Form a Multi-subunit Complex that Regulates DNA-Mediated Innate Immune Response. Mol Cell *67*, 387-399 e385.

Mourgues, S., Gautier, V., Lagarou, A., Bordier, C., Mourcet, A., Slingerland, J., Kaddoum, L., Coin, F., Vermeulen, W., Gonzales de Peredo, A.*, et al.* (2013). ELL, a novel TFIIH partner, is involved in transcription restart after DNA repair. Proc Natl Acad Sci U S A *110*, 17927-17932.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell *51*, 606-617.

Muck, F., Bracharz, S., and Marschalek, R. (2016). DDX6 transfers P-TEFb kinase to the AF4/AF4N (AFF1) super elongation complex. Am J Blood Res *6*, 28-45.

Mueller, D., Garcia-Cuellar, M.P., Bach, C., Buhl, S., Maethner, E., and Slany, R.K. (2009). Misguided Transcriptional Elongation Causes Mixed Lineage Leukemia. Plos Biology *7*.

Muhar, M., Ebert, A., Neumann, T., Umkehrer, C., Jude, J., Wieshofer, C., Rescheneder, P., Lipp, J.J., Herzog, V.A., Reichholf, B.*, et al.* (2018). SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science *360*, 800-805.

Mullen Davis, M.A., Guo, J., Price, D.H., and Luse, D.S. (2014). Functional interactions of the RNA polymerase II-interacting proteins Gdown1 and TFIIF. J Biol Chem *289*, 11143-11152.

Muniz, L., Egloff, S., and Kiss, T. (2013). RNA elements directing in vivo assembly of the 7SK/MePCE/Larp7 transcriptional regulatory snRNP. Nucleic Acids Res *41*, 4686-4698.

Muniz, L., Egloff, S., Ughy, B., Jady, B.E., and Kiss, T. (2010). Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator Tat. PLoS Pathog *6*, e1001152.

Munoz, M.J., Nieto Moreno, N., Giono, L.E., Cambindo Botto, A.E., Dujardin, G., Bastianello, G., Lavore, S., Torres-Mendez, A., Menck, C.F.M., Blencowe, B.J.*, et al.* (2017). Major Roles for Pyrimidine Dimers, Nucleotide Excision Repair, and ATR in the Alternative Splicing Response to UV Irradiation. Cell Rep *18*, 2868-2879.

Munoz, M.J., Perez Santangelo, M.S., Paronetto, M.P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J.J.*, et al.* (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. Cell *137*, 708-720.

Murphy, S., Di Liegro, C., and Melli, M. (1987). The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. Cell *51*, 81-87.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat Genet *39*, 1507-1511.

Nag, A., Narsinh, K., and Martinson, H.G. (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.

Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P.*, et al.* (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature *478*, 57-63.

Nakagawa, S., and Hirose, T. (2012). Paraspeckle nuclear bodies--useful uselessness? Cell Mol Life Sci *69*, 3027-3036.

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

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.

Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature *414*, 322-325.

Ni, Z., Xu, C., Guo, X., Hunter, G.O., Kuznetsova, O.V., Tempel, W., Marcon, E., Zhong, G., Guo, H., Kuo, W.H.*, et al.* (2014). RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. Nat Struct Mol Biol *21*, 686-695.

Nilson, K.A., Guo, J., Turek, M.E., Brogie, J.E., Delaney, E., Luse, D.S., and Price, D.H. (2015). THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. Mol Cell *59*, 576-587.

Nilson, K.A., Lawson, C.K., Mullen, N.J., Ball, C.B., Spector, B.M., Meier, J.L., and Price, D.H. (2017). Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. Nucleic Acids Res *45*, 11088-11105.

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

Nogues, G., Kadener, S., Cramer, P., de la Mata, M., Fededa, J.P., Blaustein, M., Srebrow, A., and Kornblihtt, A.R. (2003). Control of alternative pre-mRNA splicing by RNA Pol II elongation: faster is not always better. IUBMB Life *55*, 235-241.

Nojima, M., Huang, Y., Tyagi, M., Kao, H.Y., and Fujinaga, K. (2008). The positive transcription elongation factor b is an essential cofactor for the activation of transcription by myocyte enhancer factor 2. J Mol Biol *382*, 275-287.

Nojima, T., Gomes, T., Grosso, A.R., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526-540.

Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., and Carmo-Fonseca, M. (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. Mol Cell *72*, 369-379 e364.

Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R.*, et al.* (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat Struct Mol Biol *20*, 923-928.

Ohkuma, Y., and Roeder, R.G. (1994). Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. Nature *368*, 160-163.

Oksenych, V., Bernardes de Jesus, B., Zhovmer, A., Egly, J.M., and Coin, F. (2009). Molecular insights into the recruitment of TFIIH to sites of DNA damage. EMBO J *28*, 2971-2980.

Oksenych, V., Zhovmer, A., Ziani, S., Mari, P.O., Eberova, J., Nardo, T., Stefanini, M., Giglia-Mari, G., Egly, J.M., and Coin, F. (2013). Histone methyltransferase DOT1L drives recovery of gene expression after a genotoxic attack. PLoS Genet *9*, e1003611.

Okuda, M., Kinoshita, M., Kakumu, E., Sugasawa, K., and Nishimura, Y. (2015). Structural Insight into the Mechanism of TFIIH Recognition by the Acidic String of the Nucleotide Excision Repair Factor XPC. Structure *23*, 1827-1837.

Okuda, M., Nakazawa, Y., Guo, C., Ogi, T., and Nishimura, Y. (2017). Common TFIIH recruitment mechanism in global genome and transcription-coupled repair subpathways. Nucleic Acids Res *45*, 13043-13055.

Orphanides, G., LeRoy, G., Chang, C.H., Luse, D.S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. Cell *92*, 105-116.

Oven, I., Brdickova, N., Kohoutek, J., Vaupotic, T., Narat, M., and Peterlin, B.M. (2007). AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. Molecular and Cellular Biology *27*, 8815-8823.

Palangat, M., Renner, D.B., Price, D.H., and Landick, R. (2005). A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. Proc Natl Acad Sci U S A *102*, 15036-15041.

Paparidis, N.F., Durvale, M.C., and Canduri, F. (2017). The emerging picture of CDK9/P-TEFb: more than 20 years of advances since PITALRE. Mol Biosyst *13*, 246-276.

Pardee, T.S., Bangur, C.S., and Ponticelli, A.S. (1998). The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. J Biol Chem *273*, 17859-17864.

Patel, M.C., Debrosse, M., Smith, M., Dey, A., Huynh, W., Sarai, N., Heightman, T.D., Tamura, T., and Ozato, K. (2013). BRD4 coordinates recruitment of pause release factor P-TEFb and the pausing complex NELF/DSIF to regulate transcription elongation of interferon-stimulated genes. Mol Cell Biol *33*, 2497-2507.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell *125*, 703-717.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P*., et al.* (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell *161*, 774-789.

Peng, J., Zhu, Y., Milton, J.T., and Price, D.H. (1998). Identification of multiple cyclin subunits of human P-TEFb. Genes Dev *12*, 755-762.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol Cell *23*, 297-305.

Pirngruber, J., Shchebet, A., Schreiber, L., Shema, E., Minsky, N., Chapman, R.D., Eick, D., Aylon, Y., Oren, M., and Johnsen, S.A. (2009). CDK9 directs H2B monoubiquitination and controls replication-dependent histone mRNA 3'-end processing. EMBO Rep *10*, 894-900.

Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. Nature *533*, 353-358.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat Rev Mol Cell Biol *16*, 190-202.

Prasanth, K.V., Camiolo, M., Chan, G., Tripathi, V., Denis, L., Nakamura, T., Hubner, M.R., and Spector, D.L. (2010). Nuclear organization and dynamics of 7SK RNA in regulating gene expression. Mol Biol Cell *21*, 4184-4196.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.

Price, D.H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. Mol Cell Biol *20*, 2629-2634.

Price, D.H. (2018). Transient pausing by RNA polymerase II. Proc Natl Acad Sci U S A *115*, 4810-4812.

Proietti-De-Santis, L., Drane, P., and Egly, J.M. (2006). Cockayne syndrome B protein regulates the transcriptional program after UV irradiation. EMBO J *25*, 1915-1923.

Proudfoot, N. (1991). Poly(A) signals. Cell *64*, 671-674.

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

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev *25*, 1770-1782.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Puumalainen, M.R., Ruthemann, P., Min, J.H., and Naegeli, H. (2016). Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation. Cell Mol Life Sci *73*, 547-566.

Qiu, H., Hu, C., Gaur, N.A., and Hinnebusch, A.G. (2012). Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. EMBO J *31*, 3494-3505.

Qiu, Y., and Gilmour, D.S. (2017). Identification of Regions in the Spt5 Subunit of DRB Sensitivity-inducing Factor (DSIF) That Are Involved in Promoter-proximal Pausing. J Biol Chem *292*, 5555-5570.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc Regulates Transcriptional Pause Release. Cell *141*, 432-445.

Ramachandran, P., Palidwor, G.A., and Perkins, T.J. (2015). BIDCHIPS: bias decomposition and removal from ChIP-seq data clarifies true binding signal and its functional correlates. Epigenetics Chromatin *8*, 33.

Ramakrishnan, R., and Rice, A.P. (2012). Cdk9 T-loop phosphorylation is regulated by the calcium signaling pathway. J Cell Physiol *227*, 609-617.

Ramakrishnan, R., Yu, W., and Rice, A.P. (2011). Limited redundancy in genes regulated by Cyclin T2 and Cyclin T1. BMC Res Notes *4*, 260.

Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W., and Lewis, B.A. (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.

Ratner, J.N., Balasubramanian, B., Corden, J., Warren, S.L., and Bregman, D.B. (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.

Reines, D., Chamberlin, M.J., and Kane, C.M. (1989). Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene in vitro. J Biol Chem *264*, 10799-10809.

Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H., and Price, D.H. (2001). A highly purified RNA polymerase II elongation control system. J Biol Chem *276*, 42601-42609.

Riley, T., Sontag, E., Chen, P., and Levine, A. (2008). Transcriptional control of human p53-regulated genes. Nat Rev Mol Cell Biol *9*, 402-412.

Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.J., Burlingame, A.L., and Kornberg, R.D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell *166*, 1411-1422 e1416.

Rockx, D.A., Mason, R., van Hoffen, A., Barton, M.C., Citterio, E., Bregman, D.B., van Zeeland, A.A., Vrieling, H., and Mullenders, L.H. (2000). UV-induced inhibition of transcription involves repression of transcription initiation and phosphorylation of RNA polymerase II. Proc Natl Acad Sci U S A *97*, 10503-10508.

Rosado-Lugo, J.D., and Hampsey, M. (2014). The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition. J Biol Chem *289*, 33916-33926.

Sabo, A., Lusic, M., Cereseto, A., and Giacca, M. (2008). Acetylation of conserved lysines in the catalytic core of cyclin-dependent kinase 9 inhibits kinase activity and regulates transcription. Mol Cell Biol *28*, 2201-2212.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 129-143.

Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature *493*, 437-440.

Samudyata, Amaral, P.P., Engstrom, P.G., Robson, S.C., Nielsen, M.L., Kouzarides, T., and Castelo-Branco, G. (2019). Interaction of Sox2 with RNA binding proteins in mouse embryonic stem cells. Exp Cell Res *381*, 129-138.

Sanchez, A., De Vivo, A., Uprety, N., Kim, J., Stevens, S.M., Jr., and Kee, Y. (2016). BMI1-UBR5 axis regulates transcriptional repression at damaged chromatin. Proc Natl Acad Sci U S A *113*, 11243-11248.

Sano, M., Abdellatif, M., Oh, H., Xie, M., Bagella, L., Giordano, A., Michael, L.H., DeMayo, F.J., and Schneider, M.D. (2002). Activation and function of cyclin T-Cdk9 (positive transcription elongation factor-b) in cardiac muscle-cell hypertrophy. Nat Med *8*, 1310-1317.

Sano, M., and Schneider, M.D. (2004). Cyclin-dependent kinase-9: an RNAPII kinase at the nexus of cardiac growth and death cascades. Circ Res *95*, 867-876.

Sano, M., Wang, S.C., Shirai, M., Scaglia, F., Xie, M., Sakai, S., Tanaka, T., Kulkarni, P.A., Barger, P.M., Youker, K.A*., et al.* (2004). Activation of cardiac Cdk9 represses PGC-1 and confers a predisposition to heart failure. Embo J *23*, 3559-3569.

Sanso, M., Levin, R.S., Lipp, J.J., Wang, V.Y., Greifenberg, A.K., Quezada, E.M., Ali, A., Ghosh, A., Larochelle, S., Rana, T.M*., et al.* (2016). P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. Genes Dev *30*, 117-131.

Santoro, M., Nociti, V., Lucchini, M., De Fino, C., Losavio, F.A., and Mirabella, M. (2016). Expression Profile of Long Non-Coding RNAs in Serum of Patients with Multiple Sclerosis. J Mol Neurosci *59*, 18-23.

Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G.P., Heron, M., Williams, H., Soding, J., Stewart, A., and Svejstrup, J.Q. (2014). RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. Cell *157*, 1037-1049.

Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell *56*, 29-42.

Schoenherr, C.J., and Anderson, D.J. (1995). The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. Science *267*, 1360-1363.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Schroder, S., Herker, E., Itzen, F., He, D., Thomas, S., Gilchrist, D.A., Kaehlcke, K., Cho, S., Pollard, K.S., Capra, J.A*., et al.* (2013). Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. Mol Cell *52*, 314-324.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev *14*, 2435-2440.

Schuller, R., Forne, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol Cell *61*, 305-314.

Schulte, A., Czudnochowski, N., Barboric, M., Schonichen, A., Blazek, D., Peterlin, B.M., and Geyer, M. (2005). Identification of a cyclin T-binding domain in Hexim1 and biochemical analysis of its binding competition with HIV-1 Tat. J Biol Chem *280*, 24968-24977.

Schulz, D., Pirkl, N., Lehmann, E., and Cramer, P. (2014). Rpb4 subunit functions mainly in mRNA synthesis by RNA polymerase II. J Biol Chem *289*, 17446-17452.

Schwalb, B., Michel, M., Zacher, B., Fruhauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225-1228.

Schwartz, B.E., Larochelle, S., Suter, B., and Lis, J.T. (2003). Cdk7 is required for full activation of Drosophila heat shock genes and RNA polymerase II phosphorylation in vivo. Mol Cell Biol *23*, 6876-6886.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat Struct Mol Biol *16*, 990-995.

Schweikhard, V., Meng, C., Murakami, K., Kaplan, C.D., Kornberg, R.D., and Block, S.M. (2014). Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. Proc Natl Acad Sci U S A *111*, 6642-6647.

Schwertman, P., Lagarou, A., Dekkers, D.H., Raams, A., van der Hoek, A.C., Laffeber, C., Hoeijmakers, J.H., Demmers, J.A., Fousteri, M., Vermeulen, W.*, et al.* (2012). UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair. Nat Genet *44*, 598-602.

Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Mol Cell *58*, 1101-1112.

Shah, N., Maqbool, M.A., Yahia, Y., El Aabidine, A.Z., Esnault, C., Forne, I., Decker, T.M., Martin, D., Schuller, R., Krebs, S.*, et al.* (2018). Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. Mol Cell *69*, 48-61 e46.

Shao, H., Shi, S., Huang, S., Hole, A.J., Abbas, A.Y., Baumli, S., Liu, X., Lam, F., Foley, D.W., Fischer, P.M.*, et al.* (2013). Substituted 4-(thiazol-5-yl)-2-(phenylamino)pyrimidines are highly active CDK9 inhibitors: synthesis, X-ray crystal structures, structure-activity relationship, and anticancer activities. J Med Chem *56*, 640-659.

Shao, W., Alcantara, S.G., and Zeitlinger, J. (2019). Reporter-ChIP-nexus reveals strong contribution of the Drosophila initiator sequence to RNA polymerase pausing. Elife *8*.

Shao, W., and Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. Nat Genet *49*, 1045-1051.

Sharma, P., Lioutas, A., Fernandez-Fuentes, N., Quilez, J., Carbonell-Caballero, J., Wright, R.H.G., Di Vona, C., Le Dily, F., Schuller, R., Eick, D.*, et al.* (2019). Arginine Citrullination at the C-Terminal Domain Controls RNA Polymerase II Transcription. Mol Cell *73*, 84-96 e87.

Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference--a crash course. Trends Genet *21*, 339-345.

Shelton, S.B., Shah, N.M., Abell, N.S., Devanathan, S.K., Mercado, M., and Xhemalce, B. (2018). Crosstalk between the RNA Methylation and Histone-Binding Activities of MePCE Regulates P-TEFb Activation on Chromatin. Cell Rep *22*, 1374-1383.

Sheridan, R.M., Fong, N., D'Alessandro, A., and Bentley, D.L. (2019). Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5' Pause Release, Termination, and Transcription Elongation Rate. Mol Cell *73*, 107-118 e104.

Shimizu, N., Ouchida, R., Yoshikawa, N., Hisada, T., Watanabe, H., Okamoto, K., Kusuhara, M., Handa, H., Morimoto, C., and Tanaka, H. (2005). HEXIM1 forms a transcriptionally abortive complex with glucocorticoid receptor without involving 7SK RNA and positive transcription elongation factor b. Proc Natl Acad Sci U S A *102*, 8555-8560.

Shimizu, N., Yoshikawa, N., Wada, T., Handa, H., Sano, M., Fukuda, K., Suematsu, M., Sawai, T., Morimoto, C., and Tanaka, H. (2008). Tissue- and context-dependent modulation of hormonal sensitivity of glucocorticoid-responsive genes by hexamethylene bisacetamide-inducible protein 1. Mol Endocrinol *22*, 2609-2623.

Shimojima, T., Okada, M., Nakayama, T., Ueda, H., Okawa, K., Iwamatsu, A., Handa, H., and Hirose, S. (2003). Drosophila FACT contributes to Hox gene expression through physical and functional interactions with GAGA factor. Genes Dev *17*, 1605-1616.

Shopland, L.S., Hirayoshi, K., Fernandes, M., and Lis, J.T. (1995). HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. Genes Dev *9*, 2756-2769.

Shore, S.M., Byers, S.A., Maury, W., and Price, D.H. (2003). Identification of a novel isoform of Cdk9. Gene *307*, 175-182.

Shumyatsky, G.P., Tillib, S.V., and Kramerov, D.A. (1990). B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5' end. Nucleic Acids Res *18*, 6347-6351.

Simone, C., Stiegler, P., Bagella, L., Pucci, B., Bellan, C., De Falco, G., De Luca, A., Guanti, G., Puri, P.L., and Giordano, A. (2002). Activation of MyoD-dependent transcription by cdk9/cyclin T2. Oncogene *21*, 4137-4148.

Simonti, C.N., Pollard, K.S., Schroder, S., He, D., Bruneau, B.G., Ott, M., and Capra, J.A. (2015). Evolution of lysine acetylation in the RNA polymerase II C-terminal domain. BMC Evol Biol *15*, 35.

Sims, R.J., 3rd, Rojas, L.A., Beck, D., Bonasio, R., Schuller, R., Drury, W.J., 3rd, Eick, D., and Reinberg, D. (2011). The C-terminal domain of RNA polymerase II is modified by site-specific methylation. Science *332*, 99-103.

Sin, Y., Tanaka, K., and Saijo, M. (2016). The C-terminal Region and SUMOylation of Cockayne Syndrome Group B Protein Play Critical Roles in Transcription-coupled Nucleotide Excision Repair. J Biol Chem *291*, 1387-1397.

Singh, R., and Reddy, R. (1989). Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. Proc Natl Acad Sci U S A *86*, 8280-8283.

Skene, P.J., Hernandez, A.E., Groudine, M., and Henikoff, S. (2014). The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. Elife *3*, e02042.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436-439.

Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev *28*, 1384-1396.

Skreka, K., Schafferer, S., Nat, I.R., Zywicki, M., Salti, A., Apostolova, G., Griehl, M., Rederstorff, M., Dechant, G., and Huttenhofer, A. (2012). Identification of differentially expressed non-coding RNAs in embryonic stem cell neural differentiation. Nucleic Acids Res *40*, 6001-6015.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu Rev Biochem *72*, 449-479.

Smith, E., Lin, C., and Shilatifard, A. (2011a). The super elongation complex (SEC) and MLL in development and disease. Genes Dev *25*, 661-672.

Smith, E.R., Lin, C., Garrett, A.S., Thornton, J., Mohaghegh, N., Hu, D., Jackson, J., Saraf, A., Swanson, S.K., Seidel, C.*, et al.* (2011b). The little elongation complex regulates small nuclear RNA transcription. Mol Cell *44*, 954-965.

Sobhian, B., Laguette, N., Yatim, A., Nakamura, M., Levy, Y., Kiernan, R., and Benkirane, M. (2010). HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. Mol Cell *38*, 439-451.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat Rev Mol Cell Biol *19*, 262-274.

Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C., and Werner, M. (2011). Direct interaction of RNA polymerase II and mediator required for transcription in vivo. Science *331*, 1451-1454.

Spencer, C.A., and Groudine, M. (1990). Transcription elongation and eukaryotic gene regulation. Oncogene *5*, 777-785.

Starita, L.M., Horwitz, A.A., Keogh, M.C., Ishioka, C., Parvin, J.D., and Chiba, N. (2005). BRCA1/BARD1 ubiquitinate phosphorylated RNA polymerase II. J Biol Chem *280*, 24498-24505.

Steurer, B., Janssens, R.C., Geverts, B., Geijer, M.E., Wienholz, F., Theil, A.F., Chang, J., Dealy, S., Pothof, J., van Cappellen, W.A.*, et al.* (2018). Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. Proc Natl Acad Sci U S A *115*, E4368-E4376.

Stewart, H.J., Horne, G.A., Bastow, S., and Chevassut, T.J. (2013). BRD4 associates with p53 in DNMT3A-mutated leukemia cells and is implicated in apoptosis by the bromodomain inhibitor JQ1. Cancer Med *2*, 826-835.

Stiff, T., Cerosaletti, K., Concannon, P., O'Driscoll, M., and Jeggo, P.A. (2008). Replication independent ATR signalling leads to G2/M arrest requiring Nbs1, 53BP1 and MDC1. Hum Mol Genet *17*, 3247-3253.

Stiff, T., Walker, S.A., Cerosaletti, K., Goodarzi, A.A., Petermann, E., Concannon, P., O'Driscoll, M., and Jeggo, P.A. (2006). ATR-dependent phosphorylation and activation of ATM in response to UV treatment or replication fork stalling. EMBO J *25*, 5775-5782.

Sugasawa, K. (2016). Molecular mechanisms of DNA damage recognition for mammalian nucleotide excision repair. DNA Repair (Amst) *44*, 110-117.

Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol Cell *61*, 297-304.

Suh, H., Hazelbaker, D.Z., Soares, L.M., and Buratowski, S. (2013). The C-terminal domain of Rpb1 functions on other RNA polymerase II subunits. Mol Cell *51*, 850-858.

Sui, X., Sui, Y., and Wang, Y. (2018). LARP7 in papillary thyroid carcinoma induces NIS expression through suppression of the SHH signaling pathway. Mol Med Rep *17*, 7521-7528.

Sun, M., Lariviere, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). J Biol Chem *285*, 41597-41603.

Sun, Y., Liu, Z., Cao, X., Lu, Y., Mi, Z., He, C., Liu, J., Zheng, Z., Li, M.J., Li, T.*, et al.* (2019). Activation of P-TEFb by cAMP-PKA signaling in autosomal dominant polycystic kidney disease. Sci Adv *5*, eaaw3593.

Szlachta, K., Thys, R.G., Atkin, N.D., Pierce, L.C.T., Bekiranov, S., and Wang, Y.H. (2018). Alternative DNA secondary structure formation affects RNA polymerase II promoter-proximal pausing in human. Genome Biol *19*, 89.

Tahirov, T.H., Babayeva, N.D., Varzavand, K., Cooper, J.J., Sedore, S.C., and Price, D.H. (2010). Crystal structure of HIV-1 Tat complexed with human P-TEFb. Nature *465*, 747-751.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P.*, et al.* (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92-104.

Tan, J.L., Fogley, R.D., Flynn, R.A., Ablain, J., Yang, S., Saint-Andre, V., Fan, Z.P., Do, B.T., Laga, A.C., Fujinaga, K.*, et al.* (2016). Stress from Nucleotide Depletion Activates the Transcriptional Regulator HEXIM1 to Suppress Melanoma. Mol Cell *62*, 34-46.

Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. Nat Rev Mol Cell Biol *15*, 703-708.

Tettey, T.T., Gao, X., Shao, W., Li, H., Story, B.A., Chitsazan, A.D., Glaser, R.L., Goode, Z.H., Seidel, C.W., Conaway, R.C.*, et al.* (2019). A Role for FACT in RNA Polymerase II Promoter-Proximal Pausing. Cell Rep *27*, 3770-3779 e3777.

Teves, S.S., Weber, C.M., and Henikoff, S. (2014). Transcribing through the nucleosome. Trends Biochem Sci *39*, 577-586.

Teytelman, L., Ozaydin, B., Zill, O., Lefrancois, P., Snyder, M., Rine, J., and Eisen, M.B. (2009). Impact of chromatin structures on DNA processing for genomic analyses. PLoS One *4*, e6700.

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

Thompson, C.M., and Young, R.A. (1995). General requirement for RNA polymerase II holoenzymes in vivo. Proc Natl Acad Sci U S A *92*, 4587-4590.

Tian, Y., Ke, S., Chen, M., and Sheng, T. (2003). Interactions between the aryl hydrocarbon receptor and P-TEFb. Sequential recruitment of transcription factors and differential phosphorylation of C-terminal domain of RNA polymerase II at cyp1a1 promoter. J Biol Chem *278*, 44041-44048.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J., and Guigo, R. (2009). Nucleosome positioning as a determinant of exon recognition. Nat Struct Mol Biol *16*, 996-1001.

Titov, D.V., Gilman, B., He, Q.L., Bhat, S., Low, W.K., Dang, Y., Smeaton, M., Demain, A.L., Miller, P.S., Kugel, J.F.*, et al.* (2011). XPB, a subunit of TFIIH, is a target of the natural product triptolide. Nat Chem Biol *7*, 182-188.

Tornaletti, S. (2005). Transcription arrest at DNA damage sites. Mutat Res *577*, 131-145.

Tornaletti, S., Reines, D., and Hanawalt, P.C. (1999). Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA. J Biol Chem *274*, 24124-24130.

Tresini, M., Warmerdam, D.O., Kolovos, P., Snijder, L., Vrouwe, M.G., Demmers, J.A., van, I.W.F., Grosveld, F.G., Medema, R.H., Hoeijmakers, J.H.*, et al.* (2015). The core spliceosome as target and effector of non-canonical ATM signalling. Nature *523*, 53-58.

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome Res *14*, 62-66.

Tropberger, P., and Schneider, R. (2013). Scratching the (lateral) surface of chromatin regulation by histone modifications. Nat Struct Mol Biol *20*, 657-661.

Tudor, M., Murray, P.J., Onufryk, C., Jaenisch, R., and Young, R.A. (1999). Ubiquitous expression and embryonic requirement for RNA polymerase II coactivator subunit Srb7 in mice. Genes Dev *13*, 2365-2368.

Tufegdzic Vidakovic, A., Harreman, M., Dirac-Svejstrup, A.B., Boeing, S., Roy, A., Encheva, V., Neumann, M., Wilson, M., Snijders, A.P., and Svejstrup, J.Q. (2019). Analysis of RNA polymerase II ubiquitylation and proteasomal degradation. Methods *159-160*, 146-156.

Tyagi, M., Pearson, R.J., and Karn, J. (2010). Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction. J Virol *84*, 6425-6437.

Uchikawa, E., Natchiar, K.S., Han, X., Proux, F., Roblin, P., Zhang, E., Durand, A., Klaholz, B.P., and Dock-Bregeon, A.C. (2015). Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7. Nucleic Acids Res *43*, 3373-3388.

Ujvari, A., and Luse, D.S. (2006). RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. Nat Struct Mol Biol *13*, 49-54.

Van Arsdell, S.W., and Weiner, A.M. (1984). Human genes for U2 small nuclear RNA are tandemly repeated. Mol Cell Biol *4*, 492-499.

van Gool, A.J., Citterio, E., Rademakers, S., van Os, R., Vermeulen, W., Constantinou, A., Egly, J.M., Bootsma, D., and Hoeijmakers, J.H. (1997). The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex. EMBO J *16*, 5955-5965.

Van Herreweghe, E., Egloff, S., Goiffon, I., Jady, B.E., Froment, C., Monsarrat, B., and Kiss, T. (2007). Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. Embo J *26*, 3570-3580.

Van Oss, S.B., Cucinotta, C.E., and Arndt, K.M. (2017). Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation. Trends Biochem Sci *42*, 788-798.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat Rev Genet *10*, 252-263.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol *16*, 178-189.

Verma, R., Oania, R., Fang, R., Smith, G.T., and Deshaies, R.J. (2011). Cdc48/p97 mediates UV-dependent turnover of RNA Pol II. Mol Cell *41*, 82-92.

Vichi, P., Coin, F., Renaud, J.P., Vermeulen, W., Hoeijmakers, J.H., Moras, D., and Egly, J.M. (1997). Cisplatin- and UV-damaged DNA lure the basal transcription factor TFIID/TBP. EMBO J *16*, 7444-7456.

Voong, L.N., Xi, L., Sebeson, A.C., Xiong, B., Wang, J.P., and Wang, X. (2016). Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. Cell *167*, 1555-1570 e1515.

Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018a). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature *560*, 607-612.

Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018b). Structure of paused transcription complex Pol II-DSIF-NELF. Nature *560*, 601-606.

Vos, S.M., Pollmann, D., Caizzi, L., Hofmann, K.B., Rombaut, P., Zimniak, T., Herzog, F., and Cramer, P. (2016). Architecture and RNA binding of the human negative elongation factor. Elife *5*.

Voss, K., Forne, I., Descostes, N., Hintermair, C., Schuller, R., Maqbool, M.A., Heidemann, M., Flatley, A., Imhof, A., Gut, M*., et al.* (2015). Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. Transcription *6*, 91-101.

Vrouwe, M.G., Pines, A., Overmeer, R.M., Hanada, K., and Mullenders, L.H. (2011). UV-induced photolesions elicit ATR-kinase-dependent signaling in non-cycling cells through nucleotide excision repair-dependent and -independent pathways. J Cell Sci *124*, 435-446.

Wada, T., Orphanides, G., Hasegawa, J., Kim, D.K., Shima, D., Yamaguchi, Y., Fukuda, A., Hisatake, K., Oh, S., Reinberg, D*., et al.* (2000). FACT relieves DSIF/NELF-mediated inhibition of transcriptional elongation and reveals functional differences between P-TEFb and TFIIH. Mol Cell *5*, 1067-1072.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F*., et al.* (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.

Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O*., et al.* (2012). Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. Cell *150*, 1147-1157.

Walsby, E., Pratt, G., Shao, H., Abbas, A.Y., Fischer, P.M., Bradshaw, T.D., Brennan, P., Fegan, C., Wang, S., and Pepper, C. (2014). A novel Cdk9 inhibitor preferentially targets tumor cells and synergizes with fludarabine. Oncotarget *5*, 375-385.

Wang, Y., Dow, E.C., Liang, Y.Y., Ramakrishnan, R., Liu, H., Sung, T.L., Lin, X., and Rice, A.P. (2008). Phosphatase PPM1A regulates phosphorylation of Thr-186 in the Cdk9 T-loop. J Biol Chem *283*, 33578-33584.

Wang, Y., Liu, X.Y., and De Clercq, E. (2009). Role of the HIV-1 positive elongation factor P-TEFb and inhibitors thereof. Mini Rev Med Chem *9*, 379-385.

Warda, A.S., Kretschmer, J., Hackert, P., Lenz, C., Urlaub, H., Hobartner, C., Sloan, K.E., and Bohnsack, M.T. (2017). Human METTL16 is a N(6)-methyladenosine (m(6)A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. EMBO Rep *18*, 2004-2014.

Wassarman, D.A., and Steitz, J.A. (1991). Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. Mol Cell Biol *11*, 3432-3445.

Weber, C.M., Henikoff, J.G., and Henikoff, S. (2010). H2A.Z nucleosomes enriched over active genes are homotypic. Nat Struct Mol Biol *17*, 1500-1507.

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Mol Cell *53*, 819-830.

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

Weingarten-Gabbay, S., Nir, R., Lubliner, S., Sharon, E., Kalma, Y., Weinberger, A., and Segal, E. (2019). Systematic interrogation of human promoters. Genome Res *29*, 171-183.

Werner-Allen, J.W., Lee, C.J., Liu, P., Nicely, N.I., Wang, S., Greenleaf, A.L., and Zhou, P. (2011). cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. J Biol Chem *286*, 5717-5726.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522-525.

West, S., and Proudfoot, N.J. (2009). Transcriptional termination enhances protein expression in human cells. Mol Cell *33*, 354-364.

Westerling, T., Kuuluvainen, E., and Makela, T.P. (2007). Cdk8 is essential for preimplantation mouse development. Mol Cell Biol *27*, 6177-6182.

Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., and Adelman, K. (2015). Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol Cell *58*, 311-322.

Williamson, L., Saponaro, M., Boeing, S., East, P., Mitter, R., Kantidakis, T., Kelly, G.P., Lobley, A., Walker, J., Spencer-Dene, B*., et al.* (2017). UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. Cell *168*, 843-855 e813.

Wilson, M.D., Harreman, M., and Svejstrup, J.Q. (2013). Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. Biochim Biophys Acta *1829*, 151-157.

Wimmer, J., Fujinaga, K., Taube, R., Cujec, T.P., Zhu, Y., Peng, J., Price, D.H., and Peterlin, B.M. (1999). Interactions between Tat and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b. Virology *255*, 182-189.

Winter, G.E., Mayer, A., Buckley, D.L., Erb, M.A., Roderick, J.E., Vittori, S., Reyes, J.M., di Iulio, J., Souza, A., Ott, C.J.*, et al.* (2017). BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. Mol Cell *67*, 5-18 e19.

Wittmann, B.M., Fujinaga, K., Deng, H., Ogba, N., and Montano, M.M. (2005). The breast cell growth inhibitor, estrogen down regulated gene 1, modulates a novel functional interaction between estrogen receptor alpha and transcriptional elongation factor cyclin T1. Oncogene *24*, 5576-5588.

Wittmann, B.M., Wang, N., and Montano, M.M. (2003). Identification of a novel inhibitor of breast cell growth that is down-regulated by estrogens and decreased in breast tumors. Cancer Res *63*, 5151-5158.

Wittschieben, B.O., Iwai, S., and Wood, R.D. (2005). DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. J Biol Chem *280*, 39982-39989.

Wojcik, M.H., Linnea, K., Stoler, J.M., and Rappaport, L. (2019). Updating the neurodevelopmental profile of Alazami syndrome: Illustrating the role of developmental assessment in rare genetic disorders. Am J Med Genet A.

Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Mol Cell *54*, 601-612.

Wood, D.J., and Endicott, J.A. (2018). Structural insights into the functional diversity of the CDK-cyclin family. Open Biol *8*.

Wood, R.D. (1997). Nucleotide excision repair in mammalian cells. J Biol Chem *272*, 23465-23468.

Wu, L., Li, L., Zhou, B., Qin, Z., and Dou, Y. (2014). H2B ubiquitylation promotes RNA Pol II processivity via PAF1 and pTEFb. Mol Cell *54*, 920-931.

Wu, L., Zee, B.M., Wang, Y., Garcia, B.A., and Dou, Y. (2011). The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation. Mol Cell *43*, 132-144.

Xiang, K., Tong, L., and Manley, J.L. (2014). Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. Mol Cell Biol *34*, 1894-1910.

Xu, J., Lahiri, I., Wang, W., Wier, A., Cianfrocco, M.A., Chong, J., Hare, A.A., Dervan, P.B., DiMaio, F., Leschziner, A.E.*, et al.* (2017). Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. Nature *551*, 653-657.

Xue, Y., Yang, Z., Chen, R., and Zhou, Q. (2010). A capping-independent function of MePCE in stabilizing 7SK snRNA and facilitating the assembly of 7SK snRNP. Nucleic Acids Res *38*, 360-369.

Yadav, D., Ghosh, K., Basu, S., Roeder, R.G., and Biswas, D. (2019). Multivalent Role of Human TFIID in Recruiting Elongation Components at the Promoter-Proximal Region for Transcriptional Control. Cell Rep *26*, 1303-1317 e1307.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell *21*, 227-237.

Yamaguchi, T., Nagao, S., Kasahara, M., Takahashi, H., and Grantham, J.J. (1997). Renal accumulation and excretion of cyclic adenosine monophosphate in a murine model of slowly progressive polycystic kidney disease. Am J Kidney Dis *30*, 703-709.

Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., and Handa, H. (2002). Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol *22*, 2918-2927.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41-51.

Yang, C., and Stiller, J.W. (2014). Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain. Proc Natl Acad Sci U S A *111*, 5920-5925.

Yang, X.C., Sullivan, K.D., Marzluff, W.F., and 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.

Yang, Y., Eichhorn, C.D., Wang, Y., Cascio, D., and Feigon, J. (2019). Structural basis of 7SK RNA 5'-gamma-phosphate methylation and retention by MePCE. Nat Chem Biol *15*, 132-140.

Yang, Z., He, N., and Zhou, Q. (2008). Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. Mol Cell Biol *28*, 967-976.

Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol Cell *19*, 535-545.

Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature *414*, 317-322.

Yankulov, K., and Bentley, D. (1998). Transcriptional control: Tat cofactors and transcriptional elongation. Curr Biol *8*, R447-449.

Yazbeck, A.M., Tout, K.R., and Stadler, P.F. (2018). Detailed secondary structure models of invertebrate 7SK RNAs. RNA Biol *15*, 158-164.

Yeh, I.J., Song, K., Wittmann, B.M., Bai, X., Danielpour, D., and Montano, M.M. (2014). HEXIM1 plays a critical role in the inhibition of the androgen receptor by anti-androgens. Biochem J *462*, 315-327.

Yik, J.H., Chen, R., Nishimura, R., Jennings, J.L., Link, A.J., and Zhou, Q. (2003). Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. Mol Cell *12*, 971-982.

Yik, J.H., Chen, R., Pezda, A.C., and Zhou, Q. (2005). Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription. J Biol Chem *280*, 16368-16376.

Yik, J.H.N., Chen, R.C., Pezda, A.C., Samford, C.S., and Zhou, Q. (2004). A human immunodeficiency virus type 1 tat-like arginine-rich RNA-binding domain is essential for HEXIM1 to inhibit RNA

polymerase II transcription through 7SK snRNA-Mediated inactivation of P-TEFb. Molecular and Cellular Biology *24*, 5094-5105.

Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes Dev *21*, 160-174.

Yoshikawa, N., Shimizu, N., Sano, M., Ohnuma, K., Iwata, S., Hosono, O., Fukuda, K., Morimoto, C., and Tanaka, H. (2008). Role of the hinge region of glucocorticoid receptor for HEXIM1-mediated transcriptional repression. Biochem Biophys Res Commun *371*, 44-49.

Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., and Roeder, R.G. (2015). RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science *350*, 1383-1386.

Yurko, N.M., and Manley, J.L. (2018). The RNA polymerase II CTD "orphan" residues: Emerging insights into the functions of Tyr-1, Thr-4, and Ser-7. Transcription *9*, 30-40.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev *25*, 2227-2241.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet *39*, 1512-1516.

Zhang, H., Park, S.H., Pantazides, B.G., Karpiuk, O., Warren, M.D., Hardy, C.W., Duong, D.M., Park, S.J., Kim, H.S., Vassilopoulos, A.*, et al.* (2013). SIRT2 directs the replication stress response through CDK9 deacetylation. Proc Natl Acad Sci U S A *110*, 13546-13551.

Zhao, D.Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F.W., Zhong, G., Liu, K., Li, W., Moffat, J.*, et al.* (2016a). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature *529*, 48-53.

Zhao, Y., Karijolich, J., Glaunsinger, B., and Zhou, Q. (2016b). Pseudouridylation of 7SK snRNA promotes 7SK snRNP formation to suppress HIV-1 transcription and escape from latency. EMBO Rep *17*, 1441-1451.

Zheng, S., Crickard, J.B., Srikanth, A., and Reese, J.C. (2014a). A highly conserved region within H2B is important for FACT to act on nucleosomes. Mol Cell Biol *34*, 303-314.

Zheng, Y., Tipton, J.D., Thomas, P.M., Kelleher, N.L., and Sweet, S.M. (2014b). Site-specific human histone H3 methylation stability: fast K4me3 turnover. Proteomics *14*, 2190-2199.

Zheng, Z., Jia, J.L., Bou, G., Hu, L.L., Wang, Z.D., Shen, X.H., Shan, Z.Y., Shen, J.L., Liu, Z.H., and Lei, L. (2012). rRNA genes are not fully activated in mouse somatic cell nuclear transfer embryos. J Biol Chem *287*, 19949-19960.

Zhou, K., Kuo, W.H., Fillingham, J., and Greenblatt, J.F. (2009). Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. Proc Natl Acad Sci U S A *106*, 6956-6961.

Zhou, M., Lu, H., Park, H., Wilson-Chiru, J., Linton, R., and Brady, J.N. (2006). Tax interacts with P-TEFb in a novel manner to stimulate human T-lymphotropic virus type 1 transcription. J Virol *80*, 4781-4791.

Zhou, Q., Li, T., and Price, D.H. (2012). RNA Polymerase II Elongation Control. Annu Rev Biochem *81*, 119-143.

Zhou, Q., and Yik, J.H. (2006). The Yin and Yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. Microbiol Mol Biol Rev *70*, 646-659.

Zhu, B., Mandal, S.S., Pham, A.D., Zheng, Y., Erdjument-Bromage, H., Batra, S.K., Tempst, P., and Reinberg, D. (2005). The human PAF complex coordinates transcription with events downstream of RNA synthesis. Genes Dev *19*, 1668-1673.

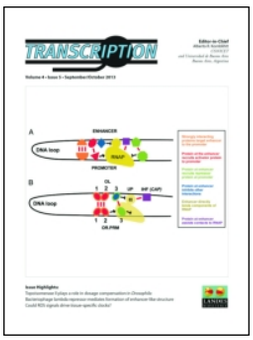
Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M.B., and Price, D.H. (1997). Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. Genes Dev *11*, 2622-2632.

Zieve, G., and Penman, S. (1976). Small RNA species of the HeLa cell: metabolism and subcellular localization. Cell *8*, 19-31.

Zieve, G.W. (1981). Two groups of small stable RNAs. Cell *25*, 296-297.

Zumer, K., Plemenitas, A., Saksela, K., and Peterlin, B.M. (2011). Patient mutation in AIRE disrupts P-TEFb binding and target gene transcription. Nucleic Acids Res *39*, 7908-7919.

Annexe



Transcription

ISSN: 2154-1264 (Print) 2154-1272 (Online) Journal homepage: https://www.tandfonline.com/loi/ktrn20

7SK small nuclear RNA, a multifunctional transcriptional regulatory RNA with gene-specific features

Sylvain Egloff, Cécilia Studniarek & Tamás Kiss

To cite this article: Sylvain Egloff, Cécilia Studniarek & Tamás Kiss (2018) 7SK small nuclear RNA, a multifunctional transcriptional regulatory RNA with gene-specific features, Transcription, 9:2, 95-101, DOI: 10.1080/21541264.2017.1344346

To link to this article: https://doi.org/10.1080/21541264.2017.1344346

Accepted author version posted online: 18 Aug 2017.
Published online: 04 Oct 2017.

Submit your article to this journal

Article views: 264

View Crossmark data

Citing articles: 4 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ktrn20

TRANSCRIPTION
2018, VOL. 9, NO. 2, 95–101
https://doi.org/10.1080/21541264.2017.1344346

Taylor & Francis
Taylor & Francis Group

POINT-OF-VIEW

Check for updates

7SK small nuclear RNA, a multifunctional transcriptional regulatory RNA with gene-specific features

Sylvain Egloff[a], Cécilia Studniarek[a], and Tamás Kiss[a,b]

[a]Laboratoire de Biologie Moléculaire Eucaryote du CNRS, UMR5099, Centre de Biologie Intégrative, Université Paul Sabatier, Toulouse, France; [b]Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary

ABSTRACT
The 7SK small nuclear RNA is a multifunctional transcriptional regulatory RNA that controls the nuclear activity of the positive transcription elongation factor b (P-TEFb), specifically targets P-TEFb to the promoter regions of selected protein-coding genes and promotes transcription of RNA polymerase II-specific spliceosomal small nuclear RNA genes.

ARTICLE HISTORY
Received 19 May 2017
Revised 14 June 2017
Accepted 14 June 2017

KEYWORDS
7SK snRNA; 7SK snRNP P-TEFb; RNA polymerase II pausing; transcription regulation

Introduction

Human 7SK RNA is an abundant 331-nucleotide-long, evolutionarily conserved small nuclear RNA (snRNA) that, in the form of 7SK small nuclear RNP (snRNP), regulates RNA polymerase II (RNAPII) transcription through controlling the activity of a major transcription elongation factor, P-TEFb. The 7SK snRNA is an RNAPIII transcript that undergoes post-transcriptional modifications. The gamma phosphate of the 5′-terminal triphosphate of the nascent 7SK RNA is methylated by the methyl phosphate capping enzyme (MePCE) and the U250 residue is converted into pseudouridine.[1] The mature 7SK snRNA folds into 4 major hairpin domains which provide docking sites for 7SK snRNP core proteins and several other transiently interacting 7SK snRNP proteins. The classical 7SK structure has been proposed to undergo conformational rearrangements to support transient association with various 7SK snRNP proteins.[2-4]

The 7SK snRNA is incorporated into the functionally active 7SK core snRNP that also contains the capping enzyme MePCE and the La-related protein 7 (Larp7).[3,5] While MePCE binds to the 5′-terminal region of the mature 7SK snRNA, Larp7 binds to its 3′-terminal hairpin and the following oligouridylate tail. Both MePCE and Larp7 are fundamental to the accumulation and stability of 7SK snRNA. Besides binding to 7SK, MePCE and Larp7 also interact with each other to form the metabolically stable 7SK core snRNP.

The 7SK core snRNP provides a scaffold for dynamic assembly of structurally and functionally distinct larger 7SK snRNP particles. The most abundant and most extensively characterized 7SK particle contains the 7SK core snRNP, a hetero- or homodimer of HEXIM1/2 and the RNAPII transcription elongation factor, P-TEFb. P-TEFb is a protein kinase composed of cyclin-dependent kinase 9 (Cdk9) and its regulatory cyclin partner CycT1 or T2. The 7SK/HEXIM/P-TEFb snRNP functions as a negative transcriptional regulatory RNP in which the kinase activity of P-TEFb is inhibited.[6] In response to various transcriptional stimuli, however, active P-TEFb can be rapidly mobilized from the transcriptionally inactive 7SK/HEXIM/P-TEFb. According to a recently emerging view, instead of simply serving as a nucleoplasmic P-TEFb reservoir, the 7SK/HEXIM/P-TEFb snRNP can directly target inactive P-TEFb to the promoter regions of protein-coding genes to promote “on site” P-TEFb mobilization.[5]

Recently, 2 novel, functionally distinct 7SK-containing transcriptional regulatory complexes have

CONTACT Tamás Kiss tamas@ibcg.biotoul.fr LBME/CNRS, Centre de Biologie Intégrative, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 09, France; Sylvain Egloff sylvain.egloff@ibcg.biotoul.fr LBME/CNRS, Centre de Biologie Intégrative, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 09, France.
© 2017 Taylor & Francis

been identified.[7,8] No evidence supports the presence of P-TEFb in these new complexes, suggesting that 7SK can also regulate RNAPII transcription through P-TEFb-independent mechanisms. Here, we discuss the recent advances on the well-established P-TEFb-dependent and the newly reported P-TEFb-independent mechanisms of transcriptional regulation by 7SK snRNP.

7SK is a master regulator of nuclear P-TEFb activity

Since P-TEFb is a major regulator of RNAPII transcription, controlling its nuclear activity is of particular importance. P-TEFb controls the conversion of transcriptionally engaged but promoter proximally paused RNAPII into elongation-competent polymerase.[5,6] Shortly after transcription initiation, the early elongating RNAPII is arrested by binding of negative transcription elongation factors, NELF and DSIF. Through phosphorylation of NELF, DSIF and the C-terminal domain (CTD) of RNAPIl at serine 2, P-TEFb allows RNAPII elongation to resume. During the past decade, P-TEFb-mediated release of paused RNAPII has been recognized as a major pervasive checkpoint of RNAPII transcription that controls expression of most RNAPII-transcribed genes with special importance in controlling stimulus-induced and developmentally-regulated genes. The nuclear activity of P-TEFb is regulated mainly by the 7SK snRNP and HEXIM1/2. Mechanistically, a hetero- or homo-dimer of HEXIM1/2 binds to conserved motifs of the long 5′-terminal hairpin of 7SK snRNA. Docking of 7SK triggers a conformational change that unveils the otherwise inaccessible P-TEFb binding surface of HEXIM proteins. The 7SK-associated "activated" HEXIM proteins bind to both subunits of P-TEFb and mask the catalytic site of Cdk9.[3,6,9]

The active and inactive forms of P-TEFb coexist in the nucleus (Fig. 1). While the active pool of P-TEFb seems to be sufficient to support basal transcription, P-TEFb rapidly mobilized from the inactive 7SK/HEXIM/P-TEFb reservoir can reinforce the increased transcriptional demand of the cell. In response to cellular stress or transcriptional blockade, the 7SK/HEXIM/P-TEFb RNP rapidly disassembles and releases ready-to-use P-TEFb, whereas the free 7SK core snRNP associates with a set of abundant heterogeneous ribonucleoproteins (hnRNPs) A1, A2/B1, R, Q and RNA helicase A.[10,11] The hnRNP proteins presumably 'capture' the newly released 7SK snRNP to prevent its re-association with HEXIM and P-TEFb. Recent RNA structural analysis revealed that stress-dependent remodeling of the 7SK snRNP is accompanied by structural rearrangements of the 7SK snRNA.[4] In the 7SK/HEXIM/P-TEFb snRNP, the 5′-terminal region of 7SK snRNA folds into the classical extended hairpin structure that supports HEXIM and P-TEFb binding. Upon 7SK/HEXIM/P-TEFb disassembly, 7SK adopts a closed conformation in which the 5′- and 3′-terminal domains are folded close to each

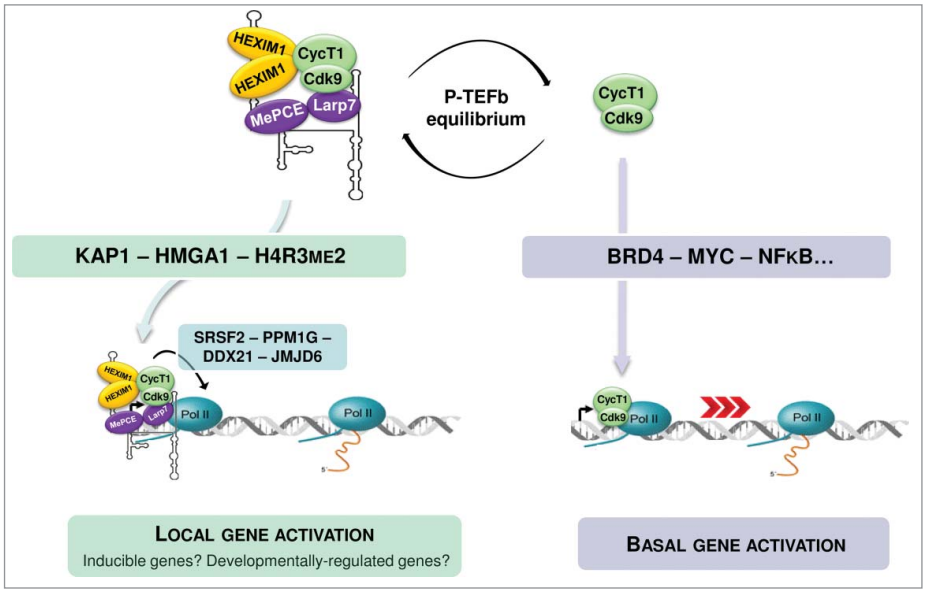


Figure 1. Both active and inactive forms of P-TEFb are recruited to their target gene promoters through association with specific transcription factors.

other. RNA helicase A may promote the structural rearrangements of 7SK and the associated hnRNP proteins may stabilize the closed conformation of 7SK snRNA that prevents HEXIM and P-TEFb binding.[11]

Diverse stimuli and signaling cascades, including the PI3K/Akt and PP1α/PP2B pathways, have been shown to induce 7SK/HEXIM/P-TEFb dissociation through addition or erasure of specific post-translational modifications on the components of the 7SK/HEXIM/P-TEFb snRNP.[5] Whether the released P-TEFb stimulates transcription of different sets of genes depending on the activated cellular pathway remains unknown. Once released from 7SK/HEXIM/P-TEFb, the free P-TEFb has to find its way to the target genes. A number of transcription factors and co-activators, including NF-κB, c-Myc, CIITA, p53, the bromodomain protein 4 (Brd4), the super elongation complexes (SECs) and the human immunodeficiency virus (HIV) transactivator Tat can directly bind and target P-TEFb to paused RNAPII.[5,6] Brd4 seems to be the major nuclear binding partner of active P-TEFb that directs it to the promoters of many primary response genes through interacting with acetylated chromatin.[12] Releasing of P-TEFb from the 7SK/HEXIM/P-TEFb snRNP concomitantly augments Brd4/P-TEFb level, suggesting that Brd4 directly captures 7SK-evicted P-TEFb for transcription stimulation.[13] More recently, P-TEFb has been identified as a constitutive component of SECs which are composed of the ELL, AFF and ENL/AF9 proteins.[14] SECs have been proposed to deliver and tether P-TEFb to promoters of protein-coding genes through association with Mediator and Integrator complexes.[15-17] It remains to be elucidated how P-TEFb is distributed among Brd4, SECs and other transcription factors under normal and stress conditions. Moreover, recent studies demonstrated that the 7SK-associated inactive form of P-TEFb can also be targeted to genes, adding a new layer to this intricate regulatory network.

The 7SK/HEXIM/P-TEFb snRNP is targeted to the chromatin

The nuclear 7SK snRNP, mostly in the form of 7SK/HEXIM/P-TEFb, accumulates mainly in the soluble fraction of the nucleoplasm. Cell fractionation experiments, however, also detected a significant fraction of 7SK snRNA being associated with chromatin.[18] This raised the possibility that the 7SK/HEXIM/P-TEFb negative transcriptional regulatory snRNP may directly interact with P-TEFb-dependent genes. The first support to this idea was provided by detection of HEXIM1, Cdk9 and Larp7 on the HIV long-terminal repeat (LTR) promoter.[19] It was proposed that tethering 7SK/HEXIM/P-TEFb to the HIV LTR promoter can promote RNAPII transcription activation through "on site and on time" release of active P-TEFb. Chromatin-proximal P-TEFb activation could rapidly increase P-TEFb concentration at the promoter regions of target genes to facilitate P-TEFb capturing by gene-specific transcription factors. Next, dissociation of chromatin-bound 7SK/HEXIM/P-TEFb in response to specific cellular stimuli was proposed to induce rapid transcriptional activation of primary response genes which are heavily loaded with promoter proximally paused RNAPII.[20] More recently, genome-wide studies suggested that chromatin-tethering of 7SK-repressed P-TEFb is a common mechanism controlling transcription elongation on most protein-coding genes.[21,22] Chromatin immunoprecipitation and deep sequencing (ChIP-seq) experiments detected Larp7, Cdk9 and HEXIM1 co-enrichments at the promoter regions of more than 15,000 human genes.[22] Likewise, chromatin isolation by RNA purification followed by deep sequencing (ChIRP-seq) analysis detected 7SK snRNA occupancy on the transcribed regions of a large set of human genes, largely resembling RNAPII distribution.[8] Widespread 7SK/HEXIM/P-TEFb anchoring to chromatin was proposed to be mediated by the Kruppel-associated box-interacting protein KAP1 that has been shown to facilitate both P-TEFb recruitment and RNAPII elongation on early-response genes upon stimulation. Moreover, KAP1 shows significant co-occupancy with components of the 7SK snRNP on most protein-coding genes controlled by RNAPII pausing.[22] Other mechanisms targeting 7SK/HEXIM/P-TEFb to various subsets of RNAPII-transcribed genes may also exist. The CTIP2 transcriptional repressor can directly interact with both HEXIM1 and 7SK snRNA to position 7SK/HEXIM/P-TEFb on a specific set of human promoters through forming a direct interaction with HMGA1.[23] Finally, the 7SK snRNA has been reported to directly bind to a specific chromatin signature, the repressive H4R3me2 mark, at a class of enhancers also occupied by Brd4 and JMJD6.[24] This interaction could tether the 7SK/HEXIM/P-TEFb snRNP to specific enhancers and allow local P-TEFb activation upon

removal of the H4R3me2 mark by the JMJD6 demethylase. On the HIV LTR promoter and primary response genes, P-TEFb release from 7SK/HEXIM/P-TEFb is triggered by recruitment of the PPM1G phosphatase that disassembles the 7SK/HEXIM/P-TEFb snRNP through dephosphorylation of the Cdk9 T-loop at Thr186. In this case, release of RNAPII pausing would require *de novo* re-phosphorylation of the Cdk9 T-loop, since this phosphorylation is absolutely required for Cdk9 kinase activity. RNA binding proteins have been also implicated in chromatin-associated P-TEFb activation. SRSF2, a serine/arginine-rich splicing factor, was found to be associated with 7SK/HEXIM/P-TEFb on several target promoters.[21] Emergence of SRSF2 binding sites on the nascent RNA transcripts captures SRSF2 and promotes 7SK/HEXIM/P-TEFb dissociation to trigger RNAPII pause release. The DDX21 DEAD-box RNA helicase has been also proposed to stimulate RNAPII transcription through promoting 'on site' disassembly of chromatin-associated 7SK/HEXIM/P-TEFb through remodeling of the 7SK secondary structure by its helicase activity.[25]

The above studies have described multiple mechanisms to position and disassemble the 7SK/HEXIM/P-TEFb snRNP directly on the chromatin encompassing protein genes. Now, a major question of the field is how widespread these regulatory mechanisms are? A great fraction of P-TEFb exists in its active free form in the nucleoplasm, and some transcription factors, including Brd4 and c-myc, can directly recruit and use active P-TEFb to stimulate global transcription elongation by RNAPII (Fig. 1). It is unclear why other transcription activators cannot directly capture and utilize transcriptionally active nucleoplasmic P-TEFb. While coordinated 'on site' P-TEFb activation can represent an apparent advantage for orchestrated activation of stimulus-inducible and developmentally-regulated genes, the benefit of such P-TEFb activation strategy is less obvious for ubiquitously-expressed genes. Nevertheless, if 7SK/HEXIM/P-TEFb delivers P-TEFb to a great fraction of genomic promoters, the loss of 7SK snRNP should lead to important transcriptional defects on RNAPII-transcribed genes. However, depletion of 7SK snRNP by RNA interference had only a limited impact on the expression of a subset of P-TEFb-dependent genes.[7] Another genome-wide study found that 7SK depletion in mouse embryonic stem cells alters expression of less than 500 genes which are mainly upregulated, pointing to a gene-specific repressor function of 7SK.[26] Further investigations are therefore required to determine whether anchoring 7SK/HEXIM/P-TEFb to chromatin is a pre-requisite for basal transcription by RNAPII or it regulates only specific cohorts of genes.

7SK snRNA regulates RNAPII transcription in a P-TEFb-independent manner

In addition to protein-coding genes, RNAPII also transcribes genes encoding small nuclear (sn) and small nucleolar (sno) RNAs. The RNAPII-specific snRNA and snoRNA genes possess specific structural and functional features distinct from protein-coding genes.[27] They display specific promoter and 3′ end processing elements which are recognized by snRNA gene-specific factors. After transcription initiation on sn/snoRNA genes, RNAPII elongation is not interrupted by promoter-proximal pausing and does not require P-TEFb activity.[28] We have recently determined the genome-wide chromatin occupancy of the 7SK snRNP protein Larp7 in human HeLa cells.[7] We detected marked Larp7 enrichments on RNAPII-transcribed sn/snoRNA genes. Other components of the 7SK core snRNP, MePCE and 7SK snRNA, were also detected on U2 snRNA genes, but not HEXIM1. We found that the 7SK snRNP interacts with an sn/snoRNA gene-specific transcription factor, the Little Elongation Complex (LEC) composed of Ice1, Ice2, ELL and ZC3H8.[29] LEC promotes RNAPII recruitment to the promoter of sn/snoRNA genes, and stimulates sn/snoRNA transcription through traveling with the elongating polymerase.[29,30]

The 7SK/LEC complex is devoid of P-TEFb, indicating that it is structurally and functionally distinct from the canonical 7SK/HEXIM/P-TEFb transcriptional regulatory snRNP (Fig. 2). Depletion of 7SK snRNP decreased both LEC and RNAPII occupancy on snRNA genes and reduced sn/snoRNA synthesis, but had no significant effect on the expression of protein-coding genes including P-TEFb-dependent ones. Interestingly, depletion of the LEC scaffold protein Ice1 resulted in similar transcriptional defects, since it also compromised RNAPII recruitment to and transcription of sn/snoRNA genes.[7,30] One intriguing possibility is that on the analogy of directing 7SK/HEXIM/P-TEFb assembly by HEXIM bound to 7SK, binding of Ice1 to the 7SK snRNA is a prerequisite for

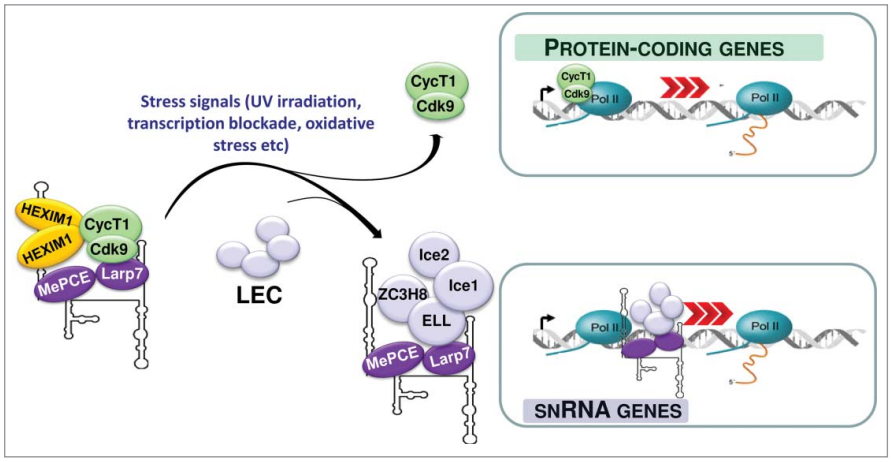


Figure 2. The 7SK/HEXIM/P-TEFb snRNP controls transcription of both protein-coding and RNAPII-specific spliceosomal snRNA genes. In response to transcriptional stress, P-TEFb and 7SK core snRNP are released from the 7SK/HEXIM/P-TEFb snRNP. While P-TEFb stimulates transcription of protein-coding genes through releasing promoter-proximal paused RNAPII, a fraction of the free 7SK core snRNP associates with the Little Elongation Complex (LEC) to facilitate snRNA genes transcription.

7SK/LEC assembly, although Ice1 contains no obvious RNA-binding domain. The mechanism targeting the 7SK/LEC snRNP to sn/snoRNA genes remains elusive. Most likely, the pre-initiation complex formed on the RNAPII-specific sn/snoRNA gene promoters specifically interacts with 7SK/LEC. Lending support to this idea, the 7SK snRNP preferentially interacts with RNAPII hyperphosphorylated at Ser5 and Ser7 that is a hallmark of RNAPII engaged in snRNA gene transcription. For instance, the snRNA-specific transcription factor SNAPc could be a possible candidate for directing 7SK/LEC recruitment to sn/snoRNA genes. The detected interaction between the LEC and MED26 suggests that the Mediator Complex might also participate in LEC recruitment, although MED26 is not specific to snRNA gene promoters.[31]

We observed strong Larp7 accumulation along the entire sn/snoRNA genes that likely reflects stable 7SK/LEC association with the transcribing RNAPII. Supporting this notion, highly similar distribution profiles were obtained for Larp7, RNAPII and LEC on sn/snoRNA genes. In contrast to RNAPII-specific sn/snoRNA genes, we failed to detect significant Larp7 association with most protein-coding genes,[7] suggesting that 7SK/HEXIM/P-TEFb association with protein-coding genes is highly transient and that fast P-TEFb activation on the chromatin may result in quick 7SK snRNP release.

Importantly, 7SK association with LEC is largely enhanced under stress conditions, such as transcriptional blockade or UV irradiation that are known to trigger P-TEFb eviction from 7SK/HEXIM/P-TEFb.[7] Stress-induced dissociation of 7SK/HEXIM/P-TEFb might transiently increase the cellular level of free 7SK snRNP available for 7SK/LEC formation. Thus, 7SK/HEXIM/P-TEFb dissociation can simultaneously augment accumulation of active P-TEFb and 7SK/LEC snRNP to synchronously boost transcription of protein-coding and snRNA genes, respectively (Fig. 2). This means that the 7SK/HEXIM/P-TEFb snRNP could be considered as a bifunctional transcriptional regulatory RNP that can coordinate pre-mRNA transcription with spliceosomal snRNA expression.

Recently, the 7SK snRNA has been also detected at active enhancers and super-enhancers that are genomic regions encompassing clustered enhancer elements controlling gene expression networks.[8] At super-enhancers, neither HEXIM1 nor P-TEFb were detected, supporting the idea that the 7SK snRNP can impact RNAPII transcription independently from P-TEFb regulation. The 7SK snRNA was shown to repress RNAPII-mediated transcription of enhancer RNAs (eRNAs) through facilitating recruitment of the BAF chromatin-remodeling complex, a known repressor of eRNA transcription. 7SK/BAF-promoted reshaping of chromatin structure at super-enhancers would restrict convergent transcription and prevent DNA damage. Of note, while 7SK/HEXIM/P-TEFb incorporates a large fraction of cellular 7SK snRNP, the 7SK/LEC and 7SK/BAF snRNPs are much less abundant in human cells. With development of sensitive RNA-based purification techniques such as ChIRP-MS, one can predict that additional low-abundance 7SK-containing complexes will be uncovered in the future.

Conclusion

Distinct 7SK core snRNP-containing particles associate with the chromatin and regulate multiple aspects of nuclear RNAPII transcription. The 7SK/BAF axis controls production of eRNAs, the 7SK/HEXIM/P-TEFb snRNP regulates RNAPII pause release and/or elongation on protein-coding genes and finally, the 7SK/LEC complex promotes transcription of sn/snoRNA genes. Moreover, recent findings suggest that 7SK also participates in proper transcription termination on a subset of genes.[26] Therefore, the 7SK snRNP plays a complex role in orchestration of global RNAPII progression along the nuclear genome. Future efforts will be required to decipher the precise mechanisms by which 7SK is able to fine-tune RNAPII transcriptional activity.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgments

We are indebted to our laboratory members P. Vitali and B. Jády for helpful discussions.

Funding

Our work was supported by grants from la Fondation ARC pour la recherche sur le cancer (to SE), l'Agence Nationale de la Recherche (to T.K.).

References

[1] Zhao Y, Karijolich J, Glaunsinger B, Zhou Q. Pseudouridylation of 7SK snRNA promotes 7SK snRNP formation to suppress HIV-1 transcription and escape from latency. EMBO Rep 2016; 17:1441-51; PMID:27558685; https://doi.org/10.15252/embr.201642682

[2] Marz M, Donath A, Verstraete N, Nguyen VT, Stadler PF, Bensaude O. Evolution of 7SK RNA and its protein partners in metazoa. Mol Biol Evol 2009; 26:2821-30; PMID:19734296; https://doi.org/10.1093/molbev/msp198

[3] Peterlin BM, Brogie JE, Price DH. 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. Wiley Interdiscip Rev 2012; 3:92-103; PMID:21853533; https://doi.org/10.1002/wrna.106

[4] Brogie JE, Price DH. Reconstitution of a functional 7SK snRNP. Nucleic Acids Res 2017; 45:6864-6880; PMID:28431135; https://doi.org/10.1093/nar/gkx262

[5] Quaresma AJ, Bugai A, Barboric M. Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. Nucleic Acids Res 2016; 44:7527-39; PMID:27369380; https://doi.org/10.1093/nar/gkw585

[6] Zhou Q, Li T, Price DH. RNA Polymerase II Elongation Control. Annu Rev Biochem 2012; 81:119-43; PMID:22404626; https://doi.org/10.1146/annurev-biochem-052610-095910

[7] Egloff S, Vitali P, Tellier M, Raffel R, Murphy S, Kiss T. The 7SK snRNP associates with the little elongation complex to promote snRNA gene transcription. EMBO J 2017; 36:934-948; PMID:28254838; https://doi.org/10.15252/embj.201695740

[8] Flynn RA, Do BT, Rubin AJ, Calo E, Lee B, Kuchelmeister H, Rale M, Chu C, Kool ET, Wysocka J, et al. 7SK-BAF axis controls pervasive transcription at enhancers. Nat Struct Mol Biol 2016; 23:231-8; https://doi.org/10.1038/nsmb.3176

[9] Kobbi L, Demey-Thomas E, Braye F, Proux F, Kolesnikova O, Vinh J, Poterszman A, Bensaude O. An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb. Proceedings of the National Academy of Sciences of the United States of America 2016; 113:12721-12726; PMID:27791144; https://doi.org/10.1073/pnas.1612331113

[10] Barrandon C, Bonnet F, Nguyen VT, Labas V, Bensaude O. The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. Mol Cell Biol 2007; 27:6996-7006; PMID:17709395; https://doi.org/10.1128/MCB.00975-07

[11] Van Herreweghe E, Egloff S, Goiffon I, Jady BE, Froment C, Monsarrat B, Kiss T. Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. EMBO J 2007; 26:3570-80; PMID:17611602; https://doi.org/10.1038/sj.emboj.7601783

[12] Jang MK, Mochizuki K, Zhou M, Jeong HS, Brady JN, Ozato K. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell 2005; 19:523-34; PMID:16109376; https://doi.org/10.1016/j.molcel.2005.06.027

[13] Yang Z, Yik JH, Chen R, He N, Jang MK, Ozato K, Zhou Q. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Molecular Cell 2005; 19:535-45; PMID:16109377; https://doi.org/10.1016/j.molcel.2005.06.029

[14] Smith E, Lin C, Shilatifard A. The super elongation complex (SEC) and MLL in development and disease. Genes Dev 2011; 25:661-72; https://doi.org/10.1101/gad.2015411

[15] Gardini A, Baillat D, Cesaroni M, Hu D, Marinis JM, Wagner EJ, Lazar MA, Shilatifard A, Shiekhattar R. Integrator regulates transcriptional initiation and pause release following activation. Mol Cell 2014; 56:128-39; PMID:25201415; https://doi.org/10.1016/j.molcel.2014.08.004

[16] Lu X, Zhu X, Li Y, Liu M, Yu B, Wang Y, Rao M, Yang H, Zhou K, Wang Y, et al. Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. Nucleic Acids Res 2016; 44:6853-67; PMID:27353326; https://doi.org/10.1093/nar/gkw571

[17] Takahashi H, Parmely TJ, Sato S, Tomomori-Sato C, Banks CA, Kong SE, Szutorisz H, Swanson SK, Martin-Brown S, Washburn MP, et al. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell 2011; 146:92-104; PMID:21729782; https://doi.org/10.1016/j.cell.2011.06.005

[18] Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C. Characterization of the RNA content of chromatin. Genome Res 2010; 20:899-907; PMID:20404130; https://doi.org/10.1101/gr.103473.109

[19] D'Orso I, Frankel AD. RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat Struct Mol Biol 2010; 17:815-21; https://doi.org/10.1038/nsmb.1827

[20] McNamara RP, McCann JL, Gudipaty SA, D'Orso I. Transcription factors mediate the enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for transcription elongation. Cell Rep 2013; 5:1256-68; PMID:24316072; https://doi.org/10.1016/j.celrep.2013.11.003

[21] Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, Xiao R, Burge CB, Fu XD. SR Proteins Collaborate with 7SK and Promoter-Associated Nascent RNA to Release Paused Polymerase. Cell 2013; 153:855-68; PMID:23663783; https://doi.org/10.1016/j.cell.2013.04.028

[22] McNamara RP, Reeder JE, McMillan EA, Bacon CW, McCann JL, D'Orso I. KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. Mol Cell 2016; 61:39-53; PMID:26725010; https://doi.org/10.1016/j.molcel.2015.11.004

[23] Eilebrecht S, Le Douce V, Riclet R, Targat B, Hallay H, Van Driessche B, Schwartz C, Robette G, Van Lint C, Rohr O, et al. HMGA1 recruits CTIP2-repressed P-TEFb to the HIV-1 and cellular target promoters. Nucleic Acids Res 2014; 42:4962-71; PMID:24623795; https://doi.org/10.1093/nar/gku168

[24] Liu W, Ma Q, Wong K, Li W, Ohgi K, Zhang J, Aggarwal AK, Rosenfeld MG. Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell 2013; 155:1581-95; PMID:24360279; https://doi.org/10.1016/j.cell.2013.10.056

[25] Calo E, Flynn RA, Martin L, Spitale RC, Chang HY, Wysocka J. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature 2015; 518:249-53; PMID:25470060; https://doi.org/10.1038/nature13923

[26] Castelo-Branco G, Amaral PP, Engstrom PG, Robson SC, Marques SC, Bertone P, Kouzarides T. The non-coding snRNA 7SK controls transcriptional termination, poising, and bidirectionality in embryonic stem cells. Genome Biol 2013; 14:R98; PMID:24044525; https://doi.org/10.1186/gb-2013-14-9-r98

[27] Egloff S, O'Reilly D, Murphy S. Expression of human snRNA genes from beginning to end. Biochem Soc Trans 2008; 36:590-4; PMID:18631122; https://doi.org/10.1042/BST0360590

[28] Medlin J, Scurry A, Taylor A, Zhang F, Peterlin BM, Murphy S. P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. EMBO J 2005; 24:4154-65; PMID:16308568; https://doi.org/10.1038/sj.emboj.7600876

[29] Smith ER, Lin C, Garrett AS, Thornton J, Mohaghegh N, Hu D, Jackson J, Saraf A, Swanson SK, Seidel C, et al. The little elongation complex regulates small nuclear RNA transcription. Mol Cell 2011; 44:954-65; PMID:22195968; https://doi.org/10.1016/j.molcel.2011.12.008

[30] Hu D, Smith ER, Garruss AS, Mohaghegh N, Varberg JM, Lin C, Jackson J, Gao X, Saraf A, Florens L, et al. The little elongation complex functions at initiation and elongation phases of snRNA gene transcription. Mol Cell 2013; 51:493-505; PMID:23932780; https://doi.org/10.1016/j.molcel.2013.07.003

[31] Takahashi H, Takigawa I, Watanabe M, Anwar D, Shibata M, Tomomori-Sato C, Sato S, Ranjan A, Seidel CW, Tsukiyama T, et al. MED26 regulates the transcription of snRNA genes through the recruitment of little elongation complex. Nat Commun 2015; 6:5941; PMID:25575120; https://doi.org/10.1038/ncomms6941

Abstract

The positive transcription elongation factor b (P-TEFb) stimulates transcription elongation through phosphorylation of negative elongation factors and the carboxyl-terminal domain (CTD) of RNA polymerase II (RNA Pol II). The cellular activity of P-TEFb is controlled by the 7SK small nuclear RNP (snRNP) composed of the 7SK snRNA, LARP7 and MePCE(Egloff et al., 2017a). Together with HEXIM, the 7SK snRNP dynamically sequesters a fraction of nuclear P-TEFb into catalytically inactive 7SK/HEXIM/P-TEFb snRNP. Outside of the 7SK snRNP, active P-TEFb is targeted to gene promoter regions to stimulate transcription elongation through its association with transcription factors and cofactors. Active P-TEFb is also rapidly released from 7SK/HEXIM/P-TEFb snRNP in response to various stress and physiological conditions which globally affect cell growth - including UV-induced DNA damage. Besides serving as a P-TEFb reservoir, the inactive 7SK/HEXIM/P-TEFb snRNP is selectively targeted to and loaded onto various gene promoters to allow 'on site' P-TEFb extraction and to support specific gene activation. Using human 7SK knock-out (KO) cell lines, we have assessed the functional consequences of 7SK snRNA loss on RNAPII transcription before and after UV-induced DNA damage. Mammalian Native Elongating Transcript-sequencing (mNET-seq) showed that under normal growth conditions the lack of 7SK snRNA has only subtle impacts on RNA Pol II transcription. However, upon UV irradiation, some key early-responsive genes which are readily induced in control cells show impaired transcriptional activation in 7SK KO cells. Since the levels of active P-TEFb in control and 7SK KO cells are highly comparable, we concluded that proper activation of these 7SK-dependent genes relies on the 7SK-associated inactive fraction of P-TEFb. Further experiments suggested that upon UV irradiation, P-TEFb is specifically extracted from the 7SK/P-TEFb snRNP and recruited to selected target genes as part of the Super Elongation Complex (SEC) in order to stimulate RNA Pol II entry into productive elongation and recruitment of RNA Pol II elongation factors. In conclusion, we have demonstrated that the 7SK snRNA is not required for cell viability under normal growth conditions but it is critical to orchestrate accurate transcriptional response upon UV-induced DNA damage.