UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

IGBMC – CNRS UMR 7104 – Inserm U 964

THESE présentée par :

Veronique FISCHER

soutenue le : **29. Septembre 2020**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Spécialité : Aspects moléculaires et cellulaires de la biologie

Functional study of the coactivator complexes SAGA and ATAC in mouse embryonic stem cells

THESE dirigée par :

M. László TORA	DR, IGBMC, Strasbourg, France

RAPPORTEURS :

Mrs. Mounia LAGHA	CR, IGMM, Montpellier, France
M. Pablo NAVARRO GIL	DR, Institut Pasteur, Paris, France

AUTRES MEMBRES DU JURY :

Mrs. Tineke LENSTRA	Group leader, Netherlands Cancer Institute, les Pays-Bas
M. Didier DEVYS	MCUPH, IGBMC et Université de Strasbourg, France
M. Patrick SCHULTZ	DR, IGBMC, Strasbourg, France

Acknowledgements

I would like to thank Dr. Mounia Lagha, Dr. Tineke Lenstra, Dr. Pablo Navarro Gil and Dr. Patrick Schultz for accepting to be members of my PhD thesis committee and for their time in reading and evaluating my thesis manuscript.

I want to thank my supervisor László Tora for giving me the opportunity to work in the exceptional scientific environment of the IGBMC for the past five years. Despite your dense and busy schedules, I could always discuss with you about my project and especially my results relating to the ATAC coactivator complex. I especially remember one of my first days as a master's student in your work group when you took the time to show and perform with me a western blot experiment thereby equipping me with all the tips and tricks for this technic that you had acquired over the years. Also, your advices in relation to the mass spectrometry experiments were crucial for their success. I also want to thank you for challenging and encouraging me and therefore bringing the best out of me.

I further would like to thank my bench neighbour and mentor Didier Devys who turned out to be a key figure in the advances of my work. Your advices and the discussions with you about the coactivator complex SAGA often resulted in ordering my thoughts and defining clear next steps. I have to admit that, at some point during my PhD, close to every discussion with you lead to the generation of a new cell line. This enabled us to acquire an overwhelming and unprecedented amount of data on the mammalian SAGA and ATAC complexes. I further admit that I respect you tremendously not only for your scientific skills and expertise, but especially also for the way you work with and motivate people of different characters. Also, as a bench neighbour, I will always remember our exchanges on various themes like the smallest dinosaur contained within an amber stone, which turned out to probably be a lizard (Xing et al., 2020; Li et al., 2020), or the physics behind unlacing of shoes (Daily-Diamond et al., 2017). Thank you for everything!

I also want to thank current and past members of the Tora team as well as colleagues and friends from IGMBC that made this past five years a pleasant experience and with whom I also shared enjoyable moments outside of the laboratory. I hope to see each one of you again in the future. I want to mention Matthieu and thank him for his help in generating a plethora of different mutant cell lines. I think at a given point we were very close in reaching an output comparable to an industrial production line with probably one new cell line every month. I want to thank especially Sascha whom I consider one of my best friends nowadays. During my first days as a master's student, you introduced me to the usage of laboratory equipment and machines especially also the cell culture rooms. You were always very patient in reexplaining where to find what. In general, I will always remember your calmness and patience. You also always had a way to cheer me up in times of frustration and setbacks. Thanks also goes to the generally cheerful and motivating members of the DUB club including Farrah, Kenny, Fang and Didier. Discussion rounds with you, frequently lasting for several hours, were always helpful and valuable.

Additionally, it was a pleasure to be able to speak German with Sascha and Kenny (the 'German' club) although while doing so I unfortunately lost a bit my Austrian accent 😉. Thanks also to Gizem, Pooja, Marta, Federica, Ivanka, Maryssa, Eli, Vincent, Bujamin, Giovanni, Pietro, Changwei, Paul, Nikolaos, and Stephane for your companionship and for being exceptional colleagues. I also want to mention Alix for enabling me to acquire first experiences in supervising and mentoring a master's student. You were an excellent student and I wish you all the best for your next steps.

I'm also grateful for the support of everyone outside of my scientific environment: My family, friends and partners that encouraged me when I was facing seemingly major hurdles or problems. I especially had the chance to be accompanied by one of the most extraordinary and probably wisest man I have ever met. Thank you, Gui!

Abstract

Development and cell differentiation rely on coordinated changes in gene expression programs. An important regulatory layer determining gene expression is thought to be the recruitment of RNA polymerase II (Pol II) to gene promoters through the assembly of the preinitiation complex (PIC). Several factors are involved in enabling PIC formation including chromatin modifying complexes such as the SAGA (Spt-Ada-Gcn5 Acetyltransferase) and the ATAC (Ada-Two-A-Containing) coactivator complexes. SAGA and ATAC share the same histone acetyltransferase (HAT) activity and deletion of genes encoding subunits of these two complexes lead to embryonic lethality in mouse. However, the importance and functions of coactivator complexes and their chromatin modifying activities in Pol II transcription is still poorly understood, as recently highlighted through two intriguing observations. First, studies in yeast indicated that the coactivator SAGA acts as a general cofactor for Pol II transcription in budding yeast unlike its previously assumed gene-specific functions (Bonnet et al., 2014; Baptista et al., 2017). Second, two other coactivator complexes were suggested to regulated Pol II transcription independently of their histone modifying activities in mouse embryonic stem cells (ESCs) (Dorighi et al., 2017; Acharya et al., 2017; Rickels et al., 2017).

To understand the roles of the SAGA and ATAC coactivator complexes for Pol II transcription in mammalian cells, we generate an unprecedented range of mouse ESC lines with inactivation for subunits of the two complexes. Surprisingly, we could observe a lethal phenotype when targeting subunits of the ATAC complex, which are believed to be required for its complex integrity. Therefore, we additionally generated mouse ESC lines which allow acute depletion of subunits of ATAC.

Using these mutant cell lines, we observed unexpected alterations of the self-renewal capacities of mouse ESCs upon inactivation or depletion of core subunits of SAGA and ATAC. This function of SAGA and ATAC for self-renewal of mouse ESCs however did not rely on their shared HAT activity. We could subsequently also reveal that the TBP-loading function of SAGA plays a major role in self-renewal of mouse ESCs. In contrast, the mechanism by which ATAC impacts the self-renewal capacities of mouse ESCs remains unclear. Through the analyses of newly synthesized RNA (4sU RNA-seq) in the mutant cell lines, we further found that SAGA and ATAC significantly regulate different genes with a potential mild global impact on Pol II transcription. In general, we could find little overlap between genes significantly regulated by SAGA or ATAC subunits, suggesting two different mechanisms by which these complexes affect the self-renewal capacities of mouse ESCs.

Overall, we found an important, HAT-independent role of the SAGA and ATAC coactivator complexes in the maintenance of mouse ESC self-renewal and growth. We also revealed that, although inactivation of the two complexes led to comparable phenotypes, SAGA and ATAC mainly regulate distinct genes.

Résumé

Introduction

Chez les eucaryotes, l'information génétique contenue dans l'ADN est compactée dans les noyaux en s'enroulant autour des protéines histones formant des nucléosomes, l'unité de base de la chromatine. La chromatine représente un obstacle majeur aux processus liés à l'ADN tels que la transcription par la Pol II (ARN polymérase II) des gènes codant pour des protéines. Les complexes coactivateurs comme les complexes SAGA (Spt-Ada-Gcn5 Acetyltransferase) et ATAC (Ada-Two-A-Containing) sont impliqués dans la stimulation de la transcription par la Pol II en servant d'intermédiaires entre les facteurs de transcription (activateurs) liés à des séquences spécifiques d'ADN et la machinerie de transcription basale.

La plupart des coactivateurs permettant l'accès de la machinerie de transcription basale à l'ADN en modifiant la structure de la chromatine. Les fonctions de modification de la chromatine des coactivateurs transcriptionnels comprennent des activités enzymatiques capables soit de modifier post-traductionnellement les protéines histones soit d'induire un remodelage du nucléosome. Les modifications post-traductionnelles d'histones, telles que l'acétylation, la méthylation ou la phosphorylation, peuvent modifier directement la compaction de la chromatine ou permettre le recrutement d'autres facteurs ayant des domaines reconnaissant spécifiquement ces histones modifiées. L'accès à la séquence d'ADN des promoteurs des gènes est crucial pour la formation du PIC (complexe de préinitiation). Le PIC est composé de la Pol II et de six GTF (facteurs de transcription généraux) et représente la première étape du cycle de transcription. Sa formation est initiée par la liaison de TBP (protéine de liaison à la boîte TATA) aux promoteurs des gènes.

Le complexe SAGA a été fortement conserve au cours de l'évolution, ayant une composition très similaire chez les levures et les mammifères. SAGA possède deux enzymes de modification des histones, une histone H2B deubiquitylase (DUB) et une histone acétyltransférase (HAT). Des études structurales récentes du complexe SAGA de levure mettent en évidence l'organisation modulaire de SAGA. Les dix-neuf sous-unités de SAGA en levure sont organisées dans quatre modules : les modules enzymatiques HAT et DUB, un module capable d'interagir avec des activateurs transcriptionnels et un module central structural. Ces études soulignent en plus la capacité de deux sous-unités du module centrale de SAGA d'interagir avec TBP et leur importance pour la déposition de TBP aux promoteurs des gènes. Les sous-unités du module central, à l'exception des sous-unités qui interagissent avec TBP, sont considérées comme cruciales pour l'assemblage et l'intégrité de SAGA.

Chez les métazoaires, l'activité HAT portée par la sous-unité GCN5 (ou son paralogue PCAF) de SAGA est partagée avec ATAC. A cote des sous-unités du module HAT, ATAC possède sept sous-unités spécifiques à ce complexe. Bien que la composition des sous-unités d'ATAC soit définie, son organisation structurale et ses fonctions sont encore très mal connues.

La transcription par la Pol II est très précisément régulée et des modifications majeures de la transcription par la Pol II ont lieu durant le développement embryonnaire d'un organisme et la différentiation des cellules. En effet, l'inactivation des gènes codant pour des sous-unités d'ATAC et de SAGA provoque une létalité embryonnaire soulignant le rôle de ces complexes dans les mécanismes de contrôle transcriptionnel au cours du développement. Néanmoins, des observations récentes soulignent combien nos connaissances sur les mécanismes et les fonctions des complexes coactivateurs sont encore parcellaires. Premièrement, des études de mon laboratoire d'accueil indiquent que SAGA agit comme un cofacteur général pour la transcription par la Pol II dans la levure, contrairement à un rôle pour la transcription d'un groupe de gènes particuliers, comme supposée précédemment. Cet effet global de SAGA sur la transcription par la Pol II n'est apparu que lors de l'analyse des quantités d'ARN nouvellement synthétisés, tandis que les taux d'ARN totaux montraient peu de changements. D'autre parte, les activités enzymatiques de modification de la chromatine des deux autres complexes coactivateurs (les complexes TIP60 et Mll3/Mll4 COMPASS-like) ne semblent pas être les activités clés pour la régulation de la transcription par la Pol II dans les cellules souches embryonnaires (ESCs) de souris.

Les ESCs de souris sont dérivées de la masse cellulaire interne des blastocystes et présentent des caractéristiques cellulaires uniques. Elles peuvent s'auto-renouveler, apparemment sans fin et, à cause de leur pluripotence, peuvent se différencier dans la plupart des lignées cellulaires. Les ESCs possèdent de plus des taux de prolifération élevés avec un cycle cellulaire caractérisé par une phase G1 raccourcie par rapport aux cellules somatiques. Les mécanismes moléculaires permettant les capacités d'auto-renouvèlement et de pluripotence des ESCs reposent sur l'action de des facteurs de transcription de pluripotence. Ceux-ci comprennent des facteurs de pluripotence centrales tels que Oct4 (*Pou5f1*) et Sox2 et d'autres facteurs de pluripotence, qui sont plus sensibles aux signaux environnementaux, tels que Nanog, Tfcp2l1, Klf4 et Esrrb. Plusieurs résultats récents indiquent que la physiologie particulière des ESCs de souris dépend également des fonctions de plusieurs complexes coactivateurs transcriptionnels.

Alors que des études précédentes ont analysé l'impact des sous-unités individuelles d'ATAC et de SAGA sur la transcription par la Pol II dans les cellules mammifères, les interprétations étaient basées sur la mesure des niveaux d'ARN totaux. Par conséquent, en s'appuyant sur nos connaissances de SAGA chez la levure et sur l'importance de l'analyse d'ARN nouvellement synthétisés, les objectifs de mon doctorat était 1) d'identifier l'importance d'ATAC et de SAGA pour la synthèse des ARNs naissants dans les cellules mammifères, 2) de révéler s'ils ont des rôles redondants à cause de leur activité HAT partagée ou s'ils agissent indépendamment de leur activité HAT et 3) de déterminer l'importance d'ATAC et de SAGA sur la physiologie unique des cellules souches embryonnaires (ESCs) de souris.

Résultats

En utilisant la technologie CRISPR-Cas9, j'ai généré des lignées cellulaires de ESCs dans lesquelles j'ai inactivé les gènes codant pour des sous-unités structurales de SAGA (*Supt7l* et *Supt20h*) (analyse Western blot des cellules *Supt7l*^-/-, Figure 1A). Pour révéler l'importance des sous-unités de SAGA pour la croissance des ESCs, j'ai effectué des tests de croissance clonale dans un milieu contenant du sérum et du LIF (facteur inhibiteur de leucémie). LIF est un facteur impliqué dans la maintenance de la pluripotence et le renouvellement des ESCs. Nous avons pu observer une réduction majeure de la croissance des cellules *Supt7l*^-/- (Figure 1B). Au contraire, les cellules *Supt20h*^-/- présentaient des phénotypes comparables au cellules contrôles.

Compte tenu de l'impact de l'inactivation de *Supt7l* sur la croissance des ESCs, nous avons voulu évaluer le rôle de SAGA sur la pluripotence des ESCs. J'ai donc réalisé des analyses clonales similaires à celles décrites ci-dessus avec une coloration supplémentaire pour révéler les niveaux de phosphatase alcaline (AP) dans les ESCs mutantes et contrôles. Les colonies indifférenciées avec haut pluripotence
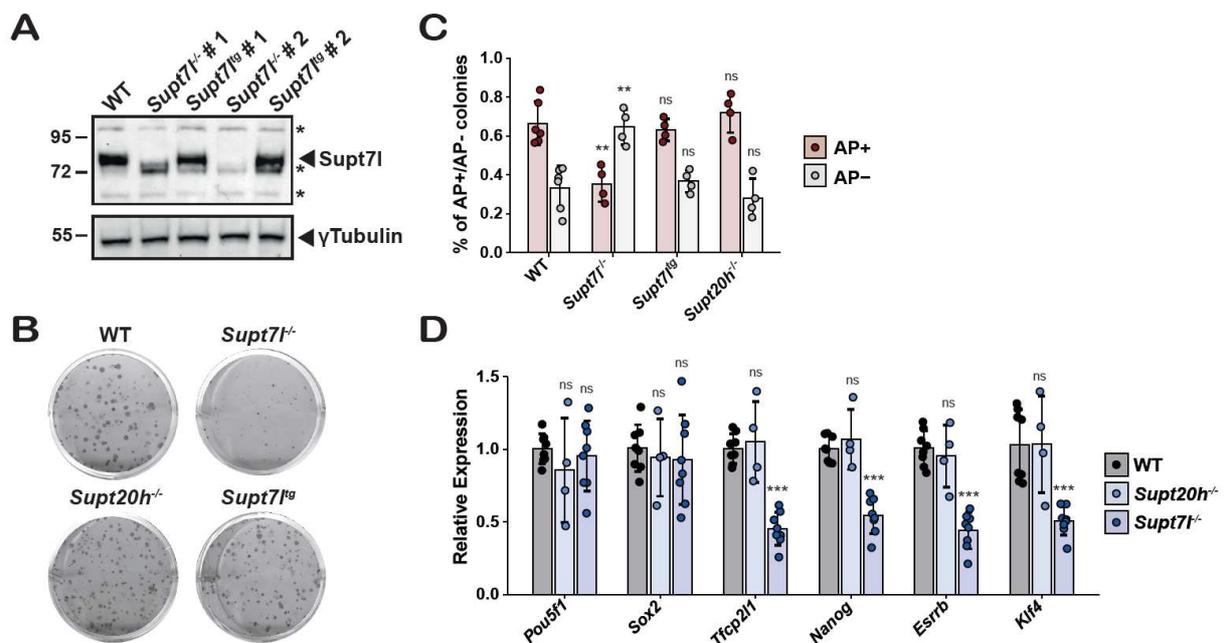
Figure 1: Le complexe coactivateur SAGA est nécessaire pour maintenir la pluripotence des cellules souches embryonnaires de souris. A. Analyse par Western blot de deux lignées indépendants des cellules *Supt7l*^-/- et dans lesquelles le gène Supt7l est ré-exprimé (*Supt7l*^tg) et des cellules contrôles (WT). La γTubulin servent de témoins de chargement. La protéine Supt7l est perdue dans les lignées cellulaires *Supt7l*^-/-, tandis que les lignées cellulaires *Supt7l*^tg présentent des niveaux comparables aux cellules WT. L'étoile indique des bandes non spécifiques. **B.** Analyse clonale des mutants SAGA dans un milieu contenant du sérum et LIF coloré au cristal violet. Une diminution du nombre de colonies peut être observée pour *Supt7l*^-/- mais pas pour *Supt20h*^-/-. Le phénotype de *Supt7l*^-/- n'est plus observé après réintroduction de la séquence codante de *Supt7l* (*Supt7l*^tg). **C.** Quantification de la coloration à la phosphatase alcaline (AP) des analyses clonales des mutants SAGA comme indiqué dans (A). Le nombre de colonies AP positives (AP+) et AP négatives (AP-) a été normalisé par rapport au nombre total de colonies évaluées par coloration au cristal violet. n = 4. Test de Wilcoxon-Mann-Whitney bilatéral. ns, non significatif ($p = 0,6095$); **, $p = 0,009524$. **D.** Les niveaux d'ARNm total de plusieurs facteurs de pluripotence dans les cellules *Supt7l*^-/- et *Supt20h*^-/- ont été évalués par RT-qPCR. Les niveaux d'ARNm total ont été normalisés à deux gènes d'ARN polymérase III (*Rpph1* et *Rn7sk*) et à des cellules control (WT). Des niveaux réduits de *Tfcp2l1*, *Nanog*, *Esrrb* et *Klf4* ont été trouvés dans les lignes *Supt7l*^-/- mais pas *Supt20h*^-/-. n = 4-8. Test de Wilcoxon-Mann-Whitney bilatéral. ns, non significatif ($p > 0,05$); ***, $p = 0,0001554$.

sont révélées par un niveau élevé d'AP et une coloration rouge (AP positives), tandis que les colonies différenciées avec une pluripotence réduite restent non colorées (AP négatives). Avec cette analyse, nous avons pu observer une réduction significative des colonies indifférenciées (AP positives) dans les lignées *Supt7l*[-/-] mais pas dans les lignées *Supt20h*[-/-] (Figure 1C). En accord, l'expression de plusieurs facteurs de pluripotence était fortement réduite dans les cellules *Supt7l*[-/-] mais pas dans les cellules *Supt20h*[-/-] (Figure 1D). La ré-expression de Supt7l dans les cellules *Supt7l*[-/-] (lignées Supt7l[tg]) a entraîné la récupération d'un phénotype sauvage (Figure 1A - C) suggérant que les effets observés dans les lignées cellulaires *Supt7l*[-/-] sont en effet dus à la perte de Supt7l.

L'implication de SAGA dans la croissance et dans la régulation de la pluripotence des ESCs de souris est en accord avec des études récentes suggérant que Taf5l et Taf6l, deux autres sous-unités du module central de SAGA, sont impliquées dans les mécanismes moléculaires important pour la maintenance de l'auto-renouvellement et la pluripotence des ESCs. Plus spécifiquement, dans les lignées cellulaires *Supt7l*[-/-] nous avons trouvé des niveaux d'ARNm réduits surtout pour les facteurs importants pour la maintenance d'une haute pluripotence, Nanog, Klf4, Esrrb et Tfcp2l1. En général, les ESCs possédant de faibles niveaux d'expression de Nanog, Klf4, Esrrb et Tfcp2l1 sont considérées comme étant destinées à la différentiation. Par conséquent, la réduction spécifique des niveaux d'ARNm pour ces facteurs dans les cellules *Supt7l*[-/-] pourrait suggérer une augmentation du taux de différentiation. Dans l'ensemble, cela pourrait impliquer que les mécanismes liés aux facteurs de pluripotence Nanog, Klf4, Esrrb et Tfcp2l1 dépendent de la fonction SAGA.

A.

	Clones screened	% +/-	# -/-
Yeats2	264	18.9	0
Zzz3	317	17.9	0
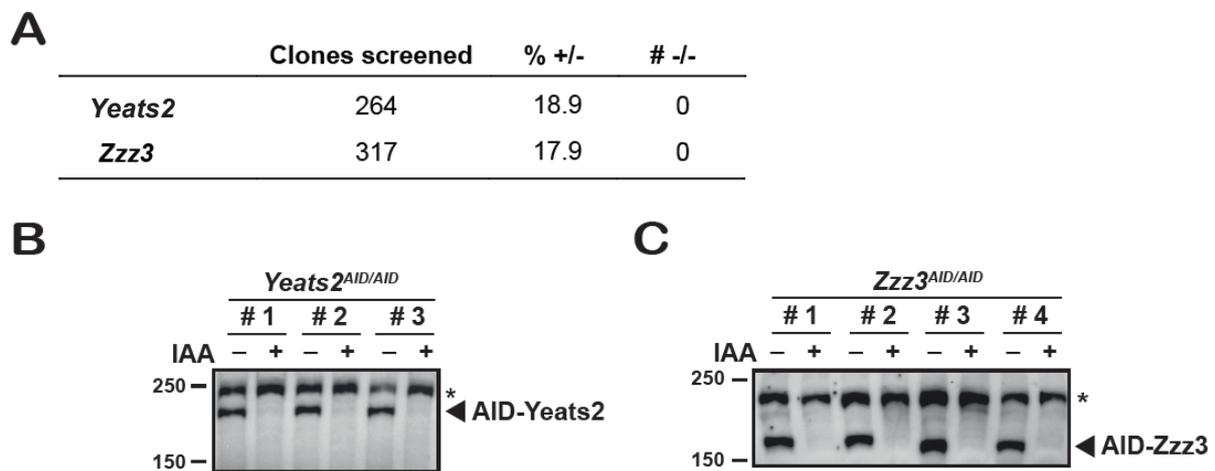
Figure 2: Le complexe coactivateur ATAC semble crucial pour la survie des cellules souches embryonnaires de souris. A. Tableau indiquant le nombre de clones criblés, le pourcentage de clones hétérozygotes (+/-) et le nombre de clones homozygotes (-/-) obtenu pour les deux gènes codant pour des sous-unités cruciales du complexe ATAC, Yeats2 et Zzz3. Aucun clone homozygote n'a pu être généré. **B.** et **C.** Des lignées cellulaires inductibles à l'auxine (AID) ont été générées pour Yeats2 et Zzz3 permettant une déplétion efficace des deux protéines après 24h de traitement à l'auxine (IAA). Trois mutants homozygotes sont présentés pour *Yeats2*[AID/AID] N-terminal (B) et deux clones homozygotes pour une fusion N-terminale (clones #1 et #2) *Zzz3*[AID/AID] ainsi que deux clones homozygotes pour une fusion C-terminale (clones #3 et #4) *Zzz3*[AID/AID] (C). Les étoiles indiquent des bandes non spécifiques.

Nous avons ensuite cherché à inactiver des sous-unités cruciales pour l'assemblage du complexe ATAC, Yeats2 et Zzz3, afin d'en évaluer la fonction et l'importance pour la physiologie des ESCs de souris. Alors que nous avons pu générer des lignes viables dans lesquelles le complexe SAGA est inactivé ($Supt7l^{-/-}$ et $Supt20h^{-/-}$), aucun clone homozygote pour les sous-unités d'ATAC n'a pu être obtenu, suggérant que ce complexe serait requis pour la survie des ESCs (Figure 2A). Ces résultats indiquent une dépendance différente des ESCs aux fonctions d'ATAC et de SAGA et suggèrent en outre que certaines fonctions potentiellement essentielles d'ATAC ne peuvent pas être compensées par SAGA.

Pour pouvoir néanmoins évaluer les fonctions d'ATAC dans les ESCs, nous avons généré des lignées dans lesquelles la déplétion des sous-unités Yeats2 et Zzz3 peut être induite en utilisant le système de dégradation inductible par l'auxine (AID). Ce système permet de dégrader les protéines Yeats2 et Zzz3 jusqu'à des niveaux indétectables après 24h de traitement par l'auxine (Figure 2B et C).
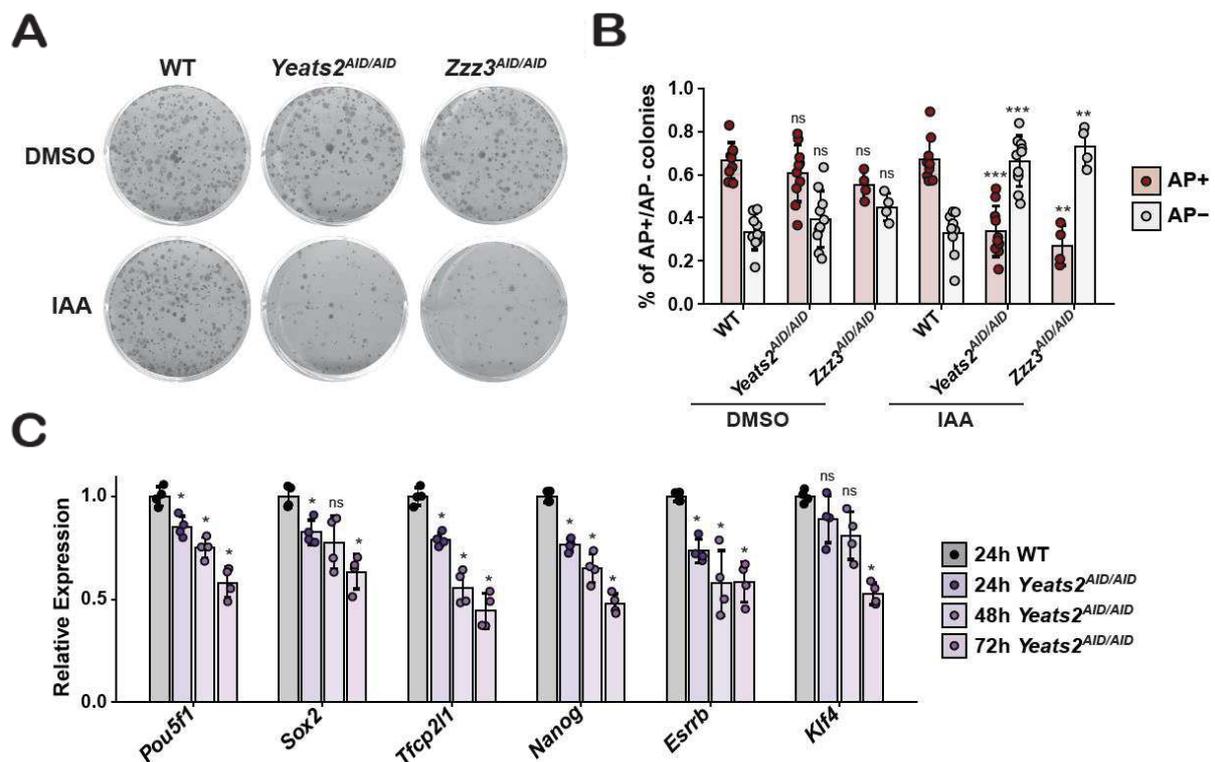
Figure 3: Le complexe coactivateur ATAC semble crucial pour la maintenance de la pluripotence des cellules souches embryonnaires de souris. A. Images représentatives de l'analyse clonal des lignées cellulaires mutantes d'ATAC dans un milieu contenant du sérum et LIF traité avec du DMSO ou de l'auxine (IAA) colorées au cristal violet. L'ajout d'IAA conduit à des densités de colonies réduites dans les lignées cellulaires $Yeats2^{AID/AID}$ et $Zzz3^{AID/AID}$. **B.** Quantification de la coloration à la phosphatase alcaline (AP) des analyses clonales comme indiqué en (A). Le nombre de colonies AP positives (AP+) et AP négatives (AP-) a été normalisé au nombre total de colonies, évalué par coloration au cristal violet. Un nombre réduit de colonies AP+ a été trouvé pour les lignées cellulaires $Yeats2^{AID/AID}$ et $Zzz3^{AID/AID}$ lors de l'ajout d'IAA. n = 4-10. Test de Wilcoxon-Mann-Whitney bilatéral. ns, non significatif ($p > 0.05$); **, $p = 0.002797$; ***, $p = 2.165 \times 10^{-5}$. **C.** Analyse des niveaux d'ARNm totaux des facteurs de pluripotence dans les lignées cellulaires $Yeats2^{AID/AID}$ comparer aux cellules control (WT) après un traitement auxine de 24h à 72h. Les niveaux d'ARNm total ont été normalisés à deux gènes d'ARN polymérase III ($Rpph1$ et $Rn7sk$). Les résultats des lignes $Yeats2^{AID/AID}$ pour chaque point temporel ont été normalisés aux cellules WT respectives des mêmes points temporels. Seuls les niveaux d'ARNm totaux des cellules WT à 24h sont montré. n = 4. Test de Wilcoxon-Mann-Whitney bilatéral. ns, non significatif ($p > 0.05$); *, $p = 0.02857$.

Pour évaluer l'impact de la déplétion de Yeats2 et Zzz3 sur la croissance des ESCs et la maintenance du renouvèlement des ESCs, nous avons effectué des expériences de croissance clonale comme décrit ci-dessus. En l'absence d'auxine, les deux lignées cellulaires AID ont montré des densités de colonies similaires à celles des cellules contrôles, tandis qu'un traitement continu d'auxine pendant six jours a entraîné une forte réduction du nombre de colonies (Figure 3A). Cela suggère que la croissance des ESCs de souris est affectée lors de la perte du complexe ATAC. En utilisant la coloration AP, nous avons aussi constaté que les capacités d'auto-renouvellement et de pluripotence des ESCs nécessitent les fonctions d'ATAC (Figure 3B). L'effet sur la pluripotence a pu être également confirmé par la mesure des niveaux d'ARNm de plusieurs facteurs de pluripotence, montrant une réduction progressive après 24 à 72 heures de traitement auxine dans les lignées cellulaires $Yeats2^{AID/AID}$ (Figure 3C). Cette analyse a montré que l'expression de tous les facteurs de pluripotence testés sont affectés par la déplétion de Yeats2, y compris les facteurs de pluripotence centrales Oct4 (*Pou5f1*) et Sox2. Ces résultats suggèrent qu'ATAC pourrait être requis pour l'expression des différents facteurs de pluripotence. Cette perte générale des facteurs de pluripotence pourrait expliquer les effets observés sur la maintenance de la pluripotence et l'auto-renouvèlement dans les lignées cellulaires mutantes d'ATAC.

Globalement, ces résultats indiquent qu'ATAC et SAGA sont nécessaires pour la maintenance de la pluripotence et pour la croissance des ESCs. Ces résultats étaient très inattendus, car l'inactivation de différentes sous-unités d'ATAC et SAGA dans l'embryon de souris cause la létalité à des stades plus tardifs.

Pour évaluer précisément les fonctions d'ATAC et SAGA sur la transcription par Pol II, nous avons analysé les ARNs naissants suivi d'un séquençage à haut débit en utilisant une méthode de marquage des ARN naissants à l'aide de 4-thiouridine (4sU) (comparaison de séquençage des ARNs totaux et naissants : Figure 4A). Cette analyse nous a permis de révéler que 1) la perte des sous-unités structurales d'ATAC ou SAGA conduit à une réduction générale, bien que modeste, de l'expression des gènes (Figure 4B) et 2) qu'ATAC et SAGA affectent significativement l'expression de groupes de gènes différents (Figure 4C). Plus précisément, dans les lignées cellulaires $Supt7l^{-/-}$, les gènes impliqués dans la réponse au LIF ont été trouvé enrichis parmi les gènes dont l'expression est réduite. Au contraire, dans les cellules où le complexe ATAC est inactivé, parmi les gènes dont l'expression est réduite ceux impliqués dans la traduction sont particulièrement enrichis.

L'effet spécifique des sous-unités d'ATAC sur l'expression des gènes impliqué dans la traduction, a pu être en outre confirmé par des analyses d'expression des RPG (gènes codant pour les protéines ribosomiques) dans les différentes lignées cellulaires (Figure 4D). Nous avons pu montrer qu'une majorité de RPG sont spécifiquement dépendants d'ATAC pour leur expression, confirmant ainsi les effets divergents d'ATAC et SAGA dans la régulation de la transcription par la Pol II.

Il est important de souligner que les effets sur les RPG et les autres gènes liés à la traduction lors de la déplétion des sous-unités d'ATAC étaient détectables sur des échantillons d'ARN nouvellement synthétisés après 24 heures de traitement auxine, alors qu'aucun défaut évident du cycle cellulaire n'avait pu être détecté. Cependant, des défauts de croissance et du cycle cellulaire étaient détectables lors d'un traitement prolongé par l'auxine. Ces résultats suggèrent que la réduction de l'expression des
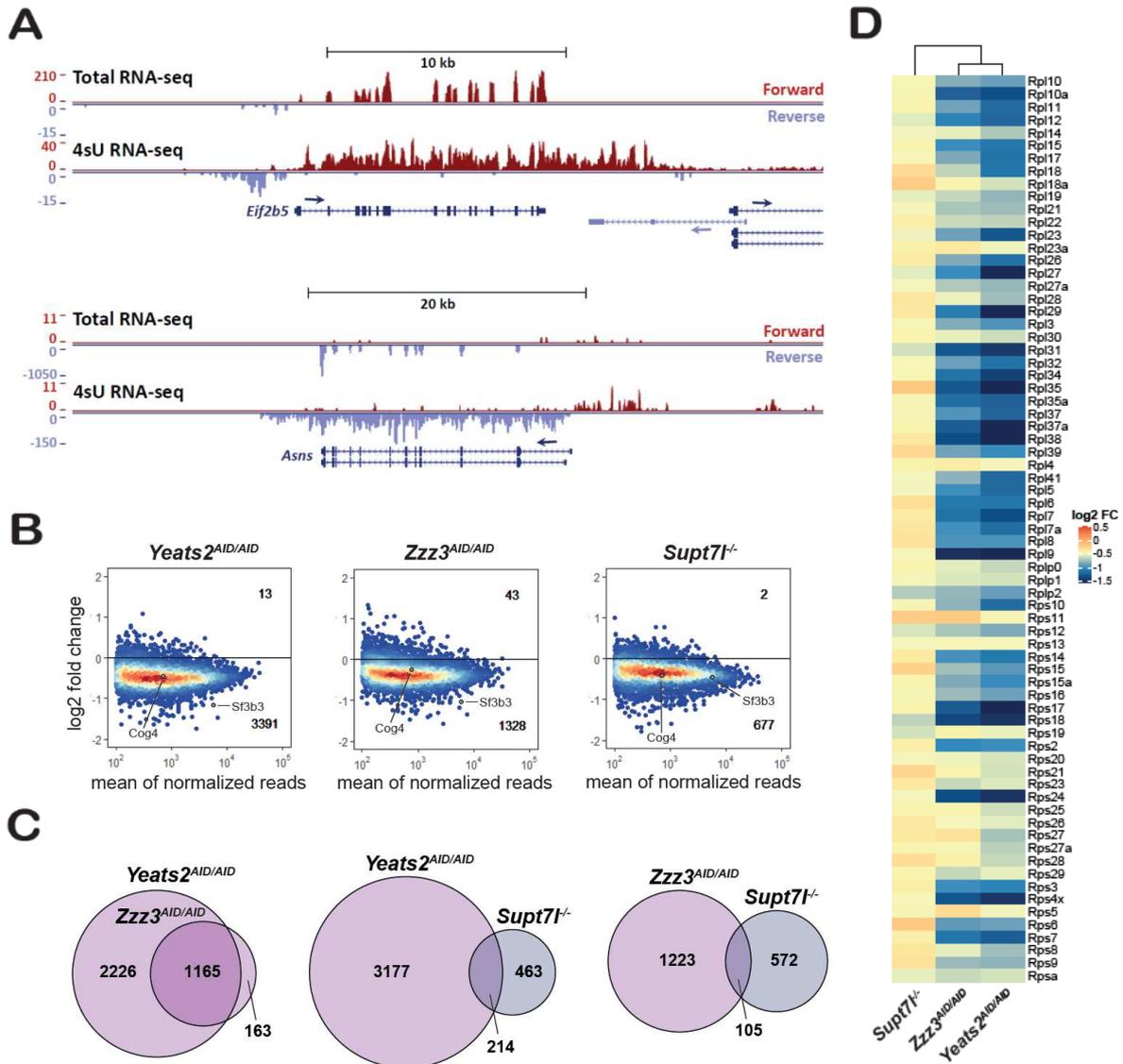
Figure 4: L'analyse d'ARN nouvellement synthétisés révèle qu'ATAC et SAGA affectent différents ensembles de gènes. A. Vues du navigateur du génome comparant les résultats d'ARN-seq total ('Total RNA-seq') et d'ARN-seq 4sU (4sU RNA-seq) pour deux gènes (en haut, *Eif2b5* et en bas, *Asns*). Les flèches indiquent la direction de la transcription. Les brins d'avant ('forward') et d'arrière ('reverse') sont illustrés. **B.** Représentation des diagrammes de densité du changement de l'expression en log2 du séquençage d'ARN nouvellement synthétisés marqués au 4sU des lignées cellulaires *Yeats2^{AID/AID}*, Zzz3^{AID/AID} et *Supt7l^{-/-}* par rapport aux cellules contrôles comparée à la moyenne des sonde normalisées. Toutes les lignées cellulaires ont été traitées pendant 24 heures avec IAA. Les nombres en haut et en bas indiquent le nombre de gènes dont l'expression est significativement augmentée ou réduite dans les différents mutants. Les gènes ont été considérés comme significativement affectés avec une valeur de *p* ajustée de <0,05 et un changement de log2 de < -0,5 ou > 0,5. Un seuil de 100 sondes a été fixé pour définir les gènes exprimés. **C.** Diagrammes de Venn comparant le chevauchement de gènes significativement réduite dans les cellules *Yeats2^{AID/AID}*, Zzz3^{AID/AID} et *Supt7l^{-/-}*. **D.** Représentation du changement de log2 (log2 FC) observé pour les gènes codant pour les protéines ribosomiques (RPG) dans les différentes lignées cellulaires mutantes. Les RPG semblent particulièrement affectés dans les cellules *Yeats2^{AID/AID}* et Zzz3^{AID/AID}.

gènes liés à la traduction pourrait entraîner des anomalies du cycle cellulaire. Ainsi les anomalies observées dans les lignées cellulaires où le complexe d'ATAC est inactivé pourraient être dues à un assemblage et une fonction altérée des ribosomes. De plus, les effets sur les gènes liés à la traduction pourraient expliquer la létalité observée après l'inactivation constitutive de Yeats2 et Zzz3, car les gènes associés à la biogenèse et à la traduction des ribosomes sont généralement considérés comme essentiels. Cependant, il ne peut être exclu que les effets sur la prolifération et la croissance cellulaires puissent être dus à des changements dans l'expression d'autres gènes, car de nombreux gènes sont affectés lors de la déplétion des sous-unités d'ATAC.

Nos résultats ne permettent pas de conclure que les gènes liés à la traduction sont des cibles directes d'ATAC dans les ESCs, mais des études récentes dans des cellules humaines ont montré que Yeats2 et Zzz3 se lient directement aux promoteurs des RPG et régulent leur expression. Ces résultats indiquent également que les effets du complexe ATAC sur l'expression des gènes liés à la traduction ne se limitent pas aux ESCs de souris mais se retrouvent également dans des cellules humaines.

L'ensemble de nos résultats indique qu'ATAC et SAGA affectent de manière significative différents ensembles de gènes et pourraient donc avoir des rôles non redondants dans la régulation de la transcription par Pol II. Néanmoins, nous n'excluons pas qu'ATAC et SAGA puissent agir de manière redondante sur certains gènes. En effet, notre analyse d'ARN nouvellement synthétisés a révélé qu'environ 200 gènes pourraient être significativement corégulés par ATAC et SAGA. Cela concorde avec des études antérieures qui suggéraient que certains gènes pourraient être transcriptionnellement dépendants ou occupés par des sous-unités des deux complexes.

Pour évaluer le rôle de l'activité HAT partagée par ATAC et SAGA, nous avons généré des lignées cellulaires avec une inactivation des gènes *Tada2a* ou *Tada2b*, codant pour deux paralogues permettant l'incorporation du HAT spécifiquement dans ATAC ou dans SAGA. Nous avons également voulu inactiver Tada3, une sous-unité du module HAT d'ATAC et de SAGA. Une étude antérieure ayant montré que cette sous-unité est nécessaire à la formation de la masse cellulaire interne, nous avons généré des lignées cellulaires permettant une déplétion inductible de Tada3 (Tada3$^{AID/AID}$). Comme pour les lignées *Yeats2$^{AID/AID}$* et *Zzz3$^{AID/AID}$*, après 24 heures de traitement à l'auxine des cellules Tada3$^{AID/AID}$, la protéine de fusion n'était plus détectable par Western blot. Étonnamment, les lignées cellulaires mutantes HAT n'ont pas montré d'effet détectable sur la croissance cellulaire ou la pluripotence (Figure 5A et B). Ces résultats indiquent pour la première fois que la fonction HAT ne parait pas être la fonction déterminante d'ATAC et SAGA dans les ESCs.

Pour comprendre le rôle de l'activité HAT partagée par ATAC et SAGA dans la transcription par Pol II, nous avons décidé de comparer les effets transcriptionnels dans les cellules après déplétion de Tada3 (*Tada3$^{AID/AID}$*) à des lignées cellulaires mutantes dans lesquelles les deux complexes sont inactivés
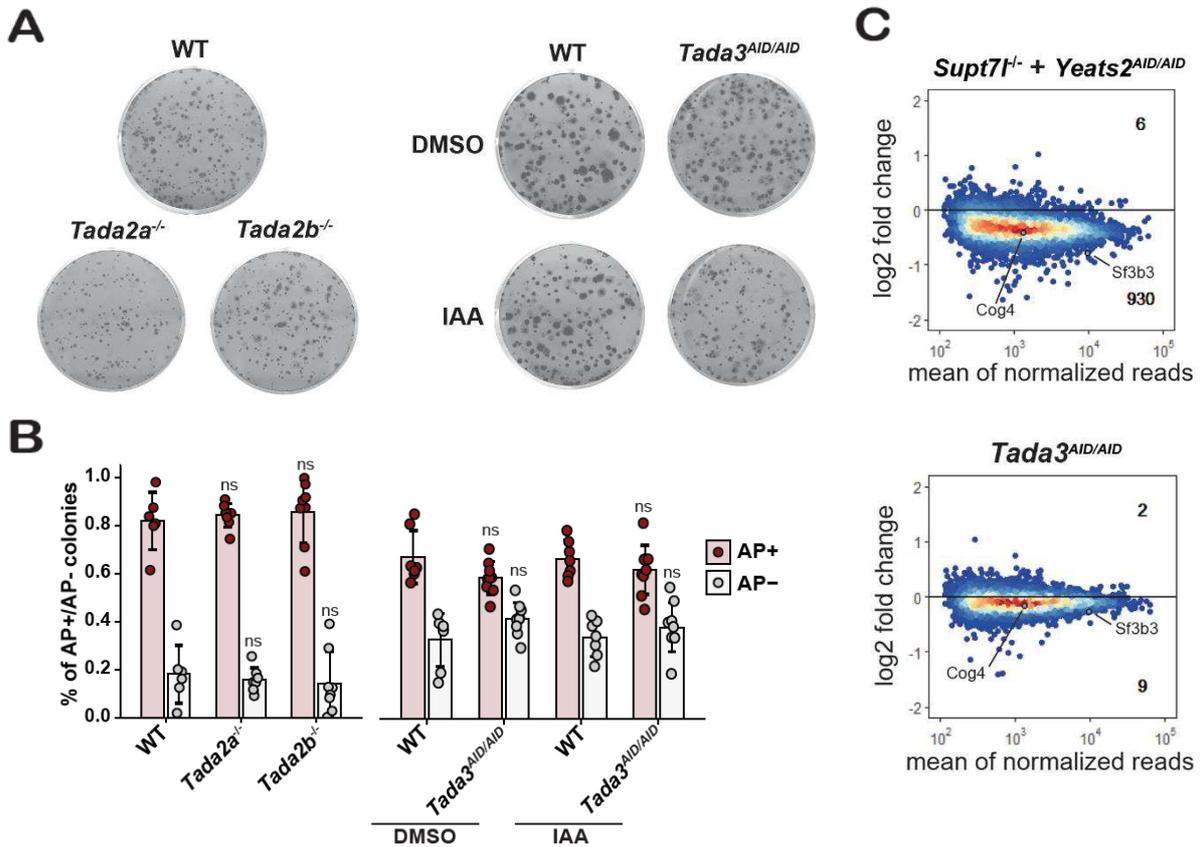
Figure 5: L'inactivation de l'activité acétyltransférase partagée entre ATAC et SAGA n'affecte pas la maintenance de la pluripotence. A. Analyse clonale des cellules mutantes du HAT dans un milieu contenant du sérum et LIF. Les cellules *Tada3$^{AID/AID}$* et control (WT) ont été traitées avec du DMSO ou de l'IAA. Aucun effet évident sur la croissance des colonies n'a été détecté pour les mutants HAT. **B.** Quantification de la coloration à la phosphatase alcaline (AP) des analyses clonales comme indiqué en (A). Le nombre de colonies AP positives (AP+) et AP négatives (AP-) a été normalisé au nombre total de colonies, évalué par coloration au cristal violet. Aucune différence majeure n'a été observée pour les mutants HAT. n = 8. Test de Wilcoxon-Mann-Whitney bilatéral. ns, non significatif ($p > 0,05$). **C.** Représentation du diagramme de densité du changement de l'expression en log2 du séquençage d'ARN nouvellement synthétisés marqués au 4sU des lignées *Tada3$^{AID/AID}$* et *Supt7l$^{-/-}$* + *Yeats2$^{AID/AID}$* par rapport aux cellules contrôles comparé à la moyenne des sondes normalisées. Toutes les lignées cellulaires ont été traitées pendant 24 heures avec IAA. Les nombres en haut et en bas indiquent le nombre de gènes ayant une expression significativement augmentée ou réduite dans les différents mutants. Les gènes ont été considérés comme significativement affectés avec une valeur de p ajustée de $< 0,05$ et un changement de log2 de $< -0,5$ ou $> 0,5$. Un seuil de 100 sondes a été fixé pour définir les gènes exprimés.

(*Supt7l$^{-/-}$* + *Yeats2$^{AID/AID}$*) (Figure 5C). Nos résultats montrent que la déplétion de l'activité HAT d'ATAC et de SAGA ne provoque pas d'effets majeurs sur la transcription, indiquant que l'activité transcriptionnelle de ces deux complexes semble majoritairement indépendante de leur activité HAT partagée.

L'indépendance des activités HAT des deux complexes pour la croissance, la pluripotence et la transcription des ESCs, indique qu'ATAC et SAGA possèdent d'autres fonctions plus importantes pour ces processus. Des résultats antérieurs basés sur l'utilisation de mutants catalytiques d'autres complexes de modification de la chromatine, suggèrent également que les complexes coactivateurs possèdent des fonctions importantes qui n'impliquent pas la modification de la chromatine. De plus, cela questionne

l'importance des modifications des histones pour la physiologie des ESCs de souris et la transcription par la Pol II.

Cependant, des études antérieures sur l'inactivation de Gcn5, l'enzyme principale du module HAT d'ATAC et de SAGA, suggèrent une importance du HAT lors de la différenciation des ESCs. Cela suggère un rôle potentiellement plus important des activités de modification des histones d'ATAC et de SAGA pendant la différenciation, ce qui est en accord avec l'importance de Gcn5 et son activité catalytique pendant le développement embryonnaire de la souris.

Enfin, nous avons pu montrer qu'une fonction de SAGA importante pour la croissance et la pluripotence des ESCs est potentiellement son interaction avec TBP par sa sous-unité, Supt3h (Figure 6A et B). Au contraire, l'activité DUB de SAGA, partagée avec d'autres complexes, semble être importante pour la croissance de ESCs, mais ne semble pas requise pour la maintenance de la pluripotence.

Cependant, l'analyse des ARNs nouvellement synthétisés dans les cellules $Supt3h^{-/-}$ a révélé que les effets transcriptionnels sont moins sévères que lors de l'inactivation de Supt7l (Figure 6C). Cela suggère que la transcription par la Pol II ne dépend pas uniquement de la fonction de déposition de TBP de SAGA dans les ESCs. En général, ces résultats sont en accord avec les résultats de levure, où la suppression de $Spt3$ (orthologue à $Supt3h$) conduit à un effet moins sévère sur la transcription par Pol II par rapport à la suppression de $Spt20$ (orthologue à $Supt20h$) ou $Spt7$ (orthologue à $Supt7l$). En général, ces résultats suggèrent une importance pour les autres fonctions de SAGA dans la transcription par la Pol II.
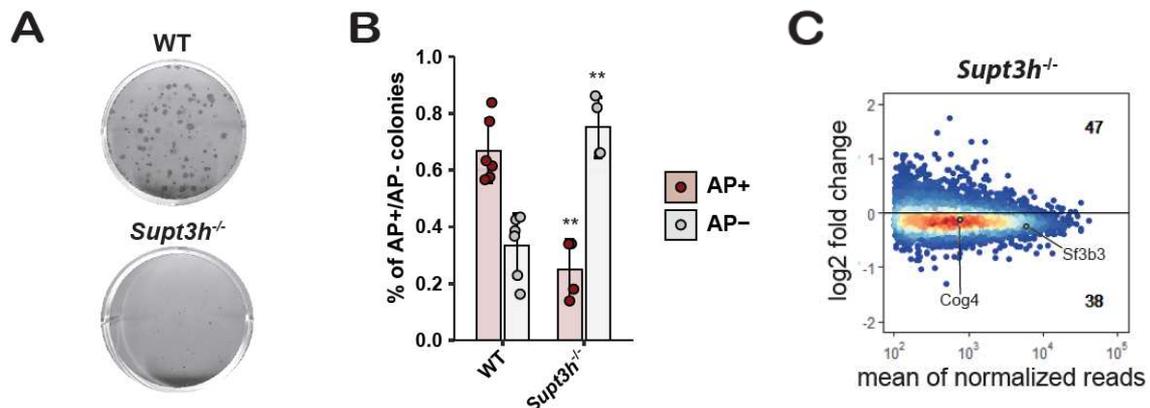
Figure 6: SAGA affecte la maintenance de la pluripotence des ESCs de souris grâce à ses interactions avec TBP. A. Images représentatives de l'analyse clonale des cellules $Supt3h^{-/-}$ dans un milieu contenant du sérum et LIF. L'inactivation de $Supt3h$ affecte la croissance cellulaire. **B.** Quantification de la coloration à la phosphatase alcaline (AP) des analyses clonales comme indiqué en (A). Le nombre de colonies AP positives (AP+) et AP négatives (AP-) a été normalisé au nombre total de colonies, évalué par coloration au cristal violet. L'inactivation de $Supt3h$ entraîne une réduction du nombre de colonies AP+. n = 4-6. Test de Wilcoxon-Mann-Whitney bilatéral. **, $p = 0{,}009524$. **C.** Représentation du diagramme de densité des changements de l'expression en log2 du séquençage d'ARN nouvellement synthétisés marqués au 4sU des lignées cellulaires $Supt3h^{-/-}$ par rapport aux cellules contrôles comparés à la moyenne des sondes normalisées. Les lignées cellulaires ont été traitées pendant 24 heures avec IAA. Les nombres en haut et en bas indiquent le nombre de gènes significativement augmente ou réduite, respectivement. Les gènes ont été considérés comme significativement affectés avec une valeur de p ajustée de < 0,05 et un changement de log2 de < -0,5 ou > 0,5. Un seuil de 100 sondes a été fixé pour définir les gènes exprimés.

Conclusions

Mon travail a permis de montrer le rôle crucial des complexes ATAC et SAGA pour la maintenance de la pluripotence des ESCs de souris. L'inactivation des complexes ATAC et SAGA modifie l'expression de groupes de gènes différents mais aboutit à un phénotype relativement similaire dans les ESCs. Le complexe ATAC joue un rôle particulièrement important pour l'expression de gènes impliqués dans la traduction. En effet, l'inactivation d'ATAC entraine une diminution de la transcription d'une majorité des gènes codant pour des protéines ribosomiques, pouvant expliquer les anomalies de croissance et de maintenance des cellules souches embryonnaires de souris. Enfin, nous avons pu montrer que les anomalies transcriptionnelles et le phénotype observé ne sont pas liées à l'activité acétyltransférase partagée entre ces deux complexes. Mes résultats indiquent que ces deux complexes ont des fonctions non redondantes et indépendantes de leurs activités acétyltransférases sur des groupes de gènes différents. Il reste à déterminer pour chaque complexe, quels sont les activités importantes pour la transcription et comment ces activités sont coordonnées. Les mécanismes précis permettant le recrutement de ces complexes sur les gènes qu'ils régulent devront également être déchiffres en utilisant des techniques innovantes. Comme cela a été proposé chez la levure, ces complexes pourraient également avoir des fonctions globales et redondantes sur la transcription par l'ARN polymérase II qui sont difficiles à caractériser avec les techniques disponibles.

Table of Content

List of Figures

List of Tables

List of protocols

List of Abbreviations

Δ	Deletion
2i	Two inhibitors, referring to PD0325901 and CHIR99021
3'SS	3' splice site
3D-SIM	3-dimensional structured illuminated microscopy
4C	chromosome conformation capture-on-chip
4sU or 4tU	4-thiouridine or 4-thiouracil
5'SS	5' splice site
AATF	Apoptosis-antagonizing transcription factor
ac	Acetylation
acetyl-CoA	Acetyl-Coenzyme A
ADA	Alteration/Deficiency in Activation; Adaptor
ADP	Adenosine diphosphate
Ahc1 or Ahc2	ADA HAT component 1 or 2
AID	Auxin-inducible degron
AP	Alkaline phosphatase
ARP4	Actin-related protein 4
ATAC	Ada-two-A-containing complex
ATAC-seq	Assay for transposase accessible chromatin followed by high throughput sequencing
ATF6α	Cyclic AMP-dependent transcription factor 6 alpha
ATP	Adenosine triphosphate
ATXN7 or ATXN7L3	Ataxin 7 or 7 like 3
BAF	Brg1/Brahma-associated factors
BioTag	Biotinylatable tag
BirA	Bifunctional ligase/repressor A
BMP	Bone morphogenetic proteins
bp	Base pair(s)
BP	Branch point
BRE	TFIIB recognition element
BRG1	Brahma protein-like 1
BRR2	Bad response to refrigeration
BTAF1	B-TATA binding protein associated factor 1 (Mot1 in yeast)
CAK	CDK-activating kinase
Cas9	CRISPR associated protein 9
CBP	CREB-binding protein
CDK	Cyclin-dependent kinase
CE	Cytoplasmic extract
CENP-A	Centromere protein A
CFI	Cleavage factor I

CFII	Cleavage factor II
CGI	CpG island
CHD1	Chromodomain helicase DNA-binding protein 1
ChIP	Chromatin Immunoprecipitation
CHIR	CHIR99021; inhibitor of GSK3 pathway
CHRAC	Chromatin accessibility complex
c-MYC	Cellular myelocytoma
CMV	Cytomegalovirus (here CMV promoter)
COMPASS	Complex of proteins associated with Set1
CpG	Cytidine and guanosine pair separated by a phosphodiester bond
CPSF	Cleavage and polyadenylation specificity factor
CR	Coactivator redundant
CRISPR	Clustered regularly interspaced short palindromic repeats
Cryo-EM	Cryogenic electron microscopy
CstF	Cleavage stimulatory factor
CTCF	CCCTC-binding factor
CTD	C-terminal repeat domain
C-ter	Carboxy-terminal or -terminus
CUL1	Cullin-1
DCE	Downstream core element
deg	Degron
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DOT1L	Dot1-like
DPE	Downstream promoter element
DR1	Downregulator of transcription 1
DRB	5,6-dichloro-1-β-$_D$-ribofuranosylbenzimidazole
DSIF	DRB sensitivity inducing factor
DTT	Dithiothreitol
DUB	Deubiquitylation
E	Embryonic day
E2F1	E2 promoter binding factor 1
Eaf1	ELL associated factor 1
eGFP	Enhanced green fluorescent protein
ENY2	Enhancer of yellow 2
EP400	E1A binding protein p400
ER	Endoplasmatic reticulum
ERK	Extracellular signal-related kinase
eRNA	Enhancer derived RNA
esBAF	Embryonic stem cell-specific BAF

ESCs	Embryonic stem cells
Esa1	Enhancer of polycomb-like protein 1
Esrrb	Estrogen related receptor beta
Ex	Exon
FACS	Fluorescence activated cell sorting
FACT	Facilitates chromatin transcription
FAT	FRAP, ATM and TRRAP
FCP1	TFIIF-associating CTD phosphatase 1
FCS	Foetal calf serum
FDR	False discovery rate
FGF4	Fibroblast growth factor 4
FGFR2	FGF receptor 2
G9a	Gene 9a in major histocompatibility complex class III region
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GATA4 or 6	GATA binding protein 4 or 6
GBX2	Gastrulation brain homeobox 2
GCN5	General control non-derepressible 5
GFP	Green fluorescent protein
GLP	G9a-like protein
GNAT	GCN5-related N-acetyltransferases
gRNA	Guide RNA
GSK3	Glycogen synthase kinase-3
GSK3i	CHIR99021; inhibitor of GSK3 pathway
GTFs	General transcription factors
H2A	Histone 2A
H2AK119ub	Histone H2A lysine 119 ubiquitylation
H2B	Histone 2B
H2BK120ub	Histone H2B lysine 120 ubiquitylation
H2BK123ub	Histone H2B lysine 123 ubiquitylation
H3	Histone 3
H3K4me1	Histone lysine 4 monomethylation
H3K4me3	Histone 3 lysine 4 trimethylation
H3K79me3	Histone 3 lysine 79 trimethylation
H3K9ac	Histone 3 lysine 9 acetylation
H3K9me3	Histone 3 lysine 9 trimethylation
H3K27ac	Histone 3 lysine 27 acetylation
H3K27me3	Histone 3 lysine 27 trimethylation
H3K36me3	Histone 3 lysine 36 trimethylation
H3K79me3	Histone 3 lysine 79 trimethylation
H3S10phos	Histone 3 Serine 10 phosphorylation

H4	Histone 4
H4K12ac	Histone 4 lysine 12 acetylation
H4K16ac	Histone 4 lysine 16 acetylation
HA-tag	Hemagglutinin tag
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDM	Histone demethylase
HEAT	Huntington, Elongation Factor 3, PR65/A, TOR
HFD	Histone-fold domain
Hi-C	Genome-wide chromosome conformation capture assay
HIT	Head interacting with Tra1
HMT	Histone methyltransferase
HP1	Heterochromatin protein 1
HPDP-biotin	N-[6-(biotinamido)hexyl]-3'-(2'-pyridyldithio)propionamide
HSA	Helicase-SANT-associated
IAA	Indole-3-acetic acid; auxin
Inr	Initiator
INO80	Inositol-requiring mutant 80
INT	Integrator complex
IP	Immunoprecipitation
IRES	Internal ribosome entry site
ISWI	Imitation switch
JAK	Janus-associated kinase
JMJD	JmJC jumonji domain containing
KAT	Lysine acetyltransferase
kb	Kilobase; 1 kb = 1000 base pairs
KD	Knockdown
kDa	Kilodalton; 1 kDa = 1000 Dalton
KDM	Lysine demethylase
Klf2 or 4	Krüppel-like factor 2 or 4
KMT	Lysine methyltransferase
KO	Knockout
LCD	Low-complexity domain
LIF	Leukaemia inhibitory factor
LisH	Lis homology
lncRNA	Long non-coding RNA
LPS	Liquid-liquid phase separation
LSD1	Lysine-specific demethylase 1
MAPK	Mitogen-activated protein kinase
Mb	Megabase; 1 Mb = 1000 kilobase = 1 million base pairs

MBIP	MAP3K12-binding inhibitory protein 1
MBP	Maltose-binding protein
mCherry	Monomeric red fluorescent protein
MDa	Megadalton; 1 MDa = 1 kilodalton = 1 million dalton; unit of protein size
MEK	Mitogen-activated protein kinase/ERK kinase
MEKi	PD0325901, inhibitor of MEK/ERK pathway
meth	Methylation
MLL	Mixed lineage leukemia
MOF	Males absent on the first
MS	Mass spectrometry
MSL	Male specific lethal
MTE	Motif ten element
MYST	MOZ Ybf2 Sas2 TIP60
NAD	Nicotinamide adenine dinucleotide
Nanog	Derives from 'Tìr nan Òg', the mythical celtic land of youth
NaOH	Sodium hydroxide
NC2	Negative cofactor 2
NDR	Nucleosome-depleted region
NE	Nuclear extracts
NELF	Negative elongation factor
NeoR	Neomycin resistance
NF-Y	Nuclear transcription factor Y
NLS	Nuclear localization signal
NMD	Non-sense mediated decay
NSL	Nonspecific lethal
N-ter	Amino- or Nitrogen-terminal or -terminus
NTC	Nineteen complex
NuA4	Nucleosome acetyltransferase of H4
NURF	Nucleosome remodelling factor
Oct4	Octamer-binding transcription factor 4
o/n	Over night
ORF	Open reading frame
P2A	Porcine teschovirus-1 2A sequence
P300	E1A binding protein P300
PAF1	Polymerase-associated factor 1
Paf1C	PAF1 complex
PAM	Protospacer adjacent motif
PAS	Polyadenylation site
PCAF	P300/CBP-associated factor
PCR	Polymerase chain reaction

PD	PD0325901, inhibitor of MEK/ERK pathway
PGK	Phosphoglycerate kinase 1 (here promoter of PGK)
PHD	Plant homeodomain
phos	Phosphorylation
PI3K	Phosphatidyl 3-kinase
PIC	Preinitiation Complex
POI	Protein of interest
Pol I	RNA polymerase I
Pol II	RNA polymerase II
Pol III	RNA polymerase III
PRC1	Polycomb repressor complex 1
PRC2	Polycomb repressor complex 2
pre-mRNA	Precursor messenger RNA
PRMT	Protein arginine methyltransferase
PROMPT	Promoter upstream transcript
PRP28	Pre-mRNA processing 28
P-TEFb	Positive transcription elongation factor b
PTM	Post-translational modification
PWAPA	PHD/WH domain in ASH2L, PHF1 and ATAC2
qPCR	Quantitative polymerase chain reaction
Rb	Retinoblastoma
RBX1	RING-box protein 1
RFP	Red fluorescent protein
R-loop	Three stranded structure of DNA and RNA; R reflects RNA
RNA	Ribonucleic acid
RNF20 or 40	Ring finger protein 20 or 40
RPB	RNA polymerase B
RPG	Ribosomal protein genes
rRNA	Ribosomal RNA
RT	Room temperature
RT-qPCR	Reverse Transcription followed by quantitative Polymerase Chain Reaction
RTR1	Regulator of transcription 1
RS	Reactive side chain
S2 cells	Schneider 2 cells from *Drosophila melanogaster*
s^{4}U	4-thiouridine
SAGA	Spt-Ada-Gcn5 Acetyltransferase complex
Sall4	Sal-like protein 4
SALSA	SAGA altered, Spt8 absent
SAM	S-adenosylmethionine
SANT	Swi3, Ada2, N-Cor and TFIIIB

SAV	Streptavidin
SC	Synthetic complete medium
SCA7	Spinocerebellar ataxia type 7
SCF	Skp, Cullin, F-box containing
SCP1-3	Synaptonemal complex protein 1
SEC	Super-elongation complex
SEP	Shp1, Eyc, P47
seq	High throughput sequencing
SET	Su(var)3-9 enhancer-of-zeste trithorax
SETD1A or 1B	SET domain containing 1A or 1B
SETDB1	SET domain bifurcated histone lysine methyltransferase 1
SETX	Senataxin
Ser2P	Phosphorylation of Serine 2 in C-terminal domain of RNA polymerase II
Ser5P	Phosphorylation of Serine 5 in C-terminal domain of RNA polymerase II
Sgf11, Sgf29 or Sgf73	SAGA-associated factor 11 or 29 or 73
sgRNA	Single guide RNA
SHL	Superhelical location
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SKP1	S-phase kinase associated protein 1
SLBP	Stem-loop binding protein
SLIK	SAGA-like complex
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
snoRNA	Small nucleolar RNA
Sox2	SRY-box transcription factor 2
S. pombe	*Schizosaccharomyces pombe*
SPT/SUPT	Suppressor of Ty
Stat3	Signal transducer and activator of transcription 3
SUV39H1 or 2	Suppressor of variegation 3-9 homolog 1 or 2
Sus1	Sucrose synthase 1
SSRP1	Structure specific recognition protein 1
SSU72	Suppressor of sua7, gene 2
SV40	Simian virus 40 (here SV40 promoter)
SWI/SNF	Switch/sucrose non-fermentable
SWIRM	Swi3, Rsc8 and Moira
TAD	Topological associated domain
TADA1, 2a, 2b or 3	Transcriptional adaptor 1, 2a, 2b or 3
TAF	TBP-associated factors
TAF5L or TAF6L	TBP-associated factors 5 like or 6 like

TAND1 or TAND2	TAF1 Nitrogen-terminal domain 1 or 2
TBP	TATA-box binding protein
Tcf3	Transcription factor 3
TCT	Polypyrimidine initiator
TET	Ten eleven translocation
TF	Transcription factor
TFBS	Transcription factor binding sites
Tfcp2l1	Transcription factor CP2 like 1
TFIIA to TFIIH	Transcription factor of RNA polymerase II A to H
TIP60	Tat interactive protein 60-kDa
TIR1	Transport inhibitor response 1
Tn5	Transposon 5 (here Tn5 transposase)
TOR	Target of rapamycin
TREX2	Transcription export 2
TRRAP	Transformation/transcription domain-associated protein
TSS	Transcription start site
TTS	Transcription termination site
uaRNA	Upstream antisense RNA
U1 to U7 RNA	Uracil-rich 1 to 7 RNA
UAS	Upstream activating sequencing
Ub	Ubiquitin
Ubp8	Ubiquitin-binding protein 8
UcH	Ubiquitin C-terminal hydrolases
USP22 or 27X or 51	Ubiquitin specific peptidase 22 or 27X or 51
UTX or UTY	Ubiquitously transcribed tetratricopeptide repeat, X or Y chromosome
VANIMA	Versatile antibody-based imaging approach
VP16	Virion protein 16
WB	Western blot
WD40	~40 amino acid long motif terminating with tryptophan (W) and aspartate (D)
WDR5	WD repeat domain 5
WT	Wildtype
XPB or XPD	Xeroderma pigmentosum type B or D
XRN2	5'-3' exoribonuclease 2
YEATS	Yaf9, ENL, AF9, Taf14 and Sas5
YEATS2	YEATS domain-containing protein 2
YPD	Yeast extract peptone dextrose medium
Zn	Zinc
ZnF	Zinc finger
ZnF-UBP	Zinc finger ubiquitin-binding domain
ZZZ3	Zinc finger ZZ-type containing 3

Introduction

Introduction

Cells represent the basic unit of living organisms. In their simplest form, as prokaryotes, they consist of viscous cytoplasm surrounded by a bilayer membrane that isolates the cell from the outside environment. In eukaryotes, the cytoplasm is speckled with several additional organelles with various crucial functions such as the cell nucleus in which most of the genetic information is stored. The genome is composed of deoxyribonucleic acid (DNA) molecules in which genes are the functional units. Information contained within genes is copied into functional ribonucleic acid (RNA) molecules during the process of transcription by DNA-dependent RNA polymerase machineries. RNA has several important functions within the cellular context and various classes of RNA molecules have been described. The most frequently studied class of RNA molecules are messenger RNAs (mRNAs) which encode for cellular proteins and originate from roughly 20.000 genes in human genomes (Lander et al., 2001). The remaining classes of RNAs are non-protein coding and are implicated in very diverse cellular processes. Examples of non-coding RNAs include ribosomal (rRNA), transfer (tRNA), small nucleolar (snoRNA), small nuclear (snRNA) and long non-coding (lncRNA).

In prokaryotes, RNA molecules of different categories are transcribed by a single RNA polymerase (Werner & Grohmann, 2011). In contrast, eukaryotic transcription in the nucleus involves three distinct RNA polymerases each transcribing different classes of genes: Class I, II and III. Transcription by RNA polymerase I (Pol I) is exclusive to the synthesis of the 45S pre-rRNA molecules occurring within its own nuclear compartments, the nucleoli. The pre-rRNA is further matured into 28S, 18S and 5.8S rRNA which are the nucleic acid components of ribosomes, key machineries of protein synthesis. RNA polymerase II (Pol II) is the enzyme responsible for the transcription of all protein-coding genes as well as the synthesis of several non-coding RNAs such as lncRNAs, most snRNAs and snoRNAs. Although,

Figure 7: Scheme of the transcription cycle of RNA polymerase II. Steps include transcription initiation with 5' capping, pausing and final activation of RNA polymerase II (Pol II) for productive elongation. Introns of the nascent mRNA molecule can be co-transcriptionally spliced. The last step is transcription termination with 3' processing including polyadenylation (black box: pA, polyadenylation signal). Pol II is assisted during the transcription cycle by numerous factors (shown in purple, red and orange). From Cramer, 2019.
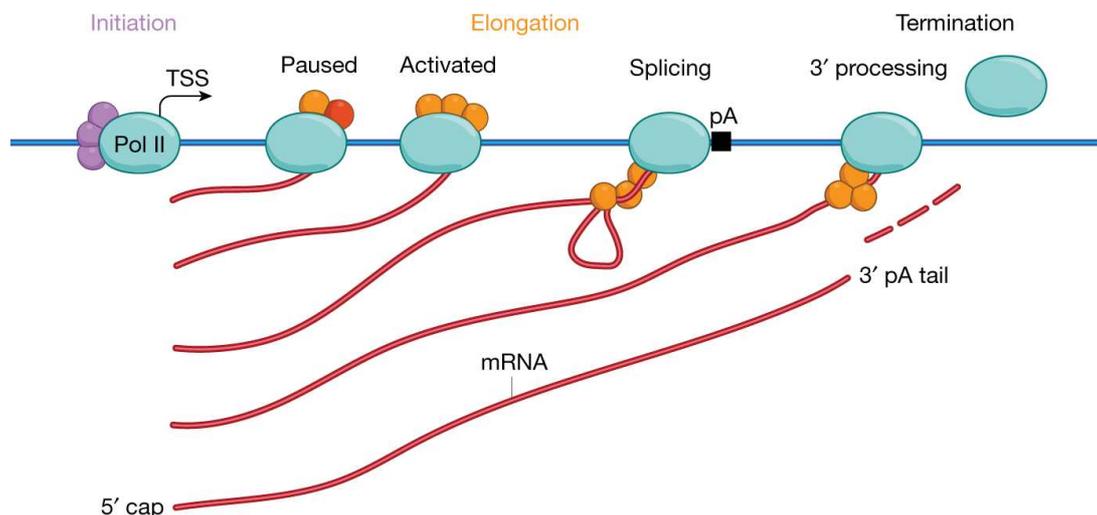
mRNA molecules represent a minor fraction (less than 5%) of all cellular RNAs, they are produced from the most diverse and complex class of genes when considering gene size, structure and organisation. Class III genes are transcribed by RNA polymerase III (Pol III) to produce tRNA molecules, 5S rRNA, and other non-coding RNAs such as the U6 snRNA or the 7SK RNA.

All RNA polymerases transcribe DNA into RNA by completing transcription cycles (shown in Figure 7 is the transcription cycle of Pol II) (Cramer, 2019). A transcription cycle is composed of three main steps: transcription initiation, elongation and termination. In eukaryotes, each of the three RNA polymerases requires specific factors to form so-called preinitiation complexes (PIC) which enable the RNA polymerases to be recruited and to initiate transcription at the class-specific genes. After initiation of transcription at transcription start sites (TSS) of genes, RNA polymerases proceed into the elongation phase during which the RNA molecule is synthesized. In metazoans, an additional phase, transcription pausing, has recently been described to occur at most genes of class II before activation into productive elongation (Levine, 2011). Additionally, introns of the nascent RNA molecule of Pol II-transcribed genes can be spliced co-transcriptionally. Transcription completes at transcription termination sites (TTS), which includes for most protein-coding genes the addition of a polyadenylation (polyA) tail during 3' processing of the nascent mRNA molecule.

The RNA polymerases are influenced and assisted by numerous factors during their transcription cycles. A major ongoing focus of research in molecular biology is the deciphering of these regulatory factors especially those controlling Pol II-mediated transcription. Regulation of Pol II transcription, and thereby notably the synthesis of protein-coding mRNAs, is crucial for the proper development of organisms composed of cells with defined and specialised functions. Once established, these specialised cells also maintain their cellular identity principally through controlling Pol II-mediated transcription. Regulation of Pol II transcription is further required for the induction of appropriate responses of cells to various external stimuli and consequently adaptation to changing environments. A key regulatory step in gene expression is the initiation of transcription by Pol II, which specifically involves the formation of the PIC composed of Pol II and six additional general transcription factors (GTFs) at gene promoters. The assembly of the PIC is thought to be especially regulated by the action of DNA-binding transcription factors (TFs) and transcriptional coactivator complexes.

In the Introduction, I will describe in detail the different phases of the Pol II transcription cycle including factors involved in their regulation with a focus on the context of mammalian cells. This also involves the brief description of important characteristics of Pol II transcription such as the establishment of enhancer-promoter contacts, transcription bidirectionality and transcription bursting. Further, I will discuss the influence of chromatin on Pol II transcription and how TFs and transcriptional coactivators can open the chromatin structure at gene promoters to enable PIC formation. As my work was mainly directed in understanding the functional roles of the chromatin modifying complexes ATAC (Ada-Two-A-containing) and SAGA (Spt-Ada-Gcn5-Acetyltransferase), I will describe in more details

the knowledge existing on these two coactivators, which share a common histone acetyltransferase (HAT) activity. To investigate the importance of SAGA and ATAC for Pol II transcription in mammalian cells, I altered the functions of these two complexes in mouse embryonic stem cells and will thus explain in the last section of the Introduction the cell biology and physiology of these pluripotent cells.

1. The RNA polymerase II transcription machinery

1.1. RNA polymerase II

Discovered in 1969, RNA polymerase II (Pol II) was soon found to be composed of twelve subunits (RPB1 to RPB12; the numbering is based on the decreasing size of the proteins in budding yeast) (Roeder & Rutter, 1969; Kedinger et al., 1970; Werner & Grohmann, 2011). The catalytic core of Pol II consists of 10-subunits and is sufficient for *in vitro* RNA transcription, while the remaining two subunits (RPB4 and RPB7) are necessary for *in vivo* transcription initiation (Cramer, 2004). Five subunits (RBP5, RPB6, RPB8, RPB10 and RPB12) are identical amongst the three nuclear eukaryotic RNA polymerases, Pol I, II and III. High-resolution structures of the eukaryotic 10-subunit core Pol II were first revealed in 2000 and 2001 (Cramer et al., 2000; Cramer et al., 2001; Gnatt et al., 2001) shortly followed by the complete structure of the 12-subunit Pol II (Figure 8A) (Craighead et al., 2002; Bushnell & Kornberg, 2003; Armache et al., 2003). These structural studies allowed the division of Pol II into domain-like regions based on their functions within the complex (Figure 8B).
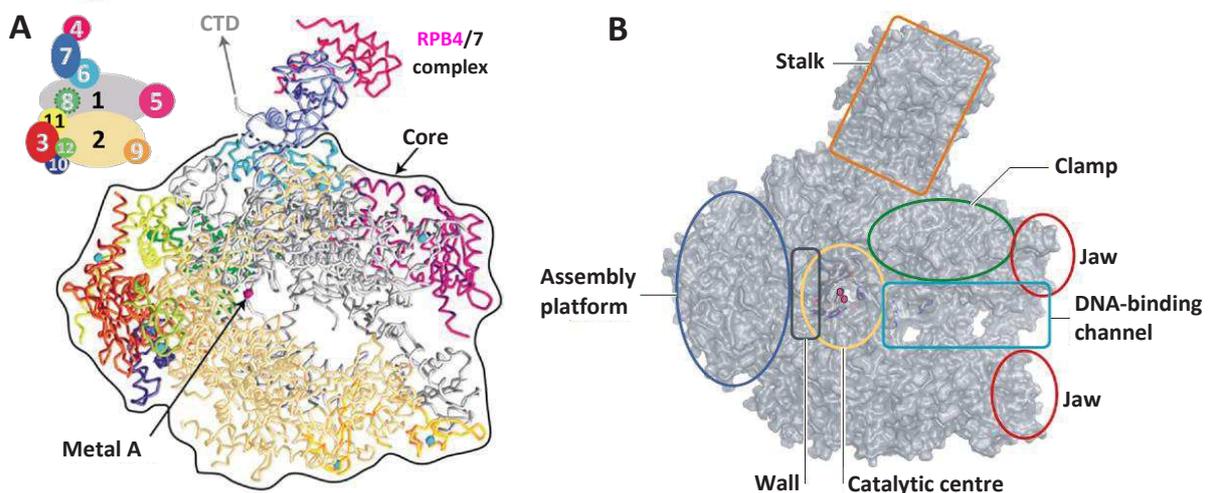
Figure 8: Architecture and structure of RNA polymerase II. A. Shown is the 12-subunit crystal structure of yeast RNA polymerase II (Pol II). Highlighted are key structures such as the 10-subunit core, the active site with Metal A, the RPB4/7 complex and the position of the Carboxy-terminal repeat domain (CTD) of Pol II. Subunits are coloured following the schematic diagram on the left. A pink sphere reflects the location of the Metal A Mg^{2+} ion of the active site. Adapted from Cramer, 2004. **B.** Scheme highlighting different important domains within Pol II. The following key features are shown: the catalytic centre (yellow) with the two Mg^{2+} ions in magenta, the jaws (red), the wall (black), the clamp (green) and the stalk (orange) domains as well as the assembly platform (darkblue) and the DNA-binding channel (lightblue). From Werner & Grohmann, 2011.

The assembly of Pol II is enabled by the RPB10 and RPB12 subunits acting as structural adaptors between RPB2 and the RPB3-RPB11 dimer forming the so-called 'assembly platform' (Werner & Grohmann, 2011). The two largest subunits, RPB1 and RPB2, form the crab claw-shaped 'cleft' which harbours the catalytic centre containing two Mg^{2+} ions, one bound permanently (Metal A) while the other seems more exchangeable (Metal B) (Cramer, 2004). The 'jaws' formed by RPB5 and RPB1/RPB9 respectively, represent the entrance to the cleft and can interact with the downstream DNA template (Cramer, 2004; Werner & Grohmann, 2011). Besides forming one of the jaws, RPB9, a subunit without homologue in prokaryotes, is further thought to influence TSS selection and transcription fidelity through interactions with other basal transcription factors such as TFIIF.

The double-stranded DNA template enters in the 'DNA binding channel' formed by the Pol II cleft until it reaches the 'wall' region formed by RPB2 which blocks the end of the cleft. RPB1 and RPB2 also constitute a mobile surface referred to as the 'clamp', which can open the cleft and close onto the duplex DNA template. During transcription elongation, distinct structural features of this domain further secure as well as separate the DNA-RNA hybrid while rising upwards from the active site perpendicular to the double-stranded DNA template. The nascent RNA transcript exits Pol II through a channel between the assembly platform and the 'stalk' domain, formed by RPB4 and RPB7 (Werner & Grohmann, 2011).

The stalk domain was reported to restrict the movement of the clamp forcing it into a closed position thereby promoting DNA strand melting during transcription initiation (Cramer, 2004; Werner & Grohmann, 2011). This domain further serves as recruitment and interaction surface for other factors especially transcription initiation factors and is connected to the remaining Pol II through a surface formed by RPB1, RPB2 and RPB6 (Hahn, 2004). Located close to the stalk domain and the RNA exit funnel is the Carboxy-terminal repeat domain (CTD) tail of RPB1 which is connected through a linker sequence to the remaining protein. Its function and role will be described in more details in the following subsection. The CTD and the linker sequence cannot be detected in Pol II crystal structures due to their highly unstructured and flexible nature (Cramer, 2004).

A secondary funnel below the active site represents the exit channel for the RNA transcript during backtracking of Pol II and also the entry channel for cleavage factors involved in resolving Pol II backtracking (described in more details in chapter 1.4.2. RNA polymerase II backtracking). This second pore also serves as entry channel for ribonucleotides. Located close to the secondary funnel, RPB8 has been suggested to interact with the nascent transcript in the backtracked complex but its definitive role remains unclear (Werner & Grohmann, 2011).

1.1.1. The Carboxy-terminal repeat domain of RNA polymerase II

In contrast to the other RNA polymerases, Pol II possesses a unique and remarkably relevant feature: the highly modifiable and repetitive Carboxy-terminal repeat domain (CTD) of its largest subunit, RPB1 (Buratowski, 2009; Hsin & Manley, 2012). The CTD of Pol II consists of tandem heptad repeats of the following amino acid consensus sequence: Tyrosine-Serine-Proline-Threonine-Serine-Proline-Serine (also represented as $Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7$ or $Y_1-S_2-P_3-T_4-S_5-P_6-S_7$,) (Buratowski, 2009; Bartkowiak & Greenleaf, 2011). In yeast species, the CTD constitutes 26-29 tandem repeats, while in *Drosophila melanogaster* (hitherto referred as *Drosophila* or fly) it is formed of 45 repeats and fish, mouse and human RPB1 possess 52 Carboxy-terminal repeats (Bartkowiak & Greenleaf, 2011; Hsin & Manley, 2012). In vertebrates, 21 of the tandem heptad repeats match the conserved consensus sequence and are mostly located in proximity to the Pol II body. The remaining, largely distal repeats are more degenerated and deviate from the canonical sequence especially at positions 7 (Hsin & Manley, 2012). An additional ten amino acid long sequence at the very end helps to stabilize the CTD (Chapman et al., 2004; Hsin & Manley, 2012). Intriguingly, although the CTD is dispensable for *in vitro* RNA synthesis, reduction of the heptad repeat number under a minimum length *in vivo* (e.g. eight repeats in yeast, 26 in mouse) leads to lethality (Nonet et al., 1987; Zehring et al., 1988; West & Corden, 1995; Phatnani & Greenleaf, 2006; Hsin & Manley, 2012). Also, a fully extended CTD of RPB1 was suggested to reach out close to 1000 Å from Pol II. This represents a distance of roughly seven times the diameter of Pol II (Cramer, 2004).

The differences in Pol II CTD length between organism were recently suggested to lead to functional differences (Boehning et al., 2018). The CTD of Pol II represents an intrinsically disordered low-complexity domain (LCD). This property was reported to allow the CTD to undergo cooperative liquid-liquid phase separation (LPS) *in vitro*. These *in vitro* findings were suggested to explain the nature of unphosphorylated Pol II clusters *in vivo* (Boehning et al., 2018). LPS properties of the CTD was described to depend on the length of the CTD: The shorter yeast CTD formed less-stable droplets and less *in vivo* Pol II clusters as did a truncated version of the human CTD in contrast to the full length human CTD (Boehning et al., 2018). However, these results should be considered carefully as another study, published around the same time, suggested that Pol II CTD by itself cannot form LPS particles *in vitro* (Lu et al., 2018).

The CTD of Pol II has several well-established roles within the transcription cycle by serving as a selective recruitment platform, at the appropriate stages of transcription, for a variety of nuclear factors involved in mRNA processing (Buratowski, 2009; Bartkowiak & Greenleaf, 2011). Upon recruitment to gene promoters and transcription initiation, the CTD of Pol II is mostly unphosphorylated (Figure 9). This changes in the course of the transcription cycle, during which the CTD is subjected to dynamic modifications of its repeat residues. The most abundantly found modifications of the CTD are phosphorylation of Serine 2 (Ser2P) and phosphorylation of Serine 5 (Ser5P). Recent, very elegant
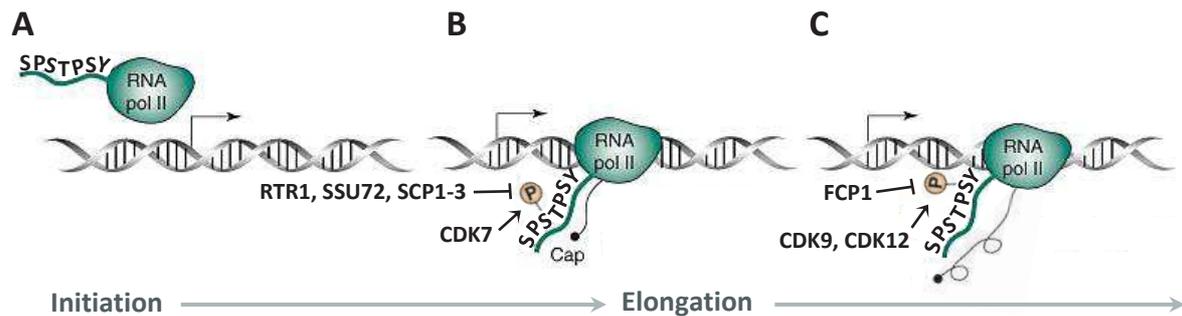
Figure 9: Simplified scheme of main phosphorylation sites and involved enzymes of the Carboxy-terminal repeat domain of RPB1. A. RNA polymerase II (RNA pol II) is recruited to gene promoters. The Carboxy-terminal repeat domain (CTD) of RPB1 consists of 26-52 heptad repeats of the amino acid sequence Y_1-S_2-P_3-T_4-S_5-P_6-S_7 and is unphosphorylated at this stage. For simplicity, only on representative repeat is depicted here. **B.** During transcription initiation, Serine 5 residues of the CTD tail get phosphorylated (Ser5P) by CDK7 of the TFIIH complex. This facilitates the recruitment of 5' RNA capping enzymes (cap shown as black sphere). RTR1, SSU72 and SCP1-3 represent Ser5P phosphatases. **C.** Upon entry into productive elongation, the CTD gets phosphorylated at its Serine 2 residues (Ser2P) by CDK9 of P-TEFb and to a lesser extent by CDK12. This enables the recruitment of several transcription-related factors such as the splicing machinery. FCP1 acts as Ser2P phosphatase. Adapted from Egloff & Murphy, 2008 and Hsin & Manley, 2012.

studies revealed that these modifications can occur generally at all heptad repeats within the CTD (Schüller et al., 2016; Suh et al., 2016). These studies further found that the CTD of Pol II does not appear to be heavily phosphorylated: Within a heptad repeat only rarely more than one residue is modified. Several other CTD modifications such as Tyrosine 1, Tyrosine 4 and Serine 7 phosphorylation are detectable beside Ser2P and Ser5P but at lower frequencies (Hsin & Manley, 2012; Schüller et al., 2016; Suh et al., 2016). Also acetylation, glycosylation, methylation, ubiquitinylation and proline isomerisation of CTD residues were reported (Hsin & Manley, 2012). The role and importance of these low frequency CTD modifications within the transcription cycle is not yet clearly defined (Hsin & Manley, 2012). The following paragraphs will therefore focus exclusively on Ser2P and Ser5P (summarized in Figure 9).

Phosphorylation of Serine 5 of the CTD has been linked to processes related to transcription initiation as it is found at high levels predominately around transcription start sites (TSS) of genes. Its presence at these regions is explained by the fact that Ser5 is phosphorylated through the CDK7 (cyclin dependent kinase 7) subunit of the general transcription factor TFIIH, which is part of the preinitiation complex required for Pol II transcription initiation (discussed in more details in chapter 1.2.2. Preinitiation complex formation). The kinase activity of CDK7 was described to be stimulated by the multiprotein complex Mediator (described in more details in a later chapter) (Sikorski & Buratowski, 2009; Plaschka et al., 2015).

Ser5P was linked to a destabilization of contacts formed by Pol II with its initiation factors and is additionally crucial for the recruitment of mRNA capping enzymes (Egloff & Murphy, 2008; Hsin & Manley, 2012). Factors involved in 5' mRNA capping were shown to bind directly to Ser5P residues of the CTD (Fabrega et al., 2003). This binding brings the capping machinery in close proximity to the

RNA exit channel of Pol II and allows for rapid 5' mRNA modification (Fabrega et al., 2003). The 5' cap of mRNAs consists of a 7-methylguanosine modification which ensures RNA stability and is critical for several downstream processes such as protein translation (Proudfoot et al., 2002). Phosphorylation of Ser5 by CDK7 was also recently implicated in dissolving LPS particles induced, in the first place, by unphosphorylated CTD (Boehning et al., 2018).

Ser5P levels start declining hundreds of nucleotides downstream of the TSS as Pol II moves away from promoters and Ser5p gets gradually removed through the action of phosphatases (Komarnitsky et al., 2000; Buratowski, 2009). The dephosphorylation of Ser5 is thought to be accomplished by the evolutionarily conserved SSU72 and RTR1 phosphatase proteins and the higher eukaryote-specific SCP1-3 phosphatases (Mosley et al., 2009; Hsin & Manley, 2012). While Ser5P levels are decreasing, Ser2P levels increase gradually as Pol II elongates into the gene body and accumulate towards the end of genes.

Ser2P has been correlated with transcription elongation several years ago (Komarnitsky et al., 2000), nevertheless elongation rates of Pol II are not directly affected by the phosphorylation of Ser2. Instead numerous interactions with other factors regulating transcription elongation are enabled through this modification. For example, Ser2P was linked to the recruitment of mRNA processing factors such as the splicing machinery and factors involved in mRNA polyadenylation and transcription termination (Mortillaro et al., 1996; Licatalosi et al., 2002; Meinhart & Cramer, 2004; Buratowski, 2009; Hsin & Manley, 2012). In mammalian cells, Ser2P of the CTD is mediated mainly by the kinase activity of P-TEFb (positive transcription elongation factor b), which consists of CDK9 and Cyclin T, and to a lesser extent by CDK12 (Egloff & Murphy, 2008; Bartkowiak et al., 2010; Bartkowiak & Greenleaf, 2011). Besides the CTD, P-TEFb also phosphorylates the transcription elongation factor DSIF and the pausing factor NELF (described in more details in section 1.3. Transcription pausing). During elongation, the opposing enzyme, the FCP1 phosphatase, dephosphorylates mainly Ser2P with a weaker activity also towards Ser5P (Cho et al., 2001; Ghosh et al., 2008; Hsin & Manley, 2012; Hsu et al., 2019a).

Recent studies also suggest that the kinases involved in Ser2 phosphorylation possess LCD and can induce LPS *in vitro* (Lu et al., 2018). The formation of LPS particles *in vivo* was proposed to allow a compartmentalization of the kinases with their substrate, the Pol II CTD, thereby leading to efficient hyperphosphorylation of the CTD (Lu et al., 2018).

1.2. Transcription initiation

Transcription initiation represents the first step of the transcription cycle. This step allows the recruitment and binding of RNA polymerases to gene promoters and the initiation of transcription from the transcription start site (TSS). It is widely accepted as the major regulatory event in defining gene

expression programs and consequently controlling the response of cells to their environment, cell differentiation and cell identity (Sainsbury et al., 2015). Several intrinsic (*cis*-regulatory) elements and extrinsic (*trans*-regulatory) factors dictate the process of transcription initiation. Among the *cis*-regulatory DNA elements are core promoters and proximal or distal regulatory elements, while *trans*-regulatory factors are very diverse from general transcription factors (GTFs), required for transcription initiation at almost every gene, to gene-specific transcription factors (TFs). The following subsections will summarize knowledge acquired over the past years on these different elements and factors with the exception of TFs which will be described in a later chapter.

1.2.1.   *Cis*-regulatory DNA elements

As mentioned earlier, two main categories of *cis*-regulatory elements encoded by DNA were defined and are involved in transcription initiation: core promoters and proximal or distal regulatory sequences, such as enhancers. They represent DNA sequences of very diverse composition and nature within the eukaryotic and especially the mammalian domain. Distal regulatory elements, commonly referred to as enhancers, are generally believed to stimulate transcription at gene promoters. A frequently encountered definition of gene promoters includes the core promoter, which will be described in more details next, within a larger promoter region containing proximal regulatory sequences, such as binding sites for TFs (Lenhard et al., 2012; Haberle & Stark, 2018).

1.2.1.1. Core promoters

Core promoters are defined as 'the minimal DNA sequence that directs accurate initiation of transcription' and encompass around 50 base pairs (bp) up- and downstream of the TSS, which is designated as the +1 position (Danino et al., 2015). Several DNA motifs have been identified within metazoan core promoters that facilitate correct positioning and orientation of the basal transcription machinery represented by the PIC (more details see chapter 1.2.2. Preinitiation complex formation). In the following, the following motifs will be described in more details: the TATA-box, the BRE (TFIIB recognition elements), the Inr (Initiator), the TCT (polypyrimidine initiator), the MTE (motif ten element), the DPE (downstream promoter element) and the DCE (downstream core element) motif (Figure 10).

In contrast to what Figure 10 might imply, these different DNA elements are not found at all promoters and seem not absolutely essential for promoter function as suggested by the fact that many human promoters lack any of these motifs (Hahn, 2004; Cramer, 2019). Instead, the varying architecture and motif composition of core promoters was proposed to result into variations in expression levels and TSS choice (Hahn, 2004; Lenhard et al., 2012; Danino et al., 2015; Schor et al., 2017). Other
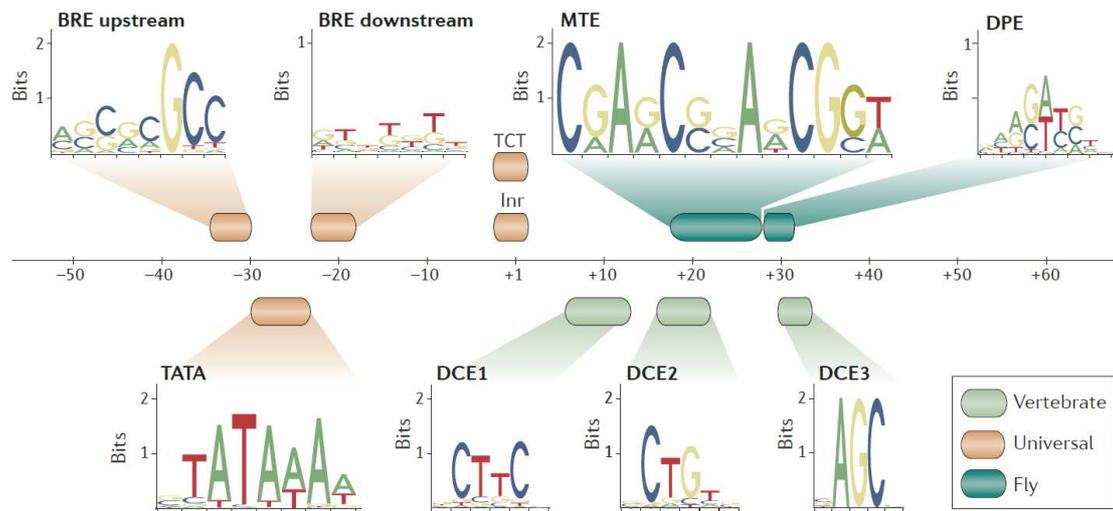
Figure 10: Sequence motif and localization of metazoan core promoter elements. Scheme shows several known core promoter motifs found in core promoters of metazoans and their localization relative to the transcription start site (TSS) at +1. BRE, TFIIB recognition element; DCE, downstream core element; DPE, DNA promoter element; Inr, Initiator; MTE, motif ten element; TATA, TATA-box; TCT, polypyrimidine initiator. Adapted from Lenhard et al., 2012.

characteristics of core promoters such as AT-content and DNA 'bendability' have further been suggested to allow promoter recognition (Levens et al., 2016; Haberle & Stark, 2018).

As mentioned above, core promoter motifs were reported to be involved in TSS selection. Three main types of transcription initiation patterns were identified by genome-wide sequencing of the 5'ends of mRNAs by for example CAGE (cap analysis of gene expression): Sharp promoters with transcription initiation from one or few TSSs in a very narrow region (also known as focused or peaked); Broad promoters with several weak TSSs within a larger region of the core promoter (also known as dispersed promoters); And mixed promoters with dispersed initiation patterns but one dominant TSS (Figure 11) (Juven-Gershon & Kadonaga, 2010; Lenhard et al., 2012; Kadonaga, 2012; Danino et al., 2015). A combination of two core promoter elements, a TATA-box at roughly 30 bp upstream of an Inr element was suggested to favour a sharp transcription initiation pattern (more details on the TATA-box and the Inr elements in next subsection) (Ponjavic et al., 2006; Lenhard et al., 2012). Promoters with a sharp TSS are often found within regulated tissue-specific genes, while broad TSS patterns are enriched at constitutive or housekeeping genes (Lenhard et al., 2012; Danino et al., 2015). In vertebrates, over 70% of all promoters were reported to have a dispersed transcription initiation pattern (Bajic et al., 2006; Danino et al., 2015).
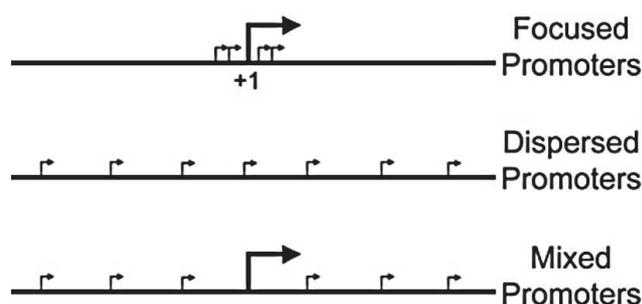
Figure 11: Classification of core promoters based on transcription initiation patterns. Illustrated are the three main types of core promoters identified based on the pattern of transcription start site (TSS) selection. TSSs are indicated by arrows, with weak and strong TSSs shown as small and big arrows, respectively. From Danino et al., 2015.

1.2.1.1.1. Core promoter elements

The first identified and most extensively studied core promoter element is the TATA-box, which was found in all studied eukaryotes, from yeast to metazoans (Danino et al., 2015). It represents the binding site of TBP (TATA-box binding protein), a factor involved in the recruitment of the basal transcription machinery to gene promoters. Its consensus sequence is TATAWAWR (where W stands for A/T and R for A/G), which is at the origin of its name, and localizes roughly 30 bp upstream from the TSS in mammals (O'Shea-Greenfield & Smale, 1992; Basehoar et al., 2004). In humans, only roughly 10% of core promoters were reported to contain a consensus TATA-box and additional 24% were found to possess a TATA-like elements, a variant TATA-box motif with up to two nucleotide substitutions (Yang et al., 2007; Danino et al., 2015). Most human promoters (> 70%) are thought to be TATA-less (Yang et al., 2007). In metazoan, core promoters with consensus TATA-box are often found at highly regulated genes such as genes involved in stress response or tissue-specific genes such as olfactory receptor or liver-specific genes (Yang et al., 2007; Lenhard et al., 2012).

The conserved BRE (TFIIB recognition elements) were found as sequences required for the binding of TFIIB, a general transcription factor described in more details in chapter 1.2.2. Preinitiation complex formation, and are often present in conjunction with a TATA-box (Deng & Roberts, 2007; Danino et al., 2015). The consensus motif of the upstream BRE (BRE^u) is SSRCGCC (where S stands for G/C), while the downstream BRE (BRE^d) consensus consists of a RTDKKKK (where D stands for A/G/T and K stands for G/T) sequence.

The Inr (Initiator) element is also present from yeast to metazoans although with differing consensus sequences between and among species (Haberle & Stark, 2018). For example, the consensus sequence of the mammalian Inr element is not clearly defined and varies between studies ranging from YYA$_{+1}$NWYY (where Y stands for C/T and N represents any nucleotide A/G/C/T) to the dinucleotide YR$_{+1}$ (Lenhard et al., 2012; Danino et al., 2015). The Inr overlaps directly with the TSS (+1) and is more abundant than the TATA-box although not universal (Yang et al., 2007; Haberle & Stark, 2018). Close to half of all human core promoters were reported to contain a consensus Inr (YYANWYY) sequence (Yang et al., 2007). Inr is thought to be bound by the TAF1 and TAF2 subunits of the TFIID complex, another general transcription factor that will be described in more details in chapter 1.2.2. Preinitiation complex formation. Intriguingly, TATA-like and consensus Inr sequences are absent in roughly 46% of human core promoters (Yang et al., 2007).

The TCT (polypyrimidine initiator) motif, similarly to the Inr, encompasses the TSS and is enriched especially at highly expressed genes related to translation such as those encoding ribosomal proteins or translation initiation and elongation factors (Parry et al., 2010; Lenhard et al., 2012). The TCT motif in combination with other core promoter elements such as a TATA-box was suggested to allow for high-level constitutive expression of these genes in all cell types (Lenhard et al., 2012). The consensus TCT motif is YC$_{+1}$TYTYY in humans (Danino et al., 2015).

Additional core promoter elements can be found downstream of the TSS. This includes the MTE (motif ten element), the DPE (downstream promoter element) and the DCE (downstream core element) (Lenhard et al., 2012; Sainsbury et al., 2015). They represent other core promoter motifs thought to be bound by the TFIID complex (Danino et al., 2015; Sainsbury et al., 2015). The DPE is located around 30 bp downstream of the TSS, while the MTE is localised immediately upstream of the DPE starting around +20 bp (Danino et al., 2015). The DPE consensus sequence is DSWYVY (V for A/C/G) and the MTE consensus was defined as CSARCSSAACGS. Both were described to be dependent on the existence of a functional Inr element and enriched in TATA-less core promoters (Danino et al., 2015; Sainsbury et al., 2015). The DCE consists of three elements with the necessary motifs CTTC (+6 to +11), CTGT (+16 to +21) and AGC (+30 to +34). This motif is often found associated with the presence of a TATA-box within the core promoter (Danino et al., 2015).

1.2.1.1.2. CpG islands

In addition, in vertebrates, more than half of the core promoters are thought to overlap with so-called CpG islands (CGI) (Jones, 2012). CpG islands represent a stretch of DNA of on average roughly 1 kb with elevated GC content and a high frequency of the dinucleotide CpG (p represents the phosphodiester bond between the nucleotides) (Gardiner-Garden & Frommer, 1987; Haberle & Stark, 2018; Greenberg & Bourc'his, 2019). Promoters of the majority of housekeeping genes were found associated with one short CGI, while promoters of developmentally-regulated genes overlap with several larger CGI (Lenhard et al., 2012; Greenberg & Bourc'his, 2019).

CGI represent targets for epigenetic regulations since the carbon-5 of cytosine within the CpG dinucleotide can be methylated (5-methylcytosine, 5mC) (Jones, 2012; Greenberg & Bourc'his, 2019). Methylation of CpGs in the vicinity of TSSs was found to correlate with transcriptional inactivation and is linked to long-term gene silencing as observed at imprinted genes, genes of the inactivated X-chromosome, germline-specific genes and transposable elements (Jones, 2012; Greenberg & Bourc'his, 2019). The mechanism of silencing remains not entirely resolved (Greenberg & Bourc'his, 2019). In general, CGIs at core promoters of actively transcribed genes are rarely methylated (Jones, 2012; Greenberg & Bourc'his, 2019).

Genome-wide CpG methylation is established *de novo* by the DNA methyltransferase enzymes DNMT3A and DNMT3B in combination with their cofactor DNMT3L (Jones, 2012; Greenberg & Bourc'his, 2019). The vast majority of CpG dinucleotides will be methylated by the *de novo* DNA methyltransferases, except those belonging to CGIs. Methylation maintenance additionally involves DNMT1 and its cofactor UHRF1, which can bind to hemimethylated CpG dinucleotides (Jones, 2012; Greenberg & Bourc'his, 2019).

5mC can be removed passively during DNA replication or actively through oxidation (Jones, 2012; Greenberg & Bourc'his, 2019). The enzymes involved in active demethylation are the three TET (TET1 to 3, Ten-eleven translocation 1 to 3) methylcytosine dioxygenases (Greenberg & Bourc'his, 2019). They act by successively oxidizing 5mC into 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine (Greenberg & Bourc'his, 2019). This oxidation promotes DNA demethylation during replication and involves the base excision repair pathway (Greenberg & Bourc'his, 2019).

1.2.1.2. Enhancers

The output of core promoters can be controlled by regulatory sequences to which transcription factors such as activators can bind and are localized either proximal or distal to the core promoter. Distal elements are termed enhancers, which can be situated downstream, upstream or in introns of the targeted or completely unrelated genes (Furlong & Levine, 2018; Schoenfelder & Fraser, 2019). Typically, human core promoters receive input from multiple enhancers, which serve as transcription amplifiers (Haberle & Stark, 2018). Compellingly, gene expression levels were found to positively correlate with the number of enhancers contacting the gene promoter (Schoenfelder et al., 2015; Schoenfelder & Fraser, 2019). In mammalian genomes, hundreds of thousands of potential enhancers were identified, clearly outnumbering promoter regions of protein-coding genes. Even more (roughly 1 million) are estimated to actually exist (Schoenfelder & Fraser, 2019).

While several enhancers have the potential to interact with various promoters, preferences among enhancer-promoter pairs have been described (Zabidi et al., 2015; Danino et al., 2015; Haberle & Stark,
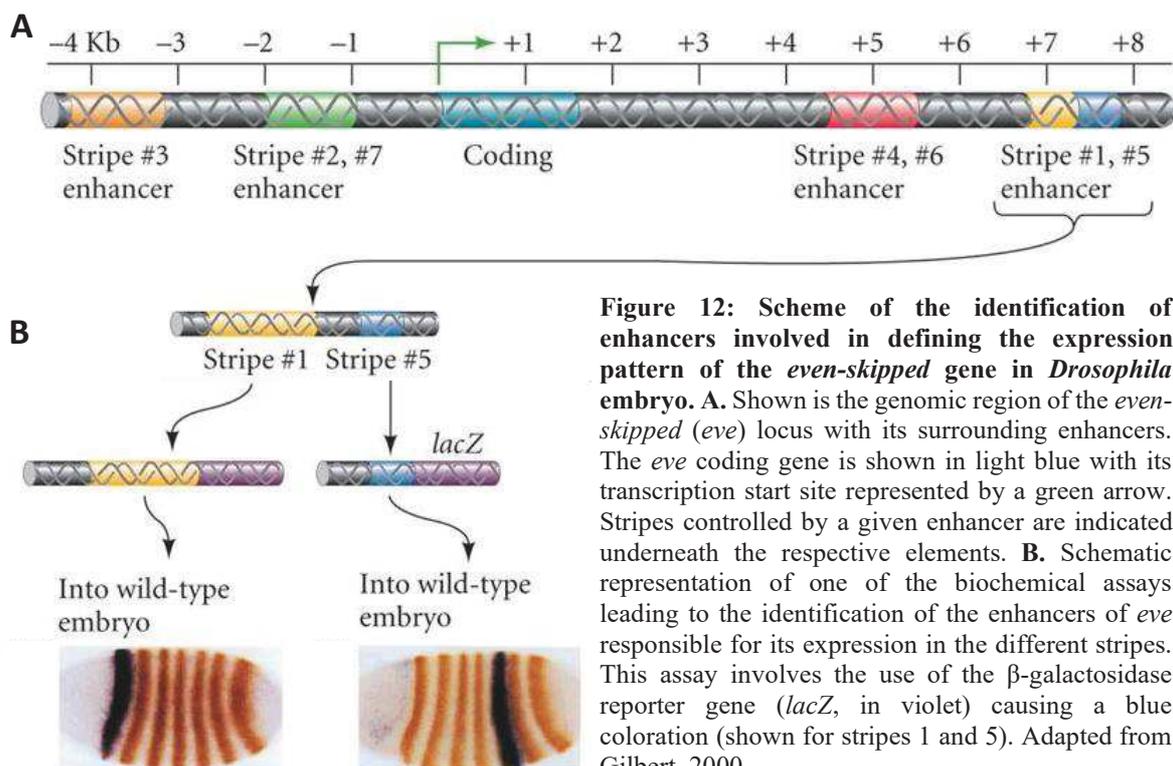
Figure 12: Scheme of the identification of enhancers involved in defining the expression pattern of the *even-skipped* gene in *Drosophila* embryo. A. Shown is the genomic region of the *even-skipped* (*eve*) locus with its surrounding enhancers. The *eve* coding gene is shown in light blue with its transcription start site represented by a green arrow. Stripes controlled by a given enhancer are indicated underneath the respective elements. **B.** Schematic representation of one of the biochemical assays leading to the identification of the enhancers of *eve* responsible for its expression in the different stripes. This assay involves the use of the β-galactosidase reporter gene (*lacZ*, in violet) causing a blue coloration (shown for stripes 1 and 5). Adapted from Gilbert, 2000.

2018; Furlong & Levine, 2018). Some enhancers have specificities by activating transcription of certain promoters but not others and similarly some promoters are exclusively induced by specific enhancers (Haberle & Stark, 2018). The mechanism behind these specificities and preferences remain unclear and likely vary between enhancer-promoter pairs (Furlong & Levine, 2018; Schoenfelder & Fraser, 2019). Individual enhancer-promoter interactions were suggested to be partially explained by the presence or absence of core promoter elements, such as TATA-box or DPE motifs, and shared transcription factor binding events (Zabidi et al., 2015; Danino et al., 2015; Haberle & Stark, 2018).

During development in metazoans, enhancers have important functions in regulating precise spatiotemporal gene expression programs (Furlong & Levine, 2018). A classic example of enhancer function in driving gene expression in space and time is the expression of the *even-skipped* (*eve*) gene in seven distinct stripes in the *Drosophila* embryo (Gilbert, 2000). Surrounding the *eve* gene locus are several enhancer units, each was found to be responsible for the induction of *eve* expression in specific stripes (Figure 12A). A powerful biochemical assay that revealed the importance of the respective enhancers in strip-patterning of *eve* was the fusion of the individual enhancer units to a β-galactosidase (*lacZ*) reporter gene (Figure 12B). This revealed, through a blue coloration caused by the expression of the *lacZ* gene, the stripes regulated by the respective enhancers. The enhancer-dependent expression of *eve* in the different stripes was further shown to depend on various activating and repressing transcription factors which bind to transcription factor binding sites (TFBS) present within the individual enhancers (Gilbert, 2000).

In vertebrates, enhancers can be located at various distances, up to millions of base pairs away with an average of roughly 20 to 50 kb, from the targeted genes (Furlong & Levine, 2018). To stimulate transcription, enhancers need to bypass intervening genes and come in proximity to their core promoter which led to the suggestion of several mechanistic models amongst which the most prominent is the formation of enhancer-promoter loops and its recent derivative the enhancer-promoter hubs (Danino et al., 2015; Furlong & Levine, 2018).

The genomes of eukaryotes were found to organize into TADs (topological associating domains) thereby bringing several enhancers and promoters in close vicinity (Furlong & Levine, 2018; Schoenfelder & Fraser, 2019). The study of the 3D organization and TADs of entire genomes has been especially enabled through methodological developments such as the Hi-C (genome-wide chromosome conformation capture assays) (Cardozo Gizzi et al., 2019). TADs are thought to be delineated by insulator proteins, such as CTCF (CCCTC-binding factor), and formed through the active extrusion of DNA loops through the ring-shaped structure of cohesin complexes until encountering the barriers formed by insulator proteins (the 'loop extrusion' model) (Furlong & Levine, 2018; Schoenfelder & Fraser, 2019).

The functional importance of TADs on enhancer-promoter associations however seems context dependent and gene specific. Several studies indicated that rearrangements of individual TADs led to

the loss of necessary enhancer-promoter interactions or gain of unwanted communications between enhancers and promoters (Furlong & Levine, 2018; Schoenfelder & Fraser, 2019). Based on these results, TADs were suggested to impose spatial restrictions or insulations required for specific and efficient enhancer-promoter interactions (Schoenfelder & Fraser, 2019). Nonetheless, recent genome-wide studies on *Drosophila* balancer chromosomes or upon acute depletion of the insulator CTCF in human cells, report that TADs seem not crucial for the expression of the majority of Pol II transcribed genes (Rao et al., 2017; Ghavi-Helm et al., 2019).

The looping model of enhancer-promoter interaction, mentioned earlier, which is potentially facilitated by the formation of TADs, is thought to involve protein-protein contacts between factors bound to enhancers and factors bound to promoters (Furlong & Levine, 2018). This protein associations were suggested to result in enhancer-promoter proximity thereby enabling transcription induction (Figure 13A). Mediator, a multiprotein complex, was suggested to enable enhancer-promoter looping by serving as an adaptor protein between activating transcription factors bound to enhancers and the Pol II transcription machinery at the core promoter (more details on this complex will be given in chapter 1.2.2.6. The Mediator complex: role in transcription initiation and initially transcribing Pol II). Some enhancer-promoter pairs are thought to be further stabilized through CTCF and cohesin. For example, in mouse embryonic stem cells, expression of the pluripotency transcription factor Sox2 was reported
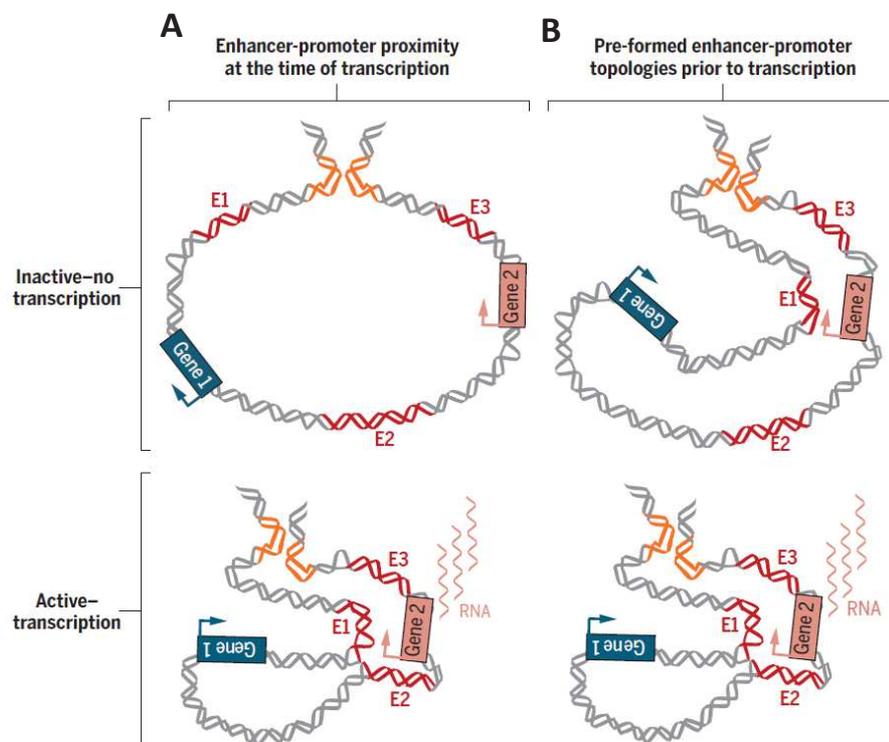
Figure 13: Scheme of looping and hub model of enhancer-promoter interactions. A. The looping model suggests that loop formation enables close enhancer-promoter contacts leading to active transcription, while the hub model **B.** has been described especially for genes at which preformed enhancer-promoter associations are found without resulting in gene expression. This model requires a secondary event to lead to transcription induction. Enhancer (E) shown in red. Insulator (e.g. CTCF) binding sequences in orange. From Furlong & Levine, 2018.

to be ensured by enhancer-promoter interactions enabled through CTCF, cohesin and Mediator (Phillips-Cremins et al., 2013; Furlong & Levine, 2018; Schoenfelder & Fraser, 2019). Intriguingly, artificially forced proximity between enhancers with their target promoters (exemplified by the β-globin promoter and its enhancer) can lead to strong transcription induction (Deng et al., 2012; Schoenfelder & Fraser, 2019).

The recently suggested hub model of enhancer-promoter interactions extends the looping model and incorporates features of other models not detailed here, such as the 'transcription factory' model (Figure 13B) (Furlong & Levine, 2018). The hub model was formulated based on recent experimental observations and suggests that at some gene promoters proximity to enhancers does not necessarily lead to transcription induction (Chen et al., 2018b; Furlong & Levine, 2018). Additional studies on long-range contact sites during *Drosophila* development or human T cell lineage differentiation reported that enhancer interactions with their target promoters are established long before induction of transcription (Spilianakis & Flavell, 2004; Ghavi-Helm et al., 2014; Furlong & Levine, 2018). For example, Ghavi-Helm et al. studied 103 developmental enhancers in *Drosophila* during mesoderm development using the 4C (chromosome conformation capture-on-chip) technology and found that 94% of these enhancers form stable contacts ahead of gene expression changes (Ghavi-Helm et al., 2014).

In the hub model, preformed loops enable the accumulation of components of the transcription machinery such as Pol II and the Mediator complex (Furlong & Levine, 2018). This accumulation might be facilitated through the recently described liquid-liquid phase separation phenomenon (Furlong & Levine, 2018). It is not clear yet how final induction of transcription from promoters within the preformed hubs is achieved. Subtle movements of enhancers or promoters within the hub, the recruitment of a specific TF or release of stalled Pol II were suggested to potentially result in ultimate transcription activation (Furlong & Levine, 2018). In this model, enhancers do not necessarily physically touch promoters but might be separated by 100 to 300 nm (Chen et al., 2018b; Lim et al., 2018; Furlong & Levine, 2018). Instead, enhancers would serve to recruit components of the transcription machinery leading to an increased local concentration of these factors close to gene promoters. An example gene that highlights the possibility of enhancer-promoter communication without proximity is the *sonic hedgehog* (*Shh*) gene (Schoenfelder & Fraser, 2019). Its expression is controlled in the developing mouse brain by several brain-specific enhancers, yet no contacts between enhancers and the target promoter could be detected by current technics.

Additionally within the mouse and human genomes, clusters of enhancers termed super-enhancers were further characterized, which differ from classic enhancers especially by their size and transcription factor density (Whyte et al., 2013; Hnisz et al., 2013). While typical enhancers extend only over several hundreds of base pairs, super-enhancers have a median size of roughly 9 kb and can span as much as 50 kb (Whyte et al., 2013). Super-enhancers were described to be especially required for the expression of

cell-type defining genes and seem to enable higher gene expression levels than regular enhancers (Whyte et al., 2013).

As a side note, in contrast to promoters, enhancers are in general thought to be CpG-poor and seem to be variably methylated (Jones, 2012). The function of 5mC in regulating enhancer activity is not yet clear, although it was suggested that CpG methylation could lead to reduced activity of enhancers (Schmidl et al., 2009; Jones, 2012).

1.2.1.3. Transcription bursting

A main feature of transcription, which has recently been emphasized through genome-wide studies and imaging technics, is transcription bursting (Raj et al., 2006; Raj & van Oudenaarden, 2008; Fukaya et al., 2016; Lenstra et al., 2016; Haberle & Stark, 2018; Larsson et al., 2019; Brouwer & Lenstra, 2019). Transcription bursting describes the phenomenon that transcription happens episodically with series of 'bursts' separated by intervals of inactivity. This is based on the stochastic nature of gene expression in which genes switch between 'on' states (active transcription) and 'off' states (no transcription) (Raj & van Oudenaarden, 2008; Lenstra et al., 2016). The level of gene output can be modulated through either the duration, the size or the frequency of transcription bursts. Burst size was defined as the number of RNA polymerases transcribing a gene per burst and is reflected in RNA molecules synthesized while the gene is in the 'on'-state (Lenstra et al., 2016; Haberle & Stark, 2018; Larsson et al., 2019). Burst frequency represents the rate at which the gene switches between active ('on') and inactive ('off') transcription states reflected in how often transcription bursts occur.

Recent studies suggest that transcription bursting depends on *cis*-regulatory elements. Core promoter elements especially the TATA-box and the Inr sequences were reported to dictate the size of transcription bursts, while enhancers were suggested to affect transcription rates through regulating burst frequencies (Fukaya et al., 2016; Larsson et al., 2019). Other factors have been implicated in regulating transcription bursting such as chromatin accessibility as well as transcription factor concentration and binding kinetics (Lenstra et al., 2016; Donovan et al., 2019; Brouwer & Lenstra, 2019).

1.2.1.4. Bidirectionality

Another key characteristic of core promoters and enhancers, highlighted by recent genome-wide studies, is the manifestation of bidirectional transcription initiation (Core et al., 2014; Duttke et al., 2015; Danino et al., 2015; Schwalb et al., 2016). These divergent transcription occurrences are thought to be caused by two head-to-head Pol II initiation events in forward and reverse orientation taking places in close proximity (Figure 14A) (Core et al., 2014; Danino et al., 2015). Bidirectional transcription was first described for adjacent protein-coding genes sharing the same promoter region, representing roughly 10-20% of mammalian genes (Orekhova & Rubtsov, 2013; Danino et al., 2015). Divergent promoters
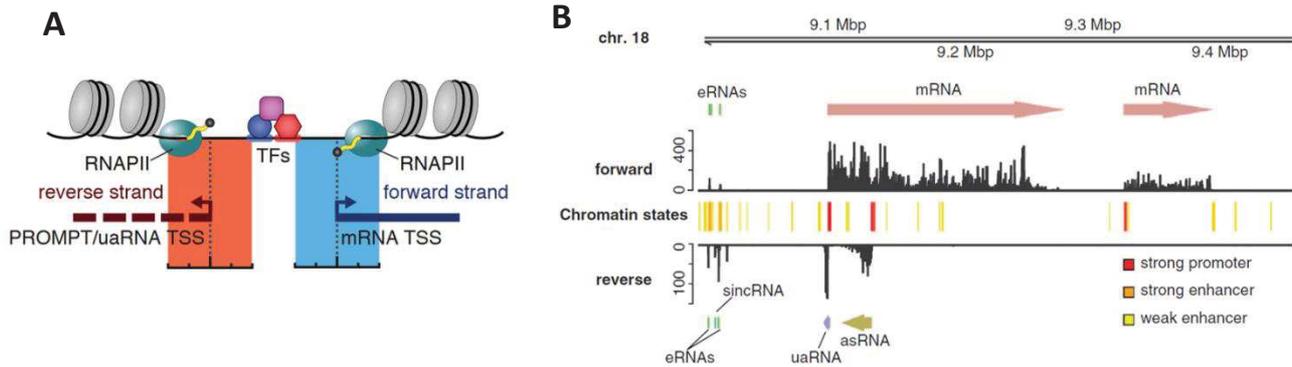
Figure 14: Divergent transcription at *cis*-regulatory regions. A. Scheme representing a bidirectionally transcribed promoter region with PROMPT (promoter upstream transcripts) or uaRNA (upstream antisense RNA) synthesized in the antisense direction (red) and mRNA synthesized in the sense direction (blue). RNAPII, RNA polymerase II; TFs, transcription factors; TSS, transcription start site. Adapted from Andersson et al., 2015a. **B.** Genome browser view on a region of chromosome 18 (chr. 18) showing examples of uaRNA (upstream antisense RNA) and enhancer derived RNAs (eRNAs). Direction of transcription is indicated by arrows. sincRNA, short intergenic non-coding RNA; asRNA, antisense RNA. From Schwalb et al., 2016.

can also result in one stable transcript in the sense direction and a non-coding transcript such as lncRNAs (long non-coding RNAs) in the antisense direction.

Other, highly unstable and shorter types of non-coding RNAs can originate from bidirectional promoters, which are referred to as PROMPTs (promoter upstream transcripts) or uaRNAs (upstream antisense RNAs) (Figure 14B). The percentage of divergent transcription at mammalian promoters varies between studies from less than half to up to 80% of all promoters (Core et al., 2014; Duttke et al., 2015; Andersson et al., 2015a; Danino et al., 2015). Bidirectional transcription was also found at active enhancers resulting in short, unstable RNAs in forward and reverse direction, which were termed eRNAs (enhancer derived RNAs) (Figure 14B). PROMPTs, uaRNAs and eRNAs are mostly short, low abundancy transcripts and are thought to be results of non-productive transcription. They are short-lived due to rapid degradation within the nucleus through the exosome, which has complicated their analysis (Preker et al., 2008; Haberle & Stark, 2018). Instability of these transcription products was further linked to the absence of 5' splice sites (U1 snRNP binding sites) and premature polyadenylation signals (Core et al., 2014; Danino et al., 2015; Haberle & Stark, 2018).

The two TSSs of enhancers and bidirectional promoters are separated by a distance that was suggested to range from 100 bp to 2 kb (Core et al., 2014; Duttke et al., 2015; Danino et al., 2015; Haberle & Stark, 2018). Further, weak or degenerated forms of core promoter elements such as a TATA-box, BRE or YR Inr motifs were reported at TSSs of enhancers and antisense TSSs of promoters. The role of bidirectional transcription and the functional importance of PROMPTs, uaRNAs and eRNAs remain however unclear (Haberle & Stark, 2018). It has been suggested that these sporadic transcription events might maintain DNA accessibility at enhancers and promoters (Mousavi et al., 2013; Haberle & Stark, 2018).

1.2.1.5. Unified architecture of *cis*-regulatory elements

Active promoters and enhancers share similarities (some mentioned in previous subsections) such as comparable frequencies in core promoter elements, divergent transcription initiation, binding of the transcription machinery, such as Pol II and other factors involved in transcription initiation, as well as features which will be described in more details in chapter 2.3. Chromatin organization into euchromatic and heterochromatic domains (Core et al., 2014; Danino et al., 2015; Haberle & Stark, 2018). In agreement, interactions between promoter pairs were recently described and were found to function in a comparable manner as enhancer-promoter pairs resulting in transcription stimulation (Li et al., 2012a; Danino et al., 2015; Haberle & Stark, 2018). Similarly, several enhancers were reported to have the potential to function as promoters especially if conserved core promoter elements are present within the enhancer region (Haberle & Stark, 2018). The shared characteristics and exchangeable functionalities of enhancers and promoters led to the recent proposition of a unified architecture and transcription initiation mechanism at these *cis*-regulatory elements challenging the established classification of enhancers and promoters as distinct entities (Core et al., 2014; Andersson, 2015; Andersson et al., 2015b).

1.2.2.   Preinitiation complex formation

RNA polymerase II (Pol II) by itself is not capable of binding to specific genomic locations (Cramer, 2019). To associate to gene promoters, to melt the DNA template and to find the transcription start site (TSS), Pol II requires the assistance of additional factors termed the general transcription factors (GTFs) namely TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. The complex formed by Pol II and the six GTFs is called the preinitiation complex (PIC), representing a remarkable cooperation of 43 proteins. Different assembly modes of the PIC were described such as the holo-enzyme model in which the PIC forms as a large complex in the nucleoplasm before being recruited to promoter DNA (Greenblatt, 1997; Myer & Young, 1998; Thomas & Chiang, 2006). The most widely accepted model of PIC assembly however is the sequential assembly model, which I will described in more details and in a linear way from the first step to the last next. Yet, within the nuclear environment a more dynamic, non-linear way of sequential PIC assembly, involving assembly and disassembly of PIC components caused by non-productive complex formation and different rate limiting steps, might better reflect this process (Sikorski & Buratowski, 2009; Hager et al., 2009).

1.2.2.1. TBP loading onto the core promoter

The first step of PIC assembly represents the binding of the GTF TFIID to the core promoter through its TBP (TATA-box binding protein) subunit. TFIID is a highly conserved, flexible complex and composed of TBP and 13 TAFs (TBP-associated factors, TAF1 to TAF13). Cryogenic electron

microscopy (cryo-EM) studies demonstrated the organization of the TFIID subunits into a three-lobed, horseshoe- or triangular-like structure (Louder et al., 2016; Kolesnikova et al., 2018; Patel et al., 2018). Six subunits (TAF4, TAF5, TAF6, TAF9, TAF10 and TAF12) are found in two copies within the complex (Kolesnikova et al., 2018). Five of these six subunits (the exception being TAF10) form the core of TFIID, which is symmetric in shape. However, this symmetry is lost upon integration of the heterodimer TAF8-TAF10 into the complex. Additional major conformational changes occur upon complete assembly of the TFIID complex (Bieniossek et al., 2013; Kolesnikova et al., 2018). Besides TBP, several TAF subunits were also suggested to have the capacity of binding to core promoter DNA elements. TAF1 and TAF2 can interact with the Inr and DPE core promoter elements, while a TAF6-TAF9 heterodimer was reported to bind to the DPE motif (Verrijzer et al., 1994; Burke & Kadonaga, 1997; Chalkley & Verrijzer, 1999; Louder et al., 2016; Patel et al., 2018). In contrast, TAF4 and TAF12 were found to interact with DNA in a relative sequence-unspecific way (Shao et al., 2005; Gazit et al., 2009; Sikorski & Buratowski, 2009; Kolesnikova et al., 2018; Patel et al., 2018). Some of these DNA binding functions of TAFs however are under debate. For example, in recent human TFIID cryo-EM structures, the TAF6-TAF9 pair was not detectably bound to DNA (Patel et al., 2018). Further, based on these recent cryo-EM structures, a mechanistic model was suggested for how TFIID binding and TBP loading onto gene promoters could occur (Patel et al., 2018). In this model, TFIID is recruited to gene promoters notably through the downstream core promoter interactions of TAF1 and TAF2. This positions TBP in proximity of upstream promoter DNA where it was suggested to search for its name-giving core promoter element, the TATA-box or its derivatives (Patel et al., 2018).

TBP interacts with the minor groove of DNA through its saddle shaped, highly conserved DNA-binding domain and causes the promoter DNA to bend in a 80- to 90-degree angle (Figure 15A and 15B) (Hahn, 2004; Alberts et al., 2010; Sainsbury et al., 2015). Structural studies revealed that TBP forms only few base-specific contacts with DNA and seems to be a relative indiscriminate DNA binder (Wong & Bateman, 1994; Coleman & Pugh, 1995; Sainsbury et al., 2015; Patel et al., 2018). This enables TBP to bind to variants of the TATA-box, containing one to two nucleotide substitutions (TATA-like motifs),
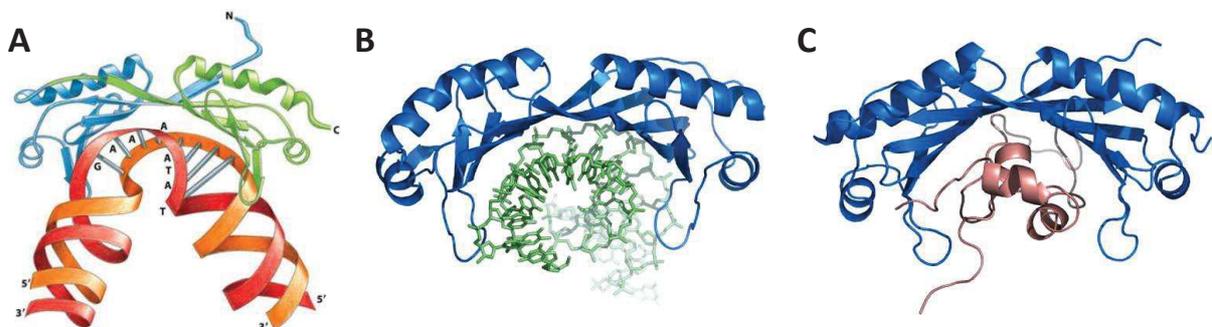
Figure 15: Structures of DNA-binding domain of yeast TBP. A. Graphical depiction of DNA bending (red-orange) caused by binding of the saddle-shaped DNA-binding domain of TBP (blue-green). From Alberts et al., 2010. **B.** and **C.** Structure of DNA-binding domain of yeast TBP (blue) bound to DNA (green) or bound to *Drosophila* TAND1 domain of TAF1 (brown), respectively. B and C from Putnam & Tainer, 2005. TBP, TATA-box binding protein.

as well as to TATA-less promoters (Hahn, 2004). From low resolution cryo-EM structures of human PIC assembled onto a TATA-less core promoter DNA, TFIID was suggested to be especially crucial to properly position TBP at these promoters through its downstream DNA interactions (Louder et al., 2016; Patel et al., 2018). Upon correct positioning, TBP is separated by roughly 30 nucleotides from the TSS in mammalian cells and released from TFIID (O'Shea-Greenfield & Smale, 1992; Carninci et al., 2006; Patel et al., 2018). Importantly in the DNA-unbound TFIID complex, the DNA-binding region of TBP is occupied by the TAND1 (TAF1 Nitrogen-terminal domain 1) of the TFIID subunit TAF1 (Figure 15C), while the TAND2 domain of the same protein interacts with the convex surface of TBP (Liu et al., 1998; Putnam & Tainer, 2005; Sainsbury et al., 2015; Patel et al., 2018). Additionally, the TAF11-TAF13 heterodimer within TFIID was found to compete with the TAND1 domain of TAF1 for the concave side of the DNA-binding region of TBP (Gupta et al., 2017). Binding of TAF11-TAF13 or the TAND1 region of TAF1 is thought to impede interactions of TBP with DNA thereby preventing aberrant PIC assembly and transcription initiation (Coleman & Pugh, 1995).

DNA-binding of TBP at gene promoters is additionally controlled by two evolutionary conserved factors that act as negative regulators (Pugh, 2000; van Werven et al., 2008). One of these factors is the SNF2-like ATPase protein BTAF1 (Mot1 in yeast), which displaces TBP from promoter DNA (Pereira et al., 2003; Moyle-Heyrman et al., 2012). The displacement of TBP by yeast Mot1 was suggested to rely on a two-step mechanism (Moyle-Heyrman et al., 2012). In the initial step, Mot1 binds to the TBP-DNA complex and causes unbending of the DNA. These conformational changes and additional ATP-dependent DNA translocation eventually leads to the dissociation of TBP from the promoter DNA. The heterodimeric complex NC2 represents the second factor involved in repressing TBP functions. NC2 is composed of two subunits, NC2α and NC2β, which dimerize through histone fold domains (more details on histone fold domains in a later chapter) (Goppelt et al., 1996; Mermelstein et al., 1996). Through its heterodimer, NC2 was found to bind to the bent DNA on the underside of the TBP-DNA complex (Kamada et al., 2001). Through additional interactions with the surface of TBP, NC2 is thought to compete with the binding of other PIC components to TBP, such as TFIIA and TFIIB (Goppelt et al., 1996; Mermelstein et al., 1996; Kamada et al., 2001). Beside this competition, NC2 was also suggested to lead to a reallocation of TBP away from the TATA-box through conformational changes in the TBP-DNA complex (Schluesche et al., 2007).

Upon core promoter binding, TFIID covers 40 to 60 bp of DNA (Hahn, 2004). These extensive DNA contacts are thought to change upon release of TBP and assembly of the remaining PIC components allowing DNA access to the other factors. These considerable promoter contacts of TFIID especially raised the question if TFIID can remain promoter-bound during complete PIC assembly and transcription initiation (Hahn, 2004; Louder et al., 2016; Patel et al., 2018; Nogales et al., 2017). Instead of being part of the growing PIC, TFIID was suggested to potentially interact with a new TBP molecule to enable subsequent rounds of PIC assembly (Patel et al., 2018).

Although TBP is necessary for transcription at all genes (including Pol I and Pol III genes), TAF subunits of TFIID were suggested to have some promoter-specific functionalities besides being required for transcription of the vast majority of Pol II genes in yeast (Sainsbury et al., 2015; Warfield et al., 2017; Donczew et al., 2020). This is supported by observations in metazoans, where several subunits of TFIID have paralogous variants that are thought to result into varying, cell type-specific subunit compositions of TFIID favouring transcription at distinct genes (Hahn, 2004). Further, TFIID was described to interact with and to be recruited to gene promoters through activating transcription factors (TFs) specific to certain gene sets. Based on these characteristics, TFIID is occasionally considered rather a transcriptional coactivator, a class of protein complexes that will be described in more details in chapter 3.2. Chromatin modifying complexes as RNA polymerase II coactivators, than a GTF (Hahn, 2004; Thomas & Chiang, 2006; Sainsbury et al., 2015; Warfield et al., 2017; Donczew et al., 2020).

1.2.2.2. TFIIA stabilizes TBP interactions with core promoter DNA

The second step of PIC assembly is the association of the GTF TFIIA to TFIID and especially TBP. TFIIA is a heterodimer consisting of two subunits (TFIIAαβ and TFIIAγ) in humans (Nogales et al., 2017). TFIIA secures TBP-DNA interactions through binding the upside of the saddle of TBP and upstream promoter DNA thereby competing with the TAND2 region of TAF1 (Hahn, 2004; Sainsbury et al., 2015). Through binding of TFIIA, TFIID changes conformation and releases TBP onto the promoter DNA (Louder et al., 2016; Patel et al., 2018). Interactions of TFIIA with TBP were suggested to enable the final positioning of TBP on DNA thereby subsequently enabling the characteristic TBP-induced bending of DNA (Figure 15) (Louder et al., 2016; Patel et al., 2018). Through genetic analysis in yeast, TFIIA was also found to interact with the TAF4 subunit of TFIID, which seems to be crucial for TFIIA-TFIID association (Layer & Weil, 2013). As mentioned above, TFIIA was additionally reported to compete *in vitro* with factors such as BTAF1 (Mot1 in yeast) and NC2 which negatively regulate TBP binding at promoter DNA (Hahn, 2004). This competition further results into stabilization of TBP-DNA binding. Occasionally and similarly to TFIID, TFIIA is considered an auxiliary factor or transcriptional coactivator for PIC formation as it is not required for basal *in vitro* transcription and interacts with activating TFs (Sikorski & Buratowski, 2009; Sainsbury et al., 2015).

1.2.2.3. TFIIB binds the TBP-DNA complex and recruits the Pol II-TFIIF complex

The third step of the canonical model of PIC formation represents the binding of the single-protein GTF TFIIB. TFIIB interacts with TBP and the surrounding DNA through its C-terminal 'core' domain and further stabilizes the TBP-DNA interaction (Werner & Grohmann, 2011). The DNA bound by TFIIB can contain the BRE core promoter elements described in an earlier chapter (1.2.1.1.1. Core promoter elements) (Sainsbury et al., 2015). The sequence differences in the upstream and downstream

BRE elements were suggested to explain the oriented formation of the PIC despite the rather symmetric TBP-DNA complex. TFIIB further interacts with Pol II through four protein domains and thereby anchors the free Pol II-TFIIF complex to gene promoters (Sainsbury et al., 2013; Sainsbury et al., 2015; Cramer, 2019). These domains include its N-terminal zinc-containing 'B-ribbon' domain, the 'B-reader' helix, the 'B-linker helix' and an N-terminal cyclin fold of the 'B-core' (Figure 16A). The B-ribbon, -reader and -linker domains enter the RNA exit tunnel of Pol II and traverse the cleft of Pol II (Figure 16B) (Sainsbury et al., 2013; Sainsbury et al., 2015). They have crucial functions at later stages of initial transcription described in more details in a following subsection.

The above described recruitment of the Pol II-TFIIF complex by TFIIB represents the fourth step of the sequential PIC assembly model. The GTF TFIIF consists of two subunits (TFIIFα and TFIIFβ) conserved between yeast to humans and forms a complex with free, DNA-unbound Pol II. TFIIF interacts with the clamp and the Rpb4/7 stalk domains of Pol II and can further bind to promoter DNA (Hahn, 2004). By forming a complex with unbound Pol II, TFIIF is thought to preclude Pol II from non-specific interactions with DNA (Sainsbury et al., 2015). Upon integration into the PIC, TFIIF further stabilizes the forming PIC especially through binding to DNA and the GTF TFIIB (Sainsbury et al., 2015). TFIIF was further reported to interact with the Pol II CTD Serine 2 phosphatase FCP1 and also has important functions in early RNA synthesis described in more details later (Sainsbury et al., 2015).
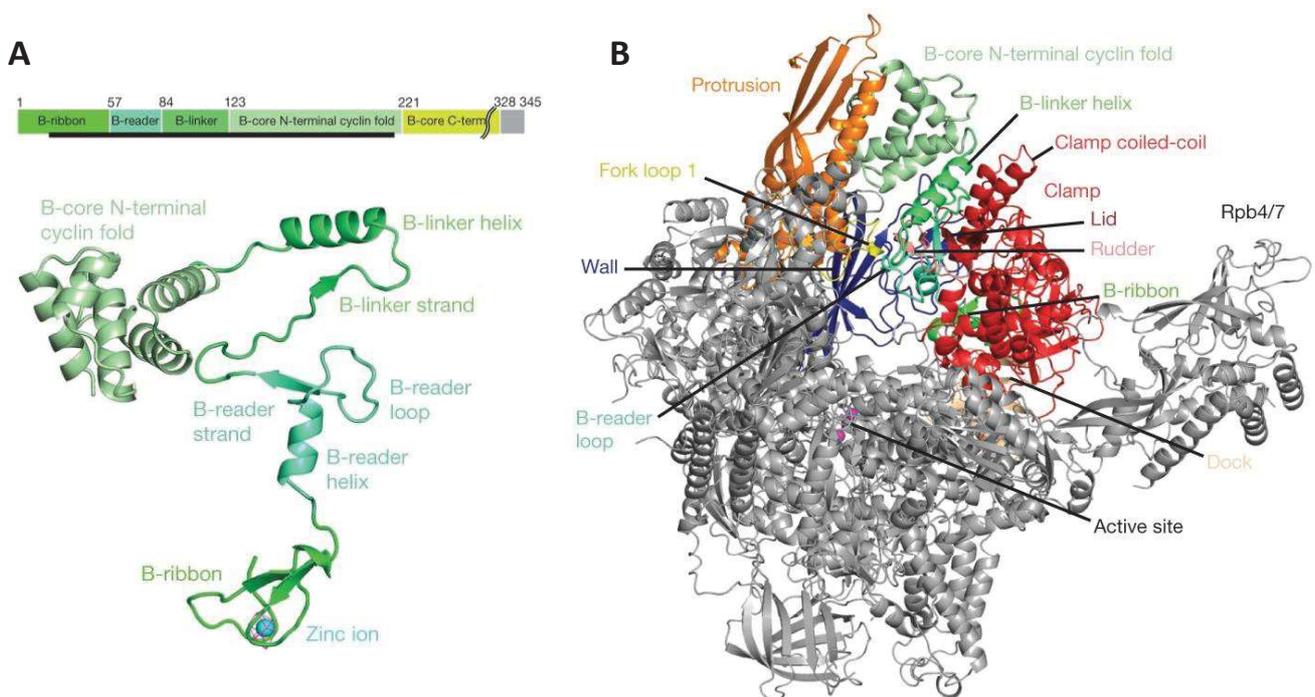
Figure 16: Crystal structure of yeast TFIIB. A. Top, scheme of yeast TFIIB with the protein fragment resolved by crystallography indicated by a black line. Bottom, side view of the crystal structure of the nearly full-length yeast TFIIB. Crucial domains involved in forming contacts with RNA polymerase II (Pol II) are indicated. **B.** Front view onto the DNA entry site of yeast Pol II (in grey) bound by TFIIB (in green). Domains of Pol II forming contacts with TFIIB are highlighted in blue, orange, yellow, salmon and shades of red. The two Mg^{2+} ions of the Pol II active side are represented by purple spheres. Adapted from Sainsbury et al., 2013.

1.2.2.4. TFIIE interacts with Pol II and core promoter DNA and recruits TFIIH

The fifth GTF, TFIIE, binds to gene promoters just upstream of the TSS and contacts Pol II between its clamp and Rbp4/7 stalk domain (Hantsche & Cramer, 2017). TFIIE is formed by an heterodimer of the subunits TFIIEα and TFIIEβ (Sainsbury et al., 2015). It positions itself onto promoter DNA and contacts Pol II on the opposite site of TFIIF (Hantsche & Cramer, 2017). TFIIE enables the recruitment of the sixth GTF, TFIIH, and forms a connection between Pol II and TFIIH (Sainsbury et al., 2015). A domain within the TFIIEα subunit is necessary for the interaction with TFIIH, while regions within TFIIEβ have been reported to bind to DNA *in vitro* (Sainsbury et al., 2015).

TFIIH with its DNA translocase subunit, XPB (Xeroderma pigmentosum group B), represents the last step of PIC assembly (Figure 17) and is crucial for the formation of an 'open promoter complex' (described in more details in next subsection). XPB uses ATP to unwind the helical DNA thereby moving away from the PIC and pushing the DNA into the cleft towards the active site of Pol II (Sainsbury et al., 2015; Cramer, 2019). This process creates DNA torsions and opens up the DNA double strand consequently facilitating the formation of the transcription bubble (Sainsbury et al., 2015; Hantsche & Cramer, 2017; Cramer, 2019). TFIIH localizes to downstream DNA close to the jaws and the head of the clamp domains of Pol II (Hantsche & Cramer, 2017). Curiously, the function of TFIIH is specific to class II transcription as Pol I and Pol III can open promoter DNA spontaneously using binding energy alone (Cramer, 2019).

TFIIH is composed of ten subunits organized into two main domains: a kinase domain, composed of three subunits, and a 'core' (Hahn, 2004; Sainsbury et al., 2015). The core domain is ring-shaped and
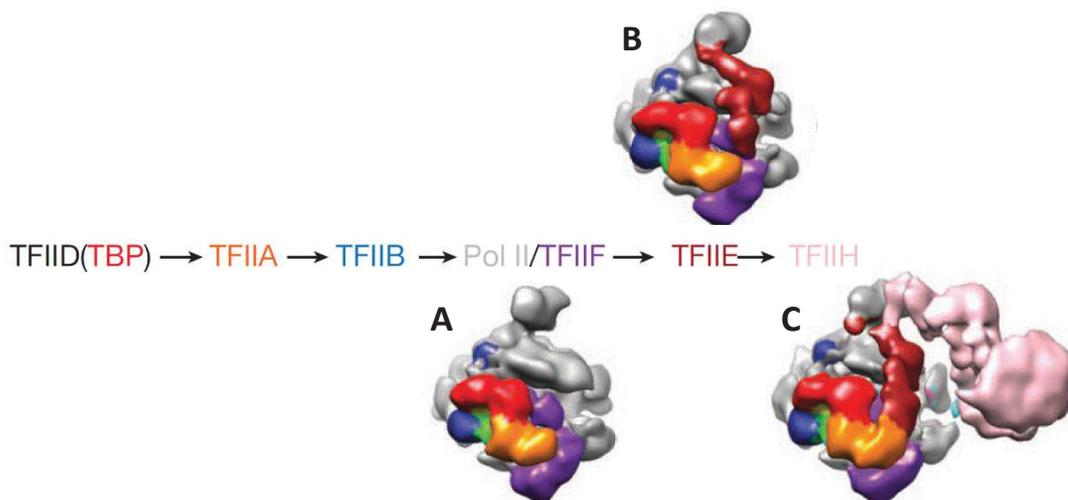
Figure 17: Sequential model of preinitiation complex (PIC) formation with illustration of cryo-EM structures of human PIC. The sequential binding order of the six GTFs is shown in the middle and also indicates the colouring of the different components in the cryo-EM structures. **A.** Low resolution complex of RNA polymerase II (Pol II) in gray with TBP (red), TFIIA (orange), TFIIB (blue) and TFIIF (purple). **B.** Same as (A) but including TFIIE (darkred). **C.** Same as (B) with TFIIH (rose). DNA in green and light blue. TBP, TATA-box binding protein. Transcription direction to the right. Adapted from He et al. 2013.

57

contains the two ATPase subunits (XPB and XPD). Double stranded DNA is thought to be accommodated within the ring of the TFIIH core domain (Hahn, 2004). The kinase domain of TFIIH termed CAK (CDK-activating kinase) contains CDK7 which can phosphorylate Serine 5 residues of the CTD of RPB1 as described in an earlier chapter (1.1.1. The Carboxy-terminal repeat domain of RNA polymerase II). Beside transcription initiation, TFIIH is further required for transcription-coupled DNA excision repair, which is thought to mainly involve its ATPase XPD (Hahn, 2004). Related to this function, mutations in the XPB and XPD genes cause the human diseases xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy (Sainsbury et al., 2015).

1.2.2.5. Functions of GTFs in the formation of the transcription bubble

The PIC is especially required to open the double strand of the template DNA. Before DNA melting and the formation of the so-called transcription bubble, the PIC is in its 'closed promoter complex' conformation (Hahn, 2004). In this closed promoter complex, the clamp region of Pol II assumes a slightly open conformation and the closed promoter DNA is positioned at the top or slightly inside the cleft of Pol II (Hantsche & Cramer, 2017). The formation of the 'open promoter complex' requires the melting of 11-15 bp of the double stranded DNA upstream of the TSS (Hahn, 2004). Subsequently, the single stranded DNA template needs to be positioned deep within the cleft of Pol II until the active site forming the transcription bubble (Hahn, 2004). During formation of the open promoter complex the flexible clamp region of Pol II is thought to swing over the cleft thereby trapping the DNA template (Gnatt et al., 2001; Bernecky et al., 2016).

Three of the six GTFs have been described to have key roles in stabilizing the unstable open promoter complex (TFIIE, TFIIF and the B-reader and B-linker helix of TFIIB) reflecting one of their most crucial functions within the PIC (Hahn, 2004; Sainsbury et al., 2015). The main function of TFIIE was suggested to be the maintaining of the opened DNA within the Pol II cleft as loading of promoter DNA into the cleft appears to be reversible. Additionally, TFIIE stimulates the ATPase and kinase functions of TFIIH required for DNA melting (Sainsbury et al., 2015). TFIIF has several roles within the open promoter complex: it stabilizes the transcription bubble, influences TSS selection and stimulates the formation of phosphodiester bonds during initial RNA synthesis through contacts to the Pol II cleft. The B-linker helix of TFIIB also has an important function in DNA melting and transcription bubble maintenance. At the same time, the B-reader of TFIIB is contacting the upstream DNA template and is thought to play a role in TSS selection through the recognition of the Inr sequence (Sainsbury et al., 2013; Sainsbury et al., 2015; Hantsche & Cramer, 2017).

Table 1: Summary of human general transcription factors (GTFs) involved in preinitiation complex formation of the RNA polymerase II transcription machinery. For each GTF the number of subunits, its functions and its occurrence in the sequential PIC assembly model are indicated. Pol II, RNA polymerase II; TBP, TATA-box binding protein; TSS, transcription start site. Based on Hahn 2004, Sikorski & Buratowski 2009, Hantsche & Cramer 2017 and Cramer 2019.

	Subunits	Functions	Sequential assembly
TFIIA	2	Stabilizes TBP and TFIID-DNA binding. Interacts with DNA. Counteracts negative cofactors of TBP.	2nd step
TFIIB	1	Binds TBP and promoter DNA and recruits Pol II-TFIIF complex. Stabilizes transcription bubble and open promoter complex. Stimulates initial RNA synthesis and TSS selection.	3rd step
TFIID	14	Binds to promoter DNA through several subunits especially TBP subunit which bends DNA by 80-90°. Serves as platform for recruitment of TFIIA and TFIIB.	1st step
TFIIE	2	Binds promoter near TSS. Essential for DNA melting and stabilisation of open promoter complex. Facilitates recruitment of TFIIH and stimulates its activities.	5th step
TFIIF	2	Forms complex with free, unbound Pol II. Prevents non-specific DNA binding of free Pol II. Is involved in Pol II recruitment to PIC and formation of transcription bubble. Aids in TSS selection. Stimulates initial RNA synthesis.	4th step
TFIIH	10	ATPase activity essential for DNA opening and stabilizes open promoter complex. Phosphorylates Serine 5 of heptad repeats of RPB1 CTD through kinase activity.	6th step

Formation of the transcription bubble and the open promoter complex allows Pol II to begin the synthesis of an RNA molecule. However, the early transcribing Pol II complex is very unstable and has to undergo several transitions before entering a stably elongating form (Cramer, 2004). One of these transitions called 'abortive cycling' involves the repeated production of short, three to nine nucleotides long, RNA molecules without Pol II dissociating from the promoter DNA (Werner & Grohmann, 2011). At these stages of initial transcription, TFIIB plays an important role by stimulating early RNA production and by maintaining the flexible clamp domain of Pol II in a closed conformation, thereby entrapping downstream DNA (Sainsbury et al., 2015).

Upon synthesis of a roughly 15 nucleotides long transcript, the RNA clashes with the B-ribbon domain of TFIIB and TFIIB is displaced (Sainsbury et al., 2013; Sainsbury et al., 2015). At this stage, the early Pol II elongation complex reaches the stage of 'promoter clearance' and moves into a stable, elongating complex (Cramer, 2004). After 30 nucleotides, Pol II is thought to loose contacts with the remaining PIC components and to move into productive elongation (Hahn, 2004). Results of *in vitro* experiments, led to the suggestion that some of the GTFs (specifically TFIIA, TFIID, TFIIE, TFIIH) remain bound to the core promoter and facilitate subsequent rounds of PIC assemblies and transcription

reinitiation (Yudkovsky et al., 2000; Hahn, 2004). The GTFs and their functional importance in PIC assembly are summarized in Table 1.

1.2.2.6. The Mediator complex: role in transcription initiation and initially transcribing Pol II

The Mediator is an evolutionary conserved, multiprotein complex consisting of 25 subunits in yeast and 33 subunits in mammalian cells (Soutourina, 2018; El Khattabi et al., 2019). It is organized into four distinct modules: 'head', 'middle', 'tail' and 'CDK8 kinase' module (Kornberg, 2005; Soutourina, 2018; El Khattabi et al., 2019). The central MED14 subunit connects head, middle and tail module by serving as a backbone for Mediator assembly. The additional subunits exclusive to mammalian Mediator increase the complexity of the tail module and were reported to strengthen the contacts of the tail with the head module, the middle module and MED14 (El Khattabi et al., 2019). A very recent structural and functional study on mammalian Mediator revealed that the majority of subunits within the head and middle module (with the exceptions of MED1, MED19 and MED20) are essential for survival of cultured mammalian cells (El Khattabi et al., 2019). In contrast, most subunits of the tail and kinase module appear to be non-essential except for MED27/3, MED28 and MED30 (El Khattabi et al., 2019).

The head and middle modules of Mediator are considered to form its 'core' and directly bind to the RBP4/7 stalk domain of Pol II (Hantsche & Cramer, 2017; Cramer, 2019). Binding of the core Mediator additionally stabilizes the PIC through contacts with the TFIIB zinc ribbon domain and was suggested to stimulate the activity of the TFIIH kinase CDK7, which phosphorylates the CTD of Pol II (Figure 18) (Kim et al., 1994; Hahn, 2004; Sikorski & Buratowski, 2009; Hantsche & Cramer, 2017).
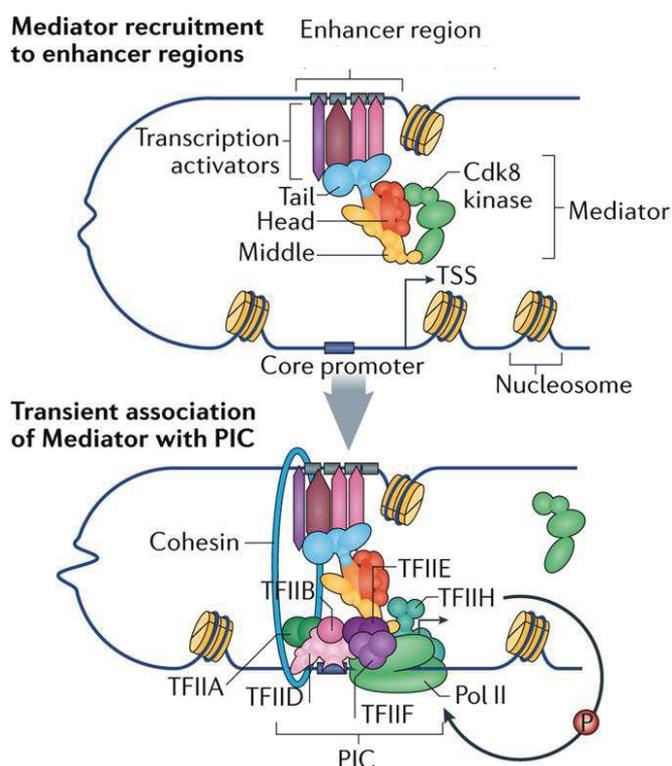
Figure 18: Schematic representation of interactions of Mediator with the preinitiation complex. Transcription activators (in shades of violet) binding to sites in enhancer regions recruit the Mediator complex through interaction with its tail module (shown in blue). The head module of Mediator is shown in red, middle module in yellow, MED14 in orange and the kinase module with Cdk8 in green. Upon interaction of Mediator with the preinitiation complex (PIC), assembled at the core promoter, the kinase module of Mediator dissociates from the remaining complex. Mediator further stimulates the kinase activity of TFIIH which is responsible for Serine 2 phosphorylation (P) of the Carboxy-terminal repeat domain of RNA polymerase II (Pol II). Cohesin has been implicated in the establishment of enhancer-promoter proximity. TSS, transcription start site. Adapted from Soutourina, 2018.

Reconstructions of Mediator and PIC interactions based on cryo-EM structures, support these contacts and reveal that Mediator positions itself close to the CAK, the kinase module of TFIIH (Plaschka et al., 2015). In yeast, Mediator was further reported to interact with the unphosphorylated CTD of Pol II (Myers et al., 1998). Several biochemical studies also imply interactions between Mediator and TFIID leading to the recent suggestion that these two complexes bind cooperatively to gene promoters (Grünberg et al., 2016; Soutourina, 2018).

The CDK8 kinase module of Mediator was actually implicated in transcription repression and dissociates from Mediator once the complex binds to the PIC (Figure 18) (Kornberg, 2005; Knuesel et al., 2009; Tsai et al., 2013; Jeronimo et al., 2016; Petrenko et al., 2016; Cramer, 2019). Indeed, the kinase module of Mediator was suggested to serve as a switch controlling Mediator-Pol II interactions. Consequently, Mediator might only form contacts with Pol II upon loss of its kinase module (Knuesel et al., 2009). The repressive function of the CDK8 kinase module was additionally linked to its ability to compete with CDK7 for the CTD of Pol II (Hsin & Manley, 2012; Tsai et al., 2013). In general, interactions between Mediator and the PIC are thought to be short-lived (Jeronimo et al., 2016; Petrenko et al., 2016; Cramer, 2019).

One of the main functions attributed to the Mediator complex is to serve as an adaptor complex bridging between activating transcription factors (TFs) at enhancers and the transcription machinery at gene promoters (Figure 18). It was therefore also implied in enabling enhancer-promoter looping (see also chapter 1.2.1.2. Enhancers) (Kagey et al., 2010; Soutourina et al., 2011; Phillips-Cremins et al., 2013; Poss et al., 2013; Grünberg et al., 2016; Petrenko et al., 2016; Soutourina, 2018; Cramer, 2019). For example in mammalian cells, Mediator subunits were found to co-localize with factors involved in 3D genome organization such as CTCF and cohesin (Kagey et al., 2010; Phillips-Cremins et al., 2013). Mediator is thought to contact the enhancer-bound activating TFs mainly through its tail domain. Yet in yeast and mammalian cells, in which the tail subunits of Mediator were deleted, a 'tailless' Mediator (equivalent to only the core of Mediator) can still bind to gene promoters and exert its functions with only minor impacts on Pol II transcription (Jeronimo et al., 2016; El Khattabi et al., 2019). Also, as mentioned earlier, subunits of the Mediator tail are mostly non-essential in contrast to the majority of its core subunits. These findings suggest that the crucial functions of Mediator lie primarily within its core and not necessarily in its enhancer-promoter adaptor functions (Plaschka et al., 2015; Jeronimo et al., 2016). This is further supported by a recent study in mammalian cells, which suggests that Mediator does not act as a physical bridge but rather indirectly results in enhancer-promoter proximity (El Khattabi et al., 2019). This study reported that combined acute depletion of Mediator and Pol II, although affecting transcription considerably, did not result into major changes in enhancer-promoter interactions in contrast to rapid depletion of cohesin as assessed through Hi-C. These results might support a recently developed model of protein hubs, which suggests that enhancers cause accumulation of transcription-

related factors by being in vicinity of their target promoters but not necessarily by being in close physical contact (more details in chapter 1.2.1.2. Enhancers) (Furlong & Levine, 2018; El Khattabi et al., 2019).

Over several years, Mediator has been the center of an ongoing debate about whether it should be considered a transcriptional coactivator or a GTF (Lewis & Reinberg, 2006; Takagi & Kornberg, 2006; Sikorski & Buratowski, 2009). Mounting evidence from yeast suggests that the Mediator complex is required for global Pol II transcription and found upstream of almost all gene promoters in yeast, which would suggest a GTF function (Andrau et al., 2006; Takagi & Kornberg, 2006; Zhu et al., 2006; Sikorski & Buratowski, 2009; Grünberg et al., 2016). Nevertheless, Mediator is not required for *in vitro* transcription and represents a point of contact for several activating TFs suggesting that, under certain conditions, it serves as a transcriptional coactivator (Lewis & Reinberg, 2006). Overall, these functional characteristics of Mediator resemble in a way the auxiliary GTFs, TFIIA and TFIID.

1.2.2.7. Visualization of the PIC organization by cryo-EM structures

Recent improvements of the cryo-EM technique provided high resolution structures of partial yeast and human PIC composed of Pol II, TFIIA, TFIIB, TFIIE, TFIIF and TBP and reflect milestones in understanding the organization of this colossal protein association (He et al., 2016; Plaschka et al., 2016; Hantsche & Cramer, 2017). These partial PIC structures in combination with lower-resolution cryo-EM structures allowed the additional location of TFIID, TFIIH and Mediator to the PIC (Figure 19) (He et al., 2013; Mühlbacher et al., 2014; Plaschka et al., 2015; Murakami et al., 2015; Nogales et al., 2017;
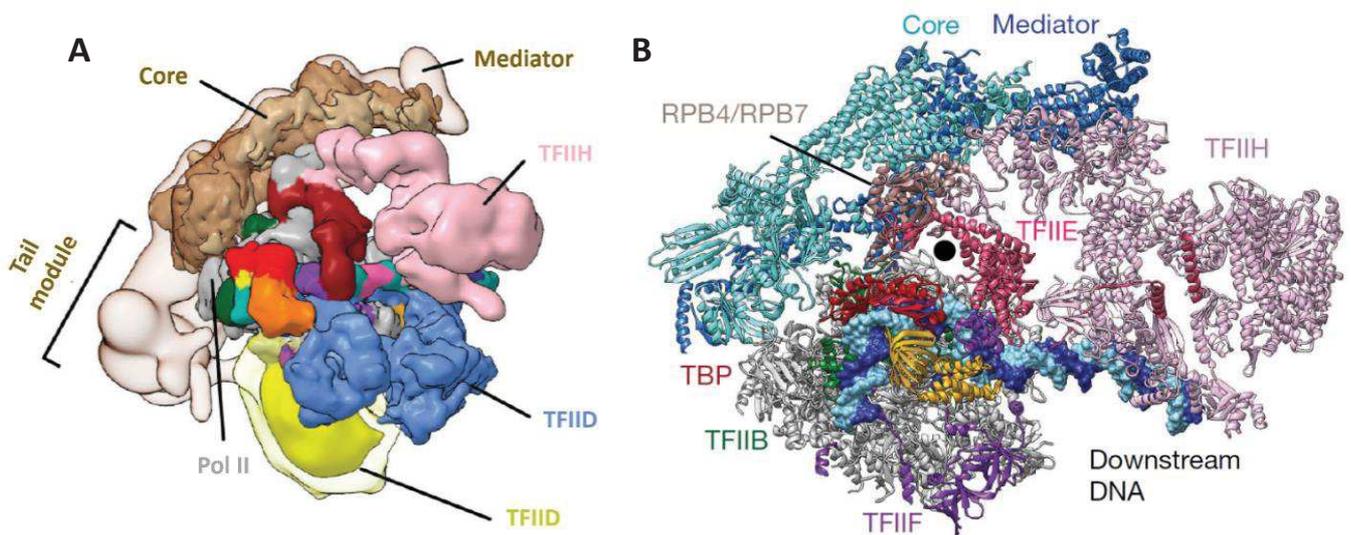
Figure 19: Cryo-EM structures of preinitiation complex. A. Reconstruction of human preinitiation complex (PIC) with TFIID, TFIIH and yeast Mediator based on cryo-EM structures. RNA polymerase II (Pol II) in grey. TBP (red), TFIIA (orange), TFIIB (green), TFIIE (darkred), TFIIF (purple), TFIIH (rose), Mediator (brown), TFIID (yellow and blue). DNA (TATA-less) is shown in turquoise, yellow and pink. Adapted from Nogales et al 2017. **B.** High resolution cryo-EM structure of yeast PIC with modelled TFIIH and Mediator positioning. RNA polymerase II (gray), TBP (red), TFIIA (yellow), TFIIB (green), TFIIE (pink), TFIIF (purple), TFIIH (rose) and Mediator (lightblue and darkblue). DNA is shown in blue and cyan. Black point highlights the CTD attachment to Pol II. Transcription direction to the right. From Cramer 2019.

Cramer, 2019). The final reconstructions revealed for example that Mediator and TFIID position themselves on opposite sides of Pol II (Figure 19A) and further allowed the visualization of contact sites among the different complexes involved in PIC formation (Nogales et al., 2017; Cramer, 2019). As PIC assembly and Mediator-PIC interactions are highly dynamic processes, the presented structures might however not represent the entire *in vivo* mechanism (Soutourina, 2018).

1.3. Transcription pausing

Recent studies further emphasized an additional rate limiting step of transcription occurring at several genes downstream of PIC assembly and transcription initiation: Promoter-proximal pausing. This phenomenon was first described for heat-shock genes in *Drosophila* (Lis et al., 2000; Levine, 2011). In their transcription inactive state, heat-shock response genes show high levels of RNA polymerase II (Pol II) 30 to 50 bp downstream of the transcription start site (TSS) (Lis et al., 2000; Levine, 2011). These stalled transcription complexes were found to be rapidly released into productive elongation upon heat-shock induction (Lis et al., 2000). Since then, about one third or even the majority of mammalian genes (depending on the stringency used to define Pol II pausing) were proposed to display stalled Pol II complexes downstream of their TSS (Guenther et al., 2007; Core et al., 2008; Rahl et al., 2010; Nechaev & Adelman, 2011; Levine, 2011; Gilchrist et al., 2012; Gilchrist & Adelman, 2012). The location of paused Pol II at *Drosophila* genes was further found to cooccur with downstream core promoter elements such as the DPE (downstream promoter element), which led to the suggestion of a potential DNA encoded mode of Pol II pausing (DPE is described in more details in chapter 1.2.1.1.1. Core promoter elements) (Nechaev et al., 2010; Levine, 2011; Nechaev & Adelman, 2011). As such also AT-rich regions were proposed to facilitate Pol II pausing (Nechaev et al., 2010; Nechaev & Adelman, 2011). A very recent study reported the additionally involvement of the GTF TFIID in *in vitro* Pol II promoter-proximal pausing (Fant et al., 2020). Curiously, however transcription pausing has not been found in yeast (Nechaev & Adelman, 2011).

1.3.1. Promoter-proximal pausing factors

Two factors have been implicated in stabilizing transcription pausing: NELF (Negative elongation factor) and DSIF (DRB sensitivity inducing factor) (Levine, 2011; Vos et al., 2018; Chen et al., 2018a; Chen et al., 2018b). These two complexes are frequently referred to as 5' pausing factors and impede the continuation of Pol II into downstream elongation (Buratowski, 2009; Levine, 2011; Nechaev & Adelman, 2011). NELF consists of four subunits (NELF-A, NELF-B, NELF-C [or an isoform of NELF-C: NELF-D] and NELF-E), while DSIF is a heterodimer of SUPT4H and SUPT5H (Figure 20A) (Nechaev & Adelman, 2011; Vos et al., 2018). NELF was found to be recruited to Pol II through two means: i) interactions with DSIF and ii) through phosphorylated Serine 5 residues of the Pol II CTD, which is mediated by CDK7, the kinase of TFIIH (Glover-Cutter et al., 2009; Nechaev & Adelman, 2011; Danino et al., 2015; Chen et al., 2018a). In general, the CTD of paused Pol II tends to be phosphorylated at Serine 5 (Levine, 2011).

Recent cryo-EM structures of the paused transcription complex reveal that binding of NELF and DSIF cause conformational changes of Pol II leading to a tilt in the DNA-RNA hybrid (Figure 20B) (Vos et al., 2018). In this configuration the association of additional ribonucleotides to the DNA template

Figure 20: DSIF and NELF in association with elongating RNA polymerase II. A. Front view on the cryo-EM structure of RNA polymerase II (Pol II, in grey) with DSIF (SUPT4H in light green, SUPT5H in dark green) and NELF (NELF-E in light blue, NELF-B in purple, NELF-C in blue and NELF-A in orange). Template and non-template DNA in blue and cyan, respectively. RNA in red. **B.** Overlay of the DNA-RNA hybrids observed in the paused transcription elongation complex of (A) (template DNA in blue and RNA in red) and the processive elongation complex (EC, in grey). Pol II bridge helix shown in green. Adapted from Vos et al., 2018.

is impaired preventing the extension of the RNA molecule and thereby blocking Pol II movement (Vos et al., 2018). DSIF was identified close to the RPB4/7 stalk domain of Pol II forming two clamps, one around the transcript through its SUPT5H subunit and another around the upstream DNA through SUPT4H (Figure 20A). Localised on the opposite site of DSIF, NELF was found to interact extensively with Pol II restraining its mobility. Through two so-called 'tentacles' of its NELF-A and NELF-E subunits, NELF was additionally found to reach out to DSIF and the exiting RNA molecule, respectively (Vos et al., 2018). Variants of NELF lacking the NELF-A tentacle, which interacts with DSIF, were unable to stabilize pausing, suggesting that cooperative binding of NELF and DSIF is required for promoter-proximal pausing (Vos et al., 2018). In contrast, loss of the NELF-E tentacle, which contacts the nascent RNA, were found to be dispensable for pause induction. These cryo-EM structures further indicated that binding of the two 5' pausing factors is incompatible with the presence of PIC components such as TFIIB, TFIIE and TFIIF. NELF and DSIF therefore were suggested to contact Pol II only once it has dissociated from the initiation factors (Vos et al., 2018).

Beside promoter-proximal pausing, NELF has further been implicated in the transcription termination processes at snRNA (small nuclear RNA) and canonical histone mRNA genes (see also chapter 1.5. Transcription termination) (Buratowski, 2009).
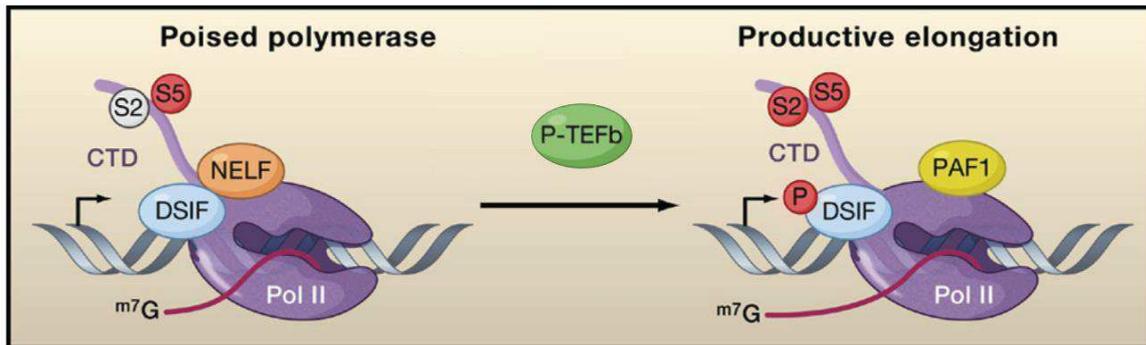
Figure 21: Factors involved in the stabilization and release of paused RNA polymerase II. RNA polymerase II (Pol II) phosphorylated at Serine 5 (red S5) residues of its CTD (Carboxy-terminal repeat domain) pauses less than 100 nucleotides downstream of the transcription start site (represented by the black arrow). Phosphorylation of DSIF (represented by red P), NELF and Serine 2 (red S2) of the Pol II CTD by P-TEFb results in the exchange of NELF with the elongation factor PAF1 allowing productive elongation. Grey S2 represents unphosphorylated Serine 2. ^{m7}G represents the 5' cap modification of mRNAs. Adapted from Price, 2010.

1.3.2. Transcription pause release

Paused Pol II is released through the action of CDK9 (cyclin-dependent kinase 9) of P-TEFb (positive transcription elongation factor b) (Price, 2010; Levine, 2011; Nechaev & Adelman, 2011; Chen et al., 2018a). Several components of the paused transcription elongation complex are subjected to modifications by CDK9: NELF, DSIF (specifically its SUPT5H subunit) and Serine 2 of the Pol II CTD (Figure 21) (Lis et al., 2000; Nechaev & Adelman, 2011; Henriques et al., 2013; Chen et al., 2018a). P-TEFb was found to associate with other proteins to form the SEC (super-elongation complex) implicated in pause release (Levine, 2011; Nechaev & Adelman, 2011; Chen et al., 2018a).

Productive transcription elongation proceeds after dissociation of NELF and binding of the elongation factor, PAF1 (polymerase-associated factor 1) (Figure 21) (more details in chapter 1.4. Transcription elongation). In contrast to NELF, phosphorylated DSIF remains associated with the elongating Pol II machinery and exerts positive functions onwards (Nechaev & Adelman, 2011). In agreement with its role in productive transcription elongation, DSIF or its homologues are found in bacteria, archaea and all eukaryotes, while NELF seems metazoan-specific (Nechaev & Adelman, 2011).

1.3.3. Transcription pausing or premature transcription termination

The exact mechanism and importance of transcription pausing remains unclear (Hsin & Manley, 2012; Lenhard et al., 2012). Promoter-proximal pausing was suggested to enable rapid transcription induction in response to external stimuli and during development as it circumvents dependencies on the slower PIC assembly step (Lis et al., 2000; Zeitlinger et al., 2007; Levine, 2011). Yet, genes without any marked transcription pausing event can be induced as rapidly as genes with paused Pol II elongation complexes (Gilchrist & Adelman, 2012; Danino et al., 2015). Transcriptional pausing was further

reported to serve as a way of integrating regulatory signals in a coordinated manner therefore allowing the synchronization of gene expression among cell populations (Zeitlinger et al., 2007; Price, 2010; Levine, 2011; Nechaev & Adelman, 2011; Henriques et al., 2013; Lagha et al., 2013). Some indications suggest that Pol II promoter-proximal pausing might serve as switch not only in inducing but especially in repressing transcription once the stimulus has ceased (Nechaev & Adelman, 2011).

An alternative explanation to promoter-proximal pausing for the accumulation of Pol II downstream of TSSs was proposed to be the inefficiency of early transcription elongation, as described in section 1.2.2.5. Functions of GTFs in the formation of the transcription bubble. Accordingly, promoter-proximal enrichment of Pol II was suggested to potentially simply reflect kinetics of abortive elongation (Buratowski, 2009; Lenhard et al., 2012; Kamieniarz-Gdula & Proudfoot, 2019). Recent studies indeed reported that less than 10% of paused Pol II molecules enter into a productive transcription elongation phase suggesting a highly dynamic turnover of Pol II at pause sites (Krebs et al., 2017; Steurer et al., 2018; Price, 2018; Kamieniarz-Gdula & Proudfoot, 2019). This high rate of early Pol II transcription termination is not necessarily in contradiction to transcription pausing. Instead, the successful release of paused Pol II through timely P-TEFb recruitment was suggested to represent the decisive step of productive elongation or premature termination (Nechaev & Adelman, 2011). External stimuli were further proposed to modulate recruitment of P-TEFb thereby regulating the relative rates of productive elongation and premature termination at pausing sites (Price, 2010; Nechaev & Adelman, 2011).

In contrast to earlier models of promoter-proximal pausing in which paused Pol II was thought to remain stably associated with DNA for a long period of time, the idea of unresolved pausing leading to premature termination and subsequent rounds of transcription initiation are recently favoured (Nechaev & Adelman, 2011; Krebs et al., 2017; Kamieniarz-Gdula & Proudfoot, 2019). Following these indications, the borders between definitions of Pol II pausing, backtracking and premature termination start to be blurry. In general, Pol II seems to encounter at several genes a promoter-proximal checkpoint that might ensure the stability and readiness of Pol II transcription complexes and further might be modulated by external signals (Price, 2010; Nechaev & Adelman, 2011; Levine, 2011; Kamieniarz-Gdula & Proudfoot, 2019).

1.4. Transcription elongation

Once RNA polymerase II (Pol II) transitions into the productive elongation phase, it represents a highly stable complex transcribing genes at an average rate of roughly 2 kb per minute without dissociating from the DNA (Singh & Padgett, 2009; Steurer et al., 2018). This processivity enables it to transcribe genes of various length with the most extreme case in the human genome being the dystrophin gene with a size of 2.4 Mb (Lander et al., 2001). The transcribing Pol II is assisted by numerous factors ensuring a stable elongation complex such as DSIF and Paf1C (PAF1 complex) (Bernecky et al., 2016; Xu et al., 2017). Paf1C consists of five subunits and accompanies Pol II from the transcription start site

to the polyadenylation site (Mayer et al., 2010; Xu et al., 2017). Interactions with the phosphorylated CTD of Pol II and the phosphorylated SUPT5H subunit of DSIF enable the recruitment of Paf1C to the elongating complex (Qiu et al., 2012; Xu et al., 2017). Paf1C directly binds to Pol II by forming extensive contacts with the RPB2 subunit (Xu et al., 2017). Genome-wide nascent RNA analysis upon deletion of Paf1C subunits further underlined its general importance for transcription revealing that the majority of Pol II-transcribed genes in yeast are downregulated upon its loss (Xu et al., 2017).

In the transcribing Pol II, the double-stranded DNA template lies within the cleft entrapped by the closed clamp domain and the RPB2 subunit (Figure 22A) (Gnatt et al., 2001). Upon encountering the 'wall' domain formed by RPB2, located behind the catalytic centre, the DNA template bends in roughly 105° upwards relative to the downstream DNA and through unwinding forms the transcription bubble (Figure 22B) (Bernecky et al., 2016). The transcription bubble is maintained from the side of the downstream DNA through the action of the RPB2 'fork loop' 2. Within the transcription bubble, the template DNA strand allocates along the bottom of the cleft and passes over the 'bridge' helix formed by the RPB1 subunit (Figure 22B) (Gnatt et al., 2001; Bernecky et al., 2016). The presence of the bridge helix leads to conformational changes in the template strand nucleotides. The base of the nucleotide in proximity to the Metal A Mg^{2+} ion (termed +1 nucleotide) is flipped and directed downwards for readout by the active site. Through this conformation change, the +1 template nucleotide lies within a pore that extends to the floor of the Pol II cleft and which serves amongst others as entrance for the ribonucleotide substrates (Figuer 22B). It also constitutes the first of eight to nine base pairs of the DNA-RNA hybrid (Gnatt et al., 2001; Westover et al., 2004).
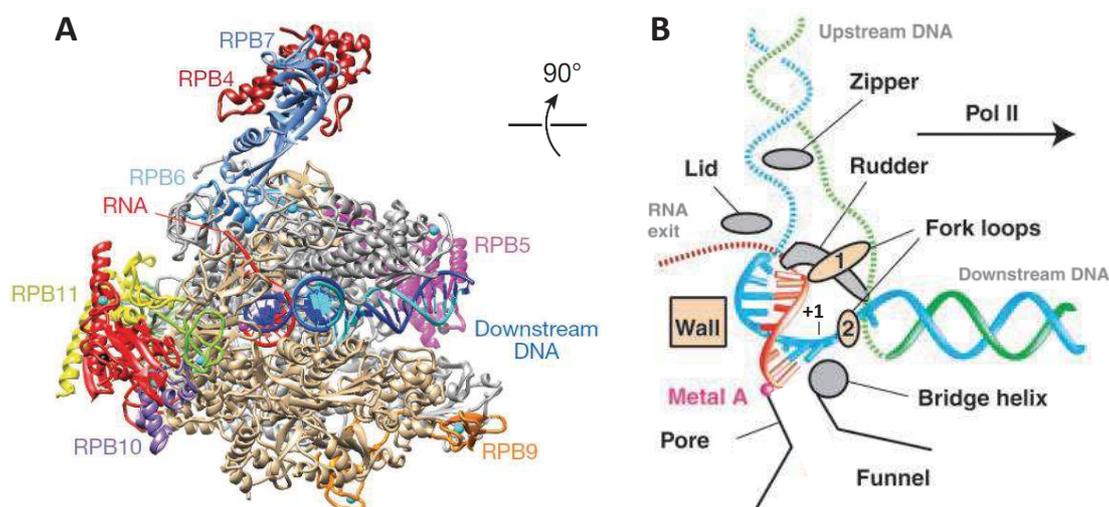
Figure 22: Structure and organization of elongating RNA polymerase II. A. Cryo-EM structure of the mammalian elongating RNA polymerase II (Pol II) complex viewed from the exit point of the upstream double-stranded DNA (top view on Pol II). RPB1 in silver and RPB2 in gold. From Bernecky et al., 2016. **B.** Schematic view from the side onto the active centre of Pol II during transcription elongation. Template DNA strand in blue. RNA in red. Metal A represents one of the two Mg^{2+} ions of the active site of Pol II. Structural elements required to maintain the transcription bubble and guide the nucleic acids at the active site such as the 'bridge' helix or the 'wall' of the RPB1 (silver) and RPB2 (gold) subunits are highlighted. The +1 template nucleotide is indicated. Arrow indicates direction of transcription. Detailed description in text. Adapted from Gnatt et al. 2001.

At the catalytic centre, Metal A is located in proximity to the ribonucleotide at the 3' end of the DNA-RNA hybrid. The second Mg^{2+} ion (Metal B) of the active site is thought to be associated with the three phosphates of the incoming substrate ribonucleotide. Upon stable base pairing of the new ribonucleotide with the template DNA, Metal B together with Metal A enables the catalysis of the phosphodiester bond thereby leading to the extension of the RNA molecule (Gnatt et al., 2001). After coupling of the new ribonucleotide to the growing nascent RNA, Pol II translocates along the template DNA by one nucleotide. This translocation is enabled through movements of the bridge helix creating an new empty site at position +1 for addition of the subsequent ribonucleotide (Gnatt et al., 2001).

Upon synthesis of a roughly 9- to 10-nucleotide long RNA molecule, the DNA-RNA hybrid must be separated to allow the exit of the nascent RNA through the exit tunnel. In addition to the 'fork loop' 1 of RPB2, three protein loops of RPB1 play an important role in the dissociation of the hybrid, guidance of the nascent RNA exit, reannealing of the template DNA and overall maintenance of the transcription bubble: the 'rudder', the 'lid' and the 'zipper' (Figure 22B) (Gnatt et al., 2001; Westover et al., 2004). Passing underneath the lid leads to the separation of the nascent RNA from the template DNA and its redirection to the RNA exit channel. The rudder further prevents the reassociation of the DNA-RNA hybrid by interacting with the single stranded template DNA, while the zipper keeps the template and non-template DNA strands separated (Gnatt et al., 2001; Westover et al., 2004). Fork loop 1 of RPB2 seems important to restrict the extent to which the DNA-RNA hybrid gets separated (Westover et al., 2004). These structural loops also form protein-protein interactions amongst each other: the lid was found to contact the rudder and the rudder can interact with fork loop 1 (Westover et al., 2004).

The DNA-RNA hybrid is almost completely buried within Pol II, secured through protein-nucleic acid interactions such as with fork loop 1 and stabilized through Watson-Crick base pairing (Westover et al., 2004). As described above, interactions between Pol II and the nucleic acids are necessary for the proper positioning of the nucleic acids within the catalytic centre and further contribute considerably to the stability of the transcription elongation complex (Gnatt et al., 2001). Importantly, these contacts are not rigid as this would interfere with the mobility of the enzyme and therefore the translocation process required for productive transcription elongation (Westover et al., 2004).

During proof-reading and backtracking of challenging template DNA, the transcription elongation machinery is aided by the TFIIS complex (more details in subsection 1.4.2. RNA polymerase II backtracking). Other factors are involved in allowing the passage of Pol II through chromatin such as the histone chaperone FACT, which will be described in more details in chapter 3.2.1. Chromatin remodellers and histone chaperones. Additionally, during transcription elongation several RNA processing factors associate with Pol II such as the splicing machinery, which enables co-transcriptional splicing of the nascent mRNA (described in more details in the following subsection).

1.4.1. Pre-mRNA splicing

In humans, a nascent mRNA molecule can contain several introns of various sizes with an average length of 3 kb but some exceeding 10 kb (Lander et al., 2001). A portion of these introns are subjected to alternative splicing events which will not be detailed here. In contrast, exons in humans tend to be small with on average roughly 150 bp (Lander et al., 2001). The longest single exon with an immense size of roughly 17 kb is found within the human titin gene, which also contains the highest number of exons (178 exons) (Lander et al., 2001).
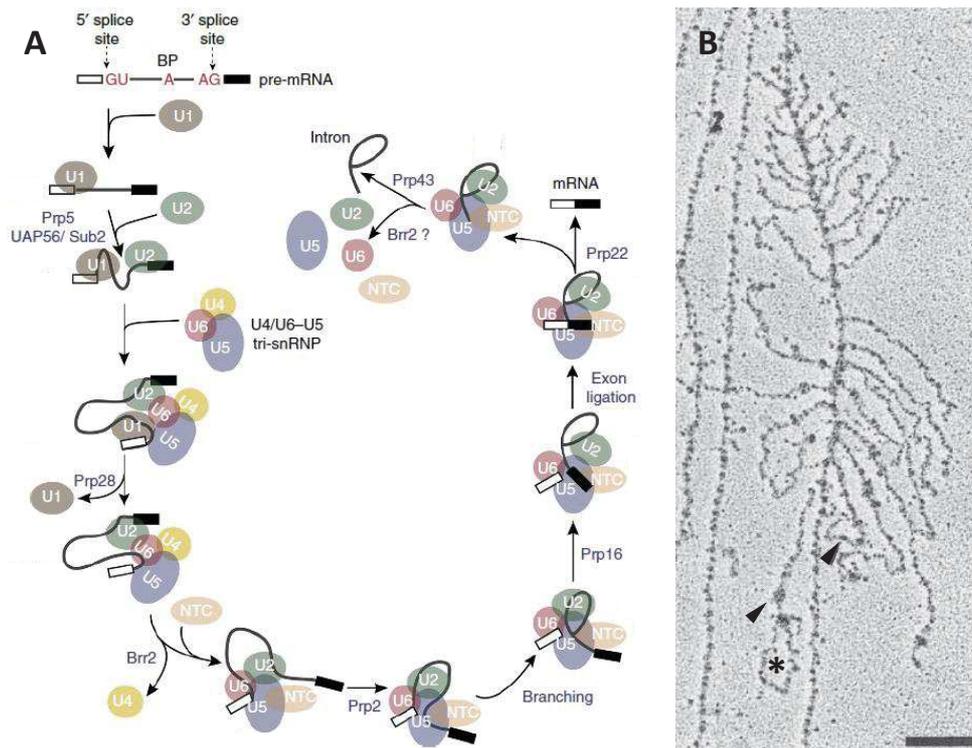
Figure 23: Mechanism of pre-mRNA splicing. A. Scheme of the mechanism of intron excision and formation of the spliceosome. White and black boxes reflect 5' and 3' exon surrounding the intron, respectively. Details see text. Additional factors involved in spliceosome function are indicated at the respective steps. BP, branch point. Adapted from Fica & Nagai, 2017. **B.** Electron micrograph of DNA from *Drosophila* embryo cells. Vertical strings represent DNA with nascent RNA spreading to the side. Black arrowheads indicate potential aggregations of factors implicated in intron excision. Star highlights the formation of a co-transcriptional intron lariat. Adapted from Beyer & Osheim 1988.

The numbers and sizes of introns encoded within mammalian genes require sophisticated mechanisms ensuring their correct excision. Three DNA sequences characterized by very short consensus sequences at or close to exon-intron junctions are needed for intron removal and exon splicing: the 5' splice site (5'SS), the branch point (BP) sequence and the 3' splice site (3'SS) (Figure 23A) (Wahl et al., 2009; Fica & Nagai, 2017). Additionally, human introns present a polypyrimidine stretch upstream of the 3'SS recognized by *trans*-acting factors involved in the assembly of the spliceosome.

Intron excision requires two transesterification reactions (Wahl et al., 2009). The first reaction involves the 5'SS and the conserved adenosine of the BP generating a free 5' exon and an intron lariat-

3'exon intermediate. Subsequently, the 5' exon chemically attacks the 3'SS resulting into the ligation of the exons and excision of the intron in form of a lariat (Wahl et al., 2009; Fica & Nagai, 2017). These reactions are catalysed through the spliceosome, consisting of five snRNPs (small nuclear ribonucleoproteins) subcomplexes. (Figure 23A). snRNPs represent associations of each of the five snRNAs (small nuclear RNAs), U1, U2, U4, U5 and U6, with variable numbers of proteins. They assemble into the spliceosome *de novo* on every intron in a stepwise manner and with the assistance of numerous other factors (Wahl et al., 2009; Fica & Nagai, 2017).

The formation of the spliceosome begins with the binding of the U1 snRNP to the pre-mRNA enabled through base pairing of the U1 snRNA with the 5'SS of the intron. Next, the U2 snRNA of the U2 snRNP complex base pairs with the BP. This interaction is further stabilized through factors recognizing the polypyrimidine tract (Wahl et al., 2009). Binding of U1 and U2 snRNP enables the formation of the complete spliceosome by recruiting the preassembled U4/U6, U5 tri-snRNP. To be catalytically active and to perform the two transesterification reactions, the spliceosome undergoes conformational and compositional rearrangements, such as the release of U1 and U4 snRNPs through the actions of the PRP28 and BRR2 ATPases, respectively, and stabilization through the NTC (nineteen complex) and related factors (Figure 23A) (Wahl et al., 2009; Fica & Nagai, 2017). Upon formation of the intron lariat and ligation of the exons, the spliceosome dissociates, releasing the U2, U5 and U6 snRNPs for subsequent rounds of splicing.

Efficient RNA splicing was found to be linked to the transcription machinery through the CTD of Pol II, which facilitates the recruitment of the spliceosome to the pre-mRNA (McCracken et al., 1997). First observed in the 1980s (Figure 23B), recent genome-wide studies have provided new insights on splicing events occurring while Pol II is still transcribing the gene, generally referred to as co-transcriptional splicing (Beyer & Osheim, 1988; Khodor et al., 2011; Oesterreich et al., 2016; Herzel et al., 2018; Drexler et al., 2020; Reimer et al., 2020). In yeast, where introns are short and rare, co-transcriptional splicing was reported to take place immediately after synthesis of the intron (Oesterreich et al., 2016; Herzel et al., 2018). Also, in *Drosophila* the majority of introns was found to be removed while transcription of the mRNA is still on-going (Khodor et al., 2011; Drexler et al., 2020). In contrast, analyses on co-transcriptional splicing in human cells are more contradictory with one study suggesting less than 50% of co-transcriptional splicing, while another reports that on average more than 80% of introns are excised before Pol II terminates transcription (Drexler et al., 2020; Reimer et al., 2020).

1.4.2. RNA polymerase II backtracking

Unfavourable DNA regions such as AT-rich stretches can cause low DNA-RNA hybrid stability and lead to a reverse movement of RNA polymerase II (Pol II), which is referred to as backtracking
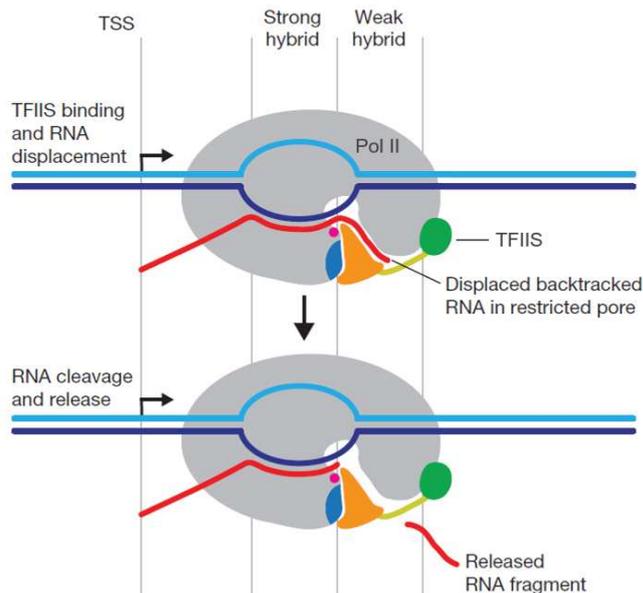
Figure 24: Backtracking and reactivation of RNA polymerase II mediated by TFIIS. Shown is a schematic representation of backtracked RNA polymerase II (Pol II, grey) bound by TFIIS (domain III of TFIIS shown in orange; interdomain linker in light green; domain II in dark green; domain I not shown). TFIIS extends into the active site of Pol II through the funnel and pore region and thereby displaces the backtracked RNA. TFIIS further facilitates the cleavage of the RNA molecule. Blue area of Pol II close to TFIIS indicates the funnel domain. DNA is shown in dark blue (template strand) and light blue (non-template strand). RNA is shown in red. Dot in magenta represents Metal A of the Pol II active site. Black arrow on DNA indicates the transcription start site (TSS). Adapted from Cheung & Cramer, 2011.

(Cheung & Cramer, 2011). This backward motion dislodges the 3' end of the nascent RNA from the DNA directing it through a pore into the funnel located underneath the catalytic site of Pol II (Figure 24). Consequently, transcription is halted as the +1 template nucleotide is made inaccessible to incoming ribonucleotides. If Pol II is backtracked by only one nucleotide, the DNA-RNA hybrid can be realigned through removal of a mono- or dinucleotide from the 3' end of the backtracked RNA by the intrinsic nuclease activity of Pol II (Kettenberger et al., 2003; Cheung & Cramer, 2011). This allows the elongation complex to continue RNA polymerisation and is often referred to as the 'proof-reading' activity of Pol II (Kettenberger et al., 2003). Intriguingly and in contrast to DNA polymerases, Pol II contains a single active site capable of catalysing both, RNA polymerisation and RNA cleavage (Kettenberger et al., 2003).

Resolving more extensive backtracking of several nucleotides, which causes transcription arrest, involves the single-protein elongation factor TFIIS (Fish & Kane, 2002; Sigurdsson et al., 2010; Cheung & Cramer, 2011). TFIIS is organized into three domains (domains I, II and III) from its N- to C-terminal end and enhances the slow intrinsic cleavage function of Pol II. Domain I of TFIIS is only weakly conserved and is not required for the reactivation of backtracked Pol II (Kettenberger et al., 2003). Domain II and the interdomain linker connecting domain II and III were shown to bind to Pol II from the RPB1/9 jaw domain until close to the rim of the Pol II funnel, while domain III reaches into the active site of Pol II through the pore and funnel domain from the underside of the enzyme (Figure 24) (Kettenberger et al., 2003; Cheung & Cramer, 2011). Two highly conserved acidic residues of a β-hairpin of domain III, which locate close to Metal A of the Pol II active site, were shown to be essential for TFIIS function (Jeon et al., 1994).

Binding of TFIIS to backtracked Pol II leads to extensive conformational changes of structural features of Pol II, such as the clamp domain and, within the active centre, the bridge helix and fork loop

72

2 (Kettenberger et al., 2003). These structural changes are thought to influence the strength of Pol II-nucleic acid interactions (Kettenberger et al., 2003). For example, insertion of domain III of TFIIS into the Pol II pore was reported to result in a displacement of the backtracked RNA thereby weakening contacts to Pol II and facilitating cleavage. The TFIIS-restricted pore however still leaves enough room for the entry of ribonucleotide substrates (Kettenberger et al., 2003). The hydrolytic cleavage reaction of the backtracked RNA induced by TFIIS involves a nucleophilic water molecule and two metal ions, one represented by Metal A of the active site of Pol II, while Metal B is thought to be contributed by TFIIS (Kettenberger et al., 2003; Cheung & Cramer, 2011). Upon cleavage and release of the backtracked RNA, transcription can resume as ribonucleotides can once more hybridize to the template DNA (Figure 24).

Proof-reading, backtracking and reactivation of Pol II occur frequently during transcription as evidenced by the fact that the intrinsic RNA cleavage activity of Pol II is essential for cell viability (Sigurdsson et al., 2010). At the same time, loss of TFIIS function is not essential, possibly because it is merely required for the stimulation of the intrinsic RNA cleavage activity of Pol II (Sigurdsson et al., 2010). TFIIS was further implicated in promoter-proximal pause release (Nechaev & Adelman, 2011). Recent cryo-EM structures however indicate that the pausing factor NELF would prevent binding of TFIIS to Pol II, as both factors interact with the funnel region of Pol II (Vos et al., 2018). Eventually, if the backtracked transcription complex cannot be resolved, transcription is thought to be prematurely terminated through polyubiquitination and degradation of Pol II by the proteasome (Somesh et al., 2005).

1.5. Transcription termination

Upon transcription termination, RNA polymerase II (Pol II) stalls and dissociates from the DNA template releasing the transcript. Transcription termination can be induced prematurely at any time during transcription elongation and plays a crucial role in restricting the extent of pervasive, unwanted transcription (see also section 1.4.2. RNA polymerase II backtracking and chapter 1.2.1.4. Bidirectionality) (Porrua & Libri, 2015). In the following, I will focus on the termination process occurring at the end of genes leading to the liberation of a full-length transcript.

Efficient termination is important to avoid conflicts with transcription processes of neighbouring genes, to liberate Pol II for subsequent rounds of transcription and, since transcription termination and 3' RNA processing are coupled, it further influences stability and functionality of RNA molecules (Kuehner et al., 2011; Porrua & Libri, 2015). Transcription termination by Pol II does not occur at a constant distance from the 3' end of the full-length transcript but instead can take place few to several thousands of nucleotides downstream (Richard & Manley, 2009). Since Pol II synthesises different classes of RNAs, such as protein-coding mRNAs, which need to be polyadenylated, or short, non-

polyadenylated RNAs, such as snRNAs or canonical histone mRNAs, the mechanisms implicated in transcription termination vary. Polyadenylation-dependent termination and termination at snRNA and histone genes will be shortly detailed in the following.
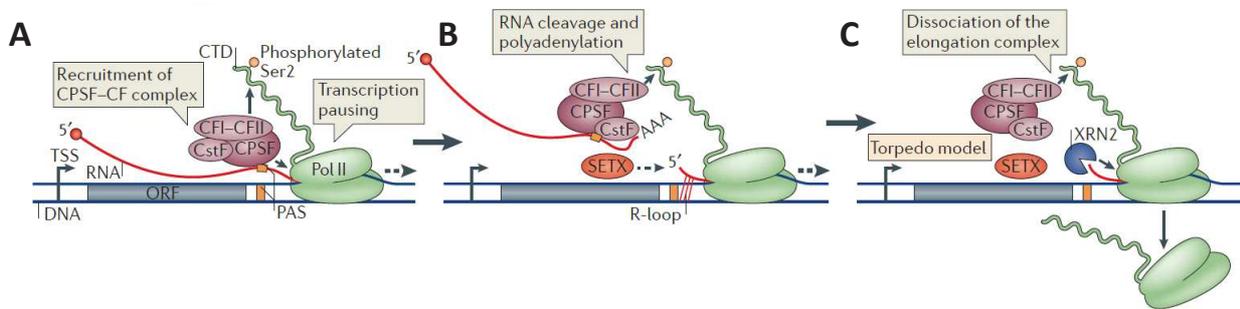
Figure 25: Schematic representation of factors implicated in polyadenylation-dependent termination at protein-coding genes in metazoans. A. Recruitment of 3' RNA processing complexes (CPSF, CstF and CFI-CFII) to the Pol II transcription machinery. Arrows indicate interactions with the modified CTD of Pol II and the Pol II body. The 3' RNA processing factors also contact sequences within the transcript such as the PAS, consequently resulting into pausing of the elongation complex. **B.** CPSF mediates endoribonucleolytic cleavage of the pre-mRNA downstream of the PAS followed by addition of a polyadenylation tail. The SETX helicase resolves potential R-loops of the remaining Pol II-bound RNA. **C.** Through the torpedo mechanism, XRN2 degrades the remaining Pol II-bound, nascent transcript and might lead, besides other factors, to the termination of Pol II. More detailed description in text. RNA polymerase II (Pol II) is shown in green. DNA template in dark blue. Nascent RNA in red. Ser2, Serine 2 residues of the Pol II CTD (Carboxy-terminal repeat domain); PAS, Polyadenylation site; TSS, transcription start site; ORF, open reading frame. From Porrua & Libri, 2015.

1.5.1. Polyadenylation-dependent termination

In polyadenylation-dependent termination, a PAS (polyadenylation signal), which gets transcribed and leads to the recruitment of RNA-binding 3' processing factors, is crucial (Richard & Manley, 2009; Kuehner et al., 2011; Porrua & Libri, 2015). Polyadenylation-dependent transcription termination represents the termination mechanism used at the large majority of protein-coding genes. Binding of 3' RNA processing factors to the PAS results in pausing of Pol II, endoribonucleolytic cleavage and polyadenylation of the nascent mRNA molecule. Seemingly a simple mechanism, polyadenylation-dependent termination involves over 14 proteins in mammals (Kuehner et al., 2011).

The two central protein complex of polyadenylation-dependent termination in metazoans, are CPSF (cleavage and polyadenylation specificity factor), which is recruited to the elongating Pol II through interaction with the Pol II core and consists of five subunits, and CstF (cleavage stimulatory factor), which forms contacts with the Serine 2 phosphorylated (Ser2P) CTD of Pol II and contains three subunits (more details on the CTD in chapter 1.1.1. Carboxy-terminal repeat domain of RNA polymerase II). Upon appearance of the conserved PAS hexanucleotide (AAUAAA) in the transcript, CPSF is thought to bind to it and induce pausing of Pol II (Figure 25A). Pausing of the transcription machinery has also been observed at termination sites of RNA polymerase I and III transcription but whether it is a requirement for transcription terminations remains unclear (Richard & Manley, 2009;

Porrua & Libri, 2015). Next, CstF interacts with the GU-rich processing signal located downstream of the PAS. Through interactions with CstF, CPSF dissociates from the Pol II core, allocates with CstF to the CTD and leads to the cleavage of the pre-mRNA (Figure 25B). Several other factors have been implicated in regulating the cleavage reaction of CPSF, such as CFI (cleavage factor I) and CFII (cleavage factor II). Cleavage of the transcript and addition of the polyadenylation tail occur roughly 18 to 30 bp downstream of the AAUAAA sequence and about 30 bp upstream of the GU-rich region (Richard & Manley, 2009; Porrua & Libri, 2015).

An additional factor that has been implicated in transcription termination is SETX (senataxin) which possesses a conserved helicase domain and is thought to resolve R-loops, a three-stranded structure of DNA and RNA (Figure 25B) (Richard & Manley, 2009; Porrua & Libri, 2015). Subsequently, XRN2 (5'-3' exoribonuclease 2) with its cofactors is believed to rapidly degrade the remaining Pol II-associated RNA from the unprotected 5' end (Figure 25C). Dissociation of Pol II from the DNA was proposed to be promoted by either conformational changes of the transcription machinery caused by factors binding to or dissociating from the paused Pol II complex (allosteric model) or collision of XRN2 with Pol II (torpedo model) (Figure 25C) (Porrua et al., 2016). Factors involved in the allosteric model include for example, Paf1, which dissociates from the elongation complex upstream of the PAS, as well as Spt4 and Spt5 (in mammals forming DSIF), which disengage from Pol II downstream of the PAS (Kuehner et al., 2011). Overall, the mechanism leading to the dissociation of Pol II from the DNA template is not yet completely understood (Richard & Manley, 2009; Kuehner et al., 2011). Alternatively, based on the presence of some termination factors at promoter regions, it was proposed that, instead of dissociating from the DNA, Pol II might be recycled back to the gene promoter through contacts of the 5' and 3' ends of genes (Richard & Manley, 2009; Kuehner et al., 2011). This model of 'gene looping' was suggested to enable subsequent rounds of transcription in rapid successions.

1.5.2. Polyadenylation-independent termination

At the short snRNA and replication-dependent histone genes, transcription termination happens in two alternative polyadenylation-independent manners.

At snRNA, a 13-16 nucleotide long 3'-box element located 9 to 19 bp downstream of the expected 3' end is required for snRNA end processing and transcription termination (Richard & Manley, 2009; Porrua & Libri, 2015). The large Integrator (INT) complex binds to this 3'-box and cleaves the snRNA through its INT9 and INT11 subunits (Figure 26) (Richard & Manley, 2009; Kuehner et al., 2011; Porrua & Libri, 2015). Integrator is thought to be recruited to mammalian snRNA-encoding genes by interacting with the phosphorylated Serine 7 residues of the Pol II CTD. Although snRNA genes share the 3'-box, the transcription termination process varies among snRNAs and the mechanistic details remain unclear (Richard & Manley, 2009).

In contrast to snRNAs, canonical histone mRNAs in metazoans accommodate two *cis*-regulatory elements required for transcription termination: a stem-loop structure recognized by SLBP (stem-loop binding protein) and a downstream sequence motif bound by the U7 snRNP (Richard & Manley, 2009). Binding of SLBP and the U7 snRNP enable the recruitment of cleavage factors processing the 3' ends of histone mRNAs. Matured replication-dependent histone mRNAs end with the stem-loop and a short single-stranded tail. The Pol II CTD was also reported to be important for histone 3' mRNA processing although the reasons are not yet understood (Richard & Manley, 2009).

NELF, a factor which is frequently related to promoter-proximal Pol II pausing (see chapter 1.3. Transcription pausing), was suggested to be involved in transcription termination following cleavage of both, snRNAs and canonical histone mRNAs (Figure 26B) (Buratowski, 2009; Kuehner et al., 2011; Porrua & Libri, 2015).
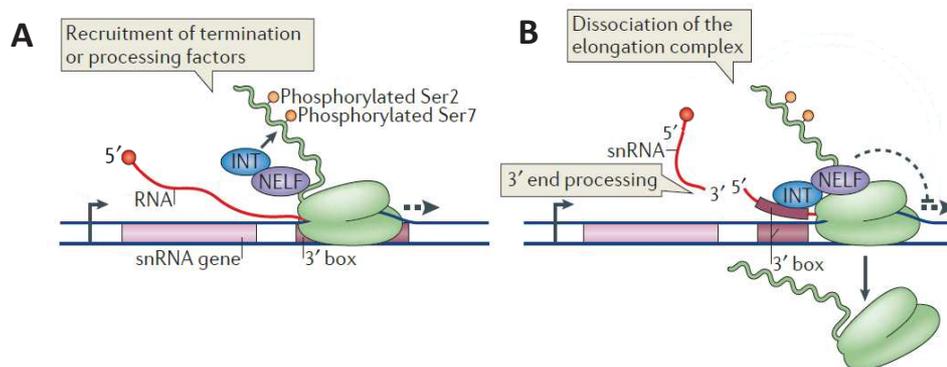
Figure 26: Scheme of mechanism and factors involved in transcription termination at small nuclear RNA genes in metazoans. A. The Integrator (INT) complex is recruited to the nascent snRNA (small nuclear RNA) through interactions with Serine 7 (Ser7) phosphorylated CTD (Carboxy-terminal repeat domain) of Pol II and through binding to the 3'-box sequence. **B.** Upon cleavage of the snRNA transcript by INT, NELF has been suggested to mediate Pol II termination. RNA polymerase II (Pol II) in green. RNA in red. Ser2, Serine 2. Adapted from Porrua & Libri, 2015.

2.   Chromatin

Around the 1880s, the chemically acidic DNA was found to be covered in basic proteins within the nucleus (Olins & Olins, 2003). In 1884, these small, basic proteins were named histones (Olins & Olins, 2003). Five different canonical histone proteins were subsequently identified: histone H1, H2A, H2B, H3 and H4 (Olins & Olins, 2003; McGinty & Tan, 2014). Nowadays, histone proteins are generally recognized as the skeleton for structures known as nucleosomes (Figure 27A). The core of nucleosomes consists of a histone octamer formed by two histone H3/H4 heterodimers (constituting a H3/H4 tetramer) each interacting with one histone H2A/H2B heterodimer (McGinty & Tan, 2014). This histone protein octamer is surrounded in roughly one and a half turns by 145-147 bp of DNA (McGinty & Tan, 2014). The fifth histone protein, the linker histone H1, is involved in higher order chromatin organization by stabilizing the DNA at the nucleosome entry and exit sites. During DNA replication in S-phase of the cell cycle, the mRNAs of canonical histone proteins are transcribed from genes found in different histone gene clusters. The human genome contains 10-20 functional copies per core histone, allowing the production of high quantities when the duplicated genome needs to be packaged (McGinty & Tan, 2014).

Interactions between dimerization partners in the core nucleosome occur through highly conserved histone fold domains (HFDs) leading to a pseudo-twofold symmetrical axis of the octamer (H2A-H2B-H4-H3-H3-H4-H2B-H2A) (McGinty & Tan, 2014). The central DNA base pair at which the nucleosome can be divided in pseudo-symmetrical halves is defined as the nucleosome dyad (arrowhead in Figure 27A) (McGinty & Tan, 2014). Each HFD consists of three α-helices, two short and one central long helix, connected by loops (McGinty & Tan, 2014). Intriguingly, HFD-like domains were also found in other proteins such as in subunits of the general transcription factor TFIID and the coactivator complex

A H3 H4 H2A H2B H1

B
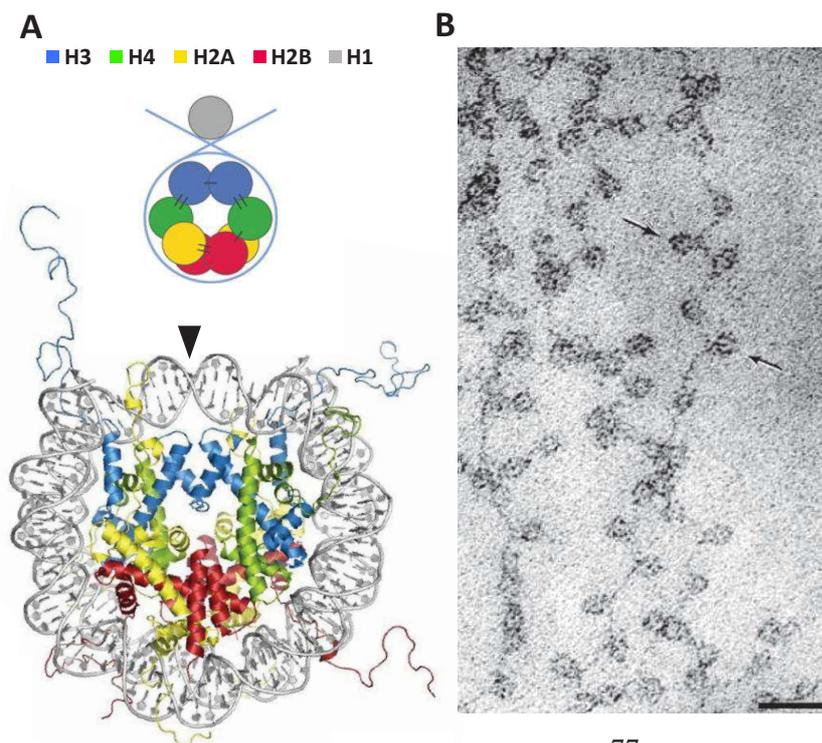
Figure 27: Chromatin organisation and nucleosome structure. A. Structure and composition of nucleosomes. Top, scheme of the nucleosome with histones represented by coloured circles and DNA represented by blue line. Bottom, crystal structure of the core nucleosome without histone H1. DNA in grey. Coloration of the histone proteins follows the scheme on top. Arrowhead indicates the nucleosome dyad. Adapted from McGinty & Tan, 2014 and Van Emmerik and van Ingen, 2019. **B.** Chromatin spread showing the 'bead on a string' organisation of chromatin. Arrows highlight nucleosome structures. Size marker: 30 nm. From Olins & Olins 2003.

SAGA leading to the suggestion that they could also form octamer-like structures (Patel et al., 2018; Kolesnikova et al., 2018; Papai et al., 2020; Wang et al., 2020).

Within the core nucleosome, every histone-fold pair was found to organize 27-28 bp of nucleosomal DNA through contacts at regular intervals, mostly through arginine side chains reaching out to the DNA phosphodiester backbone rather than the bases of the nucleotides (McGinty & Tan, 2014). Overall, the histone octamer interacts at 14 discrete places with the DNA resulting into one of the most stable DNA-protein associations (Li et al., 2007; McGinty & Tan, 2014). Contacts with the DNA backbone explain the lack of base specificity observed for histone-DNA associations. Besides the standard histone octamer, transient suboctameric stoichiometries of nucleosomes are thought to exist *in vivo* such as the hexasome and tetrasome, which lack one or both H2A/H2B dimers, respectively (McGinty & Tan, 2014).

Another key characteristic of histone proteins, beside their highly structured HFDs, are their flexible, highly modifiable N-terminal tails, which make up roughly 20 % of the entire histone octamer mass (Figure 27A) (McGinty & Tan, 2014; van Emmerik & van Ingen, 2019). These tails protrude out of the globular structure of the core nucleosome and could spread, if fully extended, to a length greater than the diameter of the nucleosome itself in the case of the 36 amino acid long histone H3 tail (McGinty & Tan, 2014).

Nucleosomes are separated from each other by linker DNA of various length resulting in a so-called 'beads on a string' chromatin organization, which was first proposed in 1974, almost a century after the discovery of histone proteins (Figure 27B) (Olins & Olins, 2003; McGinty & Tan, 2014). Chromatin was further suggested to form higher order structures such as the 30-nm fibre enabling eventually the formation of the metaphase chromosomes (McGinty & Tan, 2014). These 30-nm or other well organized and ordered fibres could however not be reproducibly detected *in vivo* (McGinty & Tan, 2014). Instead, in most cells chromatin was found to organize in rather disordered domains with a recent model suggesting that nucleosomes form so-called nucleosome 'clutches' (Fussner et al., 2012; Joti et al., 2012; Ricci et al., 2015; Ou et al., 2017). Nucleosome 'clutches' are thought to manifest along the chromatin fibre and to consist of discrete regions containing heterogeneous numbers of nucleosomes. The median number of nucleosomes per clutch was suggested to range from roughly 3 to 8 nucleosomes, separated by nucleosome-depleted regions (Ricci et al., 2015).

2.1. Nucleosome-depleted regions at *cis*-regulatory regions

Binding of proteins to DNA sequences can be impaired or restricted by nucleosomes. For example, nucleosomes represent major barriers for the RNA polymerase II (Pol II) initiation and elongation complexes (Bondarenko et al., 2006). Frequently however, active gene promoters and enhancers are found within so-called nucleosome-depleted regions (NDRs), which are flanked by a -1 (upstream) and

a +1 (downstream) highly positioned nucleosome (Lenhard et al., 2012; Hughes & Rando, 2014). NDRs are thought to facilitate the assembly and recruitment of components of the transcription preinitiation complex (PIC). Indeed, nucleosome occupancy at gene promoters was reported to negatively correlate with gene expression levels, as highly transcribed genes tend to display a stronger nucleosome depletion than lowly expressed genes (Hughes & Rando, 2014). Different types of gene promoters can further be distinguished based on distinct patters of nucleosome positioning (Lenhard et al., 2012). For example, housekeeping genes with several TSSs over a broad promoter region (broad promoters) were found to possess well-positioned -1 and +1 nucleosomes, while tissue-specific genes having one sharp TSS seem to have less ordered nucleosomes ('fuzzy' nucleosomes) and their TSS is frequently covered by a nucleosome (Rach et al., 2011; Nozaki et al., 2011; Lenhard et al., 2012).

NDRs are believed to be maintained at *cis*-regulatory elements through two main mechanisms: extrinsic factors such as chromatin modifications by chromatin remodellers or histone chaperones, which will be described in more details in chapter 3.2.1. Chromatin remodellers and histone chaperones, and intrinsic factors based on the DNA sequence composition (Hughes & Rando, 2014). For example, GC-rich regions, such as CpG islands found at several mammalian core promoters, are thought to disfavour the assembly of nucleosomes due to their higher rigidity and therefore more restricted DNA bendability compared to AT-rich regions (Deaton & Bird, 2011; Cramer, 2019). Recent studies in mouse and human cells however suggest that NDRs are not completely nucleosome free but instead are occupied by a class of 'fragile' nucleosomes (Jin et al., 2009; Voong et al., 2016; Haberle & Stark, 2018).

2.2. RNA polymerase II transcribing nucleosomal DNA

During transcription elongation, Pol II moves along the chromatinized template synthesizing RNA, while regularly encountering highly stable nucleosomes. Recent cryo-EM structures revealed how Pol II engages with the nucleosomal substrate (Figure 28) (Kujirai et al., 2018; Farnung et al., 2018; Noe Gonzalez et al., 2018; Ehara et al., 2019). Pol II was found to pause, in the presence of the elongation factor TFIIS, at multiple occasions while transcribing through the DNA bent around the histone octamer (Kujirai et al., 2018). Pausing occurred at regions where histone proteins strongly interact with the nucleosomal DNA and involved contacts between Pol II and the nucleosomal DNA (Kujirai et al., 2018; Noe Gonzalez et al., 2018). For instance, the head of the clamp domain of the RPB1 subunit of Pol II was found to contact the surface of the curved nucleosomal DNA (Figure 28) (Kujirai et al., 2018; Farnung et al., 2018; Noe Gonzalez et al., 2018).

Upon transcription and unwrapping of roughly 50 bp of the nucleosomal DNA by Pol II, one H2A/H2B heterodimer is exposed, which could allow its dissociation from the nucleosome structure. Amongst other reasons, interactions of the RPB2 lobe with this unstable H2A/H2B heterodimer were
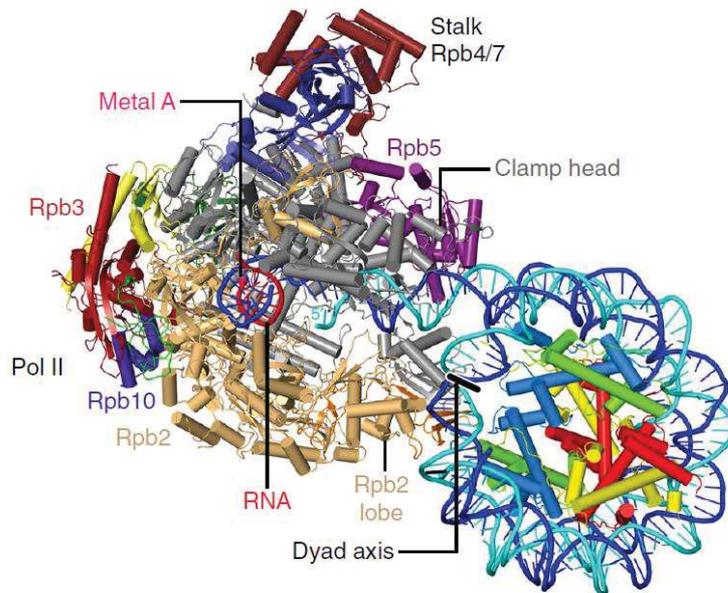
Figure 28: Yeast RNA polymerase II transcribing through a nucleosome. Subunits of RNA polymerase II (Pol II) are labelled in the structure, except for Rpb1 in silver, Rpb6 in cyan, Rpb8 in forest green, Rpb9 in orange, Rpb11 in yellow and Rpb12 in green. Histones H2A, H2B, H3 and H4 are coloured in yellow, red, blue and green, respectively. Template DNA strand shown in dark blue and complementary strand in cyan. Metal A indicates the active site of Pol II. The dyad axis represents the axis at which the nucleosome can be divided in two pseudo-symmetrical sections. From Farnung et al., 2018.

suggested to prevent the dissociation thereby preserving the octamer (Kujirai et al., 2018; Noe Gonzalez et al., 2018). After having transcribed half of the nucleosomal DNA (around 70 bp), Pol II is thought to move through the remaining DNA without obvious impediments. This is thought to be facilitated through histone transfer or eviction ahead of Pol II by chromatin remodellers, such as CHD1, or histone chaperones, such as FACT, (more details in chapter 3.2.1. Chromatin remodellers and histone chaperones) (Weber et al., 2014; Kujirai et al., 2018; Farnung et al., 2018; Noe Gonzalez et al., 2018).

2.3. Chromatin organization into euchromatic and heterochromatic domains

The nucleus is thought to be a very crowded environment, rendering its organization an essentiality to ensure efficiency of DNA-related processes such as Pol II transcription (Hancock, 2014). The first level of chromatin organization relies on the establishment and maintenance of euchromatic and heterochromatic regions within the nucleus, corresponding to transcriptionally active or inactive regions respectively. These domains are characterized not only based on their transcriptional status, but especially by post-translational modifications (PTMs) of histones and histone protein variants, which will be described next.

Different PTMs and histone variants are associated with distinct chromatin domains, turning nucleosomes to active signalling hubs (McGinty & Tan, 2014). Several studies have described numerous chromatin states with numbers ranging from fewer than 5 to more than 50 states (Baker, 2011). In the section 2.3.3. Combinations of histone PTMs and variants found at distinct chromatin domains, three of the most accepted chromatin states will be shortly described in more details based on their histone PTM signatures and histone variant incorporation: euchromatin, constitutive and facultative heterochromatin.

2.3.1. Histone post-translational modifications

Post-translational modifications (PTMs) on histone proteins were discovered more than 50 years ago and were subsequently implicated in affecting nucleosome stability and chromatin structure (Allfrey et al., 1964). Since then, numerous modifications were discovered besides the well characterized histone acetylation, methylation, phosphorylation and ubiquitylation marks, such as sumoylation, crotonylation, ADP ribosylation and deimination (Bhaumik et al., 2007; Kouzarides, 2007; Bannister & Kouzarides, 2011; McGinty & Tan, 2014). PTMs of histone proteins are frequently found on their flexible, lysine-rich N-terminal or C-terminal tails but can also occur within the globular body of histone proteins (Figure 29).

Different histone PTMs are associated with distinct transcriptional states and are further implicated in processes such as DNA damage repair, DNA replication or cell cycle transition, which will however not be detailed here (Bhaumik et al., 2007; Bannister & Kouzarides, 2011). The occurrence of histone lysine acetylation for instance is generally linked to actively transcribed regions, while histone lysine or arginine methylation can be found within transcriptionally active or inactive regions depending on the specific position of the modified residue (more details are given in the following subsections) (Bhaumik et al., 2007).

Histone modifications are thought to exert their role through two main mechanisms: First, some PTMs change the net charge of nucleosomes, which can influence the structure of chromatin by loosening nucleosomal DNA-histone, histone-histone interactions or contacts between adjacent nucleosomes (Li et al., 2007; Bannister & Kouzarides, 2011; McGinty & Tan, 2014; van Emmerik & van Ingen, 2019). Second, they can positively or negatively regulate the recruitment of downstream effector proteins (Bhaumik et al., 2007; Bannister & Kouzarides, 2011; van Emmerik & van Ingen, 2019). As such, specific histone modifications or combinations can be recognized by so-called 'reader' domains allowing the binding of proteins containing these domains to specific genomic locations

Figure 29: Post-translational modifications found on the human histone proteins H2A, H2B, H3 and H4. Represented are the positions of four key modifications occurring at N- and C-terminal tails or within the globular domain of the respective histone proteins. These four modifications include acetylation (green, ac), methylation (blue, me), phosphorylation (red, ph) and monoubiquitylation (purple, ub1). Globular domain of histone proteins shown as colored ovals. From Bhaumik et al., 2007.

marked by particular histone PTMs (Li et al., 2007; van Emmerik & van Ingen, 2019). For example, PHD fingers, chromo- and Tudor domains can recognize methylated lysines, while acetylated lysines can be bound by bromodomain-containing factors (Bannister & Kouzarides, 2011). These reader domains are frequently found within chromatin modifying complexes, which will be described in more details in a later chapter. In general, PTMs are established by dedicated 'writers' and diluted through rounds of cell divisions or actively removed by 'eraser' enzymes. In the following, histone acetylation, methylation and ubiquitinylation will be shortly described in more details.

2.3.1.1. Histone acetylation

Histone acetylation was first reported in 1964 and can occur only at lysine residues (Allfrey et al., 1964). Histone lysine acetylation is regulated through the opposing actions of two groups of enzymes: HATs (histone acetyltransferases, also referred to as lysine acetyltransferases, KATs) ('writers') and HDACs (histone deacetylases) ('erasers') (Bannister & Kouzarides, 2011). For the acetylation reaction, HATs require the cofactor acetyl-CoA, which is used to transfer an acetyl group to histone lysine side chains. Since acetylation neutralizes the positive charge of lysines, hyperacetylation of histones was suggested to decrease the affinity of the nucleosomal DNA to the histone proteins thereby promoting an open chromatin state (Bhaumik et al., 2007; Bannister & Kouzarides, 2011). Consistent with their positive impact on DNA accessibility, HATs are frequently found as subunits of transcriptional coactivator complexes (more details in chapter 3.2.2. Histone modifying complexes).

Two main classes of HATs were reported: type-A and type-B. Type-B HATs are located within the cytoplasm of cells and act on free, newly translated histone proteins (Bannister & Kouzarides, 2011). The acetylation of histones by type-B HATs is important for their later incorporation into chromatin. Type-A HATs, which act on histone proteins within nucleosomes, can be subdivided into three classes: the GNAT, MYST and CBP/p300 families (Bannister & Kouzarides, 2011). In general, enzymes of these three HAT families can modify several lysines at different sites within histones. Similarly, HDACs present also a relatively low substrate specificity and can deacetylate lysine residues antagonistic to HATs, restoring the positive charges of histone lysines. HDACs are divided into four classes (I to IV) with only class IV enzymes requiring the cofactor NAD^+ for their activity (Bannister & Kouzarides, 2011).

2.3.1.2. Histone methylation

Histone methylation can take place at lysine and arginine side chains. In contrast to histone acetylation, histone methylation is found in transcriptionally active as well as inactive regions (Bannister & Kouzarides, 2011). Histone lysines can be modified by in total three rounds of methylation from the unmethylated to mono-, di- and trimethylated forms, while arginines can be mono-, symmetrically or

asymmetrically di-methylated increasing the overall complexity of histone methylation (Bannister & Kouzarides, 2011). Methylation on lysine and arginine residues is established through HMTs (histone methyltransferases, also referred to as lysine methyltransferases, KMTs) and PRMTs (protein arginine methyltransferases), respectively.

All HMTs were found to contain a SET domain, which harbours the enzyme's active centre with one exception, the DOT1L enzyme. HMTs and PRMTs tend to modify only specific sites on histone proteins and depend on the methyl group donor SAM (S-adenosylmethionine) for the transfer of methyl-groups onto histone residues (Bannister & Kouzarides, 2011). For instance, DOT1L trimethylates specifically histone H3 lysine 79 (H3K79me3) located within the globular histone core (Bannister & Kouzarides, 2011). Further, some HMTs only modify their targets to a specific degree such as monomethylation but not di- or trimethylation, while others are capable of catalysing all three lysine methylation states. In contrast, all PRMTs can monomethylate arginine, but are classified in two groups based on their capabilities of modifying symmetric or asymmetric dimethylarginine (Bannister & Kouzarides, 2011).

HDMs (Histone lysine demethylases), acting antagonistically to HMTs, were found to be similarly sensitive to the degree of methylation and to also have high substrate specificities. Two types of HDMs were reported: LSD1 (lysine-specific demethylase 1) and JMJD (JmJC jumonji domain containing) demethylases. Arginine methylation was found to be reversed through deamination or through JMJD6 (Bannister & Kouzarides, 2011).

2.3.1.3. Histone ubiquitylation

In contrast to histone acetylation or methylation, modification of lysines with ubiquitin results in the addition of a large 76-amino acid polypeptide (Bannister & Kouzarides, 2011). Ubiquitylation is established through the sequential actions of three enzymes: E1 ('activating'), E2 ('conjugating') and E3 ('ligating') enzymes (Bannister & Kouzarides, 2011). Histones are mostly monoubiquitylated with the best described examples being histone H2A lysine 119 ubiquitylation (H2AK119ub) and histone H2B lysine 120 ubiquitylation (H2BK120ub). While H2AK119ub is implicated in gene silencing (facultative heterochromatin), H2BK120ub is found at actively transcribed genes, highlighting the divergent role the same histone modification can hold depended on its localization (Bannister & Kouzarides, 2011). Ubiquitin is removed from histones by isopeptidases termed deubiquitylation (DUB) enzymes.

2.3.1.4. Crosstalk between histone PTMs

Various histone PTMs have been implicated in so-called 'histone crosstalk' representing an additional regulatory layer (Bannister & Kouzarides, 2011). For example, competitive antagonism

between PTMs is found if the same residue can be marked by different chemical modifications such as at the histone H3 lysine 9 residue, which can be either acetylated (active transcription) or methylated (gene silencing) (Figure 29, page 81). Another example of histone crosstalk is the dependency of one modification on the deposition of another, as seen for methylation of histone H3 lysine 4 or 79 and H2BK120ub (Kim et al., 2009; Bannister & Kouzarides, 2011; Worden et al., 2019; Valencia-Sánchez et al., 2019; Hsu et al., 2019b). Interestingly, recent findings in yeast however indicate that the observed crosstalk between H2BK120ub and H3K79me3 would be rather dependent on protein interactions of the enzyme catalysing H3K79me3, Dot1 (ortholog of DOT1L), than on its methyltransferase activity (van Welsem et al., 2018).

2.3.2. Histone protein variants

Beside PTMs, nucleosome composition can also be altered by the replacement of the canonical histones with protein variants such as H3.3, H2A.Z and CENP-A, which carry few to extensive substitutions in the amino acid sequence (Li et al., 2007). Most histone variants are found for the canonical histones H3 and H2A, while only testis-specific variants are known for H2B and a variant of H4 was only recently described (Table 2) (Skene & Henikoff, 2013; Long et al., 2019).

Table 2: Human replication-independent histone variants and their specific chaperones and associated factors. Importantly, evidence for chaperones and associated factors may come from non-human species. * indicates testis-specific variants. n.d., not determined. From Skene & Henikoff 2013.

Histone	Chaperones and associated factors
H2A variants	
H2A.Z	CHZ1/NAP1/SWR1/INO80
H2A.X	FACT
macroH2A	ATRX
H2A.B forms *	n.d.
H2B variants	
H2B.1 *	n.d.
H2B.W *	n.d.
H3 variants	
CENP-A	HJURP
H3.3	HIRA/ATRX/DAXX
H3.4 *	n.d.

Specific chromatin remodellers and histone chaperones are responsible for the incorporation of histone variants into nucleosomes in exchange for canonical histone proteins (Table 2) (Skene & Henikoff, 2013). One of the main differences between histone variants and canonical histones is reflected in the moment of their chromatin incorporation: While canonical histones are only expressed and incorporated into nucleosomes during S-phase of the cell cycle, histone variants are integrated into nucleosomes in a DNA replication-independent way throughout the cell cycle (Li et al., 2007; Talbert & Henikoff, 2017). Further, mRNAs of histone variants are generally polyadenylated, contain introns and are not transcribed from clusters in contrast to canonical histone transcripts (Talbert & Henikoff, 2017).

Replacement of canonical histones with variants was suggested to change nucleosomal properties and especially affect interactions with downstream effector proteins. For example in mammals, centromeric chromatin is primarily epigenetically defined through the incorporation of CENP-A, a histone variant of canonical H3 (McKinley & Cheeseman, 2016). CENP-A is crucial for the localization of kinetochore components (McKinley & Cheeseman, 2016). Further, CENP-A incorporation is thought to result into more rigid nucleosomes and was implicated in leading to the more condensed chromatin conformations of centromeres (McKinley & Cheeseman, 2016).

In contrast, the replacement of canonical H3 by H3.3 was suggested to have few major consequences and instead to serve primarily to fill gaps in the chromatin landscape resulting from the disruption of nucleosomes especially at actively transcribed regions (Ray-Gallet et al., 2011; Talbert & Henikoff, 2017). Nevertheless, *in vivo* findings suggest that H3.3-containing nucleosomes are less stable than nucleosomes formed with canonical H3 proteins (Jin & Felsenfeld, 2007). A variant protein of the canonical H2A, H2A.Z, was additionally implicated in forming less stable nucleosomes at transcriptionally active regions (Jin & Felsenfeld, 2007; Talbert & Henikoff, 2017). Both, H3.3 and H2A.Z are thought to be components of 'fragile' nucleosomes suggested to be located within and in proximity of NDRs (Jin et al., 2009).

2.3.3. Combinations of histone PTMs and variants found at distinct chromatin domains

As mentioned in a previous section, histone PTMs and the incorporation of specific histone protein variants allow the differentiation, amongst others, of active euchromatin and silent facultative or constitutive heterochromatin, which are shortly described in the following.

2.3.3.1. Euchromatin

Euchromatin contains actively transcribed genes including their *cis*-regulatory elements and is characterised by an open chromatin structure and a general decompaction (Bannister & Kouzarides, 2011). A modification commonly associated with euchromatin is the acetylation of histone proteins. In

addition, certain regions within euchromatin such as gene bodies of actively transcribed genes, active enhancers and active promoters were found to display specific histone PTM signatures (Li et al., 2007; Bannister & Kouzarides, 2011).

Gene bodies of actively transcribed genes are enriched for histone H3 lysine 36 trimethylation (H3K36me3), H3K79me3 and histone H2BK120ub (Figure 30). Active enhancers are traditionally identified by a combination of high levels of histone H3 lysine 4 monomethylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac) marks, while active promoters are enriched for histone H3 lysine 4 trimethylation (H3K4me3) (Figure 30) (Bannister & Kouzarides, 2011; Schoenfelder & Fraser, 2019). The presence of H3K4me3 at gene promoters was suggested to provide a 'bookmark', highlighting recent transcription activities in the genome, which could facilitate subsequent rounds of transcription (Ng et al., 2003). Active forms of enhancers and promoters are further enriched for the histone variants H3.3 and H2A.Z (Pradhan et al., 2016).

Recent studies nevertheless indicate that enhancer and promoter elements cannot be strictly distinguished based on these histone modifications (Pekowska et al., 2011; Ernst et al., 2011; Core et al., 2014; Andersson, 2015; Andersson et al., 2015b). For example, some promoter elements were found to display high levels of H3K4me1, similarly some enhancers display high levels of H3K4me3 (Pekowska et al., 2011; Ernst et al., 2011; Core et al., 2014). Indeed, H3K4me3 was reported to be found in general at highly active promoters or enhancers and to correlate with Pol II accumulation at these
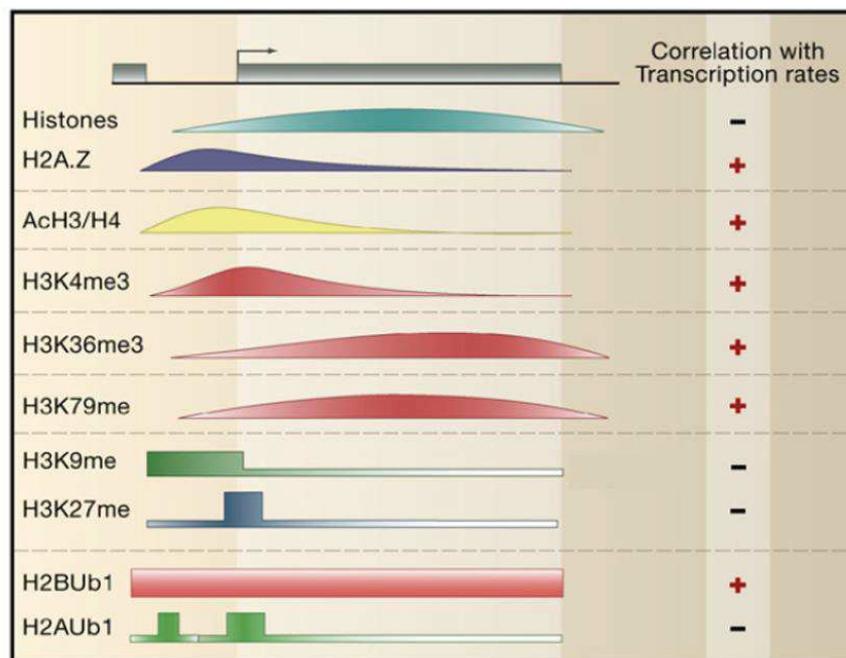
Figure 30: Distribution of histone post-translational modifications and histone variant H2A.Z. Schematic representation of profiles of histone marks and H2A.Z along an arbitrary gene. Transcription start site indicated by arrow. The correlation of the respective histone modifications and H2A.Z with transcription is indicated on the right. Adapted from Li et al., 2007.

elements, instead of distinguishing them from each other (Pekowska et al., 2011; Core et al., 2014). H3K27ac is also frequently found at active promoters (Ernst et al., 2011). These findings, amongst others, led to the suggestion of a unified architecture model of *cis*-regulatory elements (see also chapter 1.2.1.5. Unified architecture of *cis*-regulatory elements) (Andersson, 2015; Andersson et al., 2015b).

2.3.3.2. Facultative heterochromatin

Genes, which are differentially expressed in distinct cells or tissues, are found within facultative heterochromatin when they are transcriptionally inactive (Bannister & Kouzarides, 2011). This chromatin domain is associated with high levels of the histone PTMs H2AK119ub established by PRC1 (Polycomb repressor complex 1) and H3 lysine 27 trimethylation (H3K27me3) deposited by PRC2 (Polycomb repressor complex 2) (Figure 30) (Bannister & Kouzarides, 2011; Chittock et al., 2017). Beside depositing H3K27me3, PRC2 can also read this modification through one of its subunits, which enables the maintenance of H3K27me3 over broad genomic domains (Bannister & Kouzarides, 2011).

Inactivation of genes by the Polycomb group proteins and formation of facultative heterochromatin represents a plastic mode of gene silencing and allows transcriptional activation of genes upon specific cues. Another classic example for this type of chromatin domain is the inactive X-chromosome in female cells, which is additionally covered in nucleosomes containing the H2A histone variant, macroH2A (Bannister & Kouzarides, 2011; Skene & Henikoff, 2013).

2.3.3.3. Constitutive heterochromatin

Permanently silenced genes, such as centromere and telomere regions or repetitive DNA sequences, are found within constitutive heterochromatin (Janssen et al., 2018). This chromatin domain is marked by an enrichment of histone H3 lysine 9 trimethylation (H3K9me3) (Figure 30, page 86). In mammalian cells, H3K9 methylation involves five methyltransferases: Mono- and dimethylation is established by G9a and GLP (G9a-like protein), while SETDB1 (SET domain bifurcated histone lysine methyltransferase 1), SUV39H1 (Suppressor of variegation 3-9 homolog 1) and SUV39H2 (Suppressor of variegation 3-9 homolog 2) can catalyse H3K9 di- and trimethylation (Bannister & Kouzarides, 2011; Janssen et al., 2018).

Constitutive heterochromatin is compacted into condensed structures through the binding of the HP1 (heterochromatin protein 1) proteins (Janssen et al., 2018). In mammalian cells, HP1α, HP1β and HP1γ can bind to H3K9me3 via their chromodomains, and lead, through self-interactions, to the compaction of constitutive heterochromatin (Janssen et al., 2018). Centromeres are further defined by the incorporation of the histone variant CENP-A, as described earlier.

2.3.4. Correlation or causality of histone PTMs and variants at chromatin domains

Although histone PTMs and variants are highly correlated with different chromatin states, whether they are causal remains unclear (Pollex & Furlong, 2017). Recent studies indicate that individual histone modifications found at euchromatic regions such as H3.3K27ac, H3K4me1, H3K4me3 and H4 acetylation are dispensable for successful transcription at most genes and in most cells (Hödl & Basler, 2009, 2012; Rickels et al., 2017; Dorighi et al., 2017; Acharya et al., 2017; Zhang et al., 2020). These was either found by substitutions of the critical lysines with a non-modifiable amino acid (Hödl & Basler, 2009, 2012; Zhang et al., 2020) or by inactivation of the catalytic sites of the responsible effector proteins (Dorighi et al., 2017; Rickels et al., 2017; Acharya et al., 2017). Further, complete deletion of the histone variant H3.3 in *Drosophila* flies did not majorly affect viability of cells with the intriguing exception of the reproductive tissues (Hödl & Basler, 2009).

3. Regulated transcription

Cues from the cell environment frequently lead to regulated changes in gene expression, which are mediated through the action of transcription factors (TFs) possessing repressing or activating roles in RNA polymerase II (Pol II) transcription (Levine et al., 2014). The following subsections will be focused on activating transcription factors (also called activators) and chromatin modifying complexes including complexes that can act as transcriptional coactivators. Coactivator complexes interact with activating TFs and are thought to mediate the stimulating signal of these factors to the basal transcription machinery through mechanisms establishing a permissive chromatin structure at gene promoters (more details later).

3.1. Transcription factors

Transcription factors (TFs) can interact with specific DNA sequences through DNA-binding domains, which can be very variable in nature (Garvie & Wolberger, 2001). Sequence-specificity of TFs is enabled by the recognition of base pairs, exposed in the major or minor groove of DNA, most commonly through α-helix or β-sheet structures protruding into the grooves. All protein-DNA interactions further involve contacts with the negatively charged sugar-phosphate backbone of DNA (Garvie & Wolberger, 2001).

For instance, the most abundant class of DNA-binding proteins in the human genome is the zinc (Zn) finger family whose members possess several copies of the name-giving, roughly 30 amino acid long DNA-binding domain, the Zn finger (Lander et al., 2001; Garvie & Wolberger, 2001). The Zn finger is the most minimal DNA-binding domain consisting of one rather short α-helix, two antiparallel β-sheets and one core Zn^{2+} ion (Figure 31). At least two Zn fingers are required to allow DNA binding (Garvie & Wolberger, 2001). The short α-helices of the Zn fingers insert into the DNA major groove thereby recognizing 3-4 base pairs, which can result, in the case of the three successive Zn fingers of the ZIF268 protein (also known as ZNF268 or EGR1), in the tracking of the DNA major (Figure 31) (Garvie & Wolberger, 2001). Importantly, the Zn^{2+} ion possesses only structural roles by ensuring the protein fold and does not contact DNA.
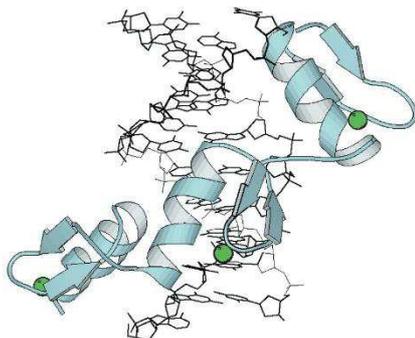
Figure 31: The zinc finger, example of a DNA-binding domain of transcription factors. Shown is the structure of the most common DNA-binding domain found within human transcription factors (TFs), the zinc (Zn) finger. Here the three Zn finger domains of the ZIF268 (ZNF268, EGR1) protein are shown. DNA helix in black. Protein structures in light blue. Zn^{2+} ions in green. The Zn fingers are tracking along the major groove of the DNA. From Garvie & Wolberger 2001.

In search for their cognate high affinity binding targets, TFs were found to unspecifically interact with genomic DNA in a random manner (Chen et al., 2014). The high-affinity DNA-binding motifs of TFs are frequently represented by specific, short sequence patterns termed TFBS (transcription factor binding sites), which are typically 6 bp long in humans (Slattery et al., 2014). TFBSs of different TFs frequently cluster within gene *cis*-regulatory regions (Slattery et al., 2014; Levo & Segal, 2014). Although the genome frequently contains several canonical TFBS for each TF, only a fraction of these TFBSs were found to be actually bound by these TFs *in vivo* as assessed by ChIP-seq (Chromatin immunoprecipitation followed by deep sequencing). Additionally, only a fraction of these binding events were found to impact gene expression (Slattery et al., 2014). As such, binding of a TF to its TFBS does not necessarily reflect in changes in gene expression.

It is unclear yet how exactly the selective and transcriptionally important binding events of TFs are achieved. Numerous additional properties besides DNA sequence can affect TF-DNA contacts and their stimulatory consequences such as for example TF dimerization or tetramerization, DNA shape, cooperative DNA-binding with other TFs, nucleosome occupancy at the TFBS and interactions with coactivators (Slattery et al., 2014; Levo & Segal, 2014).

3.1.1. Pioneer transcription factors

TFBSs are frequently found within nucleosome-depleted regions but can also be occupied by nucleosomes. DNA inaccessibility is in general thought to be a major restriction for the binding of most TFs and is therefore believed to considerably contribute to DNA-binding specificities, as indicated above (Slattery et al., 2014; Levo & Segal, 2014). Some TFs however were described to possess the ability to bind to their target TFBS within nucleosomal DNA and are believed to consequently promote chromatin accessibility at these locations (Slattery et al., 2014; Zhu et al., 2018; Dodonova et al., 2020). These TFs are referred to as 'pioneer' TFs since the destabilization of nucleosomes through their binding enables the interaction of other TFs with the now accessible DNA (Figure 32) (Slattery et al., 2014).

In an *in vitro* assay, binding to nucleosomal DNA was found to be common among human TFs and to occur in different patterns and at distinct sites along the nucleosome (Zhu et al., 2018). For example, at nucleosomal DNA, Sox2 and Sox11 showed preferential interaction with DNA roughly 20 base pairs
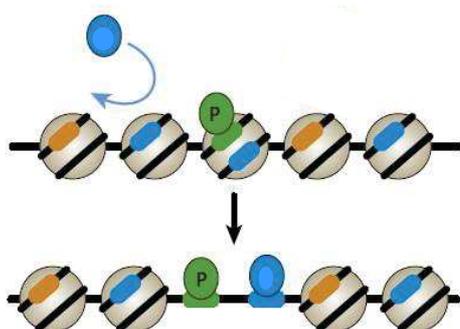
Figure 32: Pioneer transcription factors bind to nucleosomal DNA and increase chromatin accessibility. Schematic representation of the action of 'pioneer' TFs. Pioneer TFs (shown as green ovals highlighted with a P) can bind to nucleosome-associated DNA-binding sites (shown as coloured boxes), while other non-pioneer factors (blue ovals) are occluded. The binding of pioneer factors was suggested to lead to increased chromatin accessibility subsequently allowing binding of non-pioneer factors. Adapted from Slattery et al., 2014

downstream of the dyad (the central base pair of nucleosomal DNA) (Zhu et al., 2018; Dodonova et al., 2020). Binding of the Sox TFs to nucleosomes was found to distort nucleosomal DNA at their binding site, which was further suggested to enable the disengagement of terminal nucleosomal DNA from the histone proteins and repositioning of the N-terminal tails of histone H4 (Dodonova et al., 2020). Curiously, although nucleosome-disruptive, interactions of pioneer TFs with nucleosomal DNA were suggested to be relatively transient, as pioneer TFs tend to still favour free over nucleosomal DNA (Zhu et al., 2018).

Besides direct nucleosome binding by pioneer TFs, access to DNA occupied by nucleosomes can be facilitated through several other mechanisms. For instance, spontaneous unwrapping of nucleosomal DNA from the histone octamer can enable DNA access of non-pioneer TFs to nucleosome-occupied TFBS (Li et al., 2007; Slattery et al., 2014). Cooperative binding of several, unrelated TFs at these 'nucleosome breathing' sites was further suggested to result in a competition with the nucleosome for DNA binding. Indeed, interactions of pioneer TFs with nucleosomal DNA were recently reported to tend to occur at multifactor binding sites and to temporally coincide with binding of other TFs, which could facilitate the destabilization of nucleosomes (Meers et al., 2019). Nucleosomal DNA can further be exposed during DNA replication or through chromatin remodelling activities (Morris et al., 2014; Ramachandran & Henikoff, 2016). Validating pioneer functions of TFs *in vivo* and distinguishing their impact on chromatin opening from the above mentioned other processes, which allow in general access to nucleosomal DNA, remains challenging (Dufourt et al., 2018; Meers et al., 2019).

Recent findings additionally indicate that several TFs, such as for example the pioneer TF Sox2, can bind to mitotic chromosomes, structures that were previously thought to disrupt and occlude any DNA binding and gene regulatory processes (Kadauke & Blobel, 2013; Teves et al., 2016; Festuccia et al., 2016; Liu et al., 2017; Festuccia et al., 2017b; Bellec et al., 2018). This ability was termed 'mitotic bookmarking' and was suggested to enable the labelling and therefore maintenance of cell-specific transcription programs during mitosis when transcription generally halts. Mitotic bookmarking is further believed to facilitate rapid gene induction in early G1-phase once mitosis is accomplished (Kadauke & Blobel, 2013; Festuccia et al., 2016; Ferraro et al., 2016; Festuccia et al., 2017b; Bellec et al., 2018).

3.1.2. Models of integration of transcription factor signals

To influence gene expression, enhancers and proximal regulatory regions, that contain clusters of TFBS, are thought to integrate signals of a multitude of TFs. Two models were suggested to explain how this integration could occur: the 'enhanceosome' and the 'billboard' model (Lelli et al., 2012; Lorberbaum & Barolo, 2013; Slattery et al., 2014). Enhanceosome-like regulatory elements are thought to be rather rare as they are suggested to depend on an 'all-or-nothing' mechanism (Figure 33A). To be able to recruit coactivators and to stimulate gene expression according to the enhanceosome model,

multiple TFs have to cooperatively bind and form protein-protein interactions amongst each other. Enhanceosomes therefore consist of a highly constrained spacing- and orientation-specific organization of several TFBSs (Figure 33A) (Slattery et al., 2014). In contrast, the billboard model allows a more flexible architecture of TFBS sequences at regulatory elements. TFs in the billboard model collaborate to regulate gene expression output but do not act as a cooperative unit (Figure 33B).

Based on their distinct characteristics, enhanceosomes were suggested to induce rather switch-like or digital ('on'/'off') gene expression programs as required during cell differentiation and development preventing unwanted gene induction if not all activating TFs are present (Papatsenko & Levine, 2007; Lorberbaum & Barolo, 2013; Chen et al., 2014; Slattery et al., 2014). In contrast, regulatory elements following the billboard mechanism are believed to be implied in graded or analogue induction of gene expression, meaning that expression levels increase with each TF bound (Lorberbaum & Barolo, 2013; Stewart-Ornstein et al., 2013; Slattery et al., 2014). The two models detailed here are believed to represent the extremes of a continuum of possible ways how several TF binding events at gene regulatory regions can direct gene expression changes (Lorberbaum & Barolo, 2013; Slattery et al., 2014).
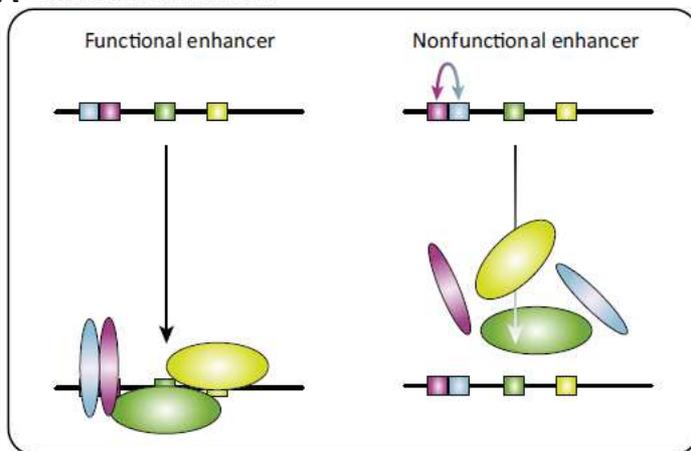
Figure 33: Integration of transcription factor binding at enhancers. A. In the enhanceosome model, binding of TFs occurs in a highly coordinated and constrained manner (left). Minor changes in the order of TF binding sites (TFBS) leads to a failure of TF assembly and consequently a nonfunctional enhancer (right). **B.** In the billboard model is based on a highly flexible TFBS order (left). Changes in position or orientation of TFBS does not affect binding of TFs and enhancer function (right). From Slattery et al., 2014.
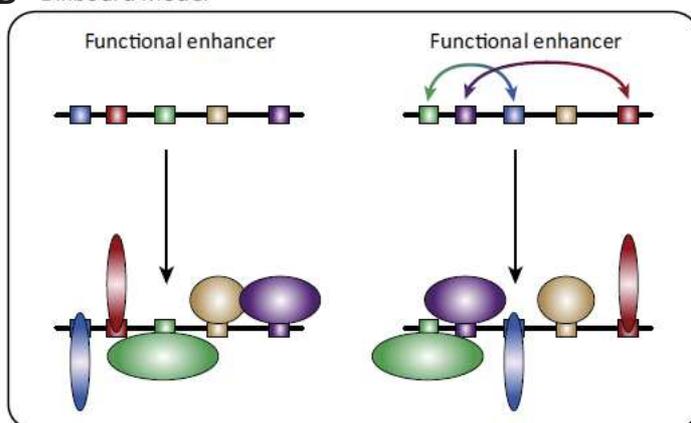
3.2. Chromatin modifying complexes as RNA polymerase II coactivators

As a consequence of genome packing into chromatin, RNA polymerase II (Pol II) transcription has to overcome the restriction imposed by the limited access to the DNA template. At the step of transcription initiation, the currently most prevailing model is that once activating transcription factors (TFs) associate to enhancers or proximal regulatory regions, they form protein-protein interactions leading to the recruitment of transcriptional coactivators, which mediate the regulatory cues to core promoters and the basal transcriptional machinery (Li et al., 2007; Sikorski & Buratowski, 2009). As transcriptional coactivators are frequently multiprotein complexes containing one or more enzymes with chromatin modifying functions, they are thought to positively stimulate transcription by rendering core promoter sequences accessible for components of the preinitiation complex (PIC) (Li et al., 2007). During transcription elongation, several other chromatin modifying complexes are involved in enabling the efficient passage of Pol II through the chromatinised DNA template serving as additional transcriptional cofactors.

The two main categories of chromatin modifying complexes are (1) chromatin remodellers such as the SWI/SNF complexes and histone chaperones such as FACT, and (2) histone modifying complexes such as the COMPASS(-like) and TIP60 complexes, which will be described shortly in the next subsections. Individual, later chapters of the Introduction will be solely dedicated to the SAGA and ATAC coactivator complexes as my thesis project was to better understand the role of these two histone modifying complexes in mammalian Pol II transcription. As mentioned in a previous chapter (see chapter 1.2.2. Preinitiation complex formation), multiprotein complexes with no obvious chromatin modifying function, such as TFIIA, TFIID and the Mediator complex, are sometimes also referred to as transcriptional coactivators as they can form protein-protein interactions with activating TFs and the transcription machinery resulting into regulated transcription. This chapter however will be restricted to the description of chromatin modifying complexes acting as transcriptional coactivators or cofactors. Intriguingly, chromatin modifying complexes are in general highly conserved throughout eukaryotic evolution, which underpins their general importance.

3.2.1. Chromatin remodellers and histone chaperones

Factors or complexes involved in nucleosome positioning and chromatin assembly or disassembly are main determinants of nucleosome-depleted regions (NDRs) at gene regulatory regions and are further crucial to enable transcription elongation through chromatin (McGinty & Tan, 2014; Hughes & Rando, 2014). These factors or complexes fall in one of two classes, either chromatin remodellers or histone chaperones. While chromatin remodelling complexes use the energy of ATP hydrolysis to transiently unwrap DNA from histones or to move nucleosomes along DNA, histone chaperones act independently of ATP and guide especially nucleosome assembly and disassembly. Both classes are

further responsible for the composition of nucleosomes by mediating the incorporation or ejection of histone variants (McGinty & Tan, 2014). Within cells, the most frequent target of chromatin remodellers and histone chaperones is thought to be the H2A/H2B dimers as they were found to very dynamically exchanged in and out of nucleosomes (Li et al., 2007).

3.2.1.1. Chromatin remodellers

Chromatin remodellers are classified based on similarities and differences in their ATPase domains into four main subfamilies: the ISWI (imitation switch), the CHD (chromodomain helicase DNA-binding proteins), the SWI/SNF (switch/sucrose non-fermentable) and the INO80 (Inositol-requiring mutant 80) subfamilies (Clapier et al., 2017). In metazoans, cell type- and development-specific subtypes for each of these subfamilies were described as well as orphan remodellers, which cannot be classified into one of these subfamilies (Clapier et al., 2017).

Although all chromatin remodellers independent of the subfamilies contain an ATPase-translocase motor, only members of the INO80 subfamily were reported to have the capacities of editing canonical nucleosomes with histone variant-containing dimers. Members of the ISWI and CHD subfamilies seem
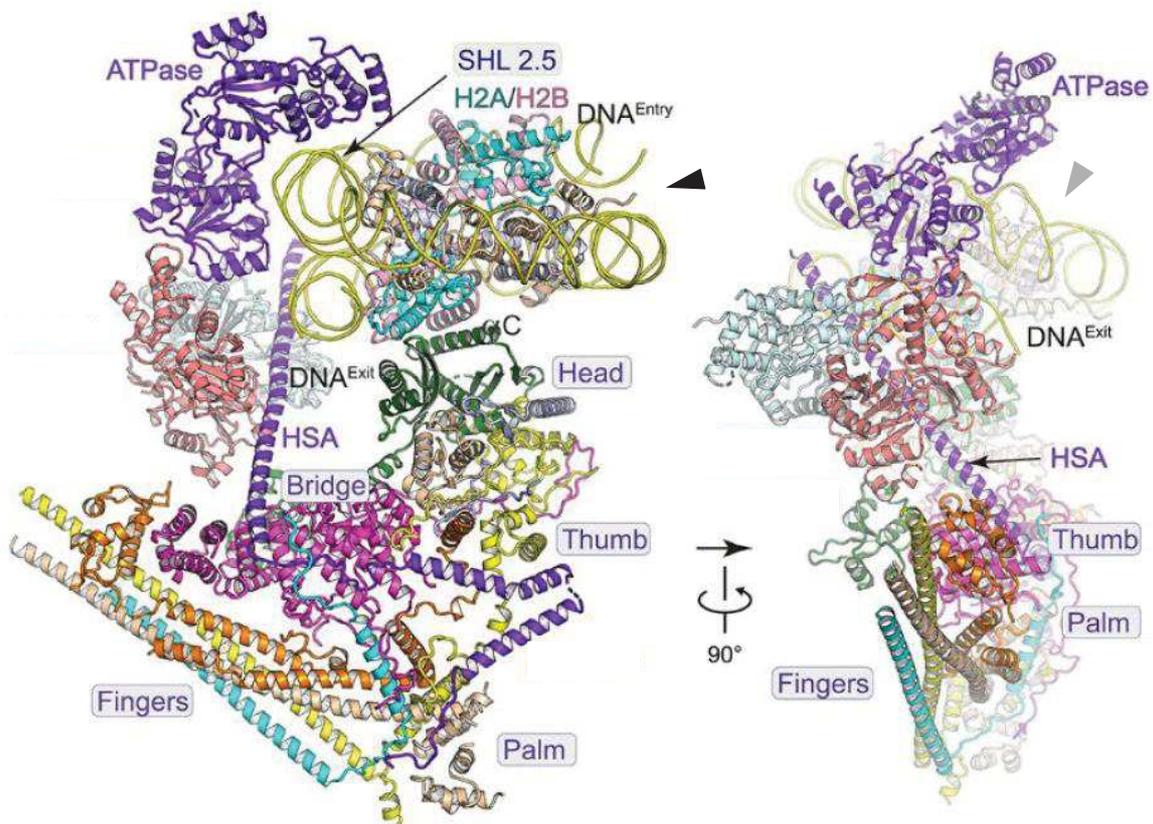
Figure 34: Cryo-EM structures of the human BAF remodelling complex bound to the core nucleosome. Cartoon model of the human BAF (Brg1/Brahma-associated factors) remodelling complex bound to the core nucleosome from two sides. Arrowheads indicate the nucleosome core. Nucleosomal DNA helix is shown in yellow. H2A in turquoise, H2B in rose, H3 in brown and H4 in light blue. The fingers, bridge, palm, thumb and head submodules are indicated as well as the DNA entry (DNA[Entry]) and exit (DNA[Exit]) sites and the ATPase module. HSA, helicase-SANT-associated. Adapted from He et al., 2020.

mainly involved in nucleosome assembly and spacing, in contrast to SWI/SNF complexes which typically promote chromatin accessibility and are thought to maintain NDRs at regulatory elements (Clapier et al., 2017).

Recent cryo-EM studies allowed the visualization of yeast and human members of the SWI/SNF subfamily revealing the intriguing manner in which they engage with the core nucleosome (human BAF complex in Figure 34) (Patel et al., 2019; Ye et al., 2019; Han et al., 2020; Wagner et al., 2020; He et al., 2020). Similarities amongst the SWI/SNF cryo-EM structures include (i) the location of the ATPase motor in proximity to the nucleosome substrate around superhelical location (SHL) +2 (roughly 20 base pairs downstream of the nucleosome dyad), (ii) contacts with the H2A/H2B acidic patch of the histone octamer through subunits containing positively charged, arginine-rich regions and located on the opposite side of the nucleosome relative to the ATPase module (nucleosome 'anchor'), (iii) the 'sandwiching' of the nucleosome by the two modules, one containing the catalytic ATPase and the other representing the nucleosome anchor (Figure 34).

Models that were proposed based on previous biochemical studies and which fit with these cryo-EM structures include models in which SWI/SNF complexes sandwich the nucleosome to enable a stable engagement of the ATPase with the nucleosomal DNA and, through ATP-dependent DNA-translocation, lead to the ejection of the nucleosome bound by the complex or, through sterical conflicts, the adjacent nucleosome (Clapier et al., 2017; He et al., 2020).

3.2.1.2. Histone chaperones

Histone chaperones are a diverse class of proteins which are in general structurally unrelated but frequently display acidic features involved in charge neutralization of the basic histone proteins (Kemble et al., 2015). The essential and highly conserved H2A/H2B histone chaperone FACT (facilitates chromatin transcription) is required for productive elongation of Pol II through the chromatin template and consists of two subunits, SUPT16 and SSRP1, in humans (Orphanides et al., 1998; Liu et al., 2020). The general requirement of this factor is highlighted by phenotypes observed in mice with homozygous inactivation of the SSRP1 subunit, which causes embryonic lethality soon after implantation with impaired growth of the inner cell mass at E3.5 (Cao et al., 2003).

FACT was shown to allow for DNA accessibility without dispersing the histones, enabling efficient reassembly of the nucleosome once Pol II moved through the DNA (Formosa, 2012). FACT therefore seems crucial to preserve chromatin integrity as it allows for Pol II transcription through the chromatinized template without loss of nucleosomes. This is enabled by FACT-mediated removal of H2A/H2B dimers and simultaneous stabilization of the H3/H4 tetramer. Recent crystal and cryo-EM structures of yeast and human FACT revealed details of the interactions of FACT with the nucleosomal
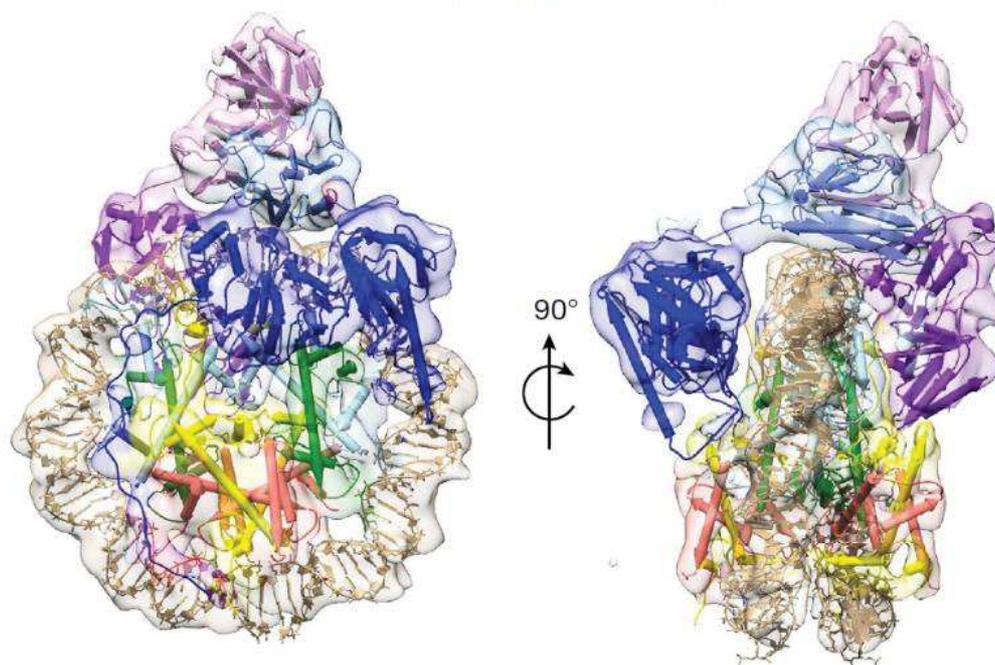
Figure 35: Cryo-EM structure of the human FACT histone chaperone bound to the core nucleosome. The SUPT16 subunit of FACT (facilitates chromatin transcription) is shown in shades of blue, while SSRP1 is displayed in shades of purple. DNA is coloured in brown. Histone H2A in yellow, H2B in orange, H3 in light green and H4 in green. Adapted from Liu et al., 2020.

DNA as well as the histone proteins and led to the suggestion of a model for the assembly and disassembly of the nucleosome by FACT (cryo-EM structure in Figure 35) (Kemble et al., 2015; Tsunaka et al., 2016; Liu et al., 2020).

In this model, FACT contacts the H3/H4 tetramer through the middle domains of its SUPT16 and SSRP1 subunits once the nucleosomal DNA is loosened by the passage of Pol II until close to the nucleosome dyad (Kujirai et al., 2018; Liu et al., 2020). The middle domains of SUPT16 and SSRP1 were found to bind to the H2A-interaction surfaces on the H3/H4 tetramer thereby competing with the H2A/H2B dimers for interaction with the H3/H4 tetramer (Tsunaka et al., 2016; Liu et al., 2020). The acidic C-terminal regions of SUPT16 and SSRP1 each further contact one H2A/H2B dimer mainly through interactions with H2B (Kemble et al., 2015). SUPT16 can stably bind a H2A/H2B dimer by preventing interactions of the histones with DNA by mimicking with its C-terminal region the acidic nature of DNA (Kemble et al., 2015; Liu et al., 2020).

The interactions of FACT with the H2A/H2B dimer and hindrance of contacts with the H3/H4 tetramer were suggested to lead to the disruption of histone-DNA associations and to nucleosome disassembly with the help of additional factors such as chromatin remodellers (Kemble et al., 2015; Tsunaka et al., 2016). Through contacts to the H3/H4 tetramer and the H2A/H2B dimers, FACT is thought to maintain the H3/H4 tetramer in the absence of DNA and to tether all the nucleosome components for subsequent reassembly (Liu et al., 2020). Upon association of DNA to the H3/H4

tetramer, FACT is believed to promote the reassembly of the nucleosome upstream of the elongating Pol II through the addition of the tethered H2A/H2B dimers (Liu et al., 2020).

3.2.2. Histone modifying complexes

In 1996, the first nuclear histone modifying protein Gcn5, a type-A HAT enzyme of the GNAT family, was identified which was rapidly followed by subsequent discoveries of other histone modifying factors (Brownell et al., 1996; Kuo et al., 1996; Bhaumik et al., 2007). Histone modifying enzymes are generally found within multiprotein complexes, for instance, one year after its discovery, Gcn5 was identified as subunit of the SAGA coactivator complex (Grant et al., 1997). Association of histone modifying enzymes within protein complexes was found to have important roles in regulating the recruitment, substrate specificity and activity of these enzymes (Grant et al., 1997; Bannister & Kouzarides, 2011). For example, by itself Gcn5 was found to be catalytically inactive against nucleosomal substrates, while within SAGA, Gcn5 efficiently acetylates lysines of nucleosomal histones (Grant et al., 1997). Besides the SAGA coactivator complex, which will be described in more details in the next chapter, two highly studied examples are the COMPASS complex, including COMPASS-like complexes in metazoans, and the TIP60 coactivator complex, which will be shortly described in the following subsections.

3.2.2.1. COMPASS and COMPASS-like complexes

The highly conserved COMPASS (complex of proteins associated with Set1) and COMPASS-like complexes are responsible for the methylation of histone H3 lysine 4 (H3K4me) in mammalian cells (Meeks & Shilatifard, 2017). Intriguingly, while in yeast only one COMPASS complex with Set1 as unique catalytic subunit exists, mammalian cells express six Set1-related methyltransferases: SETD1A (KMT2F), SETD1B (KMT2G), MLL1 (KMT2A), MLL2 (KMT2B), MLL3 (KMT2C) and MLL4 (KMT2D). These six methyltransferases represent the catalytic subunits of six distinct multiprotein
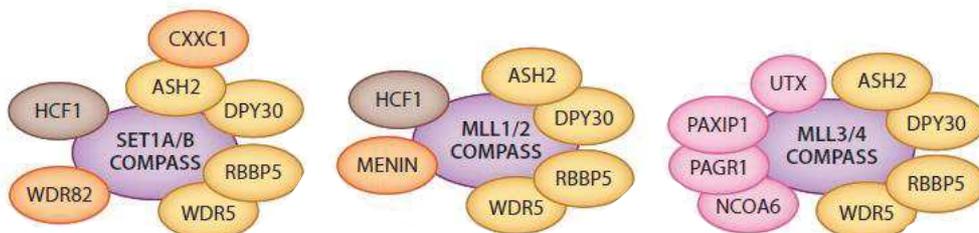
Figure 36: COMPASS and COMPASS-like histone modifying complexes. Schematic representation of the subunit composition of the mammalian SET1A- or SET1B-containing COMPASS complexes, the MLL1 or MLL2-containing COMPASS-like complexes and the MLL3- or MLL4-containing COMPASS-like complexes present in mammalian cells. Catalytic H3K4 methyltransferase subunits shown in purple. Subunits shared among the six complexes shown in yellow. Additional complex-specific subunits shown in brown, orange or rose. Adapted from Meeks & Shilatifard, 2017.

complexes: the SETD1A- or SETD1B-containing COMPASS and the MLL1-, MLL2-, MLL3- or MLL4-containing COMPASS-like complexes (Figure 36) (Meeks & Shilatifard, 2017).

These six related complexes were found to have redundant as well as non-redundant functions in the regulation of Pol II transcription. SETD1A- or SETD1B-containing COMPASS complexes are thought to mediate mainly H3K4me3 at gene promoters and were suggested to widely regulate especially housekeeping genes (Meeks & Shilatifard, 2017). These COMPASS complexes are essential for mouse development, as *Setd1a* null mouse embryos fail to gastrulate, while *Setd1b*[-/-] embryos go through gastrulation but exhibit growth defects and die by embryonic day (E)11.5 (Bledau et al., 2014).

The MLL1- or MLL2-containing COMPASS-like complexes are believed to mediate H3K4me2 and H3K4me3 at gene promoters of developmental genes, such as *Hox* genes, which are required for body segmentation (Meeks & Shilatifard, 2017). They are also believed to counteract Polycomb group protein (PcG)-mediated gene silencing. MLL1 and MLL2 are also needed for proper embryogenesis as inactivation of either the *Mll1* or *Mll2* genes leads to embryonic lethality by E10.5 (Yu et al., 1995; Glaser et al., 2009; Andreu-Vieyra et al., 2010).

The MLL3- or MLL4-containing COMPASS-like complexes were described to specifically mediated H3K4me1 at enhancer elements in a redundant manner (Meeks & Shilatifard, 2017). Inactivation of *Mll3* in mice was reported to lead to death at birth without other obvious defects, while inactivation of *Mll4* was found to cause embryonic lethality at E9.5 (Lee et al., 2013). MLL3/4 COMPASS-like complexes further contain the UTX (KDM6A) histone demethylase, which erases methylation of H3K27 and was also shown to be important for embryonic development (Welstead et al., 2012).

3.2.2.2. The TIP60 complex

The TIP60 coactivator complex is composed of at least 18 subunits in mammals and is homologous to the 13-subunit yeast NuA4 complex (Doyon et al., 2004; Sapountzi & Côté, 2011). This complex contains the histone acetyltransferase TIP60 (KAT5) (Esa1 in yeast), which belongs to the MYST family of type-A HAT and mainly modifies lysine residues of histone H2A and H4 (Sapountzi & Côté, 2011). Inactivation of *Tip60* in mice leads to pre-implantation lethality at E3.5, representing a very early stage of development (Hu et al., 2009b). Additionally, its EP400 subunit (Eaf1 in yeast) belongs to the family of INO80 chromatin remodellers and is capable of exchanging H2A/H2B dimers with dimers containing the histone variant H2A.Z (Clapier et al., 2017).

Recent low resolution cryo-EM structures revealed the overall organization of the yeast NuA4 complex (Figure 37) (Wang et al., 2018b). Strikingly, one of its subunits Tra1 (TRRAP in mammals) constitutes, due to its size, the main body of the complex and was suggested to serve with the Eaf1 subunit (EP400 in mammals) as the assembly scaffold (Wang et al., 2018b). For instance, the piccolo

module, containing the acetyltransferase subunit Esa1 (TIP60 in mammals), was found to pack predominantly against Tra1. Therefore, the incorporation of the piccolo module into the NuA4 complex might depend especially on the presence of Tra1 (TRRAP in mammals).
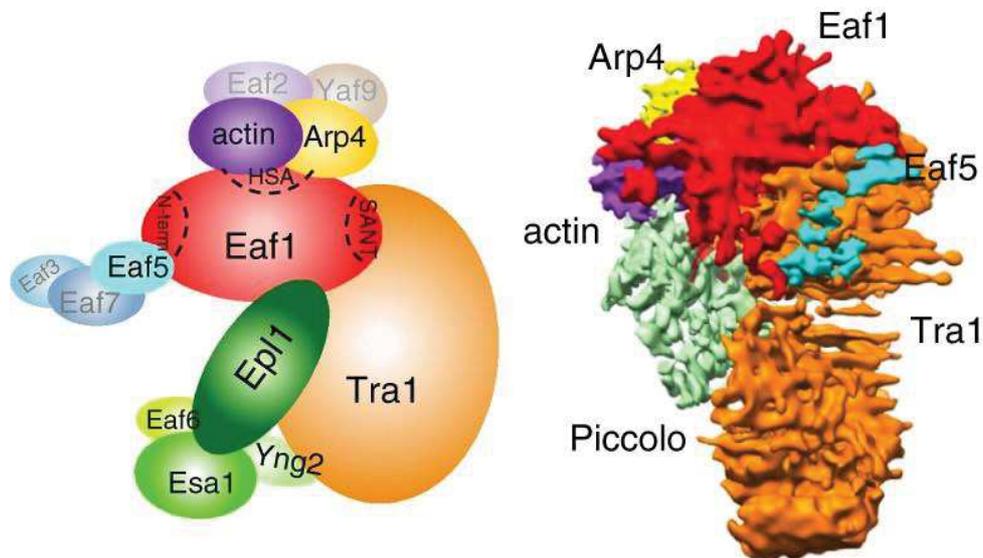
Figure 37: The TIP60/NuA4 histone modifying complex. On the left, schematic representation of the subunit composition of the yeast TIP60/NuA4 complex. Shaded subunits could not be detected in the low resolution cryo-EM structure of the TIP60/NuA4 complex shown on the right. Subunits in the cryo-EM structure follow the colour code of the scheme on the left. HSA, helicase-SANT-associated; N-term, N-terminal end. Adapted from Wang et al., 2018.

Intriguingly, several chromatin modifying complexes share subunits amongst each other. Related to the complexes that were described above, examples are the TRRAP subunit of the TIP60 complex shared with the coactivator complex SAGA or the subunit WDR5 of the COMPASS(-like) complexes shared with the coactivator complex ATAC and NSL (Helmlinger & Tora, 2017). Other subunits (ARP4 and nuclear actin) of the TIP60 complex are shared with INO80 and SWI/SNF remodelling complexes (Farrants, 2008). The importance of this phenomenon is not clear.

3.2.3. Role of chromatin modifying complexes in RNA polymerase II transcription

Besides their importance in mouse embryonic development, the fundamental roles of chromatin modifying complexes for regulation of gene expression and cell identity is further emphasized by the occurrence of mutations or translocations within genes encoding for subunits of these complexes in human malignancies (Meeks & Shilatifard, 2017). For example, chromosomal translocations involving the *Mll1* gene, encoding the catalytic subunit for one of the COMPASS-like coactivator complexes, are found in 10% of acute leukaemia and roughly 70-80% in infant acute leukaemia cases (Meyer et al., 2009). Further, roughly 20% of all cancer types are associated with a defect in BAF-related complexes (Hodges et al., 2016). Yet, mechanistic and fundamental knowledge on these complexes in the

mammalian context remain spars, largely due to the multimeric subunit composition, functional redundancies of some of these complexes and potentially diverging functions between cell types. Nevertheless, more extensive and systematic knowledge on the impact of some of these complexes on Pol II transcription exists from studies in budding yeast, which will be shortly summarized in the following.

The genome of budding yeast possesses roughly 200 genes encoding for subunits of chromatin modifying complexes (Lenstra & Holstege, 2012). Roughly a third of these genes were studied and 80% of them indeed showed to result in significant changes of gene expression levels *in vivo* as assessed by microarray analyses of total RNA in mutant strains (Lenstra & Holstege, 2012). For most chromatin modifying complexes only few hundred genes were found to be dependent on their activities (Lenstra et al., 2011). For example, deletion of components of yeast SWI/SNF remodelling complexes resulted into decreased expression of overall only 2-6% of all active yeast genes (Shivaswamy & Iyer, 2008; Lenstra et al., 2011; Rando & Winston, 2012). For coactivator complexes, this limited effect was generally in agreement with their role in regulated transcription and their interaction with gene-specific activators.

The importance of transcription coactivators for a limited set of genes was further underlined by the analysis of genome-wide binding profiles using chromatin immunoprecipitation (ChIP) experiments directed against subunits of these complexes. In general, subunits of transcriptional coactivator complexes were found to associate only to small subsets of gene promoters (Venters et al., 2011). Genes bound by a given coactivator were in consequence believed to be transcriptionally dependent on this factor. Unexpectedly however, when comparing the localization of a given coactivator at gene promoters with transcriptional dependencies of genes on the same complex, only minor overlaps were found (Lenstra & Holstege, 2012). On average only 2.5% of genes bound by subunits of a chromatin modifying complex were also transcriptionally dependent on it (Lenstra & Holstege, 2012). This indicated that a big proportion of genes bound by a given coactivator were actually not transcriptionally dependent on it. This limited overlap were suggested to be explained by potential compensatory mechanisms or gene-specific properties, such as the presence or absence of core promoter motifs (Lenstra et al., 2011; Lenstra & Holstege, 2012). Additionally, a big fraction (roughly three quarters) of genes dependent on a given coactivator for proper expression did not show binding of this factor by ChIP (Lenstra et al., 2011; Lenstra & Holstege, 2012). This was suggested to be indicative of secondary or indirect effects caused by growth defects and other phenotypes observed upon constitutive loss of several of these chromatin modifying factors.

Subsequent observations however suggested a potential technical limitation of the ChIP approach in revealing the localization of these factors, which frequently do not directly interact with DNA and are believed to be very transiently and dynamically associated to genetic elements. For instance, histone modifications and nucleosome composition or positioning, reflecting the activities of transcriptional

coactivator complexes, are commonly associated genome-wide with actively transcribed genes and therefore indirectly imply a more global recruitment of these complexes (more details see chapter 4.4.3. Gene-specific versus global functions of coactivator complexes in yeast) (Li et al., 2007).

Remarkably, the disconnection between binding and influence on gene expression was found to be a general feature of chromatin modifying complexes including also cofactors involved in transcription elongation such as Paf1, which is generally believed to be required for Pol II elongation at all genes (Lenstra et al., 2011; Lenstra & Holstege, 2012). As mentioned in a previous chapter, recent genome-wide analyses of newly synthesized RNA instead demonstrate that Paf1 is required for gene expression at the large majority of Pol II-transcribed genes (Xu et al., 2017).

4. The SAGA coactivator complex

This chapter will highlight the knowledge acquired during the past 20 years on the SAGA (Spt-Ada-Gcn5 Acetyltransferase) coactivator complex in yeast and will further focus on insights obtained for mammalian SAGA. SAGA was first discovered and extensively studied in budding yeast in which genetic, biochemical and functional studies provided a detailed characterization of its subunit composition, structural organization and functional roles not only in transcription regulation but also in DNA repair, mRNA export and aging (Rodríguez-Navarro et al., 2004; Ghosh & Pugh, 2011; García-Oliver et al., 2012; McCormick et al., 2014; Evangelista et al., 2018). This chapter however will exclusively describe recent insights on the role of SAGA in RNA polymerase II (Pol II) transcription and will focus more particularly on recent structural and functional data.

Table 3: Subunit composition of the yeast and mammalian SAGA complex. Shown are subunits of SAGA in budding yeast and their homologues in mammalian cells with alternative names in brackets. This table further shows domains found within the mammalian subunits of SAGA as well as their functional classification into distinct modules. DUB, deubiquitylation; HAT, histone acetyltransferase; TBP, TATA-box binding protein; °, shared with the ATAC complex; ‡, shared with the TIP60 complex; *, shared with the TFIID complex. Based on Spedale et al., 2012, Helminger & Tora, 2018, Wang et al., 2020 and Papai et al., 2020

Yeast SAGA	Mammalian SAGA	Domains	Function
Ubp8	USP22	ZnF-UBP, UcH	DUB module
Sgf73	ATXN7	ZnF, SCA7	
Sgf11	ATXN7L3	ZnF, SCA7	
Sus1	ENY2	EnY2	
Gcn5 (Ada4)	GCN5 (KAT2A)° or PCAF (KAT2B)°	HAT, Bromo, PCAF	HAT module
Ada2	TADA2B	ZnF, SANT, Ada boxes, SWIRM	
Ada3	TADA3°	Ada3	
Sgf29	SGF29 (CCDC101)°	Double Tudor	
Tra1‡	TRRAP‡	HEAT repeats, PI3K, FAT	TF interacting module
Spt3	SUPT3H	HFD	TBP-interacting module
Spt8	-	-	
Spt7	SUPT7L	HFD	Core module
Spt20 (Ada5)	SUPT20H (P38IP, FAM48A)	SEP	
Ada1	TADA1	HFD	
Taf5*	TAF5L	WD40, NTD	
Taf6*	TAF6L	HFD, HEAT	
Taf9*	TAF9*	HFD	
Taf10*	TAF10*	HFD	
Taf12*	TAF12*	HFD	

4.1. Identification and subunit composition of SAGA

Shortly after its discovery in 1996, the type-A HAT Gcn5 was found to be incorporated into two distinct complexes in yeast termed the ADA (Alteration/Deficiency in Activation) and SAGA complex (Kuo et al., 1996; Brownell et al., 1996; Grant et al., 1997; Horiuchi et al., 1997; Saleh et al., 1997; Eberharter et al., 1999). The 1.8 MDa large SAGA complex in budding yeast is composed of 19 subunits of which four proteins belong to the Ada group (Ada1, Ada2, Ada3 and Gcn5 (Ada4)), four belonging to the Spt group (Spt3, Spt7, Spt8 and Spt20) and five Taf proteins (Taf5, Taf6, Taf9, Taf10 and Taf12) (Table 3) (Lee et al., 2011; Helmlinger & Tora, 2017).

4.1.1. Identification of yeast Spt and Ada proteins

The Spt (Supressor of Ty) proteins were identified in budding yeast between 1984 and 1996 in genetic screens involving the Ty transposon (Winston et al., 1984a; Winston et al., 1984b; Winston et al., 1987; Fassler & Winston, 1988; Eisenmann et al., 1989; Roberts & Winston, 1996; Yamaguchi et al., 2001). This genetic screen was based on the identification of mutations in genes causing the suppression of the disruptive nature of Ty insertion on the expression of reporter genes targeted by this transposable element. Based on genetic and functional studies, the *SPT* genes were classified in two groups. The first group encodes histone H2A (Spt11) and H2B (Spt12) and proteins, which were later found to be important for transcription, such as Spt4 and Spt5 (evolutionary conserved transcription elongation factors forming DSIF in mammals), Spt6 (evolutionary conserved histone chaperone involved in transcription elongation) and a subunit of the FACT histone chaperone (Spt16) (Yamaguchi et al., 2001). The second group contains the genes encoding TBP (Spt15) as well as Spt3, Spt7 and Spt8. All Spt proteins of the TBP group, with the exception of TBP itself, were subsequently found to be components of SAGA, suggesting a functional link between SAGA and TBP deposition at promoters. Intriguingly, one SAGA subunit discovered by this screening system (Spt20) was also identified in the Ada screen (described next) as Ada5 protein (Marcus et al., 1996; Roberts & Winston, 1996).

The Ada (Adaptor) proteins were identified in yeast between 1992 and 1994 using genetic screens involving the fusion protein Gal4-VP16, formed by the DNA binding domain of the yeast transcription factor Gal4 linked to a domain of the herpes simplex virus protein VP16 (Berger et al., 1990; Berger et al., 1992; Marcus et al., 1994). Before its use in a genetic screen, findings obtained from *in vitro* transcription assays with this construct led to the postulation of the presence of so-called 'adaptor' proteins, bridging between transcription factors and the basal transcription machinery at gene promoters (Berger et al., 1990). To identify these adaptor proteins, a genetic screen was developed by taking advantage of the circumstance that ectopic expression of the Gal4-VP16 fusion protein is toxic to budding yeast cells as it sequesters the general transcription factors of the Pol II transcription machinery (Berger et al., 1992). Null mutations in genes coding for putative adaptor proteins, maintaining the association of Gal4-VP16 with the transcription machinery, were selected based on the reversion of the

toxic phenotype of Gal4-VP16. This genetic screen led to the identification of the Ada proteins Ada1, Ada2, Ada3, Ada4 (Gcn5) and Ada5 (Spt20), all subsequently identified as subunits of the SAGA complex (Table 3, page 102) (Berger et al., 1992; Piña et al., 1993; Marcus et al., 1994; Marcus et al., 1996; Roberts & Winston, 1996; Horiuchi et al., 1997; Grant et al., 1997).

These two screening systems revealed differences and similarities of the identified Ada and Spt proteins that turned out to be predictive of the organisation of SAGA into functional modules (Table 3, page 102) (Yamaguchi et al., 2001). For example, Ada2, Ada3 and Ada4 (Gcn5), subunits of the later defined HAT module, shared common phenotypes and were found to physically interact (Piña et al., 1993; Marcus et al., 1994; Sterner et al., 1999). Spt3 and Spt8, nowadays considered as the TBP-interacting module of SAGA, were also found to display similar phenotypes and to physically interact with Spt15 (TBP) (Eisenmann et al., 1992; Roberts & Winston, 1996; Sterner et al., 1999; Yamaguchi et al., 2001; Warfield et al., 2004). Similarly, mutations in Ada1 and Ada5 (Spt20) were found to cause similar phenotypes, which were more severe than those observed upon mutation of Ada2, Ada3, Ada4 (Gcn5), Spt3 or Spt8 (Marcus et al., 1996; Horiuchi et al., 1997; Sterner et al., 1999).

The subsequent insight that Ada5 and Spt20 represent the same protein and that Ada2, Ada3 and Gcn5 (Ada4) interact with Spt proteins of the TBP group, led to the identification that also mutations of Spt7 or Spt15 (TBP) could rescue Gal4-VP16 toxicity thereby revealing a connection between proteins from the Ada and Spt families (Roberts & Winston, 1996; Marcus et al., 1996; Horiuchi et al., 1997; Grant et al., 1997). Further, mutations in Ada1, Spt7 or Spt20 were found to lead to very comparable consequences and eventually were described as structural, SAGA-specific subunits required for the complex activity and stability (Table 3, page 102) (Sterner et al., 1999; Yamaguchi et al., 2001; Wu & Winston, 2002). Overall, the studies on Ada and Spt proteins gave the first hints towards a functional organisation of SAGA into modules with a HAT module composed of Ada2, Ada3 and Gcn5 (Ada4), a TBP-interacting module with Spt3 and Spt8 and a structural module containing Ada1, Spt7 and Spt20.

4.1.2. Identification of Taf proteins shared between TFIID and SAGA

The Taf (TBP-associated factors) subunits were first described for the general transcription factor TFIID and five of them were subsequently found to be shared with the yeast SAGA coactivator complex (Table 3, page 102) (Grant et al., 1998a). In SAGA, the majority of Taf subunits are implicated in the formation of a histone octamer-like structure through pairwise interactions of HFDs (histone fold domains) involving also SAGA-specific subunits (Birck et al., 1998; Gangloff et al., 2000; Gangloff et al., 2001; Selleck et al., 2001). The histone fold pairings are the following in yeast SAGA complex: Taf6 with Taf9, Taf10 with Spt7, Taf12 with Ada1 and a pseudo-dimerization of Spt3, which contains two HFDs. Together with the HFD subunits, Spt20 and Taf5 are considered to form the central core of SAGA (Table 3, page 102).

4.1.3. Identification of the transcription factor-binding subunit Tra1

Besides the HAT module, the TBP-interacting module and the core of SAGA, the large Tra1 protein was identified as additional SAGA subunit in 1998 (Grant et al., 1998b; Saleh et al., 1998). Tra1 was suggested to be targeted by transcription factors, such as Gal4 and Gcn4 in budding yeast, thereby enabling the recruitment of SAGA to gene promoters for regulated transcription induction (Grant et al., 1998b; Saleh et al., 1998; Brown et al., 2001; Fishburn et al., 2005; Reeves & Hahn, 2005). Tra1 is nowadays commonly referred to as the transcription factor (TF) interacting module and is also a subunit of the NuA4 complex (TIP60 complex in mammals) (Table 3, page 102) (Allard et al., 1999; Helmlinger et al., 2011).
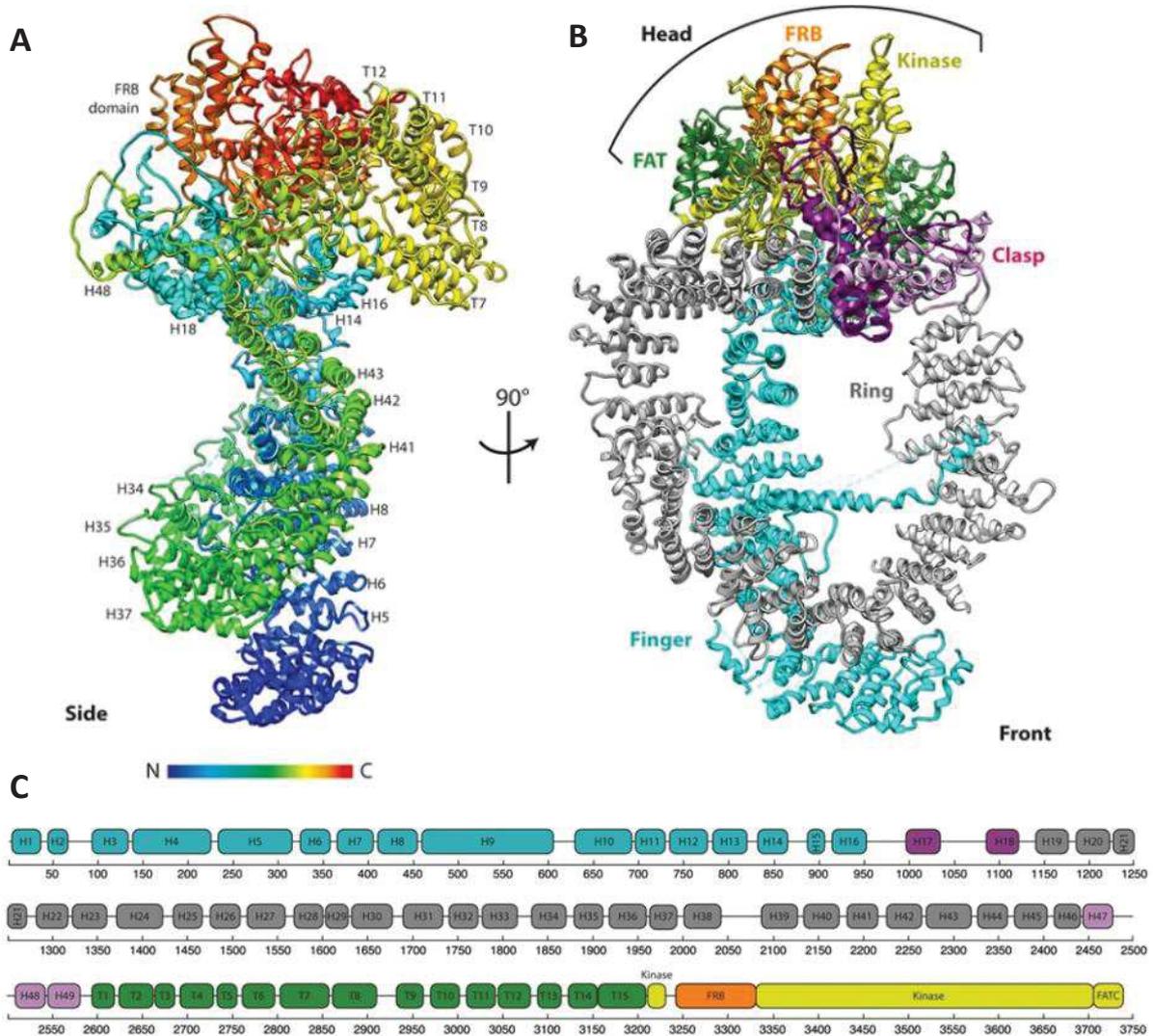
Figure 38: Cryo-EM structure of the yeast transcription factor interacting subunit, Tra1. A. and **B.** front and side view on the *S. cerevisiae* Tra1 protein. Dashed lines reflect missing residues. H, Heat repeats; T, TPR repeats. Coloration in A is shown with the N-terminus in blue transiting into red at the C-terminus. Coloration in B reflects Tra1 protein domains and topological regions as indicated. **C.** Scheme showing the protein domains of Tra1 including heat repeats (H1-H49), TPR repeats (T1-T15) of the FAT, the kinase and the FRB domain. Colouring as in (B). Adapted from Díaz-Santín et al., 2017.

Tra1 belongs to the group of phosphatidyl 3-kinase-related kinases (PI3K; another member of this family is for example TOR) and its intricated, diamond ring-like shape was recently revealed by cryo-EM studies (Figure 38) (Sharov et al., 2017; Díaz-Santín et al., 2017; Helmlinger & Tora, 2017). Although classified within a group of kinases, Tra1 was found to lack the required catalytic residues for kinase activity in the PI3K domain and is therefore considered a pseudokinase.

The 49 HEAT repeats and the FAT domain of Tra1 are highly structured through α-helices and reflect roughly 86% of the total protein (Figure 38) (Sharov et al., 2017; Díaz-Santín et al., 2017). The HEAT repeats of Tra1 were identified as the main interaction sites of several TFs (Knutson & Hahn, 2011; Setiaputra et al., 2015). Conformational changes of the FAT domain, induced by the binding of TFs to the HEAT repeats, were suggested to potentially propagate and influence activities within the SAGA or NuA4 complexes (Sharov et al., 2017).

4.1.4. Identification of the deubiquitylation module of SAGA

Additional subunits of SAGA were identified in 2002 including the fourth member of the HAT module, Sgf29, and subunits belonging to the second histone modifying activity of SAGA, its deubiquitylation (DUB) module (Table 3, page 102) (Sanders et al., 2002). Ubp8, the catalytic DUB subunit, was subsequently reported to depend on three additional, non-enzymatic subunits (Sgf11, Sgf73 and Sus1) to reach its full enzymatic capacity on its substrate, monoubiquitylated lysine 123 of yeast histone H2B (in mammals H2BK120ub) (Henry et al., 2003; Daniel et al., 2004; Ingvarsdottir et al., 2005; Samara et al., 2012). The DUB subunit, Sus1, is also part of a second complex, the TREX-2
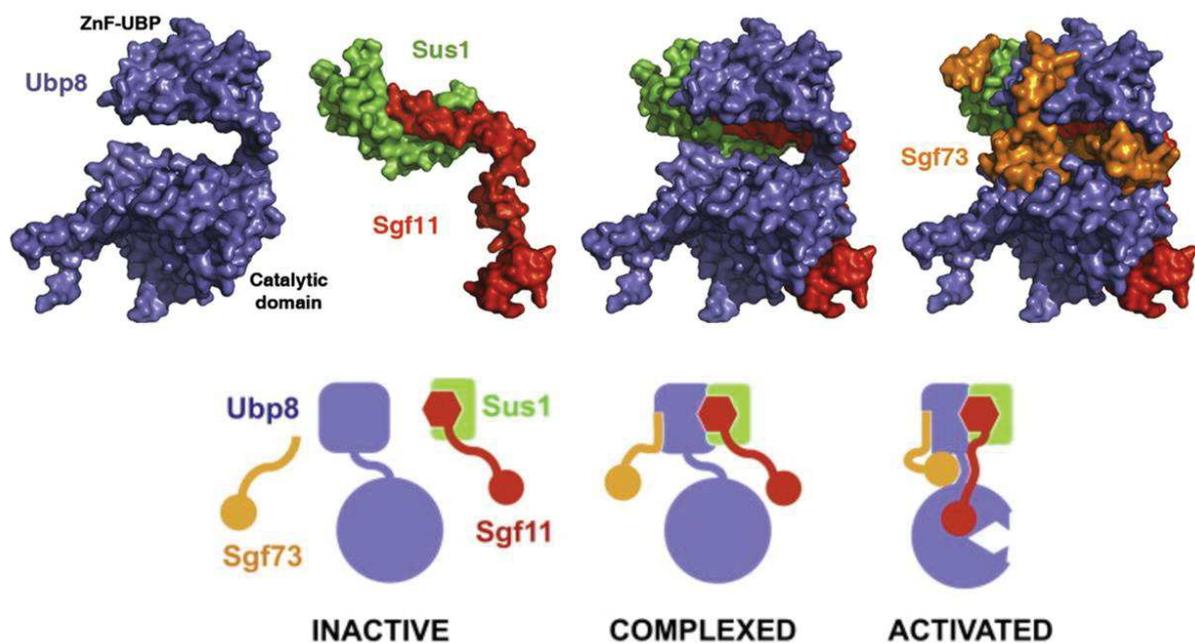
Figure 39: Crystal structure of the deubiquitylation module of yeast SAGA. Top, surface representation of the yeast SAGA deubiquitylation (DUB) subunits showing their spatial organisation. Bottom, schematic representation of the assembly of the DUB subunits as shown in (A) leading to the activation of the catalytic site of Ubp8. More details in text. Adapted from Köhler et al., 2010.

(transcription export 2) complex, which is involved in mRNA export. Through this connection, SAGA has been implicated in a phenomenon called 'gene gating', in which actively transcribed genes locate close to the nuclear pore to allow efficient mRNA export (Rodríguez-Navarro et al., 2004; Köhler et al., 2008).

Crystal structures of the DUB module revealed a two-lobed organization in which Ubp8 interacts through its ZnF-UBP (zinc finger ubiquitin-binding domain) domain with Sgf73 and Sgf11 forming the assembly lobe (Figure 39) (Köhler et al., 2010; Samara et al., 2010). The association between Sgf11 and Ubp8 in the assembly lobe is stabilized through the fourth subunit of the DUB, Sus1. At the catalytic lobe, the ZnF domains of Sgf11 and Sgf73, located at their N-terminal ends, lead to the activation of the catalytic domain of Ubp8 by inducing conformational changes. The DUB interactions through the three ZnF domains are stabilized by in total six zinc (Zn) atoms. Two additional Zn atoms are required at the active site of Ubp8 and are important for enzymatic activity (Samara et al., 2010; Köhler et al., 2010).

Recent cryo-EM structures of the DUB module binding to ubiquitylated histone H2B within the core nucleosome, further revealed the substrate engaged state of these module (Morgan et al., 2016). This exposed an additional function of the ZnF domain of Sgf11, in which conserved arginine residues form contacts with the H2A/H2B acidic patch at the lateral face of the nucleosome (Morgan et al., 2016). Besides binding the ubiquitin molecule, the catalytic domain of Ubp8 was also found to contact the histone H2B protein (Morgan et al., 2016).

4.1.5. Composition of the ADA complex

As mention briefly at the beginning of this section, Gcn5 as well as the other subunits of the SAGA HAT module (Sgf29, Ada2 and Ada3) are also part of the 0.8 MDa ADA complex in budding yeast. Besides the HAT subunits, ADA additionally possesses two complex-specific subunits, Ahc1 and Ahc2 (ADA HAT component 1 and 2) (Grant et al., 1997; Grant et al., 1999; Eberharter et al., 1999; Lee et al., 2011; Helmlinger & Tora, 2017). ADA was suggested to function as an independent HAT complex and to maintain genome-wide H3 acetylation levels (Eberharter et al., 1999; Grant et al., 1999; Lee et al., 2011; Helmlinger & Tora, 2017). Recently, a metazoan ADA complex, composed of only the four subunits of the SAGA HAT module (Ada2b, Ada3, Gcn5 and Sgf29), was described in *Drosophila* (Soffers et al., 2019).

4.2. Tri-dimensional structure of yeast SAGA complex

By electron microscopy analyses, SAGA appears divided in two lobes: one lobe containing solely the massive Tra1 subunit, while the other lobe, composed of the core and TBP-interacting subunits, connects to the two enzymatic HAT and DUB modules (Figure 40A and B) (Brand et al., 1999; Wu et al., 2004; Setiaputra et al., 2015; Sharov et al., 2017; Liu et al., 2019; Wang et al., 2020; Papai et al.,
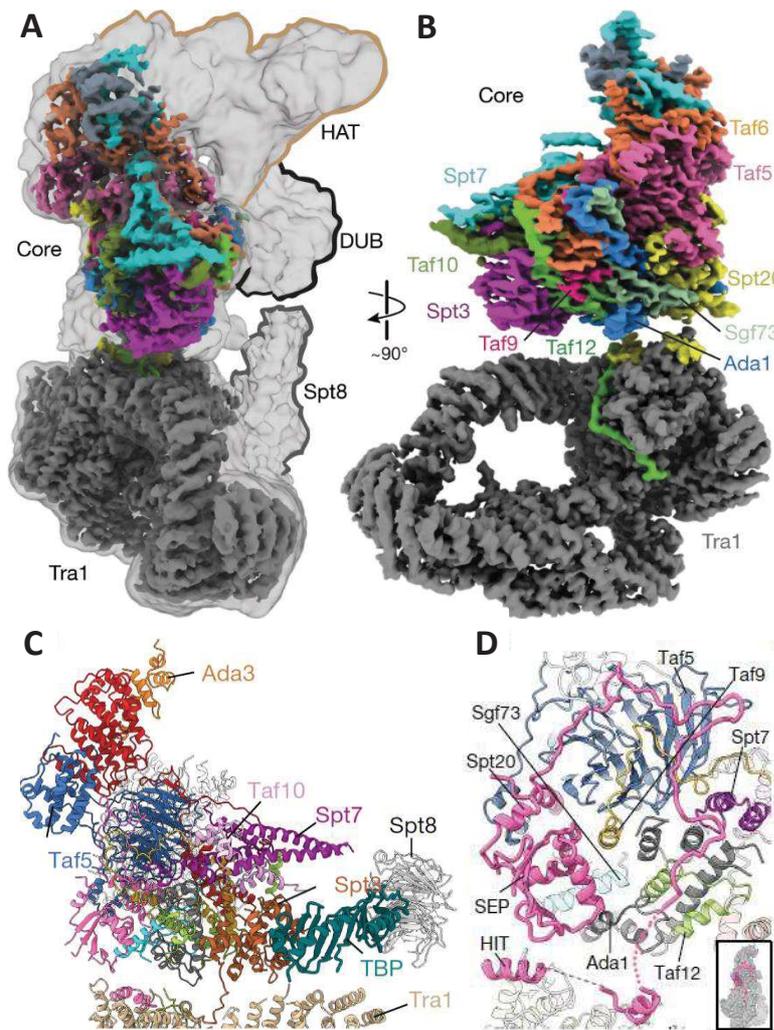
Figure 40: Cryo-EM structure of the yeast SAGA coactivator complex. A. Cryo-EM structure of the yeast SAGA complex with low resolution densities indicating the position of the HAT and DUB modules as well as the yeast-specific Spt8 subunit. **B.** High resolution cryo-EM structure of the SAGA core module (subunits shown in distinct colours) and the Tra1 TF-interacting module (in grey). **C.** Selected view on the SAGA core module highlighting the position of the histone fold pair Taf10 and Spt7 as well as the position of TBP in proximity to Spt3 (orange-brown) and Spt8. Few helices of Ada3 further indicate connections of the HAT module with the Taf6 subunit (in red). **D.** Highlight on the Spt20 subunit (in pink) contacting through a long, unstructured loop several SAGA core subunits such as Ada1, Spt7, Taf5, Taf9 and Taf12. Further, the SEP domain of Spt20 is situated in proximity to the Sgf73 subunit of the DUB module and the HIT domain of Spt20 localizes to the Tra1 subunit. Insert shows position of Spt20 in the full complex. From Wang et al., 2020 and Papai et al., 2020.

2020). The HAT and DUB modules were found to be highly flexible within the SAGA complex (Setiaputra et al., 2015; Papai et al., 2020; Wang et al., 2020). This flexibility was suggested to allow SAGA to accommodate for differences in distances when bridging between activating TFs and promoter regions at various genes (Wang et al., 2020). The very recent high-resolution cryo-EM structures of yeast SAGA allowed to reveal the detailed connections within the SAGA complex with the exception, due to their flexible nature, of the enzymatic modules (more details next) (Wang et al., 2020; Papai et al., 2020).

4.2.1. Modular organisation of SAGA and TBP-loading function of its core module

The recent cryo-EM structures highlight that SAGA is organized into four connected modules which had been predicted by genetic and biochemical studies: the HAT module, the DUB module, the Tra1 module and the 10-subunit core module comprising the TBP-interacting subunits (Figure 40A) (Papai et al., 2020; Wang et al., 2020). The 10-subunit core of SAGA is formed by the asymmetric histone octamer-like structure with a neighbouring submodule comprising Taf5, Spt20 and the C-terminus of

Taf6 (Papai et al., 2020; Wang et al., 2020). Each Taf subunit of the SAGA core was found to be present in a single copy within SAGA in contrast to the TFIID complex were several of the Tafs are found twice.

The histone octamer-like structure of SAGA, including the pairs Taf6/Taf9, Taf10/Spt7, Taf12/Ada1 and the two HFDs of Spt3, deviates from the histone octamer structure due to a 20° tilt of the Spt3 intraprotein histone fold pair (Papai et al., 2020). This deformation of the octamer liberates Spt3 from its association with the remaining histone-fold pairs allowing only few contacts between Spt3 and Taf10 (Papai et al., 2020). Intriguingly, the C-terminal tail of Spt3 was found to reach into a cavity formed by the remaining histone-fold proteins thereby establishing contacts with each histone fold pair. The integration of Spt3 within the deformed histone octamer-like structure was suggested to enable a not too rigid but at the same time not too flexible incorporation within SAGA, which might be important for TBP binding and release (Papai et al., 2020).

As proposed by prior biochemical studies, Spt3 and Spt8 were found to form contacts with TBP on opposite sides (Figure 40C): Spt3 contacts the C-terminal stirrup of TBP, while Spt8 interacts with the N-terminal half of TBP (Papai et al., 2020; Wang et al., 2020). Importantly, these contact surfaces bound by Spt3 and Spt8, respectively also represent the binding sites of the general transcription factors TFIIB and TFIIA, respectively. In the cryo-EM structures, Spt8 is flexibly tethered to SAGA through interactions with the C-terminus of Spt7, which is in accordance with biochemical studies reporting the loss of Spt8 from SAGA upon deletion of this part of Spt7 (Wu & Winston, 2002; Papai et al., 2020; Wang et al., 2020). The binding and positioning of TBP relative to SAGA is however thought to be mainly determined by Spt3 (Papai et al., 2020). Interestingly, when TBP is bound by Spt3 and Spt8, its DNA-binding domain is turned towards the SAGA core which, through steric hindrance, was suggested to prevent DNA from binding to TBP (Papai et al., 2020). The impediment of DNA binding, while being bound to SAGA, implies that TBP needs to be released from SAGA to bind to gene promoters and consequently that SAGA likely does not interact with promoter-bound TBP but instead rather serves to deliver TBP to gene promoters (Papai et al., 2020).

The following model was proposed for how TBP could be released from SAGA: Small movements ('breathing') of TBP are enabled by the flexible association of Spt8 to SAGA potentially allowing transient access of DNA or regulatory factors to TBP (Papai et al., 2020). For example, TFIIA could compete with Spt8 for the binding sites on TBP thereby fully displacing Spt8 from SAGA. Subsequent conformational changes within Spt3 would lead to a tilting of TBP and allow binding of DNA. This interaction of TBP with DNA was found to be facilitated by the presence of a consensus TATA-box or TATA-like element within the promoter DNA (Papai et al., 2020). Final release of TBP from SAGA might involve additional changes in conformation of Spt3.

4.2.2. Flexible tethering of the remaining modules to the core of SAGA

The cryo-EM structures further revealed how the HAT, DUB and Tra1 modules can be flexibly but still stably retained to the SAGA core. Tra1 is contacted by two to three distinct bridges formed by i) Spt20 reaching out with its HIT domain for the FAT domain of Tra1 (Figure 40D, page 108), ii) a remarkably long loop of Taf12 also interacting with the FAT domain of Tra1 (Figure 40B, page 108) and iii) potentially Spt3 associating with Tra1 between two HEAT repeats (partially visible in Figure 40C, page 108) (Papai et al., 2020; Wang et al., 2020). Of these three Tra1 contact sites, deletion studies indicate that Tra1 incorporation into SAGA is mainly achieved through Spt20 (Elías-Villalobos et al., 2019b).

The DUB module, which is almost completely detached from SAGA, is connected to the core through its Sgf73 subunit. While contacting the remaining DUB subunits through its N-terminal end, the C-terminal part of Sgf73 traverses the core of SAGA forming contacts amongst others with Spt20 (Figure 40D) (Papai et al., 2020; Wang et al., 2020). The importance of the C-terminal region of Sgf73 in integrating the DUB module within SAGA had been previously suggested biochemically (Kamata et al., 2013; Han et al., 2014). In contrast, no subunits of the HAT module were found embedded within the SAGA core with only parts of Ada3 contacting the HEAT repeats of Taf6 (Figure 40C) (Papai et al., 2020). Instead, subunits of the core of SAGA could potentially reach out to the HAT module, similarly as observed for Tra1, ensuring its association with SAGA. Intriguingly, the HAT and DUB module were reported to be situated on the same side of SAGA, the side which is thought to face towards the promoter DNA (Wang et al., 2020).

4.2.3. Intertwined contacts of the SAGA core module

The high-resolution cryo-EM structures further enabled to show the incredible intertwined network of contacts within the SAGA core with for example the WD40 propeller of Taf5 forming on one side contacts with Taf6, while on the other side connecting to the histone octamer-like structure thereby associating with at least 11 different protein domains (Wang et al., 2020; Papai et al., 2020). Besides forming a histone fold pair with Taf9, Taf6 was found to contribute one β-strand to the WD40 propeller of Taf5, suggesting the obligatory formation of a heterodimer also with Taf5 (Papai et al., 2020; Wang et al., 2020).

Two other example of extensive connections among subunits of the SAGA core are represented by i) the C-terminal tail of Taf9, which associates with the four histone fold pairs and additional six other protein domains, and ii) a long unstructured loop of Spt20 forming contacts with several subunits while traversing the surface of the SAGA core (Spt20 loop shown in Figure 40D, page 108) (Papai et al., 2020).

4.3. Subunit composition of mammalian SAGA

The SAGA complex is highly conserved from yeast to humans with only few differences, which are described in the following subsections (Spedale et al., 2012; Helmlinger & Tora, 2017). In general, protein domains described for yeast SAGA subunits are also found in their mammalian orthologs and most of the domains have well-defined structural and functional roles. This includes the histone-fold domains (HFD) in SUPT3H (yeast Spt3), SUPT7L (yeast Spt7), TADA1 (yeast Ada1), TAF6L (ortholog to yeast Taf6, see below), TAF9 (yeast Taf9), TAF10 (yeast Taf10) and TAF12 (yeast Taf12); the HEAT domain of TAF6L involved in forming potential contacts to TADA3 (yeast Ada3, a subunit of the HAT module); the SEP domain of SUPT20H (yeast Spt20) interacting with several subunits of the SAGA core; the WD40 repeats of TAF5L (ortholog to yeast Taf5, see below) forming a propeller structure; the N-terminal domain (NTD) of TAF5L, which in yeast Taf5 is involved in forming contacts to Taf6; the ZnF domains of ATXN7L3 (yeast Sgf11), ATXN7 (yeast Sgf73) and USP22 (yeast Ubp8) crucial for the organization and activity of the DUB module as well as the HEAT repeats, the FAT and the PI3K domains of TRRAP likely sharing their functions with their yeast Tra1 counterparts (subunits of the human SAGA with protein domains excluding HAT subunits shared with the ATAC coactivator complex shown in Figure 41).
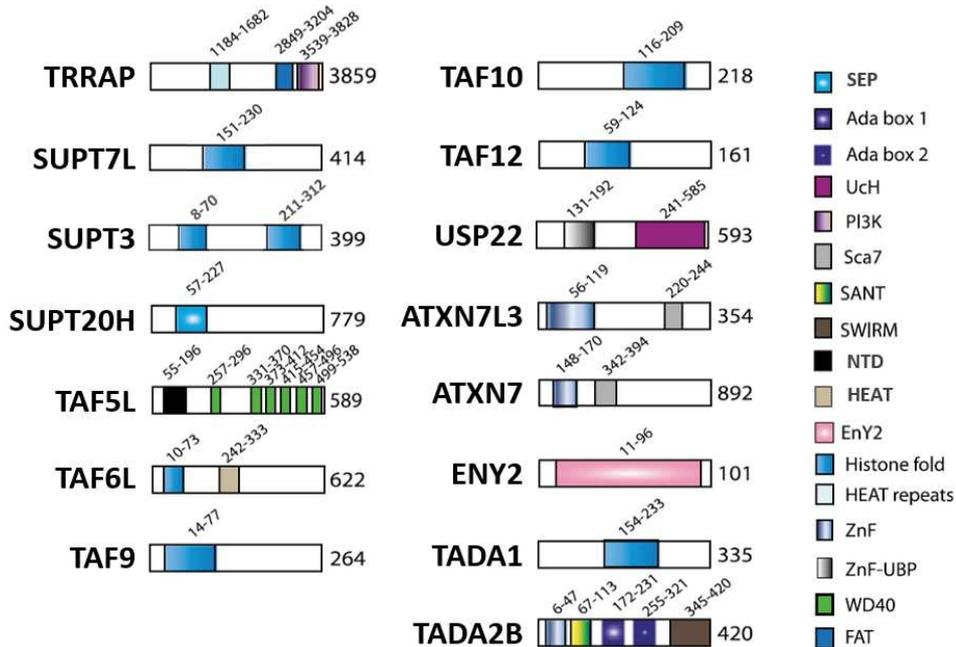
Figure 41: Schematic representations of protein domains found within subunits of the human SAGA coactivator complex. Domains containing enzymatic activities are the UcH and PI3K domains. However, the PI3K domain of TRRAP is catalytically inactive. The SANT, Sca7 and SWIRM domains reflect DNA or histone binding domains, while the histone fold, WD40, SEP, NTD, FAT, HEAT (repeats), ZnF-UBP and ZnF domains are thought to be protein interaction domains. The Ada boxes and the EnY2 domain represent homology domains of unknown function. Subunits of the HAT module are not shown as they will be described in more details in a separate chapter. Adapted from Spedale et al., 2012 and based on Helmlinger & Tora, 2017.

4.3.1. Distinctions of yeast and mammalian SAGA complex

As mentioned above, SAGA is highly conserved from yeast to humans. However, metazoan genomes lack a homolog to the *Spt8* gene of budding yeast, therefore the TBP-interacting module in metazoan SAGA only consists of the SUPT3H subunit (yeast Spt3) (Table 3, page 102). Loss of Spt8 in budding yeast was reported to weaken interactions of SAGA with TBP, although Spt3 was suggested to be the main determinant of TBP binding in yeast SAGA (Wieczorek et al., 1998; Spedale et al., 2012; Papai et al., 2020). A less stable binding of TBP to metazoan SAGA caused by the absence of Spt8 could explain the following finding: While TBP can be detected when purifying SAGA from yeast cells, it does not co-purify with human SAGA (Wieczorek et al., 1998; Martinez et al., 1998; Ogryzko et al., 1998). Further, metazoan SUPT7L is truncated in its C-terminus compared to yeast Spt7, therefore effectively missing the Spt8 interacting surface (Nagy et al., 2009; Spedale et al., 2012).

Metazoan SAGA resembles the yeast SALSA (SAGA altered, Spt8 absent) also known as SLIK (SAGA-like complex) complex, in which Spt8 is missing and Spt7 is lacking its C-terminal region interacting with Spt8 (Belotserkovskaya et al., 2000; Pray-Grant et al., 2002; Sterner et al., 2002; Wu & Winston, 2002; Nagy et al., 2009; Spedale et al., 2010). Recent results however strongly suggest that loss of the C-terminal region of yeast Spt7 found within the SALSA/SLIK complex might occur during cell lysis and SAGA purification upon liberation of the Pep4p protease leading to an artificial cleavage of Spt7 and therefore artificial formation of the SALSA/SLIK complex (Spedale et al., 2010). This is supported by recent purifications of SAGA from yeast cells, which showed stoichiometric amounts of all subunits, including Spt8, arguing against the cellular existence of the SALSA/SLIK complex (Wang et al., 2020; Papai et al., 2020).

4.3.2. Increased complexity of mammalian SAGA due to gene duplication events

Although SAGA is highly similar between yeast and human, mammalian genomes encode for SAGA-specific versions of the yeast Taf5 and Taf6 proteins, which are shared between SAGA and TFIID in yeast, termed TAF5L and TAF6L (Ogryzko et al., 1998; Spedale et al., 2012). In contrast to yeast, mammalian TAF5 and TAF6 represent specific subunits of the TFIID complex. Interestingly, TAF5L is missing the LisH domain, found within yeast Taf5 as well as mammalian TAF5, which was recently reported to be involved in forming contacts with Spt20 and potentially Sgf73 in yeast SAGA (Spedale et al., 2012; Wang et al., 2020; Papai et al., 2020). The consequences of loss of this domain of TAF5L for the structure of mammalian SAGA are not clear.

Similarly, to the duplication of the yeast *Taf5* and *Taf6* genes, two *Ada2* paralogs are found in mammalian genomes encoding the TADA2A or TADA2B proteins. These two proteins associate with the remaining HAT module subunits similarly to the yeast Ada2 protein but in a mutually exclusive manner. TADA2B is specifically incorporated into SAGA, while the TADA2A-containing HAT module

is part of another transcriptional coactivator, the ATAC complex (described in more details in chapter 5. The ATAC coactivator complex) (Muratoglu et al., 2003; Kusch et al., 2003; Spedale et al., 2012). TADA2B possesses five domains: two Ada boxes, a ZnF, a SANT and a SWIRM domain (Figure 41, page 111) (Muratoglu et al., 2003; Gamper et al., 2009). While the functions of the Ada boxes are unknown, the ZnF and SANT domain are thought to mediate contacts with the catalytic enzymes (GCN5 or PCAF, more details below), while the SWIRM domain contacts TADA3.

Additional gene diversification events resulted into the expansion of the possible subunit composition of mammalian SAGA, such as the duplication of yeast *Taf9* in *Taf9* and *Taf9b* or the duplication of yeast *Gcn5* in *Gcn5* and *Pcaf* in mammals. TAF9 and TAF9B were suggested to have overlapping as well as distinct functions and to incorporate into SAGA in a mutually exclusive manner (Frontini et al., 2005; Spedale et al., 2012). Similarly, GCN5 and PCAF can be part of the HAT module of SAGA in a mutually exclusive manner (more details in chapter 6. The shared HAT function of SAGA and ATAC) (Koutelou et al., 2010; Nagy et al., 2010; Spedale et al., 2012). Another example is the DUB subunit, Sgf73, which was triplicated in mammalian genomes encoding for ATXN7, ATXN7L1 and ATXN7L2. These three paralogous proteins were further suggested to incorporate into the DUB module of SAGA in a mutually exclusive way (Vermeulen et al., 2010; Helmlinger & Tora, 2017). Also, the yeast *Sgf11* gene is duplicated in mammalian genomes in two genes coding for ATXN7L3 and ATXN7L3B (Li et al., 2016; Helmlinger & Tora, 2017). In contrast to the above mentioned examples, ATXN7L3B is not part of the SAGA complex and is detectable primarily in the cytoplasm, while ATXN7L3 is predominantly nuclear (Li et al., 2016). As ATXN7L3B was found to interact with the DUB subunit ENY2 (yeast Sus1), it might limit the assembly of functional DUB modules by sequestering ENY2 in the cytoplasm (Li et al., 2016; Helmlinger & Tora, 2017).

Besides USP22, the catalytic subunit of the DUB module of SAGA (ortholog to yeast Ubp8), mammalian genomes further encode for two additional, very similar USP proteins, USP27X and USP51, (Kobayashi et al., 2015; Atanassov et al., 2016; Helmlinger & Tora, 2017). Curiously, USP27X and USP51 were reported to interact in a mutually exclusive manner with ATXN7L3 and ENY2 but not with ATXN7, forming independent DUB modules. Since the C-terminus of ATXN7 is crucial for the connection of the DUB module to SAGA, these USP27X- or USP51-containing DUB modules are thought to not integrate within SAGA (Atanassov et al., 2016; Helmlinger & Tora, 2017). Additionally, these mammalian SAGA-independent DUB modules were found to act on monoubiquitylated H2BK120 and the non-histone substrate, Hes1, in a redundant manner to the USP22-containing DUB of SAGA (Kobayashi et al., 2015; Atanassov et al., 2016). For example, depletion of ATXN7L3 or ENY2 led to the expected increase of H2BK120ub levels, due to the loss of the three DUBs (USP22, USP27X and USP51) activities. In contrast, depletion of USP22, the SAGA-specific DUB enzyme, did not, which implies compensatory functions of USP27X- and USP51-containing DUB modules (Atanassov et al.,

2016). Nevertheless, the functional relevance of the three distinct DUB modules remains unclear (Helmlinger & Tora, 2017).

In mammalian cells, the DUB module subunits ATXN7L3 and ATXN7, homologs of yeast Sgf11 and Sgf73, respectively, were additionally found to contain an atypical zinc finger, the Sca7 domain whose function is not fully understood (Figure 41, page 111) (Bonnet et al., 2010; Spedale et al., 2012). This domain is absent in yeast Sgf11 but present in yeast Sgf73. In humans, polyglutamine expansions at the N-terminus of ATXN7, but not ATXN7L3, were reported to cause an inherited neurodegenerative disease called SCA7 (spinocerebellar ataxia type 7) (Koutelou et al., 2010; Wang & Dent, 2014). The mechanisms of how polyglutamine expansions in ATXN7 affects functions of the DUB module of SAGA and how it relates to the disease phenotypes are still unclear.

4.4. Role of SAGA in RNA polymerase II transcription

4.4.1. Yeast SAGA has gene-specific functions predominantly on stress-inducible genes

In agreement with its function as a transcriptional coactivator complex, deletions of SAGA subunits in budding yeast caused reduced expression of only some hundred genes, as had been found for other chromatin modifying complexes (see previous chapter) (Lee et al., 2000; Huisinga & Pugh, 2004; Lenstra et al., 2011). Overall, only around 5-12% of all active genes showed a considerable dependency on SAGA and were consequently termed SAGA-dependent genes (Lee et al., 2000; Huisinga & Pugh, 2004; Lenstra et al., 2011). Surprisingly, these analyses revealed considerable variability between the subunits tested as, for example, Gcn5 was suggested to be required for the expression of nearly all yeast genes, while Spt3 was found to be required at only 10% of genes. In contrast, inactivation of subunits of the general transcription factor TFIID resulted in decreased transcription of 84% of the analyzed genes, which were therefore termed TFIID-dependent (Huisinga & Pugh, 2004). As combined deletion of SAGA and TFIID subunits resulted into near complete loss of transcription, it was further suggested that the two complexes act partially redundantly.

The higher dependencies of some genes on one of the two complexes as well as the partial redundant functions of SAGA and TFIID were attributed to their role in delivering TBP to gene promoters (Lee et al., 2000; Huisinga & Pugh, 2004). Overall, the yeast genome was divided into 10% of Pol II-transcribed genes requiring especially SAGA for transcription, while the remaining 90% would majorly depend on TFIID functions (Huisinga & Pugh, 2004). Subsequent studies found that SAGA-dominated genes were enriched for genes containing a consensus TATA-box in their promoters and responsive to stresses, while TFIID-dependent genes were enriched for genes with housekeeping functions and TATA-less or TATA-like containing promoters (Basehoar et al., 2004; Huisinga & Pugh, 2004; Rhee & Pugh, 2012).

Similarly, when analyzing the genome-wide localization of SAGA by performing chromatin immunoprecipitation (ChIP) experiments against different subunits in budding yeast, SAGA was

suggested to be recruited at a small subset of genes (van Werven et al., 2008; Venters et al., 2011). Surprisingly however, as reported for numerous other chromatin modifying complexes, the majority of genes transcriptionally affected upon loss of SAGA were not bound by SAGA and conversely, only a small fraction of genes bound by SAGA were transcriptionally affected upon deletion of genes encoding SAGA subunits (Lenstra & Holstege, 2012).

4.4.2. Yeast SAGA may have a more general role in RNA polymerase II transcription

Adding to these discrepancies, changes in histone modifications upon deletion of the histone modifying enzymes of yeast SAGA implied a rather genome-wide recruitment of SAGA (Henry et al., 2003; Bian et al., 2011; Bonnet et al., 2014). Deletion of Gcn5, the catalytic enzyme of the HAT module, in budding yeast resulted in a genome-wide reduction of H3K9ac and H3K18ac (Bian et al., 2011; Bonnet et al., 2014). Similarly, deletion of Ubp8, the enzymatic subunit of SAGA's DUB module, caused a global increase in H2B ubiquitylation (Henry et al., 2003; Bonnet et al., 2014). The general correlation of H3K9ac levels with actively transcribed gene promoters and H2Bub levels with actively transcribed gene bodies (as discussed in chapter 2.3.3.1. Euchromatin) further indicated that SAGA acts at all actively transcribed genes (Bonnet et al., 2014). Overall, these findings collectively revealed a remarkable lack of overlap between transcription effects, binding profiles and enzymatic activities.

Several reasons could explain these discrepancies, as mentioned in an earlier chapter, such as compensation by other coactivator complexes, requirement for simultaneous, cooperative binding of chromatin modifying complexes or methodological restrictions (Lenstra et al., 2011; Sun et al., 2012). For example, although SAGA binds to several genes, not all of these genes might be transcriptionally dependent on it due to compensatory functions by other coactivator complexes, which could explain the discrepancy between binding profiles and transcription effects. This discrepancy might also be explained by gene-specific properties. For instance, some genes might be more sensitive to loss of SAGA function due to the specific architecture of their gene promoters such as differences in nucleosome positioning, core promoter elements or transcription factor binding site densities at regulatory sequences.

Also, more recent findings suggested that ChIP approaches may not properly reveal binding sites of transcriptional coactivators. Coactivators are thought to very transiently interact with chromatin and only indirectly localize to DNA, which may be problematic for ChIP approaches (Zentner et al., 2015; Grünberg et al., 2016). Adding to this, recent studies further suggested that global transcriptional effects could be compensated by changes in mRNA decay rates in budding yeast, resulting in a general stabilization of steady-state mRNA levels although Pol II transcription is majorly affected (Munchel et al., 2011; Sun et al., 2012). Due to this buffering mechanism, global effects on Pol II transcription can only be properly monitored by analysing newly synthesized RNA levels (Sun et al., 2012). Several methods have been developed to assess newly synthesized or nascent mRNA molecules including

labelling with 4-thiouracil (4tU) (Wissink et al., 2019). This approach is based on the labelling of newly synthesized RNA by specific incorporation of 4tU into RNA during transcription. The thiol group of 4tU subsequently allows the specific purification and analysis of the labelled, newly synthesized RNA (Duffy et al., 2019).

Using 4tU labelling and other improved technics, recent studies in yeast revealed that SAGA is required for global Pol II transcription and also binds to the vast majority of actively transcribed genes
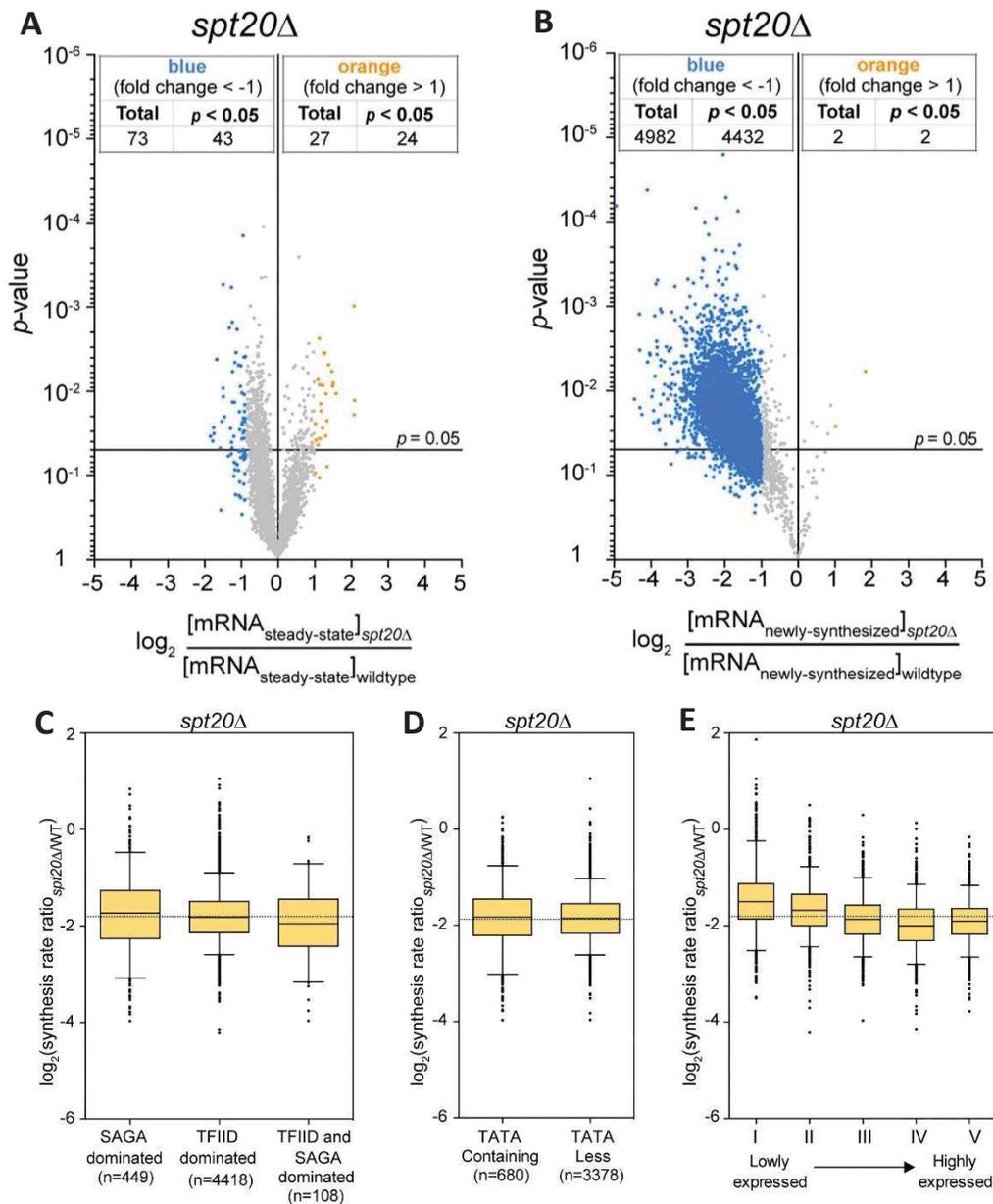
Figure 42: SAGA acts as a general cofactor for RNA polymerase II transcription in yeast. A. and **B.** Volcano plot representation of log2 fold changes of steady-state mRNA levels (A) or newly synthesized mRNA levels (B) relative to their statistical significance (p-value) observed in *spt20Δ* strains compared to wildtype strains. Downregulated genes in blue, upregulated genes in orange. **C., D.** and **E.** Box plot representation of changes in newly synthesized RNA levels between *spt20Δ* and wildtype cells for different gene classifications. No obvious difference in degree of transcription changes could be observed whether genes were classified as SAGA- or TFIID-dominated (C) or if their promoters were TATA-box containing or TATA-less (D). In contrast, more highly expressed genes tended to have a more strongly reduced expression upon loss of Spt20 compared to lowly expressed genes (E). Adapted from Baptista et al., 2017.

(Bonnet et al., 2014; Baptista et al., 2017). Deletion of either *Spt7* or *Spt20*, two subunits of the SAGA core, were found to lead to a roughly 4-fold change in global Pol II transcription levels (results for *Spt20* deletion shown in Figure 42B) (Baptista et al., 2017). This major effects on Pol II transcription were only apparent when analyzing newly synthesized RNA, while steady-state RNA levels only showed few changes (results for *Spt20* deletion shown in Figure 42A (steady-state) and 42B (newly synthesized)).

Interestingly, the general function of SAGA for Pol II transcription was found to be especially dependent on synergistic effects of its subunits Gcn5 (HAT module) and Spt3 (TBP loading). Combined deletion of these two subunits led to a 10-fold reduction of Pol II transcription. In contrast, loss of Ubp8 (DUB module) and Spt8 (TBP-interacting) did not cause obvious effects on transcription (Baptista et al., 2017). Intriguingly, SAGA was found to be equally important for transcription of the previously defined SAGA- and TFIID-dominated gene classes as a similar decrease in transcription could be observed for both gene categories (results for *Spt20* deletion shown in Figure 42C) (Baptista et al., 2017). Transcription effects upon SAGA loss were also found to be independent of the presence of a TATA-box in gene promoters (results for *Spt20* deletion shown in Figure 42D) (Baptista et al., 2017). Instead, effects on transcription were suggested to correlate with gene expression levels, with the most highly transcribed genes displaying the strongest reduction of mRNA synthesis upon loss of SAGA (results for *Spt20* deletion shown in Figure 42E) (Baptista et al., 2017). Overall, SAGA seems to function as a general cofactor for Pol II transcription in yeast instead of influencing only a limited set of genes (Baptista et al., 2017).

4.4.3. Gene-specific versus global functions of coactivator complexes in yeast

We summarized and discussed these new insights and findings described above in more details, including the relevant methods enabling the assessment of the general role of yeast SAGA on Pol II transcription, in our point-of-view on the 'Global role for coactivator complexes in RNA polymerase II transcription' published online on the 9th of October 2018 in *Transcription* (see next pages).

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS   Check for updates

Global role for coactivator complexes in RNA polymerase II transcription

Veronique Fischer [a,b,c,d*], Kenny Schumacher[a,b,c,d*], Laszlo Tora [a,b,c,d], and Didier Devys [a,b,c,d]

aInstitut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France; bCentre National de la Recherche Scientifique, UMR7104, Illkirch, France; cInstitut National de la Santé et de la Recherche Médicale, Illkirch, France; dUniversité de Strasbourg, Illkirch, France

ABSTRACT

SAGA and TFIID are related transcription complexes, which were proposed to alternatively deliver TBP at different promoter classes. Recent genome-wide studies in yeast revealed that both complexes are required for the transcription of a vast majority of genes by RNA polymerase II raising new questions about the role of coactivators.

Introduction

RNA-polymerase II (Pol II)-mediated transcription is a highly regulated process that determines cellular function and cell identity through the accurate synthesis of mRNAs. Although its regulation occurs at all stages of transcription, regulation at the stage of initiation is a key mechanism to control gene expression. For initiation, the basal transcription machinery composed of Pol II and the general transcription factors (GTFs), nucleates pre-initiation complex (PIC) formation on gene promoters [1,2]. TFIID, composed of the TATA-box binding protein (TBP) and TBP-associated factors (TAFs) is the first GTF that binds promoter sequences. Once bound to the different promoter elements, GTFs enable correct positioning of Pol II relative to the transcription start site (TSS) and facilitate the transition to productive elongation. However, the compact structure of chromatin has been shown to act as a barrier for PIC formation. Thus, other transcription factors are required to specifically modulate the chromatin landscape at proximity of promoters for productive PIC assembly.

Coactivators are recruited to the vicinity of gene promoters through their interaction with gene-specific activators bound to mammalian enhancers or yeast Upstream Activating Sequences (UAS). Different activities facilitating transcription were found associated with coactivators, namely chromatin remodelers, histone modifiers or adaptors that link activators to the transcription machinery. While most coactivator complexes were initially thought to regulate specific subsets of genes, some were reported to have a more global role in transcription. One important example is the Mediator complex which was described as an integral part of the basal transcription machinery, required for nearly all Pol II mediated transcription [3–6]. In this point-of-view, we will summarize recent insights into the role of two coactivator complexes, TFIID and SAGA (Spt-Ada-Gcn5-acetyltransferase), in global Pol II transcription in *S. cerevisiae*. Furthermore, we will discuss potential global functions for other coactivators and whether similar mechanisms exist in metazoans.

Role of SAGA and TFIID in pol II transcription

An extensively characterized coactivator is the evolutionary conserved SAGA complex organized in distinct functional and structural modules (reviewed in [7]). SAGA activates transcription through histone modifying activities (acetylation and deubiquitination) and by recruiting TBP to promoters. Early genome-wide analyses of SAGA function in Pol II transcription in budding yeast by Pugh and colleagues, showed that upon deletion of the TBP-interacting subunit Spt3, the steady-state RNA levels of ~10% of genes were decreased by more than 2-fold [8]. Meanwhile, 90% of the genes were affected upon

---

conditional depletion of the TFIID subunit Taf1, also involved in TBP-recruitment to promoters. Although this seminal study cautiously concluded that "TFIID and SAGA make overlapping contribution to the expression of all genes", the proposed classification of genes as either SAGA-dominated or TFIID-dominated was oversimplified over time, categorizing each gene as dependent exclusively on one or the other coactivator. It was further shown that the SAGA-dominated genes were highly enriched in stress-regulated genes containing consensus TATA elements in their core promoters, while the TFIID-dominated genes tended to be more constitutively expressed and lack a strong consensus TATA [9]. In good agreement with these findings, early studies emphasized that SAGA is recruited to its target genes through the interaction of its Tra1 subunit with a set of activators predominantly stimulating stress-responsive genes, Gcn4 and Gal4 among others [10,11]. Together, these observations pointed towards a specific role for SAGA in the transcription of highly regulated genes. Importantly, these findings suggested that genes can be differentially regulated depending on their promoter sequence by utilizing specific sets of transcription factors and coactivators.

TFIID is a general transcription factor, composed of TBP and several TAFs. Only its TBP subunit but not TAFs is necessary and sufficient for PIC assembly and transcription *in vitro*. TAFs are targeted by several activators and potentiate their activities suggesting that TFIID has coactivator functions [12]. Unlike SAGA, TFIID directly contacts DNA and interacts with other components of the basal transcription machinery. TBP, as part of the TFIID complex, tends to bind promoters lacking a consensus TATA-box sequence, whereas TATA-containing promoters are bound by TBP but are relatively depleted of TAFs [13]. These observations further supported a predominant role for TFIID in the regulation of TATA-less genes, enriched with housekeeping functions.

Integration of binding profiles and transcriptional effects

The analysis of the genome-wide localization of SAGA by chromatin immunoprecipitation (ChIP) indicated that SAGA is recruited to a limited subset of genes, in agreement with its requirement for the expression of only ~10% of the yeast genome [14]. However, subsequent comparison of localization and expression studies showed a weak correlation between chromatin binding sites and transcriptional effects [15]. Similar findings were made for several other coactivators in *S. cerevisiae*. Indeed, transcriptome analyses of mutant strains for different chromatin modifiers including Set1, Set2 and Dot1, catalyzing H3K4, H3K36 and H3K79 methylation respectively, revealed limited effects of these enzymes on transcription. These results were surprising as the histone marks deposited by these factors are localized at nearly all active genes, suggesting a more global role for Pol II transcription [16]. Similar observations for SAGA showed that its histone acetylation and deubiquitination activities act on the entire transcribed genome [17]. Thus, the contrast between broad enzymatic activities and restricted transcriptional effects appears as a general feature of chromatin altering complexes.

As a very large number of different proteins are known to bind each promoter, a functional redundancy between these factors is likely to explain at least partially, the discrepancy between factor location and expression effect. However, it seems possible that for some factors, the observed differences could instead result from limitations of the methodologies used. Indeed, ChIP approaches are antibody dependent and may be insensitive to transient chromatin interactions which might be problematic for coactivators with low ChIP efficiency [18,19]. In addition, steady-state mRNA analyses might be inaccurate to measure Pol II activity. Indeed, several studies revealed that a global decrease in Pol II transcription is compensated by a simultaneous and global decrease in mRNA decay, thereby buffering steady-state mRNA levels [20–23]. The use of improved methodologies was highly warranted to re-examine the role of SAGA and TFIID in Pol II transcription, in light of recent observations showing that TFIID is equally recruited at promoters of both SAGA- and TFIID-dominated genes and that SAGA inactivation decreases Pol II recruitment at both classes of genes [17,18].

SAGA and TFIID are generally required for pol II transcription

Two recent studies aimed at providing a more detailed analysis of the genome-wide occupancy

and the role in Pol II transcription of the SAGA and TFIID complexes [24,25]. They used chromatin endogenous cleavage coupled with high-throughput sequencing (ChEC-seq), a formaldehyde- and antibody-independent approach to determine the binding profiles of dynamic factors such as coactivators [19]. ChEC-seq was previously used to clarify the genome-wide binding profile of Mediator, revealing an association with the UASs at a majority of genes, whereas TFIID was recruited at core promoters, to which it binds cooperatively with Mediator [18]. Using ChEC-seq SAGA was exclusively detected at the UASs of both SAGA- and TFIID-dominated genes, in agreement with the idea that SAGA is recruited to UASs by sequence-specific DNA-binding transcription factors [24].

To quantify nascent transcription upon inactivation of SAGA or TFIID, these studies used native Pol II ChIP or metabolic labeling with 4-thiouracil (4tU) followed by quantification of the purified newly transcribed mRNAs. These analyses were done on SAGA deletion strains or using inducible depletion systems (auxin-inducible degradation or anchor away technology) and revealed that nearly all Pol II transcribed genes are dependent on TFIID, SAGA and Mediator. Importantly, TATA-containing and TATA-less genes were similarly affected upon inactivation of either of these three complexes.

These analyses of nascent transcription indicate that SAGA, TFIID and Mediator make important contributions to Pol II transcription. Each of these coactivators appears to be absolutely required for gene expression as an inducible depletion of Mediator subunits or TAFs caused a dramatic decrease in genome-wide Pol II recruitment by about 8-fold for Med14 or by about 3 to 4-fold for different TAFs [25]. Similarly, nascent mRNA transcription was reduced by about 10-fold in a double SAGA mutant strain (*SPT3* and *GCN5* deletions) [24]. Such large transcriptional effects indicate that the activities of these three coactivators on Pol II transcription are not functionally redundant but rather suggest that SAGA, TFIID and Mediator function at different rate-limiting steps. Earlier work indicated a reciprocal dependency in genome-wide recruitment of TFIID and Mediator and cooperativity between SAGA and Mediator has also

been suggested [1,18]. The broad genome-wide recruitment of SAGA, Mediator and TFIID at most active genes support the idea that different coactivators work cooperatively to assemble the PIC. However, a more detailed description of coactivator interactions and the mechanisms of cooperativity remain to be elucidated.

Interestingly, the average Pol II occupancy was decreased to a comparable extent, by about 3- to 4-fold, upon depletion of four different TFIID subunits [25]. However, the depletion of these TAFs did not significantly alter the complex architecture, indicating that most TAFs are individually important for TFIID function. These observations strikingly contrast with the results for SAGA, in which suppression of different activities results in highly variable effects in Pol II transcription. The mRNA synthesis rates were unaffected by the loss of the deubiquitinase Ubp8, or the TBP-interacting protein Spt8, but were significantly decreased upon loss of Spt3 (2-fold change) or the histone acetyltransferase Gcn5 (1.5-fold change). Strikingly, global mRNA synthesis was decreased by about 10-fold in a *SPT3* and *GCN5* double deletion strain, suggesting that the functional modules of SAGA make different contributions but act in a synergistic manner on Pol II transcription.

The intriguing observation that some coactivators are recruited at most expressed genes and have a global contribution to Pol II transcription, raises multiple interesting considerations. Beyond TFIID and SAGA, are other coactivators globally required for Pol II transcription? If these coactivators each occupy the regulatory regions of most genes, which factors define the specific expression levels for each gene? How do coactivators contribute to gene expression changes in response to variations in transcription factor recruitment? How do SAGA and TFIID mediate transcription from TATA-containing and TATA-less promoters? Do these complexes have similar broad distributions and functions in metazoans?

Do other coactivators have a global role in pol II transcription?

As a global role in Pol II transcription is now proposed for three different coactivator complexes (Mediator, SAGA and TFIID), it is tempting to speculate that other coactivators might also have a

broader function than anticipated. Particularly, histone-modifying complexes such as NuA4 or COMPASS/Set1C deposit marks (H4 acetylation and H3K4 trimethylation) that are enriched at most active promoters. These factors may have broader effects on transcription that were overlooked when analyzing steady-state mRNA. Chromatin remodeling complexes are also expected to have broad genome-wide activities. Indeed, the RSC complex was previously shown to act at a majority of yeast promoters to slide or evict nucleosomes thus positioning nucleosomes flanking the nucleosome depleted region (NDR) [26,27]. Similarly to our observations on SAGA, RSC was shown to be required for global transcription, although localization studies revealed a limited number of RSC binding sites [28,29]. The authors suggested that the interaction of RSC with many binding sites might be too transient to be readily detected by ChIP. A re-analysis using ChEC-seq may reveal more a widespread localization of RSC across the genome. Similarly, at the majority of yeast promoters, the first nucleosome found downstream of the NDR (+1 nucleosome) often contains the histone variant H2A.Z where it is deposited by SWR complex [27]. Here again, the broad distribution of H2A.Z and of the SWR complex contrasted with limited gene expression changes detected by transcriptome studies in the corresponding mutant strains [14,30]. Although these differences might be explained by factor redundancy or by gene-specific features accounting for dependencies on certain

factors only, it would be important to analyze the role of the above-mentioned factors through a direct characterization of Pol II activity.

Differential gene sensitivity to specific coactivators

Although SAGA, TFIID and Mediator can be considered as general cofactors for Pol II in yeast, it does not imply that each complex makes equal contributions to the expression of every individual gene. Although genome-wide Pol II occupancy or mRNA synthesis rates were consistently decreased upon depletion of these coactivators, a range of gene expression changes was observed in each mutant strain suggesting variable requirements of genes on certain coactivator complexes [24,25,31]. These variable dependencies on coactivators are indicative of gene-specific properties of varying importance that would determine the relative contribution of each coactivator complex to PIC formation. Gene-specific features are diverse in nature, corresponding to DNA sequence elements (e.g. presence and position of a consensus TATA-box, number and diversity of transcription factor binding sites) or specific chromatin architecture (e.g. nucleosome positioning and occupancy around the promoter, patterns of histone modifications or transcription factor occupancy) [32,33].

Genes are often divided in two categories corresponding to different strategies for transcriptional regulation (Figure 1). Housekeeping genes are constitutively expressed with little influence from
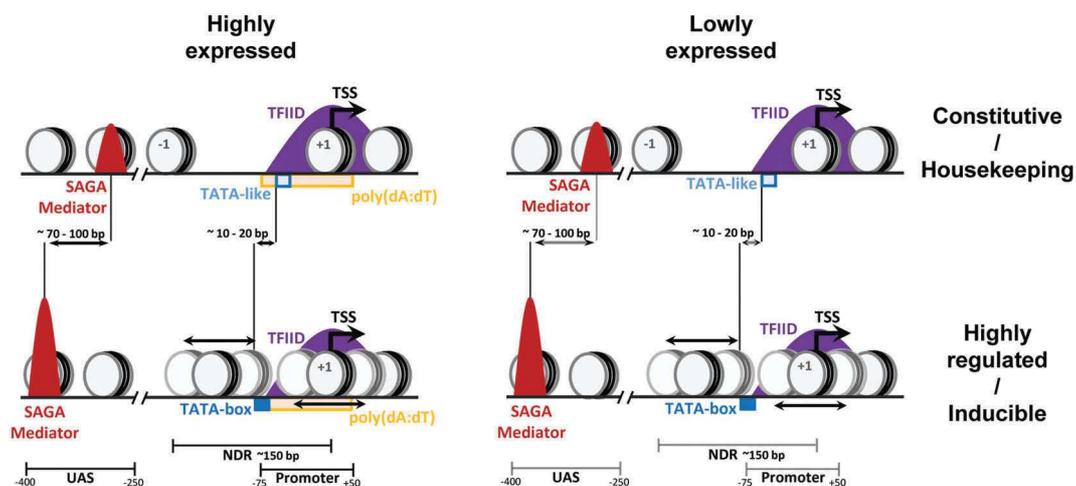
Figure 1. Different combinations of gene-specific features in constitutively active and stress responsive genes lead to the characterization of distinct mechanisms for transcription regulation of these two gene categories.

external or internal signals and were proposed to be more dependent on TFIID. In contrast, highly regulated genes have a higher transcriptional plasticity and were suggested to be more dependent on SAGA. However, the above-described observations clearly demonstrate that these two classes are equally sensitive to the loss of TFIID or SAGA and thus cannot be distinguished by their dependency on these coactivators. Nevertheless, these two gene classes can be differentiated by gene specific features including promoter organization or chromatin architecture [27].

Promoters with a consensus TATA-box (TATAWAWR sequence), are more often found in highly regulated genes, whereas housekeeping genes are predominantly lacking a TATA-box in their promoters [9]. Further analyses identified TATA-like elements, having 1 or 2 mismatches from the consensus, at the sites of PIC assembly, in the majority of TATA-less promoters in yeast [13]. Interestingly, the localization of SAGA and Mediator analyzed by ChEC-seq revealed a higher occupancy of these complexes at the UASs of TATA-containing genes than at TATA-less genes [18,24]. In addition, the average location of both SAGA and Mediator was found more upstream (by 70–100 bp) relative to the TSS at TATA-containing than at TATA-less genes [24]. As these two gene categories are equally sensitive to the loss of SAGA or Mediator subunits, the PIC formation is likely differently regulated at promoters containing either a consensus TATA-box or a TATA-like element. These gene classes also differ by their respective distances between the TATA-element and the TSS, being 10–20 bp longer at TATA-containing than at TATA-less promoters [13]. In contrast, ChEC-seq signals for Taf1 were highly similar at both promoter categories suggesting that TFIID is similarly recruited at the NDR of both TATA-containing and TATA-less promoters [18].

It is not clear whether the presence of a consensus TATA-box or a TATA-like element is the only sequence element which determines the PIC architecture at these two gene classes. Along these lines, it was recently shown that TAFs interact with downstream promoter elements to facilitate transcription re-initiation [34]. Downstream binding events occurred specifically at TFIID-dominated genes although these genes did not have higher TAF occupancy. These observations suggest that the ability of a promoter to drive TAF-dependent re-initiation events might better define which genes are more sensitive to TAF mutation. Another recent study used an *in vitro* system to assess transcription from TATA-containing and TATA-less promoters and revealed that both promoter classes are TFIID-dependent, in agreement with *in vivo* observations described earlier [35]. On TATA-containing promoters, TBP could complement the loss of TFIID only *in vitro*, but not *in vivo*. These data together suggest that most PICs assembled *in vivo* contain the TFIID complex. This work also revealed other promoter sequence features in addition to TATA-elements that distinguish TATA-less from TATA-containing promoters. Several studies indicated that T-richness upstream and A-richness downstream of the TSS distinguish highly from lowly expressed genes [27,33,36,37]. Indeed, highly expressed genes were found to be more sensitive to SAGA mutations and displayed higher Mediator occupancy than lowly expressed genes [18,24]. Although it is unclear how the T- and A-richness would be mechanistically linked with the requirement for certain coactivators, A/T rich sequences are known to negatively influence nucleosome occupancy, thereby potentially reducing the requirement of chromatin regulators [32].

Housekeeping/TATA-less and highly regulated/ TATA-containing genes were also associated with differential chromatin organization and sensitivity to chromatin regulators [27,32]. Constitutively expressed genes often display a broad NDR with well positioned flanking nucleosomes. At these promoters, transcription factor binding sites lie within the NDR which may explain their lower sensitivity to disruption of chromatin regulators. In contrast, promoters of highly regulated genes have higher nucleosome occupancy upstream of the TSS with less defined positioning. At these gene promoters, transcription factor binding sites are more distal relative to the TSS and are often occupied by nucleosomes. A putative competition between nucleosomes and transcription factors at these promoters may account for their higher sensitivity to chromatin regulation, in agreement with their higher histone turnover. Although SAGA, TFIID and Mediator similarly affect the expression

of these two gene classes, different functions of these coactivators could be used depending on promoter elements and chromatin architecture characteristic of housekeeping genes or genes with high transcriptional plasticity.

Openings and perspectives

Improvements in genome-wide approaches to analyze gene expression and chromatin binding have conciliated conflicting data concerning the role of transcriptional coactivators. The arising findings challenge the established paradigm stating that coactivators act on specific gene subsets and start to shift it towards a more global role of many coactivators in Pol II transcription in *S. cerevisiae* [18,24,25,31]. These results suggest that the combinatorial activities of these factors are necessary for accurate transcription. However, to match the various transcriptional requirements of all genes, coactivators might act to different extents to facilitate the expression of each individual gene. Thus, the combination of activators binding to UASs, nucleosome occupancy, promoter architecture and sequence elements in core promoters like TATA elements or others, seem to participate in the fine-tuning of transcription by properly coordinating the activities of each coactivator on every gene according to its expression needs.

These discoveries raise new questions regarding the mechanisms of coactivator recruitment to all active genes. Either each coactivator can interact with a wide variety of activators or recruitment can be also mediated through activator-independent interactions. Interestingly, the loss of the Tra1 subunit which is expected to mediate SAGA interaction with activators, has limited phenotypic and transcriptional effects in *S. pombe* [38]. Thus, SAGA recruitment likely relies on other subunits that can interact with either DNA-bound transcription factors or directly to chromatin. Indeed, many coactivators have been shown to contain a variety of protein domains that recognize histone marks found at most active promoters [39]. For example, SAGA contains a Tudor domain and bromodomains reported to interact with methylated and acetylated histones which could stabilize the binding of SAGA to active gene promoters [40].

The increased complexity of gene expression programs and regulation in mammalian cells is linked with an expansion of activators and coactivators repertoire and further diversification of gene promoter and enhancer features. Most coactivators were highly conserved through evolution, with similar complexes found in yeast and metazoans. Nevertheless, several duplication events led to the expansion of the coactivator repertoire, potentially due to the increased complexity of multicellular organisms. For example, the HAT module of the SAGA complex in yeast is incorporated in both the SAGA and ATAC coactivator complexes in metazoans [7]. Similarly, the yeast COMPASS complex diverged into at least seven different complexes in mammalian cells [41]. Thus, the presence of related activities in different complexes with redundant functions, complicates the analysis of coactivator requirements for Pol II transcription in mammalian cells. However, a recent study using quantification of newly synthesized mRNAs, proposed that the bromodomain-containing protein BRD4 acts as a general coactivator for Pol II transcription, demonstrating that this approach is also feasible in mammalian cells [42]. However, deciphering the genome-wide action of each coactivator will likely require sophisticated experimental set-ups and approaches.

Disclosure statement

Funding

ORCID

Veronique Fischer http://orcid.org/0000-0002-6512-2960
Laszlo Tora http://orcid.org/0000-0001-7398-2250
Didier Devys http://orcid.org/0000-0001-9655-3512

References

[1] Hahn S, Young ET. Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics. 2011;189:705–736.

[2] Sainsbury S, Bernecky C, Cramer P. Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol. 2015;16:129–143.

[3] Allen BL, Taatjes DJ. The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol. 2015;16:155.

[4] Takagi Y, Kornberg RD. Mediator as a general transcription factor. J Biol Chem. 2006;281:80–89.

[5] Holstege FC, Jennings EG, Wyrick JJ, et al. Dissecting the regulatory circuitry of a eukaryotic genome. Cell. 1998;95:717–728.

[6] Thompson CM, Young RA. General requirement for RNA polymerase II holoenzymes in vivo. Proc Natl Acad Sci USA. 1995;92:4587–4590.

[7] Helmlinger D, Tora L. Sharing the SAGA. Trends Biochem Sci. 2017;42:850–861.

[8] Huisinga KL, Pugh BF. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. Mol Cell. 2004;13:573–585.

[9] Basehoar AD, Zanton SJ, Pugh BF. Identification and distinct regulation of yeast TATA box-containing genes. Cell. 2004;116:699–709.

[10] Brown CE, Howe L, Sousa K, et al. Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. Science. 2001;292:2333–2337.

[11] Bryant GO, Ptashne M. Independent recruitment in vivo by Gal4 of two complexes required for transcription. Mol Cell. 2003;11:1301–1309.

[12] Thomas MC, Chiang CM. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol. 2006;41:105–178.

[13] Rhee H, Pugh FB. Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature. 2012;483:295–301.

[14] Venters BJ, Wachi S, Mavrich TN, et al. A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. Mol Cell. 2011;41:480–492.

[15] Lenstra TL, Holstege FC. The discrepancy between chromatin factor location and effect. Nucleus. 2012;3:213–219.

[16] Pokholok DK, Harbison CT, Levine S, et al. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell. 2005;122:517–527.

[17] Bonnet J, Wang CY, Baptista T, et al. The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes Dev. 2014;28:1999–2012.

[18] Grunberg S, Henikoff S, Hahn S, et al. Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. EMBO J. 2016;35:2435–2446.

[19] Zentner GE, Kasinathan S, Xin B, et al. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. Nat Commun. 2015;6:8733.

[20] Munchel SE, Shultzaberger RK, Takizawa N, et al. Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol Biol Cell. 2011;22:2787–2795.

[21] Rodriguez-Molina JB, Tseng SC, Simonett SP, et al. Engineered covalent inactivation of TFIIH-kinase reveals an elongation checkpoint and results in widespread mRNA stabilization. Mol Cell. 2016;63:433–444.

[22] Sun M, Schwalb B, Pirkl N, et al. Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol Cell. 2013;52:52–62.

[23] Sun M, Schwalb B, Schulz D, et al. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res. 2012;22:1350–1359.

[24] Baptista T, Grünberg S, Minoungou N, et al. SAGA is a general cofactor for RNA polymerase II transcription. Molecular Cell. 2017;68:130–14300000.

[25] Warfield L, Ramachandran S, Baptista T, et al. Transcription of nearly all Yeast RNA polymerase II-transcribed genes is dependent on transcription factor TFIID. Mol Cell. 2017;68:118–129.e5.

[26] Hartley PD, Madhani HD. Mechanisms that specify promoter nucleosome location and identity. Cell. 2009;137:445–458.

[27] Rando OJ, Winston F. Chromatin and transcription in Yeast. Genetics. 2012;190:351–387.

[28] Parnell TJ, Huff JT, Cairns BR. RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. EMBO J. 2008;27:100–110.

[29] Ng HH, Robert F, Young RA, et al. Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex. Genes Dev. 2002;16:806–819.

[30] Lenstra TL, Benschop JJ, Kim T, et al. The specificity and topology of chromatin interaction pathways in yeast. Mol Cell. 2011;42:536–549.

[31] Plaschka C, Larivière L, Wenzeck L, et al. Architecture of the RNA polymerase II-mediator core initiation complex. Nature. 2015;518:376–380.

[32] Levo M, Segal E. In pursuit of design principles of regulatory sequences. Nat Rev Genet. 2014;15:453–468.

[33] Lubliner S, Keren L, Segal E. Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Nucleic Acids Res. 2013;41:5569–5581.

[34] Joo Y, Ficarro SB, Soares LM, et al. Downstream promoter interactions of TFIID TAFs facilitate transcription reinitiation. Genes Dev. 2017;31:2162–2174.

[35] Donczew R, Hahn S. Mechanistic differences in transcription initiation at TATA-less and TATA-containing promoters. Mol Cell Biol. 2017;38:17.

[36] Lubliner S, Regev I, Lotan-Pompan M, et al. Core promoter sequence in yeast is a major determinant of expression level. Genome Res. 2015;25:1008–1017.

[37] Maicas E, Friesen JD. A sequence pattern that occurs at the transcription initiation region of yeast RNA polymerase II promoters. Nucleic Acids Res. 1990;18:3387–3393.

[38] Helmlinger D, Marguerat S, Villén J, et al. Tra1 has specific regulatory roles, rather than global functions, within the SAGA co-activator complex. EMBO J. 2011;30:2843–2852.

[39] Hassan AH, Prochasson P, Neely KE, et al. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell. 2002;111:369–379.

[40] Musselman CA, Lalonde ME, Cote J, et al. Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol. 2012;19:1218–1227.

[41] Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem. 2012;81:65–95.

[42] Muhar M, Ebert A, Neumann T, et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science. 2018;360:800–805.

4.4.4. New insights on the functions of yeast SAGA in RNA polymerase II transcription

A very recent study in budding yeast readdressed the importance of SAGA for Pol II transcription using genome-wide analysis in auxin-inducible degradation strains of SAGA and TFIID subunits (Figure 43) (Donczew et al., 2020). Auxin-inducible degradation (AID) allows for a very rapid depletion of proteins tagged with the AID sequence through the addition of the plant-specific hormone auxin (IAA) (Verma et al., 2020). This approach is frequently referred to as degron system and enables efficient and almost complete depletion of the AID-tagged proteins within short periods of time especially in budding yeast. For example, using this system, targeted SAGA and TFIID subunits were efficiently depleted with only roughly 10% of protein remaining after 30 minutes of induction (Donczew et al., 2020).

Curiously, in contrast to constitutive deletion strains, rapid depletion of SAGA subunits only resulted into moderate effects on transcription (Figure 43). Newly synthesized RNA analyses revealed that only roughly 13% of expressed genes were downregulated in the SAGA degron strains as opposed to the majority of genes in deletion strains (Baptista et al., 2017; Donczew et al., 2020). This limited set of genes was also found to be modestly dependent on TFIID subunits leading the authors to suggest that both complexes act redundantly at these genes referred to as 'coactivator redundant' (CR) gene set (Figure 43) (Donczew et al., 2020). Upon combined acute depletion of SAGA and TFIID subunits (last
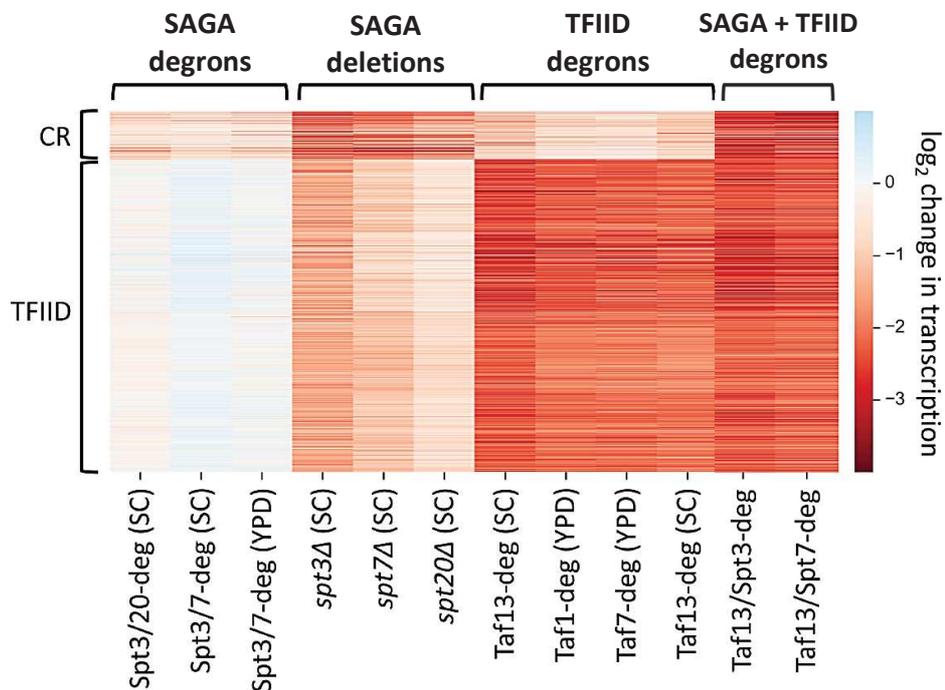
Figure 43: Roles of the SAGA and TFIID complex for RNA polymerase II transcription in budding yeast. Heatmaps of expression changes of genes actively transcribed by RNA polymerase II (Pol II) as assessed by newly synthesized RNA analyses shown as mean log$_2$ changes of replicated experiments in different mutant yeast strains (from left to right): three auxin-inducible degradation yeast strains (degrons) for SAGA subunits, three deletion strains of SAGA subunits, four degron strains for TFIID subunits and yeast strains were a SAGA as well as a TFIID subunit were tagged with the auxin-inducible degron sequence. SC, synthetic complete medium; YPD, yeast extract peptone dextrose medium; deg, degron; Δ, deletion; CR, coactivator redundant. Details in text. Adapted from Donczew et al., 2020.

two columns of Figure 43), gene expression of the CR genes is more severely affected compared to the individual SAGA or TFIID degron strains, indeed implying an overlapping function of the two complexes at these genes.

At the remaining roughly 87% of protein-coding genes, conditional loss of TFIID subunits caused a more drastic decrease in transcription compared to the CR genes, while little changes were observed in the SAGA degron strains (Figure 43). In agreement with previous reports, constitutive deletion of SAGA subunits showed striking effects on Pol II transcription, which intriguingly are more drastic than those observed in the SAGA degron strains (Bonnet et al., 2014; Baptista et al., 2017; Donczew et al., 2020). The discrepancies between SAGA subunit degron and deletion strains was suggested to be explained by several means (Donczew et al., 2020). For instance, in contrast to observations in deletion strains, histone modifications mediated by SAGA are not affected upon rapid depletion and only gradually decrease with prolonged auxin treatment (Donczew et al., 2020). This could suggest that transcriptional effects might be enhanced in the deletion strains due to the loss of histone modifications through slowly acting histone deacetylases or during cell divisions. Other secondary or indirect effects in the constitutive deletion strains such as decreased expression of components of the basic transcription machinery could further account for the more severe effect on Pol II transcription (Donczew et al., 2020). It is also not clear, if the remaining 10% of proteins in the degron strains could enable residual formation and activities of the SAGA and TFIID complexes.

The raison for the dependency of the CR gene class on both complexes, in contrast to the remaining mainly TFIID-dependent genes, also remains unclear, but was proposed to potentially involve differences in core promoter sequences or specific chromatin architecture (Donczew et al., 2020). Interestingly, differences between the two gene classes did not seem to be caused by specific binding preferences of SAGA or TFIID as no bias in promoter binding for one or the other gene class was detectable (Donczew et al., 2020). In general, these findings highlight that the exact role of yeast SAGA and its functions in Pol II transcription are not completely resolved.

4.4.5. Functions of SAGA in RNA polymerase II transcription in metazoans

The transcriptional functions of SAGA are much less studied in metazoans than in yeast and its function remains largely enigmatic. Metazoan SAGA was reported to be required for the transcriptional regulation of stimulus- or stress-responsive genes (Hardy et al., 2002; Zhao et al., 2008; Zhang et al., 2008; Pijnappel & Timmers, 2008; Nagy et al., 2009; Spedale et al., 2012). For example, the SAGA subunits SUPT20H, SUPT3H and ATXN7L3 were found to be recruited to promoters of genes induced upon ER (endoplasmatic reticulum)-stress but not upon activation of the p38 MAPK (mitogen-activated protein kinase) pathway (Nagy et al., 2009). Similarly, depletion of *Supt20h* using siRNA impairs the induction of responsive genes upon ER-stress but not p38 MAPK activation (Nagy et al., 2009). The

dependency of ER-stress response on SAGA was further reported following shRNA-mediated depletion of several DUB module subunits (Lang et al., 2011). Congruently, earlier studies indicated that human SAGA associates with the NF-Y transcription factor implicated in ER-stress response (Schröder & Kaufman, 2005). Studies on SAGA in *Drosophila* further revealed a function of its DUB module in transcription regulation of nuclear receptor-dependent and ecdysone-responsive genes (Zhao et al., 2008; Weake et al., 2008). In human cells, reduction of USP22, the catalytic subunit of the DUB module of SAGA, through shRNA-mediated knockdown was found to affect the activation of genes dependent on the transcription factors c-MYC and p53 (Zhang et al., 2008). Indeed, several studies indicate direct interactions of SAGA subunits with c-MYC including the subunits TRRAP (shared with TIP60), GCN5 (shared with ATAC), TAF5L, TAF9 and SUPT3H (McMahon et al., 1998; Liu et al., 2003). TRRAP and GCN5 were further found to interact with E2F transcription factors involved in the regulation of cell cycle progression (McMahon et al., 1998; Lang et al., 2001; Murr et al., 2007).

4.4.6. Role of SAGA during metazoan embryonic development

Subunits of the SAGA complex are essential during fly and mouse development. For example, the HAT functions of SAGA are crucial for fly development as the loss of Ada2b in *Drosophila* leads to lethality at the late-pupa stage (Pankotai et al., 2005). By comparing the genome-wide gene expression levels in these Ada2b mutant flies relative to wildtype flies, several hundred genes (roughly 400 genes) were found to be downregulated (Zsindely et al., 2009).

In mice, hypomorphic mutations in the gene encoding SUPT20H (alternative names: P38IP, FAM48A) resulted in growth defects at gastrulation, around E6.25 (Zohn et al., 2006). By E9.5 and E10.5, the mutants displayed several abnormalities such as general developmental delay, misshaped heads and frequently failed to form rostral somites leading to defects in the axil skeleton (Zohn et al., 2006; Warrier et al., 2017). The deficiencies in somite formation were linked to the misexpression of *Hox* genes and suggested to resemble hypomorphic mutants of GCN5, one of the HAT enzymes of mammalian SAGA (Warrier et al., 2017).

Inactivation of *Supt3h*, the gene encoding the potential TBP-interacting subunit of mammalian SAGA, was reported to be lethal by E14.5 in mice, while displaying placental defects by E9.5 (Perez-Garcia et al., 2018). Loss of ATXN7L3, subunit of the SAGA-dependent as well as SAGA-independent DUB modules, led to embryonic lethality by E11.5 with severe growth defects from E7.5 onwards (El-Saafin, Wang et al. manuscript in revision, Annexe 2). Intriguingly, homozygous deletion of the *Usp22* gene, encoding the enzymatic subunit of the SAGA DUB module, also caused embryonic lethality but only by E14.5 with very mild phenotypes compared to wildtype littermates (Lin et al., 2012; Koutelou et al., 2019; El-Saafin, Wang et al. manuscript in revision, Annexe 2). These differences in phenotypes between loss of ATXN7L3 and USP22 support findings indicating that USP27X and USP51 form

functional DUB modules with ATXN7L3 (Atanassov et al., 2016). This might further imply that USP27X and USP51 have important roles during mouse development or that they can compensate for the loss of the SAGA DUB module.

Opposed to the rather late embryonic phenotypes observed for the above described mouse mutants of SAGA subunits, inactivation of the gene encoding TRRAP, the subunit thought to serve as interaction platform for several activating TFs, results into very early embryonic lethality (Herceg et al., 2001). Homozygous null embryos for *Trrap* were found to die by E7.5 and outgrowth of *Trrap*$^{-/-}$ blastocysts did not show any proliferation of the inner cell mass (Herceg et al., 2001). TRRAP therefore seems required at a very early stage in mouse embryonic development. Besides SAGA, TRRAP is further incorporated into the TIP60 complex (see also chapter 3.2.2. Histone modifying complexes) for which it might be important for complex assembly, based on structures of yeast NuA4 (equivalent to the mammalian TIP60 complex) (Wang et al., 2018b). Indeed, these very early defects in mouse embryonic development observed upon loss of TRRAP are very similar to phenotypes observed upon inactivation of *Tip60,* which encodes the catalytic HAT subunit of the TIP60 complex (more details see chapter 7.4. Roles of chromatin modifying complexes in mouse embryonic stem cells) (Hu et al., 2009b). Therefore, the earlier embryonic lethality caused by loss of TRRAP, as compared to loss of other SAGA subunits, might be explained by a disruption of functions of TIP60 rather than SAGA, highlighting the difficulties encountered when studying subunits shared by several complexes.

5. The ATAC coactivator complex

In metazoans, the HAT module found in the SAGA complex can further be incorporated in another complex, the ATAC (Ada-Two-A containing) complex. The differential incorporation of the HAT module is enabled through the complex-specific interaction with one of two paralogs of the yeast Ada2 subunit. HAT modules containing the TADA2A protein are specifically incorporated into the ATAC complex, while formation of the HAT with the TADA2B subunit anchors the enzymatic module into SAGA (Spedale et al., 2012; Helmlinger & Tora, 2017).

5.1. Identification of two Ada2 proteins in metazoans

The two protein homologues of yeast Ada2 (Ada2a and Ada2b) were first identified in *Drosophila* in 2003 simultaneously by two independent laboratories (Muratoglu et al., 2003; Kusch et al., 2003). These studies also indicated that the paralogous proteins were incorporated into two distinct complexes, one being the SAGA complex. Subsequently in *Drosophila* cells, Ada2a was found to be part of ATAC, a novel histone acetyltransferase complex, whose composition was further defined (Guelman et al., 2006; Nagy & Tora, 2007; Suganuma et al., 2008). This was soon followed by the description and purification of the ATAC complex from human and mouse cells (Wang et al., 2008; Guelman et al., 2009; Nagy et al., 2010; Krebs et al., 2010; Spedale et al., 2012). Additional biochemical studies on *Drosophila* Ada2a and Ada2b reported that the incorporation of these two proteins in either SAGA or ATAC and their functions seemed mediated through their C-terminal regions (Vamos & Boros, 2012).

Table 4: Subunit composition of the mammalian ATAC complex. Shown are the subunit composition of the ATAC complex with alternative names in brackets, protein domains contained within the subunits and their functional classification. HAT, histone acetyltransferase; °, shared with SAGA; *, shared with NC2 complex; ‡, shared with several other complexes. Based on Spedale et al., 2012.

ATAC subunits	Domains	Function
GCN5 (KAT2A) ° or PCAF (KAT2B) °	HAT, Bromo, PCAF	
TADA2A	ZnF, SANT, Ada boxes, SWIRM	HAT module
TADA3 °	Ada3	
SGF29 (CCDC101) °	Tudor	
ZZZ3 (ATAC1)	ZZ-type ZnF, SANT	
YEATS2	YEATS, HFD	
WDR5‡	WD40	Other subunits
NC2β (DR1)*	HFD	
MBIP	-	
ATAC2 (KAT14, CSRP2BP)	PWAPA cassette, ZnF, HAT?	

5.2. Subunit composition of ATAC

Beside the four HAT subunits (TADA3, SGF29 and GCN5 or PCAF) shared with the SAGA co-activator complex (described in more details in chapter 6. The shared HAT function of SAGA and ATAC), ATAC is composed of five complex-specific subunits (TADA2A, ZZZ3, YEATS2, MBIP and ATAC2) and two additional subunits shared with other chromatin modifying complexes (WDR5 and NC2β) (Table 4).

The ATAC-specific HAT module subunit TADA2A possesses three domains that have been implicated in protein-protein interactions: the zinc finger (ZnF), the SANT and the SWIRM domain (Figure 44) (Muratoglu et al., 2003; Gamper et al., 2009; Zhang et al., 2019). The SWIRM domain was reported to form contacts with the TADA3 subunit, while the ZnF and SANT domains were suggested to contact GCN5 (or PCAF) (Gamper et al., 2009). TADA2A further possesses two Ada homology boxes of unknown functions and was also suggested to interact with DNA through its SWIRM domain (Muratoglu et al., 2003; Qian et al., 2005; Spedale et al., 2012).

The YEATS2 subunit of ATAC was shown to contain a HFD (histone fold domain), which was reported to dimerizes with the HFD of NC2β thereby forming an interface for the binding of TBP (Figure 44) (Wang et al., 2008). Beside the ATAC complex, NC2β, is also part of the heterodimeric NC2 complex composed of NC2α and NC2β (Spedale et al., 2012). The NC2 complex has been implicated in the dissociation of TBP from gene promoter DNA thereby negatively regulating transcription (more details in section 1.2.2.1. TBP loading onto the core promoter) (van Werven et al., 2008; Spedale et al., 2012). The role of NC2β within the ATAC complex, aside from being the HFD dimerization partner of YEATS2, is not yet understood. It has been proposed, that by dimerizing with YEATS2, NC2β forms a TBP-interacting NC2-like module in ATAC, which might act in a repressive manner on basal transcription (Wang et al., 2008). YEATS2 was also found through biochemical experiments to interact
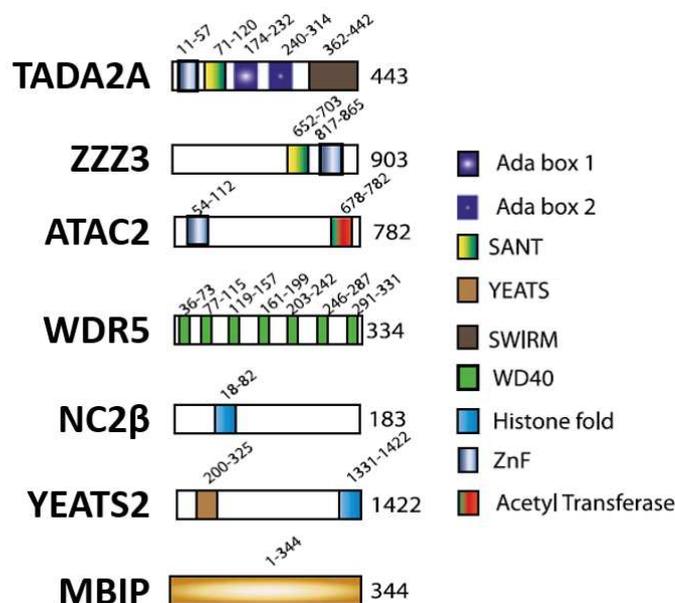
Figure 44: Schematic representation of protein domains found within subunits of the ATAC coactivator complex. The YEATS domain reflects a histone binding domain, while the histone fold, WD40, SANT, SWIRM and ZnF domains are thought to be protein-protein interaction domains. The Ada boxes represent homology domains of unknown function. More details in text. Subunits of the HAT module are not shown as they will be described in more details in a separate chapter. Adapted from Spedale et al., 201 and based on Helmlinger & Tora, 2017.

with the majority of the other ATAC subunits and was therefore suggested to be a key scaffold subunit of the complex (Wang et al., 2008). YEATS2 also contains a YEATS domain, which was recently proposed to enable the binding to histone H3 lysine 27 modified through acetylation or acylation, such as crotonylation or propionylation (Zhao et al., 2016; Li et al., 2017; Mi et al., 2017).

After YEATS2, ZZZ3 (ATAC1) is the second largest ATAC-specific subunit. Little is known about this subunit and its role for ATAC structure or assembly. ZZZ3 contains a ZZ-type ZnF domain, which was recently suggested to serve as a reader domain by binding to the histone H3 tail through recognition of the first four amino acids (Mi et al., 2018; Zhang et al., 2019; Spedale et al., 2012). The ZZ-type ZnF domain of ZZZ3 was further reported to promote acetylation of histone H3 by the ATAC complex and to be required for the recruitment of ATAC to gene targets (Mi et al., 2018). ZZZ3 also possesses a SANT domain of unknown function (Figure 44).

ATAC2 (KAT14, CSRP2BP) represents also one of the biggest ATAC-specific subunits and was suggested to have scaffolding functions within the complex (Guelman et al., 2009). It possesses a ZnF domain in its N-terminal region, a domain which is generally thought to enable protein-protein interactions (Figure 44) (Nagy et al., 2010). Its function within ATAC2 is however not yet clear. The ATAC2 subunit further contains a PWAPA (PHD/WH domain in ASH2L, PHF1 and ATAC2) cassette, which was recently proposed to potentially serve to recognize DNA and histone PTMs, such as methylation (Callebaut & Mornon, 2012). Additionally, in *Drosophila*, ATAC2 was described to act as a second HAT within the ATAC complex and to acetylate preferentially histone H4 (Suganuma et al., 2008). In *Drosophila* embryos, mutation of the gene encoding ATAC2 led to loss of acetylation specifically of histone H4 lysine 16 (H4K16ac) (Suganuma et al., 2008). Nevertheless, the HAT function of mammalian ATAC2 is controversial as *in vitro* acetylation assays suggest only little to no acetyltransferase activity on histone H4 (Wang et al., 2008; Guelman et al., 2009; Nagy et al., 2010; Spedale et al., 2012).

The MBIP subunit was also suggested to possess a scaffold function within the ATAC complex (Guelman et al., 2009). Some studies indicate that through its MBIP subunit ATAC could have a role outside of transcription in the regulation of iron-responsive translation initiation (Suganuma et al., 2016). However, little else is known about this subunit within ATAC.

The WDR5 subunit of ATAC represents a subunit shared with other chromatin modifying complexes including the COMPASS(-like) and the NSL coactivator complexes (Spedale et al., 2012; Dias et al., 2014; Meeks & Shilatifard, 2017). Through its WD40 repeat domain, WDR5 is thought to form a propeller-like structure, which could serve as a protein association platform similar to the WD40 propeller of the Taf5 subunit of yeast SAGA (Figure 44) (Xu & Min, 2011; Spedale et al., 2012).

Although some knowledge has been acquired on the subunits of ATAC, the structural organization and importance of the different subunits remains unclear. For example, PCAF was suggested to potentially exist in a dimeric state within the ATAC complex *in vivo* (Shi et al., 2014).

5.3. Functional roles of ATAC

5.3.1. Role of ATAC during metazoan embryonic development

The ATAC complex was found to have important functions during development in metazoans (Pankotai et al., 2005; Guelman et al., 2009). For example, loss of the ATAC-specific HAT subunit Ada2a in *Drosophila* was reported to be essential for embryonic development causing lethality at late-larva stages of development (Pankotai et al., 2005). Knockout of the ATAC-specific ATAC2 subunit in mice was further shown to lead to embryonic lethality at E8.5 to E11 by affecting cell cycle progression (Guelman et al., 2009).

5.3.2. Role of ATAC in cell cycle regulation

Defects in cell cycle transition were further reported upon depletion of ATAC subunits in human cells (Nagy et al., 2010; Orpinell et al., 2010). For instance, depletion of the ATAC-specific HAT subunit TADA2A using siRNA mediated knockdown was found to lead to G2/M arrest causing mitotic abnormalities and to coincide with reductions in global phosphorylation levels of histone H3 serine 10 phosphorylation (H3S10phos), a histone mark linked to the transition through the cell cycle (Nagy et al., 2010; Orpinell et al., 2010). This influence on H3S10phos was also found in Ada2a mutant flies and linked to histone H4 modification activities of ATAC (Ciurciu et al., 2008). Beside histone proteins, ATAC was further suggested to regulate mitosis through the acetylation of non-histone proteins such as PLK4 and the CyclinA/CDK2 complex, which is part of the cell cycle-regulating machinery (Orpinell et al., 2010; Fournier et al., 2016).

5.3.3. Role of ATAC in RNA polymerase II transcription

Additional loss-of-function or depletion studies in cultured cells indicated the involvement of metazoan ATAC in the regulation of specific sets of Pol II-transcribed genes during development, such as genes involved in ecdysone biosynthesis in *Drosophila*, or in response to stimuli, such as ER stress response or MAPK signalling in human cells (Pankotai et al., 2010; Suganuma et al., 2010; Nagy et al., 2010; Sela et al., 2012; Suganuma et al., 2016).

Further studies in *Drosophila* suggested that transcription of roughly 50 genes was commonly found to be reduced upon loss of either Gcn5, Ada2a or Nurf301, a subunit of the ISWI remodelling complex NURF (Carré et al., 2008). At these genes, access of ATAC to the promoter was proposed to dependent

on the activities of the NURF remodelling complex (Carré et al., 2008). Early *in vitro* studies using yeast ISWI or SWI/SNF complexes and chromatinized DNA templates also suggested a potential role of fly ATAC in stimulating nucleosome sliding activities (Suganuma et al., 2008). This impact of ATAC on nucleosome remodelling might however be attributed to a purification contamination with the CHRAC14 protein, subunit of the ATP-dependent CHRAC (chromatin accessibility complex), which by itself can stimulate nucleosome sliding (Corona et al., 2000; Eberharter et al., 2001; Hartlepp et al., 2005). Additionally, ATAC was reported to associate with the Mediator complex in mouse embryonic stem cells suggesting a potential molecular assembly of these two complexes at gene promoters (Krebs et al., 2010). In human cells, ATAC was further suggested to localize to enhancers, which are not bound by p300 (Krebs et al., 2011).

Very recent studies indicate that the human YEATS2 and ZZZ3 subunits of ATAC are involved in the transcriptional regulation of genes related to ribosome biogenesis, DNA replication and cell cycle (Mi et al., 2017; Mi et al., 2018). These two subunits were further found by ChIP-seq to co-localize at gene promoters with histone modifications including H3K9ac and H3K4me3 (Mi et al., 2018). Besides transcription effects, shRNA-mediated depletion of *Yeats2* in human cancer cell lines was found to cause reduced cell growth and survival of these cells (Mi et al., 2017). Additionally, genome-wide levels of H3K9ac as well as binding of ZZZ3 to gene promoters were affected. Importantly, the defects in growth, H3K9ac levels and transcription observed upon depletion of YEATS2 could be restored by the overexpression of WT YEATS2. In contrast, overexpression of mutant YEATS2 carrying point mutations in the YEATS domain could not rescue the detected defects. This led to the suggestion that the YEATS histone mark reader domain of YEATS2 is critical for ATAC functions in transcription regulation (Mi et al., 2017). Similarly, the ZZ-type ZnF domain of ZZZ3 was reported to be critical for its recruitment to gene promoters as well as for ATAC-dependent H3K9 acetylation and transcription regulation (Mi et al., 2018). In human cells with shRNA-mediated depletion of ZZZ3, overexpression of ZZZ3 containing loss-of-function mutations in the ZZ-type ZnF domain did not enable binding of ZZZ3 to gene promoters and was unable to restore the reduced H3K9ac levels or the transcriptional effects induced by depletion of endogenous ZZZ3 (Mi et al., 2018). In general, these two studies suggest an important function of the YEATS and ZZ-type ZnF domains of YEATS2 and ZZZ3, respectively, in transcription regulation and recruitment of ATAC to gene promoters.

6. The shared HAT function of SAGA and ATAC

This chapter will provide a more detailed summary on the composition, activities and functions of the HAT module shared by the SAGA and ATAC complexes. As most subunits of the HAT are shared by both complexes, it is challenging to attribute specific functions observed for this enzymatic module to either SAGA or ATAC. The last subsection of this chapter will be oriented towards highlighting non-overlapping functions of SAGA and ATAC in regulating transcription including also functions outside of the HAT module. Importantly, recent reports indicate that the SAGA HAT module can further act as a metazoan ADA complex (Soffers et al., 2019), which will however not be discussed here as the importance of this complex in mammalian cells is not yet clear.

6.1. Subunit composition of the HAT module

In yeast, the catalytic subunit of the histone acetyltransferase (HAT) module of SAGA is Gcn5, while the mammalian genome contains two paralogous genes encoding for two Gcn5-related proteins, GCN5 and PCAF. GCN5 and PCAF are highly similar, sharing roughly 75% of protein identity, and are incorporated into mammalian SAGA or ATAC HAT modules in a mutually exclusive way (Ogryzko et al., 1998; Krebs et al., 2010; Nagy et al., 2010; Spedale et al., 2012). No preference for either GCN5 or PCAF has been reported to date. Consequently, in mammalian cells, the HAT module of SAGA consists of TADA2B, TADA3, SGF29 and GCN5 or PCAF, while in ATAC it contains TADA2A, TADA3, SGF29 and GCN5 or PCAF.
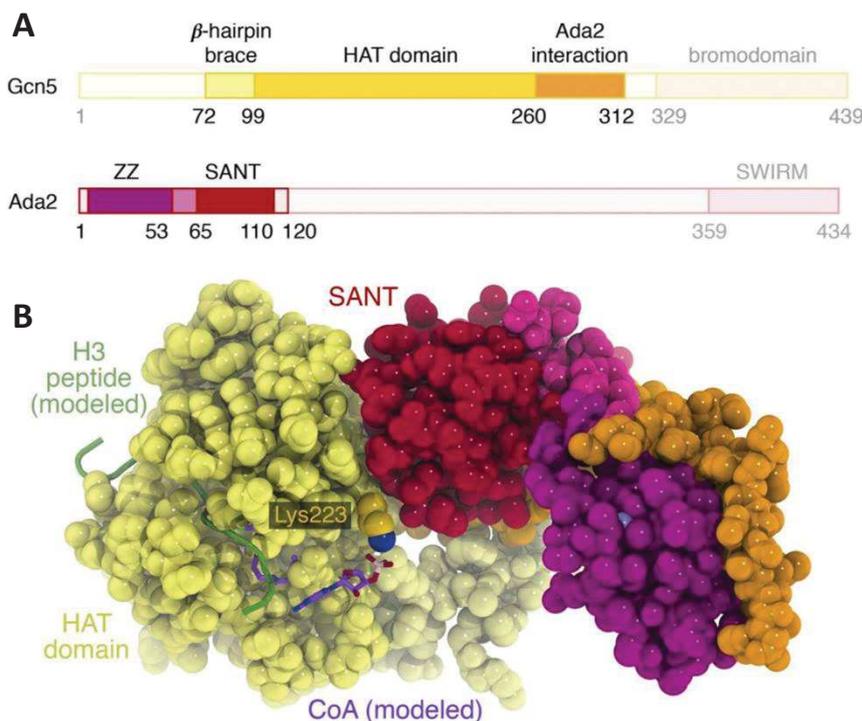
Figure 45: Crystal structure of the yeast Gcn5 and Ada2 assembly. A. Scheme of yeast Gcn5 and Ada2 proteins indicating protein domains. Regions of proteins observed in the crystal structure are shown in saturated colours, while the remaining regions are represented by shaded colours. **B.** Space-filling representation of the crystal structure of Gcn5 bound to Ada2 with modelled H3 peptide and acetyl-coenzyme A (CoA). The SANT domain of Ada2 creates a wall thereby constraining the lysine 223 (Lys223) side chain of Gcn5 enhancing the interaction with CoA. Same colour code as in (A). Adapted from Sun et al., 2018.

The subunits Sgf29/SGF29, Ada3/TADA3 and the Ada2/TADA2 proteins were found to be required for the stimulation of the enzymatic HAT activity in yeast and mammals (Grant et al., 1997; Balasubramanian et al., 2002; Spedale et al., 2012; Riss et al., 2015; Sun et al., 2018). For instance, recent crystal structures of the HAT domain of yeast Gcn5 bound to the ZZ and SANT domains of yeast Ada2 revealed that the two proteins form extensive contacts and might assemble cooperatively into a stable structure (Figure 45) (Sun et al., 2018). Also, the SANT domain of Ada2, which is conserved in the mammalian TADA2A and TADA2B proteins, was found to stimulate Gcn5 HAT activity by enhancing the binding to acetyl-CoA (Sun et al., 2018).

Additionally, other domains in the HAT subunits were suggested to interact with chromatin. For example, SGF29 contains a double Tudor domain which was reported to recognize H3K4me2 and H3K4me3 modifications (Figure 46) (Vermeulen et al., 2010; Bian et al., 2011). Curiously, pull-down experiments with H3K4me3-modified histone peptides indicated a selective interaction with SAGA, but not ATAC, although both complexes contain SGF29 (Vermeulen et al., 2010). Additionally, GCN5 and PCAF contain a bromodomain which was reported to recognize acetylated lysine residues (Figure 46) (Mujtaba et al., 2007).
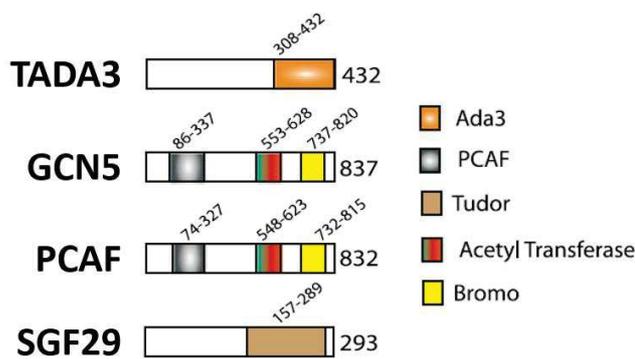
Figure 46: Schematic representation of protein domains contained within subunits of the HAT module shared by the SAGA and ATAC coactivator complexes. The Ada3 and PCAF domains represent homology domains of unknown functions, while the Bromo and Tudor domains reflect chromatin binding regions. The Acetyl Transferase domain contains the catalytic site of GCN5 and PCAF. Adapted from Spedale et al., 2012.

6.2. Acetylation substrates of the HAT module

6.2.1. Histone substrates

In vitro studies of the HAT activities of purified endogenous human SAGA or ATAC complexes on histone tail peptides mainly resulted in the acetylation of lysine 14 of histone H3 (H3K14ac) but also acetylation of H3K9, H3K23, H3K27, H3K36, H4K5 and H4K8 (Riss et al., 2015). Intriguingly, no major complex-specific differences in residue preferences were detected on histone tail peptides or on histone octamers (Riss et al., 2015). In agreement with previous findings in yeast, *in vitro* assays on histone tail peptides, histones and nucleosomes, showed that GCN5 alone has a weak HAT activity, which is however highly stimulated by its incorporation into SAGA and ATAC (Grant et al., 1999; Wang et al., 2008; Riss et al., 2015). Incorporation of the HAT module into human SAGA or ATAC was estimated to stimulate catalytic GCN5 HAT activity by 6 to 10 times, when compared to GCN5 alone (Riss et al., 2015).

In vivo findings on histone residues targeted by and dependent on SAGA and ATAC for acetylation are less clear. Nevertheless, in mammalian cells a reproducible *in vivo* acetylation target affected by loss of SAGA and ATAC HAT functions seems to be histone H3 lysine 9 (H3K9ac) (Kikuchi et al., 2005; Guelman et al., 2009; Nagy et al., 2010; Jin et al., 2011; Spedale et al., 2012; Mi et al., 2017; Mi et al., 2018). Indeed, depletion of TADA3, a subunit shared by SAGA and ATAC HAT, using shRNA in human cells was found to result into a general loss of H3K9ac at gene promoters of all actively transcribed genes (Figure 47) (Bonnet et al., 2014). The HAT module of SAGA and ATAC might also acetylate other histone H3 lysine residues such as H3K14 *in vivo*. However, upon loss of SAGA and ATAC HAT activities, other acetyltransferases seem to act redundantly on these residues, maintaining their acetylation levels. Potential *in vivo* HAT activities of the mammalian ATAC complex against histone H4 remain controversial (Wang et al., 2008; Guelman et al., 2009; Nagy et al., 2010; Orpinell et al., 2010; Riss et al., 2015).
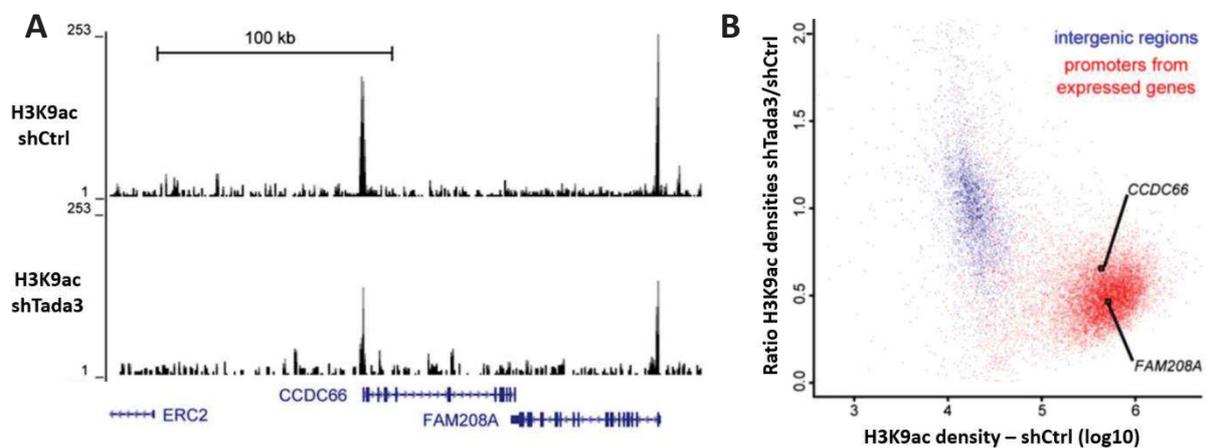
Figure 47: Loss of TADA3 of the shared histone acetyltransferase module of SAGA and ATAC leads to globally reduced levels of H3K9 acetylation. A. Genome browser shots of the *CCDC66* and *FAM208A* genes comparing H3K9ac peaks found in HeLa cells treated with control shRNAs (shCtrl) to cells treated with shRNAs targeting the mRNA of the HAT module subunit TADA3 (shTada3). **B.** Scatter plot showing a general reduction of H3K9ac levels on promoters of actively transcribed genes (red) upon depletion of TADA3 in contrast to intergenic control regions (blue). Adapted from Bonnet et al., 2014.

6.2.2. Non-histone substrates

Besides histone proteins, numerous other, non-histone targets of GCN5 and PCAF were identified including the apoptosis and cell cycle regulator p53 and transcription factors such as E2F1 and c-MYC involved in the regulation of cell cycle progression and proliferation (Liu et al., 1999; Marzio et al., 2000; Martínez-Balbás et al., 2000; Barlev et al., 2001; Liu et al., 2003; Patel et al., 2004; Di Stefano et al., 2005; Nagy & Tora, 2007; Spedale et al., 2012; Wang & Dent, 2014). Based on their activities also on non-histone substrates, GCN5 and PCAF are sometimes more specifically referred to as KATs (lysine acetyltransferases) instead of HATs.

Acetylation of non-histone proteins was frequently found to cause a stabilization of the targets. For instance, acetylation of E2F1 by PCAF was reported to lead to an increased DNA binding activity and protein stability (Martínez-Balbás et al., 2000; Marzio et al., 2000). Similarly, acetylation of the oncogenic c-MYC transcription factor by GCN5 and PCAF was found to lead to a dramatically increased stability and seems required for transcription induction of genes targeted by c-MYC (Liu et al., 2003; Patel et al., 2004). Also, acetylation of the tumour suppressor p53 by GCN5 and PCAF was suggested to promote transcription induction of p53-responsive genes upon DNA damage (Liu et al., 1999; Barlev et al., 2001; Di Stefano et al., 2005). The mechanism of p53 stimulation through acetylation remains unclear, but was proposed to involve the regulation of its DNA-binding or coactivator recruitment capacities (Barlev et al., 2001; Nagy & Tora, 2007).

6.3. Functional roles of the shared HAT module

6.3.1. Pathways and cellular processes regulated by GCN5 and PCAF

As indicated above, the HAT activities of SAGA and ATAC were implicated in the transcription regulation of numerous pathways involving transcription factors such as p53, c-MYC and E2F1 (Liu et al., 1999; Barlev et al., 2001; Liu et al., 2003; Di Stefano et al., 2005; Guo et al., 2011; Chen et al., 2013). Through their associations with p53 and E2F1, these HATs were also implied in DNA damage repair, which will however not be detailed here (Liu et al., 1999; Di Stefano et al., 2005; Guo et al., 2011). Additionally, HAT subunits were suggested to interact with AATF (apoptosis-antagonizing transcription factor) (Caliskan et al., 2017). Through its SGF29 subunit, the HAT module was also implicated in enabling the survival of human cells upon ER-stress through the regulation of stress-response genes (Schram et al., 2013).

Based on their interactions with c-MYC, the HAT activities of SAGA and ATAC were further implicated in cancer biology (Wang & Dent, 2014; Qiao et al., 2018; Farria et al., 2019). For example, inhibition of the catalytic activity of GCN5 through a small molecule inhibitor was reported to reduce cell viability and proliferation of Burkitt lymphoma cells driven by MYC overexpression (Farria et al., 2019). These phenotypes were suggested to be caused by transcriptional downregulation of MYC target genes and genes downstream of the B cell receptor signalling pathway (Farria et al., 2019). GCN5 was further reported to be required for DNA synthesis and cell cycle progression in cells expressing the human papillomavirus oncoprotein E7 (Qiao et al., 2018). In this study, the mechanism of GCN5-mediated acetylation of c-MYC described earlier was implicated in regulating E2F1 protein levels and thereby mediating E7-induced cell proliferation (Qiao et al., 2018). Although alterations in GCN5 and also PCAF levels have been found in cancer cells, it remains unclear if these changes play a leading role in cancer development (Wang & Dent, 2014).

6.3.2. Phenotypes caused by *Gcn5* and *Pcaf* inactivation in mice and other cellular models

During mouse development, homozygous inactivation of genes encoding either GCN5 or PCAF were shown to lead to distinct consequences (Yamauchi et al., 2000; Xu et al., 2000). While inactivation of *Gcn5* caused embryonic lethality by E10.5, *Pcaf*$^{-/-}$ mice were born and did not show any obvious abnormal phenotype. *Gcn5*-null mice displayed severe growth defects at E8.5 including loss of dorsal mesoderm due to increased cell death (Xu et al., 2000). Intriguingly, GCN5 levels were drastically increased in several tissues of the *Pcaf*-null mice suggesting a potential functional compensation for loss of PCAF. Further, PCAF was also found to be expressed only at later stages during development compared to GCN5, which could additionally explain the lack of phenotype of *Pcaf*$^{-/-}$ mice. Subsequent studies reported behavioural changes in *Pcaf*-null mice, especially memory deficiencies, suggesting that Gcn5 can not compensate for all functions of Pcaf (Maurice et al., 2008). Combined inactivation of *Gcn5* and *Pcaf* resulted into embryonic lethality around E7.5 to E9.5 displaying more drastic phenotypes than *Gcn5*-null embryos, which suggested that GCN5 and PCAF indeed act partially redundant during embryogenesis (Xu et al., 2000; Yamauchi et al., 2000).

These results in mouse embryos could be recapitulated in cultured chicken cells (Kikuchi et al., 2005). While inactivation of *Gcn5* delayed growth rates and affected the transcription of cell cycle- and apoptosis-related genes, such as E2F protein genes and the *c-Myc* gene, cell growth was unaffected by loss of PCAF in chicken cells (Kikuchi et al., 2005). Curiously, in this cellular model, PCAF levels were increased in *Gcn5*$^{-/-}$ cells further suggesting potential compensatory mechanisms. Later studies further revealed that loss of the catalytic activity of GCN5 through point mutations also caused embryonic lethality in mice but did not fully recapitulate phenotypes observed in *Gcn5*-null mice (Bu et al., 2007). For example, while *Gcn5*-null embryos died by E10.5 and showed defects in mesodermal lineages, embryos with catalytically dead Gcn5 survived until E16.5 and displayed only slightly reduced sizes as well as some defects in neural tube closure (Bu et al., 2007). These results implied that Gcn5 possesses functions required for mouse embryonic development, which are independent of its HAT activity and were suggested to potentially involve its bromodomain or interactions with the Tada2 proteins (Bu et al., 2007; Spedale et al., 2012).

6.3.3. Loss of the shared Tada3 subunit causes embryonic lethality before implantation

In complement to the analyses of the catalytic subunits, the effects caused by loss of the shared Tada3 subunit were also assessed during mouse embryonic development (Mohibi et al., 2012). Inactivation of *Tada3* was also found to be embryonic lethal, but surprisingly was shown to induce a drastically more severe phenotype. No homozygous *Tada3*-null embryos could be found at E8.5 and subsequent analysis indicated that blastocysts from E3.5 were lacking an inner cell mass suggesting very early developmental defects (Mohibi et al., 2012). Conditional deletion of *Tada3* in mouse embryonic

fibroblast also caused major proliferation defects as well as delays in cell cycle progression. These defects were linked to impaired Tada3 recruitment to the *c-Myc* gene promoter and to decreased *c-Myc* expression levels (Mohibi et al., 2012). The causes for the strong differences in mouse embryo phenotypes between inactivation of *Tada3* and the combined inactivation of *Gcn5* and *Pcaf* remain unclear and are remarkable as both are thought to lead to loss of the same HAT functions.

6.4.   Non-overlapping roles for SAGA and ATAC in transcription regulation

Although, SAGA and ATAC have very similar HAT activities on nucleosome substrates *in vitro*, several studies indicated that they act as transcriptional coactivators for distinct sets of genes in *Drosophila* and human cells (Pankotai et al., 2005; Nagy et al., 2010; Pankotai et al., 2010; Krebs et al., 2011; Riss et al., 2015). One of the first indications came from deletion studies of the *Ada2a* (ATAC) and *Ada2b* (SAGA) homologues in *Drosophila* embryos (Pankotai et al., 2005). Phenotypes observed in fly embryos upon loss of the Ada2a protein could not be compensated by overexpression of *Ada2b* or vice versa, suggesting that these two proteins are required for distinct and independent HAT functions in *Drosophila* (Pankotai et al., 2005). Subsequent analysis of gene expression changes in these *Drosophila* mutants, found that different sets of genes were transcriptionally affected upon loss of Ada2a or Ada2b (Pankotai et al., 2010). More specifically, the expression of roughly 40% of all genes were altered in *Ada2a Drosophila* mutants compared to only roughly 3% upon loss of the SAGA-specific subunit Ada2b (Pankotai et al., 2010). This was in agreement with an earlier embryonic lethality caused by loss of Ada2a compared to loss of Ada2b. Many of the genes affected by loss of Ada2a were found to correspond to genes of the ecdysone biosynthesis pathway and therefore suggested a specific role of ATAC, but not SAGA, in regulating transcription of ecdysone-related genes (Pankotai et al., 2010).

SAGA and ATAC were further reported to be required for the induction of the transcriptional response for different signalling pathways (Nagy et al., 2010; Spedale et al., 2012). For example, ATAC, but not SAGA, was found to be specifically required for induced gene expression following the activation of the protein kinase C signalling pathway in *Drosophila* and human cells (Nagy et al., 2010). For instance, transcription sites induced by protein kinase C activation on polytene chromosomes in *Drosophila* were bound by the ATAC-specific subunit Mbip, but not the SAGA-specific subunit Ada2b (Nagy et al., 2010). Similarly, in human cells, the ATAC-specific subunits ZZZ3 and ATAC2 were found to be indispensable for transcription induction following protein kinase C activation, in contrast to the SAGA-specific subunit SUPT20H (Nagy et al., 2010). Another example, mentioned in a previous section, is gene expression induced by p53, which was suggested to be regulated by GCN5, and which was subsequently found to involve the SAGA complex rather than ATAC in human cells (Gamper et al., 2009). While the SAGA-specific subunit TADA2B was recruited to p53-response genes, the ATAC-specific subunit TADA2A was not (Gamper et al., 2009).

Functional differences between SAGA and ATAC were also found in unstressed human cells (Krebs et al., 2011). In this study, SAGA and ATAC subunits (ZZZ3 and SUPT20H, respectively) were found to bind and to affect transcription of different sets of genes (Krebs et al., 2011). Overall, roughly 400 high confidence binding sites were identified for either ZZZ3 or SUPT20H with a subset of overlapping binding sites. ZZZ3 was found at promoters as well as enhancers, while SUPT20H was generally found at gene promoters (Krebs et al., 2011). Genes bound by either ZZZ3 or SUPT20H did not show enrichment for genes of any particular pathway and were found equally likely at ubiquitously expressed or tissue-specific genes (Krebs et al., 2011). More recent studies on ZZZ3 and YEATS2 functions in human cells, described in more details in an earlier chapter (see section 5.3.3. Role of ATAC in RNA polymerase II transcription), suggest however specific dependencies of ribosome biogenesis, DNA replication and cell cycle-related genes on ATAC subunits (Mi et al., 2017; Mi et al., 2018).

In general, these studies indicate that, although SAGA and ATAC share the same HAT module, distinct and rather small gene sets are dependent on or bound by their subunits. These functional differences were suggested to be based on characteristics of the remaining, complex-specific subunits such as additional chromatin interaction motifs, specific associations with other chromatin modifying complexes or interactions with distinct TFs (Spedale et al., 2012). Importantly, genome-wide studies suggested that some genes are transcriptionally dependent on or occupied by subunits of both complexes (Pankotai et al., 2010; Krebs et al., 2011). Also, in human cells, the transcription factor ATF6α was reported to recruit, besides the Mediator complex, also both, SAGA and ATAC, to enhancer elements of ER-stress responsive genes (Sela et al., 2012). Curiously, the interactions of Mediator, SAGA and ATAC with ATF6α was found to be mediated through overlapping binding sites on this TF (Sela et al., 2012). Nevertheless, this example suggests that SAGA and ATAC might be recruited to the same genetic elements through TFs, which can interact with both complexes.

7. Mouse embryonic stem cells

Mouse embryonic stem cells (ESC), derived from the inner cell mass of mouse blastocysts at E3.5 or E4.5, were first maintained outside of the embryo in 1981, which represented a major milestone (Martin, 1981; Evans & Kaufman, 1981; Surani et al., 2007; Evans, 2011; Martello & Smith, 2014). Mouse ESCs are thought to resemble the naïve epiblast, also known as the primitive ectoderm, which reflects the founder cell population of the embryo (Surani et al., 2007; Martello & Smith, 2014). No more than a few dozen cells giving rise to ESCs are thought to exist in normal early mouse embryonic development (Evans, 2011).

7.1. Mouse embryonic stem cell properties

A key characteristic of ESC is their pluripotency based on which they can resume development when reintroduced into early embryos, contributing to every germ layer even after multiple generations in culture (Bradley et al., 1984; Surani et al., 2007; Martello & Smith, 2014). Upon exposure to defined cues, pluripotent ESCs can also give rise to very diverse populations of differentiated cells when cultured outside of the embryo. Another key characteristic of mouse ESCs is their seemingly endless potential to self-renew, which allows them to proliferate in theory indefinitely.

Since normal embryonic development progresses very dynamically, ESCs correspond to a very transient state *in vivo*, which gives their self-renewal capacities a seemingly paradoxical notion (Surani et al., 2007; Martello & Smith, 2014). However, embryonic development can be paused by delaying blastocysts implantation, which is referred to as 'diapause' (Martello & Smith, 2014). In this period, developmentally blocked blastocysts are viable for days up to weeks and were reported to display cell turnover in the naïve epiblast, which implies the existence of self-renewal (Martello & Smith, 2014). Diapause was therefore suggested to be the evolutionary cause for the self-renewal potential of mouse ESCs (Martello & Smith, 2014).

Compared to somatic cells, ESCs possess additional distinguishing characteristics such as rapid cell cycle transitions, distinct epigenetic features such as low DNA methylation levels and bivalent domains, and a generally accepted open chromatin landscape described below in more details (Surani et al., 2007; Festuccia et al., 2017a).

7.1.1. Cell cycle characteristics in mouse embryonic stem cells

A key hallmark of the cell cycle of mouse ESCs is their short G1-phase, which is thought to be enabled through the omission of the G1/S checkpoint (Festuccia et al., 2017a). This characteristic allows for a large proportion of actively replicating cells in mouse ESC cultures and permits a cell division roughly every 12 hours (Festuccia et al., 2017a). The fast doubling time and cell cycle transition of

mouse ESCs compared to somatic cells are thought to be based on three adjustments (Festuccia et al., 2017a). First, Rb (retinoblastoma), a key regulator of G1/S-phase transition, is hyperphosphorylated in mouse ESCs allowing seemingly unrestricted entry into S-phase (Savatier et al., 1994; Festuccia et al., 2017a). Second, cell cycle inhibitors are lowly expressed in ESCs and third, the characteristic fluctuations of protein levels of cyclins and their dependent kinases occurring during cell cycle in somatic cells are strongly damped in mouse ESCs (Savatier et al., 1996; Stead et al., 2002; Faast et al., 2004; Fujii-Yamamoto et al., 2005; White et al., 2005; Festuccia et al., 2017a).

Importantly, differentiation of ESCs was found to coincide with a prolonged G1-phase. Further, artificial manipulation of the duration of G1-phase was reported to influence the differentiation likelihood of mouse ESCs with a longer G1-phase favouring differentiation (Coronado et al., 2013; Festuccia et al., 2017a). Based on these insights, G1-phase is believed to be the cell cycle phase in which ESCs are most susceptible and most responsive to differentiation cues. In consequence, the abbreviated G1-phase of mouse ESCs relative to somatic cells might thus be important to allow the efficient maintenance of the pluripotent state and reduce the initiation of differentiation (Festuccia et al., 2017a).

7.1.2. Chromatin organization in mouse embryonic stem cells

In the ground state of pluripotency (more details in a following section) mouse ESCs exhibit a substantial demethylation of their genomes, very similar to the preimplantation epiblast (Ficz et al., 2013; Habibi et al., 2013; Leitch et al., 2013). This circumstance was suggested to be caused by low expression levels of the DNA methyltransferases Dnmt3a and Dnmt3b and is thought to contribute to a transcriptionally permissive chromatin environment in mouse ESCs (Ficz et al., 2013; Leitch et al., 2013; Habibi et al., 2013).

Also, histone proteins were found to display hyperdynamic behaviours in the nucleus of mouse ESCs compared to differentiated cells (Meshorer et al., 2006). Therefore, histone proteins are believed to be only loosely bound to DNA, which led to the suggestion that chromatin in ESCs is in a mostly active, 'breathing' state (Meshorer et al., 2006). In ground state pluripotency, mouse ESCs also present low levels of histone marks associated with heterochromatin such as H3K9me3 or H3K27me3 and higher levels of euchromatic marks such as H3 and H4 acetylation and H3K4me3, suggesting a mainly active chromatin with only few transcriptionally repressed regions (Meshorer et al., 2006; Niwa, 2007; Efroni et al., 2008; Festuccia et al., 2017a).

In general, this open chromatin landscape is thought to enable a globally transcriptionally active genome and to facilitate the transcriptional activation of lineage-determining genes during induction of differentiation (Meshorer et al., 2006; Meshorer & Misteli, 2006; Niwa, 2007; Efroni et al., 2008; Martello & Smith, 2014).

7.1.3. Bivalent chromatin domains in mouse embryonic stem cells

An additional feature of mouse ESCs, shared with cells of the inner cell mass, is the cooccurrence of active and inactive histone modifications at some *cis*-regulatory regions referred to as 'bivalent' domains (Azuara et al., 2006; Bernstein et al., 2006; Surani et al., 2007; Festuccia et al., 2017a). Although bivalent domains possess the euchromatic H3K4me3 mark, they are not transcriptionally active. Transcription at these domains is thought to be prevented by the simultaneous existence of the histone modification H3K27me3 deposited by PRC2 and related to transcriptional repression (Azuara et al., 2006; Bernstein et al., 2006). Interestingly, in ground state pluripotency low levels of H3K27me3 are observed also at bivalent genes without however causing induction of expression of these genes (Marks et al., 2012; Festuccia et al., 2017a). An alternative way of regulating transcription at bivalent regions, besides the counteracting histone modifications mentioned above, was suggested to be at the level of promoter-proximal pausing of Pol II (Stock et al., 2007; Festuccia et al., 2017a).

During differentiation, bivalent regions were found to resolve in a lineage-specific manner into either active or inactive domains accompanied by the loss of either H3K27me3 or H3K4me3, respectively (Mikkelsen et al., 2007). Bivalent domains are therefore believed to allow gene silencing of developmental genes in ESCs while simultaneously poising them for future activation (Bernstein et al., 2006; Mikkelsen et al., 2007; Surani et al., 2007; Festuccia et al., 2017a). Intriguingly, as mentioned in an earlier chapter, mammalian Mll3/Mll4 COMPASS-like complexes contain, besides their H3K4 methyltransferase activity, also Utx (Kdm6a), a demethylase acting on H3K27me3. This could allow the Mll3/Mll4 COMPASS-like complexes to methylate H3K4, while simultaneously demethylating H3K27me3 (Herz et al., 2012; Meeks & Shilatifard, 2017). In agreement, Utx is not required for proliferation or maintenance of ESCs but its loss affects differentiation (Morales Torres et al., 2013). However, the effects of loss of Utx on differentiation was reported to be independent of its catalytic activity and to be potentially compensated by the functions of two alternative demethylases, Uty and Jmjd3 (Morales Torres et al., 2013).

7.2. Major signalling pathways in mouse embryonic stem cells

In the early days of mouse ESC research, ESCs were commonly cultured in the presence of calf serum on a layer of so-called 'feeder' cells, represented nowadays by mitotically inactivated fibroblasts. The coculturing with feeder cells allowed to maintain ESCs in a pluripotent state, often referred to as ground state pluripotency. In 1988, the first pluripotency-mediating component of these cocultures was identified and was found to be the cytokine LIF (leukaemia inhibitory factor) (Smith et al., 1988; Williams et al., 1988). The presence of LIF was reported to inhibit differentiation of cultured mouse ESCs and allowed, in combination with serum, the long-term culturing of mouse ESCs on gelatine-coated dishes in the absence of any feeder cells. On a side note, LIF is primarily localized to the cell

surface of feeder cells and not secreted, however mouse ESCs also recognized unbound LIF as used in culturing medium (Rathjen et al., 1990).

Stat3 (signal transducer and activator of transcription 3) was subsequently identified as the main effector of LIF-mediated self-renewal (Boeuf et al., 1997; Surani et al., 2007; Martello & Smith, 2014). Binding of LIF to a bipartite receptor leads to the activation of JAKs (Janus-associated kinases), which in turn activate Stat3 through phosphorylation events (Martello & Smith, 2014). Phosphorylation of Stat3 enables its dimerization, its relocation to the nucleus and eventually gene activation of key pluripotency transcription factors (Figure 48). Interestingly, the MEK/ERK (mitogen-activated protein kinase ERK kinase/extracellular signal-related kinase) pathway, which also lies downstream of activated JAKs, is acting antagonistically to self-renewal (more details below) (Martello & Smith, 2014).

Mouse ESCs cultured in medium containing only LIF without serum spontaneously differentiate mainly into neural precursors and neurons. The antagonistic effects of serum on neuronal specification was found to depend on BMPs (bone morphogenetic proteins) (Ying et al., 2003; Martello & Smith, 2014). In combination with LIF, BMP can maintain the self-renewal capacities of mouse ESCs and replace serum. Curiously, during embryonic development, pluripotential cells of the epiblast are not exposed to LIF signalling (Martello & Smith, 2014). However, LIF-mediated self-renewal was reported to be essential to maintain the epiblast cell population during diapause explaining the responsiveness of cultured mouse ESCs to LIF (see also previous section) (Nichols et al., 2001).
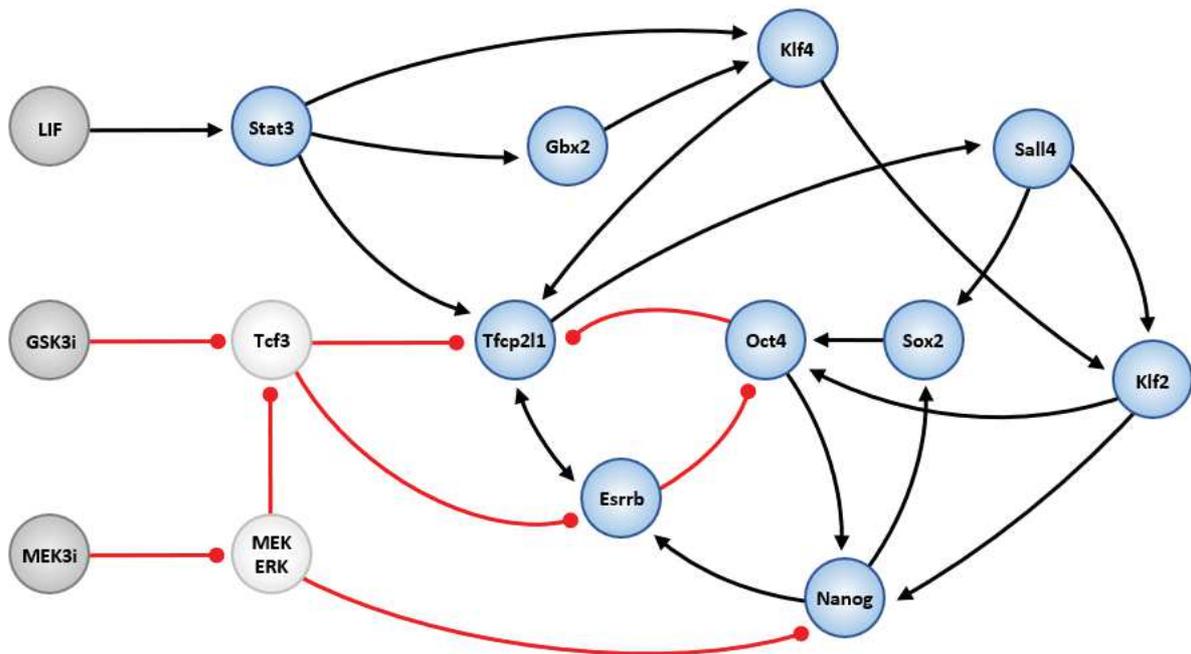
Figure 48: Simplest computationally predicted model based on experimental evidences of a possible interaction network among key pluripotency transcription factors governing ground state pluripotency of mouse embryonic stem cells. The three environmental inputs (LIF and two inhibitors, GSK3i and MEKi) present in the culture medium are shown in dark grey nodes. Light grey nodes highlight differentiation-inducing factors implicated in the exit from the ground state pluripotency. Red circle-headed lines reflect negative regulation, while black arrows serve to indicate positive regulation. LIF, leukaemia inhibitory factor; GSK3i, GSK3 inhibitor; MEKi, MEK inhibitor. Adapted from Martello & Smith, 2014.

In most cell types, MEK/ERK signalling is essential, yet, as mentioned earlier, in mouse ESCs MEK/ERK activation is undesirable as it acts antagonistically to self-renewal (Martello & Smith, 2014). The main activator of MEK/ERK is a common compound of ESC culture medium, Fgf4 (fibroblast growth factor 4). Fgf4 binds to Fgfr2 (FGF receptor 2), which is expressed by mouse ESCs, leading to MEK/ERK activation, which in turn counteract self-renewal by repressing pluripotency transcription factors, especially *Nanog* expression, and by favouring differentiation (Figure 48) (Kunath et al., 2007; Martello & Smith, 2014).

Beside MEK/ERK, GSK3 (glycogen synthase kinase-3) activities were further identified to antagonise ESC self-renewal capacities mainly via the destabilization of β-catenin (Sato et al., 2004; Martello & Smith, 2014). In ESCs, β-catenin is required to prevent the repressive effects of the DNA-binding factor Tcf3 (transcription factor 3), which directly impedes expression of key pluripotency factors (Figure 48) (Wu et al., 2012; Martello et al., 2012; Shy et al., 2013).

Combined utilisation of MEK/ERK and GSK3 inhibitors in mouse ESC cultures revealed to be highly potent in sustaining self-renewal of ESCs (Sato et al., 2004; Ying et al., 2008; Martello & Smith, 2014). Medium containing LIF and the two small-molecule inhibitors PD0325901 and CHIR99021, counteracting MEK/ERK and GSK3 signalling respectively, is frequently referred to as LIF+2i medium (Martello & Smith, 2014; Navarro, 2018). Culturing of mouse ESCs in LIF+2i medium supports the growth of a homogeneous, undifferentiated cell population. As indicated earlier, ESCs cultured in these conditions are considered to be in ground state pluripotency, thereby displaying robust self-renewal and resembling the naïve epiblast of preimplantation embryos (Silva & Smith, 2008; Wray et al., 2010; Martello & Smith, 2014; Boroviak et al., 2014; Navarro, 2018). In contrast to cultures of ESCs maintained in LIF+2i medium or on feeder cells, mouse ESCs grown on gelatine-coated plates in medium containing only LIF and serum, are morphologically more heterogeneous (more details below) (Chambers et al., 2007; Wray et al., 2010; Martello & Smith, 2014; Navarro, 2018).

7.3. Pluripotency factors and interaction networks in mouse embryonic stem cells

7.3.1. Identification of Nanog as a pluripotency factor

A pluripotency factor is defined 'as a gene product that directly and specifically supports the maintenance of ESC identity' (Martello & Smith, 2014). Several pluripotency factors were identified through their specific expression in the early embryo and in mouse ESCs such as Nanog (Mitsui et al., 2003; Surani et al., 2007; Silva et al., 2009; Martello & Smith, 2014). Additionally to its specific expression profile, the inner cell mass of *Nanog*[-/-] embryos fails to produce an epiblast and instead results in the generation of parietal endoderm-like cells (Mitsui et al., 2003; Silva et al., 2009). *Nanog*-null mouse ESCs display reduced self-renewal efficiencies in culture without losing their pluripotency.

Indeed, *Nanog*-null ESCs can contribute to all germ layers upon injection into blastocysts (Chambers et al., 2007; Hastreiter et al., 2018).

Additional insights from ESCs allowed the description of an interconnected network of pluripotency transcription factors governing the ESC state. The simplest version of a computationally predicted model integrating experimental evidences of various cell culturing conditions with as little interactions as possible amongst factors is shown in Figure 48 (Martello & Smith, 2014). Importantly some of the connections are predicted and not experimentally validated such as the repression of Oct4 by Esrrb. Also, some experimental evidences for functional links between these TFs are not represented by this model (Martello & Smith, 2014).

7.3.2. The interaction network of the pluripotency transcription factors

The core activities of the pluripotency network are thought to be Oct4 and Sox2, with Oct4 (encoded by the *Pou5f1* gene) being the first identified in 1990 (Schöler et al., 1990; Okamoto et al., 1990; Avilion et al., 2003; Elling et al., 2006; Martello & Smith, 2014). Deletion of *Pou5f1* leads to loss of inner cell mass pluripotency with differentiation being restricted to the extraembryonic trophoblast lineage (Nichols et al., 1998). Sox2 physically interacts with Oct4 and together they bind to Oct/Sox elements, which enables the positive regulation of gene expression such as *Pou5f1* itself (Martello & Smith, 2014).

Interestingly, although essential, the expression of the core pluripotency TFs Oct4 and Sox2 is not restricted to pre-implantation lineages, such as in the naïve epiblast or mouse ESCs, but also found at post-implantation stages (Surani et al., 2007; Martello & Smith, 2014). In contrast, expression of *Nanog* is more restricted to cells in naïve pluripotency, therefore Nanog is considered a naïve-specific pluripotency factor (more details in a later section) (Silva et al., 2009). Other transcription factors specifically expressed in naïve epiblast and mouse ESCs are Tfcp2l1, Esrrb, Klf4, Klf2, Gbx2 and Tbx3. Similar to *Nanog*, inactivation of the genes encoding these naïve-specific pluripotency factors can be tolerated by mouse ESCs, with the exception of inactivation of *Esrrb*, which greatly impairs self-renewal (Martello & Smith, 2014). These findings highlight potential redundancy among pluripotency transcription factors.

The expression of the naïve-specific pluripotency factors *Nanog*, *Klf2* and *Esrrb* was reported to be directly repressed by Tcf3 and are consequently upregulated upon inhibition of GSK3 in LIF+2i medium (Figure 48, page 145) (Martello & Smith, 2014). Interestingly, Nanog was found to autorepress itself forming a negative feedback loop (Navarro et al., 2012). In contrast, LIF-mediated activation of Stat3 induces the transcription of naïve-specific pluripotency TFs, such as Klf4 and the abundant Tfcp2l1 transcription factor (Figure 48, page 145). Indeed, Tfcp2l1 was suggested to be the central player of LIF-mediated self-renewal (Ye et al., 2013; Martello et al., 2013; Martello & Smith, 2014). *Tfcp2l1* also represents a target of Tcf3, thus LIF signalling and inhibition of GSK3 converge on regulating Tfcp2l1

levels (Figure 48, page 145) (Martello et al., 2013; Ye et al., 2013; Martello & Smith, 2014). Intriguingly, expression of the core pluripotency TFs Oct4 and Sox2 are not directly regulated by either 2i or LIF (Figure 48, page 145) (Martello & Smith, 2014).

7.3.3. Variable expression of naïve-specific pluripotency factors and mouse embryonic stem cell differentiation potential

In the absence of 2i, mouse ESCs display highly variable expression levels of Nanog and other naïve-specific pluripotency TFs, while levels of the core pluripotency TFs, such as Oct4, largely remain constant (Figure 49) (Martello & Smith, 2014; Navarro, 2018). These fluctuations of levels of naïve-specific pluripotency TFs were associated with the stochastic differentiation potential observed in mouse ESCs grown in LIF and serum (Figure 49) (Chambers et al., 2007; Kalmar et al., 2009; Wray et al., 2010; Martello & Smith, 2014). For example, mouse ESCs with reduced levels of Nanog in FCS + LIF medium were suggested to be inclined to differentiate and to express a variety of differentiation markers at low levels (Filipczyk et al., 2013; Abranches et al., 2014; Navarro, 2018). Therefore, cultures of ESCs in FCS + LIF medium consist of a heterogenous population of cells with some cells maintaining high levels of naïve-specific pluripotency TFs, such as Nanog, while others transiently display reduced levels of naïve-specific TFs thereby seemingly being more sensitive for differentiation.
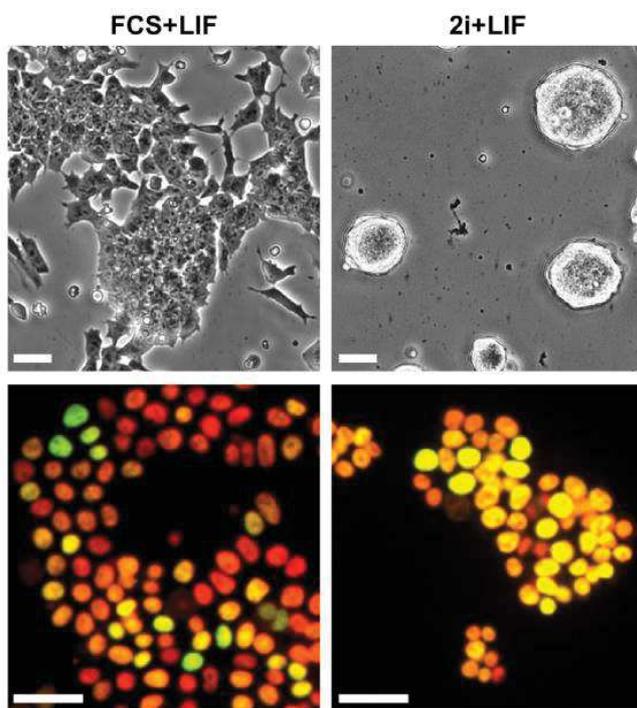
Figure 49: Morphological characteristics and differences in Nanog expression of mouse embryonic stem cells cultured in LIF and serum or 2i and LIF medium. Top shows bright-field microscopy images of mouse ESC colonies cultured either in medium containing serum (FCS) and LIF (left) or 2i and LIF (right). Bottom shows immunofluorescent staining of Oct4 (red) and Nanog (green). LIF, leukaemia inhibitory factor. Scale bars indicate 30 μm. From Navarro, 2018.

Fluctuations of naïve-specific pluripotency factors were reported to be caused by their downregulation through the Fgf4 and MEK/ERK pathway (Silva et al., 2009; Lanner & Rossant, 2010). Tcf3 represents an additional potent intrinsic negative regulator, which could also be at the root of the observed fluctuations of naïve-specific TFs (Martello & Smith, 2014). The downregulated naïve-

specific pluripotency TFs can spontaneously regain expression, which led to the notion of metastability or dynamic equilibrium of pluripotency (Hayashi et al., 2008; Silva & Smith, 2008). These notions suggest that the heterogeneity of mouse ESCs grown in FCS + LIF medium could reflect transient cell states induced through responses to incoherent external inputs (Smith, 2013; Kalkan & Smith, 2014). On the other hand, the autorepression of *Nanog*, mentioned earlier, was suggested to be sufficient in establishing a heterogenous population of cells with fluctuating *Nanog* expression levels (Navarro et al., 2012).

Pluripotency in FCS + LIF medium was also described as being a precarious balance in which naïve-specific pluripotency TFs and rivalling lineage specifiers continually compete against each other (Loh & Lim, 2011). In contrast, in medium containing 2i, mouse ESCs represent a rather homogenous population with every cell presenting detectable levels of the naïve-specific pluripotency TFs (Nanog levels shown in Figure 49) (Wray et al., 2010; Martello & Smith, 2014). Also, ESCs exposed to 2i display reduced to undetectable levels of most lineage markers in contrary to ESCs grown in LIF + serum medium (Marks et al., 2012).

The maintenance of a homogenous population of ESCs with high levels of naïve-specific TFs in medium containing 2i was reported to be enabled through two main mechanisms (Hastreiter et al., 2018; Navarro, 2018). First, 2i culturing condition induce the expression of naïve-specific pluripotency transcription factors, such as *Nanog*, *Tfcp2l1* and *Esrrb*. For example, a recent study indicates that cells expressing low levels of Nanog in LIF and serum conditions can increase transcription of *Nanog* almost immediately following 2i addition to the medium (Hastreiter et al., 2018; Navarro, 2018). Second, differentiating cells or cells expressing Nanog at low levels display a selective growth disadvantage in 2i conditions and are generally eliminated through increased cell death rates (Hastreiter et al., 2018; Navarro, 2018).

7.3.4. Exit from ground state pluripotency and early differentiation of mouse embryonic stem cells

Mouse ESCs are believed to exit self-renewal before lineage commitment and differentiation by first loosing expression of naïve-specific factors, such as Nanog, Tfcp2l1 and Esrrb, (mentioned above) (Martello & Smith, 2014). Subsequently, expression of lineage-specification markers is thought to be induced, leading eventually to loss of the expression of the core pluripotency regulators (Oct4 and Sox2) (Martello & Smith, 2014). For example, overexpression of several lineage-specific transcription factors, such as Gata4 and Gata6, in mouse ESCs cultured in LIF and serum conditions were found to cause a loss of self-renewal and to induce differentiation (Shimosato et al., 2007). During differentiation, the G1-phase of the cell cycle extends and chromatin structures become more compact, forming

heterochromatic domains and limiting overall transcription (Aoto et al., 2006; Meshorer & Misteli, 2006; Festuccia et al., 2017a).

The continued expression of the core pluripotency regulators during the early differentiation processes of cultured mouse ESCs, when expression of naïve-specific pluripotency factors is lost and ESCs enter a primed state of pluripotency (epiblast-like stem cells), is consisted with the expression of these factors in the post-implantation epiblast of mouse embryos (Thomson et al., 2011; Trott & Martinez Arias, 2013; Martello & Smith, 2014). Entry of mouse ESCs into differentiation is believed to be a continuous transition. However, until a given transition point, at which the ESC identity is permanently lost, the initial exit from ground state pluripotency was found to be reversible (Kalkan & Smith, 2014).

7.4. Roles of chromatin modifying complexes in mouse embryonic stem cells

Several evidences link the pluripotency network and its signalling with chromatin modifying complexes, which have diverse roles in mouse ESC derivation and maintenance (Surani et al., 2007; Festuccia et al., 2017a). Self-renewal and survival of mouse ESCs were found to be especially sensitive to the loss of chromatin modifying complexes linked to active transcription and open chromatin, such as the TIP60 complex, Setd1a-containing COMPASS complex, Brg1-containing esBAF (embryonic stem cell-specific BAF) complex and INO80, while repressive chromatin modifiers seem to play inferior roles with the exception of Setdb1 (summarized in Table 5, page 154) (Surani et al., 2007; Festuccia et al., 2017a).

7.4.1. Role of histone methyltransferases in mouse embryonic stem cells

Mouse ESCs cannot be derived from blastocysts with inactivation of Setd1a, a catalytic subunit of the mammalian COMPASS complex responsible for H3K4me3 (see also chapter 3.2.2. Histone modifying complexes) (Bledau et al., 2014; Fang et al., 2016). Further, inducible deletion of Setd1a was reported to affect self-renewal and proliferation of mouse ESCs and to lead to transcriptional misregulation of especially Oct4 regulated genes (Bledau et al., 2014; Fang et al., 2016). Further, Setd1a was suggested to directly interact with Oct4 suggesting that Setd1a is a coactivator of Oct4 (Fang et al., 2016).

In contrast, loss of Mll2, another catalytic subunit of mammalian COMPASS-like complexes, was found to lead to only few changes in gene expression in mouse ESCs. In agreement, defects in *Mll2*[-/-] embryos appeared at later stages in embryonic development with growth retardation from E6.5 and lethality observed by E10.5 (Glaser et al., 2006; Glaser et al., 2009). Similarly, Mll3 and Mll4, which are responsible for the deposition of H3K4me1 at enhancers, were found to be dispensable for ESC

maintenance (Wang et al., 2016). Overall, Setd1a was suggested to be the major H3K4 methyltransferase in mouse ESCs (Bledau et al., 2014).

7.4.2. Role of chromatin remodellers in mouse embryonic stem cells

In agreement with the general open chromatin landscape observed in mouse ESCs, several chromatin remodelling complexes involved in nucleosome sliding or removal were found to be important for ESC physiology. For example, a cell type-specific version of the SWI/SNF remodeller BAF was described for mouse ESCs, esBAF with Brg1 as its ATPase subunit (Ho et al., 2009b; Kadoch & Crabtree, 2015). Brg1 is required for early mouse development as its loss leads to lethality at the peri-implantation stage and impairs growth of the inner cell mass (Bultman et al., 2000; Alfert et al., 2019). In cultured mouse ESCs, loss of Brg1 impairs the self-renewal capacities and reduces expression of pluripotency transcription factors including Oct4, Sox2 and Nanog (Ho et al., 2009b; Kidder et al., 2009). Brg1 was further found to colocalise with these three pluripotency TFs, suggesting that esBAF represents a component of the pluripotency network (Kidder et al., 2009; Ho et al., 2009a). An additional remodeller, which was found to be essential for self-renewal and proliferation of mouse ESCs, is the INO80 complex (Wang et al., 2014). Depletion of INO80 was reported to lead to reduced expression of pluripotency TFs such as Oct4, Sox2 and Nanog and to strongly impaired blastocysts growth (Wang et al., 2014).

7.4.3. Role of histone acetyltransferase-containing complexes in mouse embryonic stem cells

Several subunits of the TIP60 complex were further identified in a siRNA-based screen as important factors for proliferation and self-renewal of mouse ESCs (Fazzio et al., 2008). In agreement, loss of the catalytic HAT subunit of the TIP60 complex and TRRAP caused embryonic lethality before implantation and cell death within blastocysts outgrowth (Herceg et al., 2001; Hu et al., 2009b).

Recently, a CRISPR-Cas9 loss-of-function genetic screen identified subunits of the SAGA complex as epigenetic regulators of the mouse ESC state (Seruggia et al., 2019). Two SAGA-specific subunits, Taf5l and Taf6l, were suggested to be required for self-renewal and proliferation of mouse ESCs by regulating c-Myc and Oct4 responsive genes (Seruggia et al., 2019). H3K9ac deposition and recruitment of c-Myc were proposed to be the molecular mechanisms by which Taf5l and Taf6l regulate gene expression in mouse ESCs (Seruggia et al., 2019). In contrast, the acetyltransferase subunit of SAGA Gcn5 was found to be especially required during differentiation of mouse ESCs in embryonic body formation assays and not for self-renewal of mouse ESCs, which is in agreement with its essentiality at post-implantation stages of mouse embryonic development (Xu et al., 2000; Yamauchi et al., 2000; Wang et al., 2018a). Although, Gcn5 was reported to colocalize at genomic elements with Myc and TFs of the E2f family involved in the regulation of cell cycle-related genes, *Gcn5*-null mouse ESCs did not

display defects in morphology or growth comparable to inactivation of either E2f4 or Myc (Lin et al., 2007; Smith et al., 2010; Varlakhanova et al., 2010; Hirsch et al., 2015; Scognamiglio et al., 2016; Hsu et al., 2019a). In contradiction with these findings however, chemical inhibition of Gcn5 suggested that loss of the enzymatic activity of Gcn5 affects self-renewal capacities and destabilizes the pluripotency networks of mouse ESCs (Moris et al., 2018).

Additionally, the histone acetyltransferase Mof (Kat8) was found to be required for blastocyst development and for mouse ESC self-renewal and proliferation (Gupta et al., 2008; Thomas et al., 2008; Li et al., 2012b). Mof is part of the MSL and NSL complexes and especially responsible for histone H4 lysine 16 acetylation (Smith et al., 2005; Li et al., 2009; Cai et al., 2010). Mof was further suggested to be a key component of the pluripotency network by acting upstream of Nanog especially through the MSL complex (Li et al., 2012b; Ravens et al., 2014).

7.4.4. Roles of chromatin modifying factors involved in heterochromatin formation in mouse embryonic stem cells

Loss of factors involved in the formation of heterochromatic domains are in general associated with milder effects on self-renewal and growth of mouse ESCs with the exception of the catalytic subunit of PRC1, required for H2AK119ub, and Setdb1, required for H3K9me2 and H3K9me3 (Surani et al., 2007; Festuccia et al., 2017a). For example, individual loss of several PRC2 subunits, the complex responsible for H3K27me3 at facultative heterochromatin, did not cause overt differentiation of mouse ESCs (Montgomery et al., 2005; Chamberlain et al., 2008; Shen et al., 2008). In contrast, loss of the ubiquitinylation enzymes of PRC1 was found to cause differentiation and impaired cell viability of mouse ESCs (Leeb & Wutz, 2007; Endoh et al., 2008; van der Stoop et al., 2008).

Inactivation of G9a, GLP and Suv39h1/ Suv39h2, representing methyltransferases acting on H3K9, did not impair survival of mouse ESCs, in line with post-implantation phenotypes in the respective mutant mouse embryos (Peters et al., 2001; Tachibana et al., 2002; Lehnertz et al., 2003; Tachibana et al., 2005; Festuccia et al., 2017a). Also, simultaneous impairment of several complexes involved in heterochromatin formation was reported to have little effect on the self-renewal capacities of mouse ESCs (Walter et al., 2016; Festuccia et al., 2017a). For example, triple inactivation of Suv39h1, Suv39h2 and a subunit of PRC2 leads to a global reduction of H3K27me3 (facultative heterochromatin) and H3K9me3 (constitutive heterochromatin) but ESCs were viable, although they displayed reduced proliferation rates (Walter et al., 2016).

In stark contrast, mouse ESCs cannot be generated from $Setdb1^{-/-}$ blastocysts (Dodge et al., 2004). Further, inducible depletion or knockdown of Setdb1 in mouse ESCs impedes viability and differentiation (Yuan et al., 2009; Matsui et al., 2010; Karimi et al., 2011). This extreme phenotype of loss of Setdb1 in contrast to the other H3K9 methyltransferases seems astonishing as Setdb1 is thought

to only di- and trimethylate H3K9 and to be dependent on the actions of G9a and GLP for initial monomethylation of H3K9, which might suggest methyltransferase-independent functions of Setdb1 in mouse ESCs (Bannister & Kouzarides, 2011; Janssen et al., 2018).

7.4.5. Confounding effects in the analysis of multisubunit protein complexes

Detailed comparison of phenotypes and effects between subunits of a given chromatin modifier is frequently problematic due to several reasons. For instance, loss of different subunits of a specific complex might have variable consequences on the structure and consequently on the function of this complex. Varying effects of loss of different subunits on multimeric complex organization were comprehensively described for the yeast SAGA complex or mammalian Mediator complex (Lee et al., 2011; El Khattabi et al., 2019). Additionally, as mentioned earlier, some proteins are shared among different chromatin modifying complexes such as the Trrap subunit shared between the SAGA and TIP60 complexes or Wdr5 shared between ATAC, NSL, COMPASS and COMPASS-like complexes. For example, depletion of Wdr5 was reported to affect self-renewal through reduced expression of pluripotency transcription factors including Nanog, Sox2 and Esrrb (Ang et al., 2011). Wdr5 was further found to colocalize with Oct4, which was suggesting to reflect overlapping functions between these two factors (Ang et al., 2011). Due to similarities with phenotypes observed upon loss of Setd1a, Wdr5 loss was suggested to reflect defects of the COMPASS complex (Ang et al., 2011; Fang et al., 2016). However, effects on the ATAC, NSL or COMPASS-like complexes cannot be ruled out (Bledau et al., 2014). Similarly, phenotypes caused by the loss of Trrap are thought to mainly reflect inactivation of the TIP60 complex as its catalytic subunit Tip60, but not Gcn5 show comparable effects (Lin et al., 2007; Fazzio et al., 2008).

As mouse ESCs can be maintained in various culture conditions affecting their properties, comparison of phenotypes between studies is often complicated (Festuccia et al., 2017a). For instance, mouse ESCs grown in LIF+2i medium display very different chromatin environments and dependencies on pluripotency transcription factors. As such, culture conditions with 2i enhance mouse ESC self-renewal capacities and lead to epigenetic changes such as the erasure of DNA methylation and loss of H3K9me3 and H3K27me2 (Walter et al., 2016; Festuccia et al., 2017a). How culturing conditions modify the phenotypes observed in mouse ESCs is best highlighted by the following recent observation. Self-renewal in LIF+2i medium seems to be partially independent of Nanog as demonstrated by the fact that *Nanog*-deficient ESCs can maintain self-renewal in 2i conditions (Chambers et al., 2007; Martello & Smith, 2014; Hastreiter et al., 2018). Curiously however, the addition of serum to the LIF+2i medium was found to dramatically impair this Nanog-independent self-renewal capacity (Hastreiter et al., 2018; Navarro, 2018).

Considering the above described difficulties, Table 5 presents a highly simplified summary of phenotypes, which were reported upon perturbation of chromatin modifying complexes involved in heterochromatin or euchromatin states in mouse ESCs.

Table 5: The importance of several chromatin modifying complexes in mouse embryonic stem cell self-renewal and survival. Simplified description of the effects of loss of functions of chromatin modifying complexes on self-renewal and cell survival with corresponding references. Shown are chromatin modifying complexes, involved in either gene repression and heterochromatin formation or active gene transcription. n.d., not determined. Based on Surani et al., 2007 and Festuccia et al. 2017a.

	Self-renewal	Cell survival	References
involved in gene repression and heterochromatin formation			
PRC2	unaffected	viable	Montgomery et al., 2005; Chamberlain et al., 2008; Shen et al., 2008
PRC1	impaired	impaired	Leeb & Wutz, 2007; van der Stoop et al., 2008; Endoh et al., 2008
G9a	n.d.	viable	Tachibana et al., 2002
GLP	n.d.	viable	Tachibana et al., 2005
Setdb1	impaired	not viable	Dodge et al., 2004; Yuan et al., 2009; Matsui et al., 2010; Karimi et al., 2011
Suv39h1/Suv39h2	n.d.	viable	Peters et al., 2001; Lehnertz et al., 2003
involved in active gene transcription and euchromatin			
COMPASS	impaired	not viable	Bledau et al., 2014; Fang et al., 2016
COMPASS-like	unaffected	impaired	Glaser et al., 2006; Glaser et al., 2009; Wang et al., 2016
esBAF	impaired	not viable	Bultman et al., 2000; Ho et al., 2009a; Ho et al., 2009b; Kidder et al., 2009
INO80	impaired	not viable	Wang et al., 2014
TIP60	impaired	not viable	Herceg et al., 2001; Fazzio et al., 2008; Hu et al., 2009b
SAGA	impaired	impaired	Seruggia et al., 2019
MSL/NSL	impaired	impaired	Li et al., 2012b; Ravens et al., 2014

7.4.6. Catalytic-independent functions of chromatin modifying complexes in mouse embryonic stem cells

As chromatin modifying complexes were identified based on their enzymatic subunits, they are in general assumed to impact transcription through their chromatin modifying activities. However, recent studies in mouse ESCs indicate that the chromatin modifying complexes TIP60 and Mll3/Mll4 COMPASS-like complexes function in manners independent of their histone modifying activities (Dorighi et al., 2017; Acharya et al., 2017; Rickels et al., 2017).

As described earlier, loss or depletion of the Tip60 acetyltransferase impairs mouse ESC self-renewal and proliferation, in agreement with pre-implantation defects observed for *Tip60*-null mouse embryos (Fazzio et al., 2008; Hu et al., 2009b). In contrast, amino acid substitutions destroying the catalytic site of Tip60, were recently found to have a minimal impact on ESC self-renewal and caused only post-implantation defects during mouse embryo development (Acharya et al., 2017). Similarly, depletion of Ep400, the ATPase remodelling subunit of the TIP60 complex, dramatically altered the self-renewal capacities of mouse ESCs, while catalytic dead Ep400 had minimal effects (Acharya et al., 2017). These drastic functional differences between depletion of the entire protein or catalytic
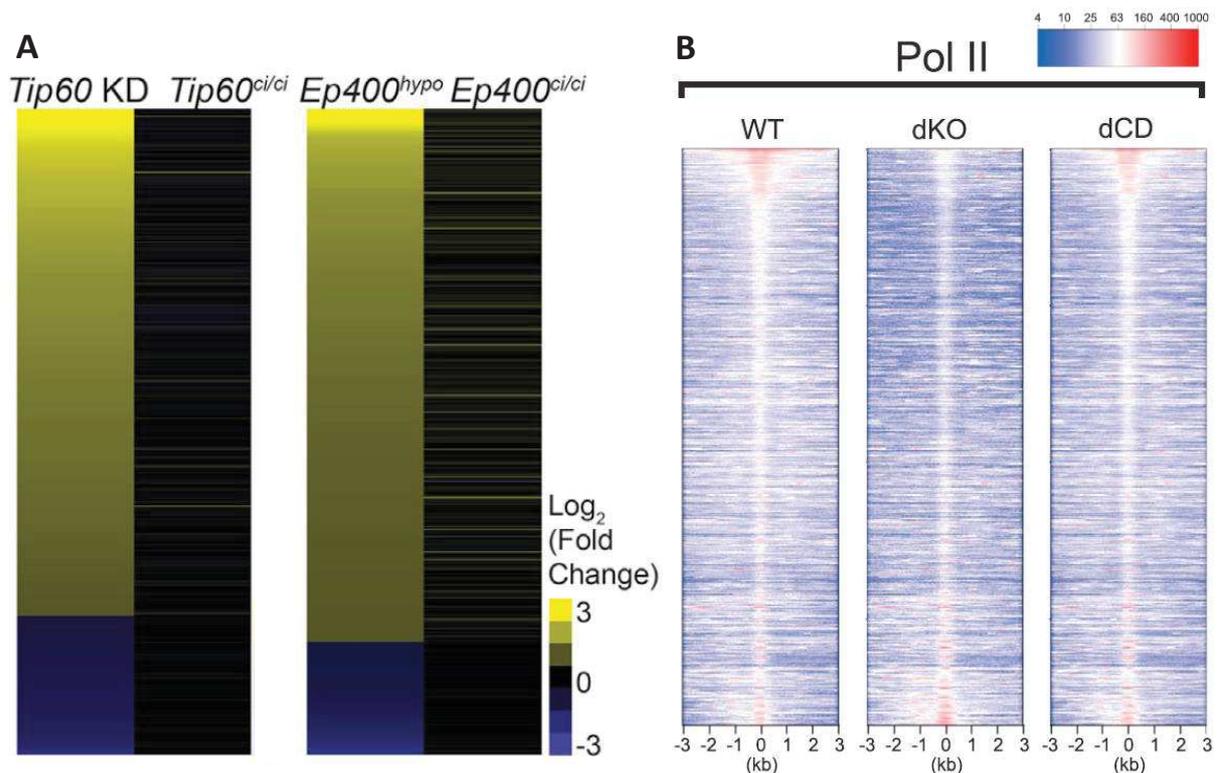
Figure 50: Catalytic-independent functions of the chromatin modifying complexes TIP60 and Mll3/Mll4 COMPASS-like complexes in mouse embryonic stem cells. A. Heatmap showing log2 fold changes of genes differentially expressed in mouse embryonic stem cells with either *Tip60* knock-down (*Tip60* KD), catalytically inactive Tip60 (*Tip60*[ci/ci]), hypomorphic Ep400 (*Ep400*[hypo]) or catalytically inactive Ep400 (*Ep400*[ci/ci]). Adapted from Acharya et al., 2017. **B.** Heatmap showing RNA polymerase II (Pol II) ChIP-seq signals at enhancers of wildtype (WT) mouse ESCs compared to cell lines with double knockout of *Mll3* and *Mll4* (dKO) or catalytic dead Mll3 and Mll4 (dCD). The heatmaps are sorted based on signal from Mll3/Mll4 ChIP-seq from top, high Mll3/Mll4 binding, to bottom, low Mll3/Mll4 binding. Adapted from Dorighi et al., 2017.

inactivation of Tip60 or Ep400 were further apparent on mRNA expression levels (Figure 50A) (Acharya et al., 2017). Depletion of Tip60 and Ep400 caused the significant misregulation of thousands or hundreds of genes, respectively, while loss of their catalytic activity caused only minor (25 genes) or no significant effects on mRNA levels, respectively. In contrast, analysis of the differentiation potential of mouse ESCs lacking the catalytic activity of Tip60 showed that its acetyltransferase activity seems important to ensure proper establishment of mesoderm and endoderm lineages (Acharya et al., 2017). Overall, these results led the authors to conclude that Tip60 acts in a catalytic-independent manner to ensure mouse ESC self-renewal and proliferation, while its acetyltransferase activity is required for proper differentiation (Acharya et al., 2017).

Likewise, combined loss of the catalytic activities of Mll3 and Mll4, enzymes of the COMPASS-like complexes responsible especially for H3K4me1 at enhancers, was found to cause much milder effects compared to combined inactivation of *Mll3* and *Mll4* (Figure 50B) (Dorighi et al., 2017). Combined catalytic inactivation by point mutations in the respective SET domains of Mll3 and Mll4 caused a major reduction of H3K4me1 levels comparable to that observed in *Mll3* and *Mll4* double knockout (KO) cell lines (Dorighi et al., 2017). However, RNA polymerase II (Pol II) recruitment and transcription in mouse ESCs with catalytic dead Mll3 and Mll4 were largely unchanged compared to wildtype cells, in stark contrast to double *Mll3* and *Mll4* KO cell lines (Figure 50B) (Dorighi et al., 2017). These findings led the authors to suggest that H3K4 methylation at enhancers has only limited importance for the functions of Mll3/Mll4-containing COMPASS-like complexes (Dorighi et al., 2017).

These results are largely in agreement with another study in which the entire catalytic SET domain from both, Mll3 and Mll4, was deleted (Rickels et al., 2017). Interestingly, alkaline phosphatase staining of mouse ESCs was generally unaffected upon deletion of the Mll3 and Mll4 SET domains, in contrast to previously reported double *Mll3* and *Mll4* KO cell lines (Wang et al., 2016; Rickels et al., 2017). Alkaline phosphatase staining is commonly used to assess self-renewal capacities of mouse ESCs, as the intensity of staining correlates with the level of pluripotency such that undifferentiated ESCs show the strongest staining. Although combined inactivation of Mll3 and Mll4 was found to cause loss of alkaline phosphatase staining, this was reported to be based on the reduced expression of *Alpl*, the gene encoding alkaline phosphatase, instead of reflecting reduced self-renewal capacities as neither proliferation nor expression levels of pluripotency factors were affected (Wang et al., 2016).

Finally, a recent report suggests that H3K27ac, a histone modification frequently associated with active enhancers, is also dispensable for Pol II transcription in mouse ESCs (Zhang et al., 2020). These findings together highlight the missing understanding in how chromatin modifications and their effectors regulate Pol II transcription.

Overall, mouse ESCs represent an unique cellular model as they are rapidly dividing and have the potential to self-renew endlessly in culture (Martello & Smith, 2014). Importantly, they have euploid karyotypes without gross genomic alterations and are permissive to genetic manipulations (Evans, 2011; Martello & Smith, 2014). Additionally, they can be differentiated in culture into various cell types facilitating the analysis of very diverse developmental processes (Martello & Smith, 2014).

Aims

Aims

Regulation of gene expression is an essential requirement to allow cells to properly maintain their cell identity and to respond to external cues, such as during differentiation and development. Coactivator complexes represent key components involved in regulating RNA polymerase II (Pol II) transcription by serving as connections between stimulus-induced, DNA-bound transcription factors (TFs) and the basal transcription machinery. Coactivators frequently possess chromatin modifying activities, which are thought to allow them to increase DNA accessibility and thereby facilitate the assembly of the transcription preinitiation complex (PIC) at gene promoters.

The importance and functions of these highly dynamic, multiprotein complexes and their chromatin modifying activities in Pol II transcription is however still poorly understood, as recently highlighted through two intriguing observations. First, the chromatin modifying functions of two transcriptional coactivator complexes were reported to not be their most crucial function in regulating Pol II transcription in mouse embryonic stem cells (ESCs) (Dorighi et al., 2017; Acharya et al., 2017; Rickels et al., 2017). Second, studies from my host laboratory indicated that the coactivator SAGA acts as a general cofactor for Pol II transcription required for the expression of all actively transcribed genes in budding yeast, unlike its previously assumed specific functions (Bonnet et al., 2014; Baptista et al., 2017).

SAGA represents an evolutionarily highly conserved coactivator, which contains two histone modifying enzymes, a histone H2B deubiquitylase (DUB) and a histone acetyltransferase (HAT), and possesses a TBP-loading function. Several subunits of SAGA are however shared with other complexes (Helmlinger & Tora, 2017). In metazoans, a HAT module with the enzymatic subunits Gcn5 or Pcaf is found in both, SAGA and ATAC. Also, two subunits of the DUB module of SAGA are shared with SAGA-independent DUBs and its TRRAP subunit, which interacts with several TFs, is shared with the TIP60 complex.

Based on this prior knowledge, the aims of my PhD thesis were to…

a. …explore the role of the SAGA coactivator for global Pol II transcription in mammalian cells.
b. …reveal if, due to their shared HAT module, the SAGA and ATAC coactivators have redundant functions in ensuring Pol II transcription.
c. …uncover the importance of the HAT activities of SAGA and ATAC in Pol II transcription.
d. …explore the role of the TBP-loading function of SAGA in Pol II transcription.
e. …understand the importance of functions of SAGA and ATAC for mouse ESC proliferation and self-renewal.
f. …reveal the importance of the TRRAP subunit, shared by the SAGA and TIP60 coactivators, for Pol II transcription in human cells.

Results

Results

1. HAT-independent functions of the SAGA and ATAC coactivator complexes in RNA polymerase II transcription.

Regulation of RNA polymerase II (Pol II) transcription is generally thought to involve the specific recruitment of transcription coactivator complexes to gene promoters through the recognition of gene-specific, DNA-binding transcription factors. Coactivator complexes consequently are believed to facilitate formation of the preinitiation complex (PIC) at gene promoters by allowing access to the chromatinized promoter DNA. In general, coactivators are thought to enable access to DNA through modifying histone tails or by remodelling nucleosomes. In agreement with recruitment through specific DNA-recognizing transcription factors, coactivators were frequently found to influence only a limited set of genes in yeast and mammalian cells.

Recent findings however revealed that the SAGA coactivator acts on all actively transcribed genes in yeast, in contrast to its previously assumed gene-specific functions (Bonnet et al., 2014; Baptista et al., 2017; Donczew et al., 2020). The usage of improved technics, such as newly synthesized RNA analysis, allowed to reveal that SAGA binds at the vast majority of active genes, modifies histone residues at most genes and acts as a general cofactor for Pol II transcription in yeast. As SAGA is evolutionary conserved from yeast to humans, the question arose if SAGA acts as a general cofactor for Pol II transcription also in mammalian cells.

SAGA possess two histone modifying activities, a histone acetyltransferase (HAT) and a deubiquitylase (DUB), and is involved in depositing TBP, the factor nucleating PIC formation, at gene promoters. In mammalian cells, subunits of the HAT module of SAGA are shared with another coactivator complex, ATAC. Besides subunits of the HAT module, SAGA and ATAC are composed of distinct subunits. Although little is known about ATAC, its HAT function is generally believed to be the only activity with which it regulates Pol II transcription.

Interestingly, depletion of a shared subunit of the SAGA and ATAC HAT modules was found to lead to genome-wide reduction of acetylation of H3K9 in human cells (Bonnet et al., 2014). This finding suggested that SAGA and ATAC might act on all actively transcribed genes in mammalian cells. In contrast, other studies in human cells indicated that SAGA and ATAC act on specific genes and only overlap at few genes (Nagy et al., 2010; Krebs et al., 2011).

Based on these earlier findings and to address **aim a)** and **aim b)** about understanding the roles of SAGA and ATAC for Pol II transcription in mammalian cells including potential redundant functions, we individually inactivated subunits of SAGA and ATAC, which are thought to be crucial for complex integrity, in mouse ESCs using CRISPR-Cas9. To more specifically assess potential redundant functions of SAGA and ATAC and to reveal the importance of functions of their shared HAT modules (**aim b)**

and **aim c))**, we generated double mutant cell lines in which both, an ATAC and a SAGA core subunit, were targeted for inactivation as well as cell lines with inactivation of genes encoding subunits of the HAT modules. We especially focused on the functions of the shared HAT module, as this shared activity might explain potential overlapping or redundant functions of SAGA and ATAC. To identify in an unbiased way the genes affected upon inactivation of SAGA or ATAC and to identify potential overlapping function of SAGA and ATAC, we assessed newly synthesized RNA levels on a genome-wide level in the above described mutant cell lines.

Further, to answer **aim e)** on the importance and functions of SAGA and ATAC for self-renewal and proliferation of mouse ESCs, we assessed our mutant mouse ESC lines for their growth potential and self-renewal capacities.

Among the different results obtained, we found that SAGA and ATAC are required for self-renewal and proliferation of mouse ESCs. Also, we found that SAGA and ATAC seem to significantly affect distinct Pol II-transcribed genes. We could observe however a mild but general downward shift of Pol II transcription upon inactivation of structural subunits of SAGA and ATAC, which might suggest a genome-wide function of these two coactivator complexes. We also found that the impact on mouse ESC physiology and Pol II transcription was independent of the HAT functions of SAGA and ATAC, suggesting that SAGA and ATAC possess other important functions.

Overall, we found that SAGA and ATAC are involved in maintaining the self-renewal capacities of mouse ESCs by regulating distinct genes in a way which is mainly independent of their shared HAT modules. The finding of a largely HAT-independent function of SAGA and ATAC in Pol II transcription and mouse ESC physiology is in agreement with recent reports suggesting that the chromatin modifying function of two other coactivator complexes may not be their most crucial function (Dorighi et al., 2017; Acharya et al., 2017; Rickels et al., 2017).

The following manuscript represents a preliminary manuscript of these results.

Two related coactivator complexes SAGA and ATAC control embryonic stem cell self-renewal through two different acetyltransferase-independent mechanisms

Veronique Fischer[1,2,3,4], Damien Plassard[1,2,3,4,5], Bernardo Reina San Martin[1,2,3,4], Matthieu Stierle[1,2,3,4], Laszlo Tora[1,2,3,4], Didier Devys[1,2,3,4]*

1 Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch CEDEX, France

2 Centre National de la Recherche Scientifique, UMR7104, 67404 Illkirch CEDEX, France

3 Institut National de la Santé et de la Recherche Médicale, U964, 67404 Illkirch CEDEX, France

4 Université de Strasbourg, 67000 Strasbourg, France

5 Plateforme GenomEast, infrastructure France Génomique

* Corresponding author : devys@igbmc.fr

Author list footnotes
Lead author contact
Didier Devys
1 Rue Laurent Fries,
67404 Illkirch CEDEX
France
E-mail: devys@igbmc.fr

Summary

Acetylation of chromatin at actively transcribed genes is thought to stimulate transcription. The requirement of the related histone acetyltransferase (HAT) complexes SAGA and ATAC for RNA polymerase II transcription is however not well understood. We generated a series of mouse embryonic stem cell (ESC) lines in which SAGA or ATAC subunits were inactivated and/or depleted. We show that the SAGA subunit Supt7l or the ATAC subunits Yeats2 and Zzz3 are required for complex assembly, cell growth, and mouse ESC self-renewal. Additionally, ATAC, but not SAGA subunits are required for ESC viability by regulating the transcription of translation-related genes. Surprisingly, depletion of a shared or specific HAT module subunits caused a global decrease in histone H3K9 acetylation, but did not result in significant phenotypic or transcriptional defects. Thus, our results indicate that SAGA and ATAC are differentially required for viability and self-renewal of mouse ESCs by regulating transcription through different pathways, but in a HAT-independent manner.

Keywords

Mouse embryonic stem cells, coactivator complexes, SAGA, ATAC, Pol II transcription, 4sU labelling, newly synthesized RNA, histone acetyltransferase, HAT-independent function.

Introduction

Mouse embryonic stem cells (ESCs) are derived from the inner cell mass of blastocysts and present unique cellular characteristics (Evans and Kaufman, 1981; Martello and Smith, 2014; Martin, 1981). They can self-renew seemingly endlessly and, based on their pluripotency, can differentiate into most cell lineages. Mouse ESCs further possess high proliferation rates with a shortened G1 cell cycle phase, compared to somatic cells (Festuccia et al., 2017). The molecular mechanism underlying the self-renewal and pluripotency capacities of mouse ESCs are dpendent on key pluripotency transcription factors (TFs). This includes core pluripotency TFs such as Oct4 (encoded by *Pou5f1*) and Sox2 and other naïve-specific pluripotency TFs, which are more responsive to environmental cues, such as Nanog, Tfcp2l1, Klf4 and Esrrb (Martello and Smith, 2014; Young, 2011). Several recent findings indicate that the physiology of mouse ESCs is also dependent on several histone modifying complexes (Acharya et al., 2017; Festuccia et al., 2017; Li et al., 2012; Seruggia et al., 2019; Young, 2011).

Transcriptional coactivator complexes are involved in regulating RNA polymerase II (Pol II) transcription *in vivo* by modulating the chromatin environment and thereby enabling the access of the transcription machinery to the template DNA (Kouzarides, 2007; Li et al., 2007; Young, 2011). Coactivators contain enzymatic activities to either deposit/remove post-translational modifications of histone proteins, or to mobilize core nucleosomes through ATP-dependent remodelling functions. Through their actions on chromatin, coactivators are thought to be crucial for the assembly at gene promoters of the preinitiation complex, composed of Pol II and six general transcription factors. Preinitiation complex formation is nucleated by the binding of the TFIID complex, containing the TATA-box binding protein (TBP) and TBP-associated factors (TAFs) (Sainsbury et al., 2012; Thomas and Chiang, 2006).

The transcriptional coactivator complex SAGA (Spt-Ada-Gcn5 acetyltransferase) is highly conserved throughout evolution and contains two histone modifying enzymes, a histone acetyltransferase (HAT) and a histone H2B deubiquitylase (DUB) (Grant et al., 1997; Helmlinger and Tora, 2017; Henry et al., 2003; Weake and Workman, 2012). Recent cryo-electron microscopy studies revealed that the yeast SAGA complex is organized in four structural and functional modules (Helmlinger et al., 2020; Papai et al., 2020; Wang et al., 2020). The core module of SAGA, located at the center of the complex contains a histone octamer-like structure composed by the assembly of histone fold (HF) domain heterodimers. Seven SAGA subunits form four HF heterodimers, namely Taf6-Taf9, Taf10-Spt7, Taf12-Ada1 and two HF domains of Spt3. The octamer-like structure, located at the periphery of the core module was shown to bind TBP and was proposed to be important for

TBP delivery at promoters (Papai et al., 2020). The rest of the core module, made of Spt20, Taf5 and other domains of HF-containing subunits, connects the octamer-like with the two enzymatic modules and the activator-binding module composed of the single large Tra1 subunit. In *Saccharomyces cerevisiae*, the SAGA complex cannot assemble in the absence of the core subunits Spt20, Spt7 or Ada1, suggesting that most SAGA functions are lost when these critical structural subunits are missing (Han et al., 2014; Lee et al., 2011; Sterner et al., 1999; Wu and Winston, 2002). The role of other core subunits in the assembly of SAGA could not be studied as Taf5, Taf6, Taf9, Taf10 and Taf12 are also components of the essential TFIID complex. In metazoan, Taf5l and Taf6l, two paralogs of Taf5 and Taf6, are exclusive components of SAGA, whereas Taf5 and Taf6 are specifically incorporated within TFIID. Recently, Taf5l and Taf6l were identified as epigenetic regulators required for maintenance of mouse ESC state, although their role on SAGA complex integrity has not been determined (Seruggia et al., 2019). In mammals, two paralogous acetyltransferases Gcn5 (Kat2a) or Pcaf (Kat2b) associate in a mutually exclusive manner with three adaptor proteins, Sgf29, Tada3 and Tada2b to form the SAGA HAT module (Nagy et al., 2009; Yang et al., 1996). Interaction with the three adaptor proteins was shown to increase Gcn5 activity and to define its substrate specificity (Balasubramanian et al., 2002).

The functional roles of SAGA in Pol II transcription have been extensively studied with recent findings indicating that SAGA acts as a general cofactor for Pol II transcription in yeast (Baptista et al., 2017; Bonnet et al., 2014; Donczew et al., 2020). This major importance of SAGA for Pol II transcription in yeast was only apparent when analyzing newly synthesized RNA levels, while total RNA levels showed few changes. In mammalian cells, subunits of SAGA have been linked to the regulation of transcription induction following specific stress responses and signaling pathways (Nagy et al., 2009; Spedale et al., 2012; Wang and Dent, 2014). Further, mutation of complex-specific subunits of SAGA in mouse embryos were reported to cause lethality after the gastrulation stage (Perez-Garcia et al., 2018; Zohn et al., 2006).

In metazoans, three subunits of the HAT module of SAGA (Gcn5 or Pcaf, Tada3 and Sgf29), are shared with another coactivator complex, the ATAC (Ada-Two-A-Containing) complex. The shared subunits and Tada2a form the HAT module of ATAC, and mammalian ATAC possesses six additional subunits (Yeats2, Zzz3, Atac2, Mbip, Wdr5 and Nc2β) (Helmlinger and Tora, 2017; Spedale et al., 2012). Although the subunit composition of ATAC is defined and well conserved in metazoans, little is known about its structural organization. Similarly to SAGA, ATAC has been implicated in the regulation of the transcription response following certain stresses or signaling pathways in mammalian cells (Nagy et al., 2009; Spedale et al., 2012). However, recent genome-wide studies in human cells imply ATAC in the transcription regulation of genes involved in house-keeping functions,

such as ribosome protein coding genes (RPGs) (Mi et al., 2017, 2018). In mouse embryos, inactivation of the Atac2 subunit of ATAC was reported to lead to lethality around the gastrulation stage (Guelman et al., 2009). Loss of the catalytic HAT subunits, Gcn5 and Pcaf, shared by SAGA and ATAC were found to cause embryonic lethality in mice also around the gastrulation stage (Bu et al., 2007; Xu et al., 2000; Yamauchi et al., 2000).

In this study, we aimed to address the functional importance of the coactivators SAGA and ATAC and their shared HAT activity in mouse ESC physiology and Pol II transcription regulation. A major hurdle to studying these two coactivator complexes is their multiprotein composition and enzymatic redundancies. We therefore individually targeted several complex-specific or shared subunits of SAGA and/or ATAC using CRISPR-Cas9 mediated genome editing. These cell lines allowed us to reveal that, while loss of core subunits of SAGA and ATAC severely affected mouse ESC growth and self-renewal capacities, loss of HAT subunits did not cause apparent phenotypes. By analyzing genome-wide newly synthesized RNA levels in these cell lines, we found that loss of SAGA and ATAC subunits mainly affected different sets of genes. Thus, our data suggest that SAGA and ATAC play important, but distinct roles in the maintenance of mouse ESC self-renewal capacities mainly through non-redundant, HAT-independent functions.

Results

The SAGA core is required for ESC growth but is not essential for ESC survival

To understand the role of HAT-containing coactivator complexes in a relevant physiological model, we first inactivated genes encoding either Supt7l or Supt20h, subunits of the central core module of the SAGA complex, using CRISPR-Cas9 genome editing in mouse ESCs (Figure 1A and S1A). We obtained several lines with a homozygous deletion in either *Supt7l* or *Supt20h* (Figure 1B). *Supt7l* or *Supt20h* mRNA analyses in the respective ESC lines, confirmed the deletion of an out-of-frame exon and the resulting degradation of mRNA containing a premature stop codon (Figure S1B). Western blot analysis for Supt7l, for which a specific antibody was available, revealed undetectable levels of Supt7l (Figure 1C). These results indicate that *Supt7l* and *Supt20h* are not essential for mouse ESC survival, when cultured in medium containing foetal calf serum (FCS), leukaemia inhibitory factor (LIF) and two inhibitors of MEK/ERK and GSK3b pathways (hereafter referred to as FCS+LIF+2i medium).

To assess the importance of *Supt7l* and *Supt20h* for ESC growth and self-renewal, we performed clonal assays in FCS+LIF medium and observed that the loss of Supt7l had a major impact on colony formation of mouse ESCs (Figure 1D). When clonal assays were performed in FCS+LIF+2i medium, known to promote self-renewal and to block differentiation, we observed that the *Supt7l*[-/-] line

produced significantly smaller colonies (Figure S1C and S1D). In contrast, inactivation of *Supt20h* did not affect ESC growth which appeared similar to that of wild-type cells in both media. To rule out off-target or clonal effects of *Supt7l* deletion during cell growth, *Supt7l* coding sequence was reintroduced in *Supt7l*$^{-/-}$ ESCs (*Supt7l*^{tg}), resulting in the synthesis of *Supt7l* mRNA and encoded protein at levels comparable to that of the endogenous wild-type protein (Figure 1C and S1E). *Supt7l* transgene expression fully rescued the growth phenotype observed in a *Supt7l* null background, which confirmed that the growth defect is caused by the loss of Supt7l (Figure 1D, S1C and S1D). These results together show that Supt7l is required for ESC growth, while Supt20h is not (Figure S1F).

Supt7l is required to preserve the integrity of the SAGA core, but Supt20h is not

The growth defect observed in *Supt7l*$^{-/-}$, but not *Supt20h*$^{-/-}$, suggested that the loss of Supt7l in mouse ESCs may have a more severe impact on SAGA structure than the loss of Supt20h. To verify this hypothesis, we developed a double immuno-precipitation strategy to purify SAGA complexes from wild-type, *Supt7l*$^{-/-}$, *Supt20h*$^{-/-}$ or *Supt7l*^{tg} cells and analysed their composition by mass spectrometry or western blotting. First a Taf7 immunoprecipitation was performed to clear nuclear extracts from TFIID, which shares with SAGA Taf10 and Taf12 proteins. Then the SAGA complex was immuno-purified using either anti-Taf10 or anti-Taf12 antibodies. Mass spectrometry analysis of complexes purified from *Supt7l*$^{-/-}$ cells did not retrieve any SAGA subunits, with the exception of remaining TFIID subunits (Taf9, Taf10 and Taf12), suggesting a profound disorganization of SAGA in the absence of Supt7l (Figure 1E, S1G and S1H). Using the same purification scheme in *Supt7l*^{tg} cells, SAGA subunits were detected at levels similar to that observed in wild-type cells (Figure 1E and S1G). In contrast, loss of Supt20h in ESCs was less deleterious, as all subunits of the SAGA core module were found in the Taf10 immuno-purified fractions, in amounts comparable to that of fractions purified from control wild-type cells (Figure 1E and S1G).

These results demonstrate that, in good agreement with structural studies (Papai et al., 2020; Wang et al., 2020), Supt7l has a major role for preserving SAGA integrity. Consequently, the loss of Supt7l in ESCs induced a growth phenotype suggesting that the SAGA core is required for cell growth. Importantly, expression of *Supt7l* transgene in the *Supt7l* null background (*Supt7l*^{tg}) rescued cell proliferation defects and SAGA structure integrity, indicating a direct role for SAGA in mouse ESC physiology. In contrast, in the absence of Supt20h, partial SAGA complexes containing most SAGA core subunits can still assemble and maintain cell proliferation.

SAGA core is required for mouse ESC self-renewal

Using the above-characterized ESC lines, we further investigated the impact of SAGA on ESC self-renewal. Clonal assays in FCS+LIF medium coupled with alkaline phosphatase (AP) staining allow to

distinguish undifferentiated colonies having high levels of AP (AP positive), from differentiated colonies which remain unstained (AP negative). By counting the number of AP positive and AP negative colonies relative to the total amount of colonies, we could determine the proportion of undifferentiated cells as a proxy for ESC self-renewal. These analyses indicated a major impact of *Supt7l* inactivation, but not that of *Supt20h* on the maintenance of ESC self-renewal, with a dramatic reduction of the number of AP positive undifferentiated colonies in the *Supt7l*[-/-] lines when compared to *Supt20h*[-/-], or control wild-type cells (Figure 1F). Re-expression of *Supt7l* in *Supt7l*[tg] cells fully restored ESC self-renewal (Figure 1F). This observation together with the above structural results further suggest that a fully assembled SAGA core is required for ESC self-renewal.

We further quantified the steady-state mRNA levels of several pluripotency factors in *Supt7l*[-/-], *Supt20h*[-/-] and wild-type cells. In *Supt7l*[-/-] cells, we observed reduced mRNA levels for four out of the six analysed pluripotency factors, namely *Tfcp2l1*, *Nanog*, *Esrrb* and *Klf4*, while *Pou5f1* (encoding for Oct4*)* and *Sox2* mRNA levels were unaffected (Figure 1G). In agreement with preserved self-renewal in *Supt20h* cells, the mRNA levels for all tested pluripotency factors were not significantly changed when compared to that in wild-type cells (Figure 1G). Thus, our clonal assays and pluripotency TF expression analyses concordantly support a reduction of undifferentiated, pluripotent ESCs upon loss of Supt7l, but not following Supt20h loss. These differences between *Supt7l*[-/-] and *Supt20h*[-/-] cell lines further suggest an important role for the octamer-like structure of SAGA in mouse ESC self-renewal.

Core ATAC subunits Yeats2 and Zzz3 are crucial for the structural integrity of ATAC and are essential for mouse ESC survival

Although the coactivator ATAC was shown to be important for fly and mouse development, the structure and the functions of this complex are much less defined than those of SAGA and its potential functions in mouse ESCs has not been examined. To investigate the role of ATAC in mouse ESC physiology, we targeted the *Yeats2* and *Zzz3* genes encoding two distinct ATAC subunits (Figure 1A), which were shown to be important for ATAC-mediated histone acetylation and gene expression (Mi et al., 2017, 2018). By using the above-described gene editing strategy (Figure S2A), we obtained heterozygous clones in the same proportion as for SAGA subunits, indicating that the CRISPR-Cas9 worked with comparable efficiency at the different targeted genomic loci (Figure 1A and 2A). However, we could not isolate any homozygous mutant clone, suggesting that *Yeats2* or *Zzz3* have essential functions for mouse ESC survival and that these functions cannot be compensated by SAGA or other coactivator complexes.

To be able to further investigate the functions of ATAC in mouse ESCs, we employed the auxin-inducible degron (AID) system to generate cell lines in which Yeats2 and Zzz3 proteins can be

depleted in an inducible manner. Upon addition of the plant-specific hormone auxin (IAA), the exogenously expressed Tir1 protein recognizes the AID peptide sequence fused to the protein of interest, leading to polyubiquitination and degradation of the fusion protein (Natsume et al., 2016). We thus first created an ESC line expressing the Tir1 and BirA proteins and then by using CRISPR-Cas9, we inserted the AID and biotin-tag sequences on both alleles of the *Yeats2* (*Yeats2*[AID/AID]) or *Zzz3* (*Zzz3*[AID/AID]) genes. The encoded fusion proteins had the expected size and were rapidly degraded. AID-Yeats2 and AID-Zzz3 protein levels were highly reduced from 4 hours of auxin treatment (Figure S2B) and were undetectable after 24 hours of treatment (Figure 2B and 2C).

Using our inducible cell lines, we assessed how the loss of these two ATAC-specific subunits affect the complex integrity by performing immuno-purification experiments using an antibody targeting Atac2 (also called Csrp2bp, or Kat14), an ATAC-specific subunit. Western blot analyses using antibodies that recognize mouse ATAC subunits, showed that Atac2 co-purified with Tada3, Wdr5 or Sgf29 from wild-type cells and from *Yeats2*[AID/AID] or *Zzz3*[AID/AID] cell lines in the absence of auxin treatment, indicating that the fusion of the AID sequence to Yeats2 or Zzz3 did not affect the complex assembly (Figure 2D). In contrast, upon auxin-induced depletion of AID-Yeats2 or AID-Zzz3 for 24 hours, the tested subunits could no longer be detected in the Atac2-containing complexes, showing that the loss of Yeats2 or Zzz3 caused a major disorganization of ATAC (Figure 2D).

To further evaluate ATAC complex integrity upon loss of Yeats2 or Zzz3, we generated ESCs expressing endogenously HA-tagged Tada2a in *Yeats2*[AID/AID] or *Zzz3*[AID/AID] backgrounds. Upon depletion of AID-Yeats2 or AID-Zzz3 in these cells, the levels of Tada2a-HA were found decreased in whole cell extracts (Figure S2C). This observation suggests a degradation of Tada2a as a consequence of ATAC complex disassembly and thus, jeopardized our anti-HA immune-purification strategy. Our data together suggest that Yeats2 and Zzz3 are crucial for the structural integrity of the ATAC complex and are required for ESC survival.

ATAC core subunits are required for mouse ESC maintenance

To assess the impact of ATAC inactivation on mouse ESC growth and self-renewal, we performed clonal assays with or without auxin in the *Yeats2*[AID/AID], *Zzz3*[AID/AID] and control cell lines cultured either in FCS+LIF or FCS+LIF+2i medium. In the absence of auxin, the size of colonies was comparable between the two AID cell lines and control cells. Upon continuous auxin treatment for six days, the size of *Yeats2*[AID/AID] or *Zzz3*[AID/AID] colonies were strikingly reduced, when compared to control cells (Figure 2E, S2D and S2E). In addition, growth curve analyses based on the number of viable cells cultured in FCS+LIF+2i further indicated reduced proliferation rates of *Yeats2*[AID/AID] or *Zzz3*[AID/AID] cells treated with auxin, in agreement with the smaller colony area measured in these lines (Figure S2E

and S2F). These results indicate that the growth of mouse ESCs is seriously affected upon depletion of ATAC core subunits.

AP staining on *Yeats2^(AID/AID)* or *Zzz3^(AID/AID)* cells cultured in FCS+LIF and treated with auxin for 6 days, evidenced a reduced fraction of undifferentiated AP positive colonies and an increase in differentiated AP negative colonies, when compared to control cells (Figure 2F). In the absence of auxin, *Yeats2^(AID/AID)* and *Zzz3^(AID/AID)* cells were comparable to wild-type cells. In agreement with impaired self-renewal, mRNA levels of all tested pluripotency factors progressively declined upon depletion of AID-Yeats2 (Figure 2G). These results together indicate that the self-renewing capacities of mouse ESCs require ATAC functions.

Earlier studies suggested a role for ATAC subunits in cell cycle regulation in mammalian cells (Fournier et al., 2016; Guelman et al., 2009; Orpinell et al., 2010). To determine whether the depletion of ATAC subunits affects cell cycle transition, *Yeats2*^(AID/AID), *Zzz3*^(AID/AID) or control cells grown in FCS+LIF+2i medium were treated with auxin and stained with propidium iodide for cell cycle analyses. Following 24 hours of auxin treatment of *Yeats2*^(AID/AID) or *Zzz3*^(AID/AID) cells, cell cycle distribution was not significantly affected (Figure S2G). In contrast, after 48 hours of auxin treatment, an increase in cells in G1 could be observed in the ATAC mutant cell lines relative to control cells (Figure S2H). In agreement with the reduced self-renewal observed following inactivation of the ATAC complex, these changes in cell cycle distribution could reflect an increased occurrence of cell differentiation.

Newly synthesized RNA quantification reveals non-overlapping roles for SAGA and ATAC in Pol II transcription

As the loss of SAGA core or the depletion of ATAC core subunits altered mouse ESC growth and maintenance, we next asked whether these two coactivators, sharing a similar HAT module, could regulate a common set of genes. To determine the role of SAGA and ATAC in Pol II transcription, we used 4-thiouridine (4sU) labelling of newly synthesized RNA coupled with sequencing of the labelled RNA (4sU-seq) (Rädle et al., 2013; Schwalb et al., 2016). Two independent clones of *Yeats2^(AID/AID)*, *Zzz3^(AID/AID)* and *Supt7l^(-/-)* lines as well as control cells in FCS+LIF+2i were treated for 24 hours with auxin prior to 4sU labelling, followed by purification and quantification of the labelled RNA by sequencing. By using 25 minutes 4sU labelling, we observed throughout all samples a very similar enrichment in intronic reads, when compared to total steady-state RNA-seq, indicating that the method allows the enrichment of unspliced transcripts in a highly reproducible manner (Figure S3A). The efficient purification of newly synthesized RNA was further evidenced when comparing 4sU-seq profiles with that of total RNA-seq at individual representative genes. 4sU-seq profiles revealed the presence of

mainly unprocessed transcripts with a high density of reads in introns and downstream of the polyadenylation signal, as well as unstable transcripts such as upstream antisense RNAs (Figure 3A and S3B).

In *Supt7l* null cells, the newly synthesized mRNA levels of 677 genes (out of 8208 expressed protein coding genes) were significantly decreased, when applying a threshold of -0.5 log2 fold change and an adjusted *p* value below 0.05 (Figure 3B). For many other genes the mRNA levels were only slightly decreased or did not reach statistical significance (Figure 3B). In contrast, only two genes were found upregulated in *Supt7l*[-/-] cells. Similarly, upon depletion of the ATAC subunits, Yeats2 or Zzz3, we observed a significant decrease in newly synthesized mRNA levels for a large number of genes (3391 and 1328, respectively), whereas very few genes were found upregulated when a 0.5 log2 fold change threshold was used (Figure 3B). Observation of these MA plots clearly indicated a slight downward shift for a majority of genes suggesting that ATAC may have global effects on Pol II transcription, although for most genes it did not reach statistical significance (Figure 3B). Importantly, the comparison of gene expression changes induced by Yeats2 or Zzz3 depletion revealed a strong correlation between the two datasets (Pearson correlation coefficient = 0.74), with about 87 % of genes significantly downregulated by Zzz3 depletion being also significantly affected upon Yeats2 depletion (Figure 3C and 3D). In contrast, no correlation was observed between changes in mRNA levels in the SAGA mutant line (*Supt7l*[-/-]) and those in ATAC mutant lines (*Yeats2*[AID/AID] or *Zzz3*[AID/AID]) (Figure 3C). When comparing genes significantly downregulated in these different cell lines, we could find only little overlap between genes affected upon inactivation of Supt7l and those downregulated by the loss of ATAC subunits (Figure 3D). As an example, at the *Sf3b3* gene shown in Figure 3A, the number of reads was reduced to a similar extent in *Yeats2*[AID/AID] or *Zzz3*[AID/AID] cell lines, but was only slightly affected in *Supt7l*[-/-] cells without reaching statistical significance.

These data together suggest that, contrary to our hypothesis, SAGA and ATAC are of particular importance for the expression of different sets of genes suggesting that their effects on mouse ESC self-renewal may be caused by different mechanisms. In addition, our observations suggest that these two HAT-containing coactivator complexes have a broad, but moderate effect on Pol II transcription.

ATAC regulates the expression of translation-related genes

To better understand how SAGA and ATAC influence differently the proliferation and self-renewing capacities of mouse ESCs, we further asked which gene categories are affected by the inactivation of SAGA or ATAC subunits. By using Gene Set Enrichment Analysis (GSEA), we searched for gene ontology (GO) biological processes which are enriched in our newly synthesized RNA datasets. We

identified 'cytoplasmic translation' as a GO term which was significantly downregulated upon depletion of either AID-Yeats2 or AID-Zzz3 (Figure 4A). In contrast, 'cytoplasmic translation' was not revealed by GSEA analysis in *Supt7l*[-/-] cells, but rather genes involved in the 'response to LIF' were enriched, which may be related to the phenotype observed in these SAGA mutant cells. The specific effect of two different ATAC subunits on translation-related genes was further confirmed by the visualization of newly synthesized mRNA levels of ribosome protein genes (RPGs) in the different mutant cell lines. These analyses revealed that many, but not all RPGs were highly and significantly downregulated in either *Yeats2*[AID/AID] or *Zzz3*[AID/AID] cell lines upon auxin addition, but not in *Supt7l*[-/-] cells, further highlighting the divergent effects of ATAC or SAGA inactivation on Pol II transcription (Figure 4B and 4C). We also found that genes of the 'response to LIF' GO category were downregulated to a similar extend between *Supt7l*[-/-], *Yeats2*[AID/AID] and *Zzz3*[AID/AID] cells, suggesting that inactivation of ATAC not only impacts translation but also affects the pluripotency network (Figure S3C). Thus, the two related HAT-containing coactivator complexes, SAGA or ATAC, appear to be particularly important for the expression of different, but also common groups of genes. Most notably, ATAC but not SAGA has a crucial function for the expression of RPGs and genes involved in translation.

We then asked whether within the groups of genes that we found to be regulated by SAGA or ATAC, the co-binding of some specific TFs could explain the differential recruitment/function of the two complexes to/at their target genes. To search for such putative TFs involved in SAGA or ATAC recruitment, we analysed the significantly downregulated genes identified in the respective mutant cell lines for overlaps with ChEA and Encode ChIP datasets using the Enrichr database. TFs enriched in genes regulated by the SAGA subunit Supt7l include several pluripotency TFs such as Oct4 (encoded by *Pou5f1*), Nanog and Sall4, in agreement with our GO analyses (Figure 4A and S3D). Among the identified ATAC-regulated genes, we observed an enrichment for Gcn5 (Kat2a) bound genes, thereby validating this approach (Figure S3D). Myc and the E2F family member E2F4 were also identified as TFs bound to ATAC-regulated genes, and are thus potentially recruiting ATAC to regulate the expression of these genes (Figure S3D). Both Myc and E2F4 are involved in cell cycle regulation, which may partially explain the cell cycle defects observed upon inactivation of ATAC in mouse ESCs (Fagnocchi and Zippo, 2017; Hsu et al., 2019; Scognamiglio et al., 2016). Thus, the preferential dependence of genes on SAGA, or ATAC inactivation appears to depend on the recruited TFs eventually leading to higher recruitment of either SAGA or ATAC.

Inactivation of the HAT modules of SAGA and ATAC do not impact mouse ESC proliferation or self-renewal

Loss of the SAGA core subunit Supt7l, or depletion of ATAC core subunits altered ESC growth and self-renewal, but primarily affected the expression of different groups of genes, raising the question of the role of their shared HAT subunits in both gene expression and ESC properties. To determine whether the observed phenotypes are caused by the loss of their HAT activities, we generated mouse ESCs lacking either Tada2a, or Tada2b, specific HAT subunits of ATAC or SAGA, respectively (Figure 1A, 5A and S4A) (Helmlinger and Tora, 2017; Spedale et al., 2012). mRNA analyses confirmed the inactivation of the corresponding genes in homozygous $Tada2a^{-/-}$ and $Tada2b^{-/-}$ cells (Figure S4B).

In addition, to supress the HAT activities of both SAGA and ATAC, we further targeted Tada3 (Figure 1A). As Tada3 was reported to be required for inner cell mass formation (Mohibi et al., 2012), we generated mouse ESC lines with homozygous insertion of the AID sequence at the $Tada3$ locus. In these $Tada3^{AID/AID}$ cells, the Tada3 protein levels were dramatically reduced from 4 hours of auxin treatment and depletion was almost complete after 24 hours (Figure 5B and S4C).

Characterization of complexes immuno-purified from $Tada2a^{-/-}$ and parental cells indicated that only Gcn5 or Pcaf HAT enzymes were lost from ATAC when Tada2a is missing (Figure 5C). In $Tada2a^{-/-}$ cells, we observed a minor reduction of H3 lysine 9 acetylation (H3K9ac), a reported *in vivo* substrate of SAGA and ATAC (Jin et al., 2011) (Figure 5D). The reduction of H3K9ac levels in $Tada2a^{-/-}$ cells was comparable to that seen upon depletion of core ATAC subunits, Zzz3 or Yeats2 (Figure S4D). In contrast, the whole HAT module dissociated from the core SAGA upon loss of Tada2b (Figure 5E), but no obvious changes in H3K9ac could be detected in $Tada2b^{-/-}$ cells (Figure 5D).

In $Tada3^{AID/AID}$ cells treated with auxin for 24 hours, H3K9 acetylation levels were significantly reduced, although the AID-Tada3 fusion protein was still detectable within the purified SAGA and ATAC complexes (Figure 5D, S4E and S4F). Thus, we asked whether this residual AID-Tada3 fusion protein would sustain a partial HAT activity within SAGA and ATAC. To this end, we generated a double $Tada2a^{-/-}+Tada2b^{-/-}$ ESC line in which the HAT enzymes (Gcn5 or Pcaf) are not supposed to incorporate in the respective complexes (Figure 5C and 5D). In this ESC line, the H3K9ac levels were reduced to a similar extent than upon depletion of Tada3 (Figure S4H), suggesting an almost complete loss of SAGA and ATAC HAT activities in $Tada3^{AID/AID}$ upon auxin treatment.

To determine whether the HAT activities of SAGA or ATAC play a role in ESC proliferation and maintenance, we performed clonal assays and AP staining using cells in which the HAT activities of SAGA ($Tada2b^{-/-}$), or ATAC ($Tada2a^{-/-}$), or of both complexes ($Tada3^{AID/AID}$) were suppressed. The loss of any of these activities did not cause any detectable effect on mouse ESC growth or self-renewal (Figure 5F, 5G and S5A-D). These results indicate that the phenotypes observed in ESCs depleted for Supt7l, Zzz3 or Yeats2 are related to functions of either SAGA or ATAC, but are essentially

independent of their HAT activities. This suggests important HAT-independent functions for SAGA and ATAC cores in mouse ESC maintenance.

SAGA and ATAC influence Pol II transcription mostly independent of their HAT modules

Although our gene expression studies pointed to different roles in Pol II transcription for the related SAGA and ATAC coactivator complexes, it still remained conceivable that these transcriptional effects may depend on their HAT activities that would be recruited at different target genes. Thus, we aimed to compare Pol II transcription in cells in which the HAT activities of both SAGA and ATAC are suppressed with that of cells inactivated for basically all SAGA and ATAC functions. Therefore, we quantified newly synthesized mRNA in *Tada3*^{AID/AID} cells and in a double mutant cell line, *Supt7l*^{-/-} +*Yeats2*^{AID/AID} treated with auxin for 24 hours. A comparable decrease of H3K9ac levels was observed in both cell lines, indicating that the HAT activities of the two complexes are similarly affected (Figure S6A). In contrast with the normal growth of *Tada3*^{AID/AID} cells treated with auxin (Figure 5 and S5), the *Supt7l*^{-/-}+*Yeats2*^{AID/AID} cells displayed dramatically reduced growth upon auxin treatment as assessed by clonal assay and growth curve analyses in FCS+LIF+2i medium (Figure S6B and S6C).

Newly synthesized RNA analyses in two independent *Tada3*$^{AID/AID}$ or *Supt7l*$^{-/-}$+*Yeats2*$^{AID/AID}$ clones revealed that very few genes were downregulated upon Tada3 depletion. In contrast, 930 protein-coding genes were significantly downregulated when core subunits of SAGA and ATAC were lost (Figure 6A and 6B). All subsequent analyses of the *Supt7l*$^{-/-}$+*Yeats2*$^{AID/AID}$ datasets identified features already characterized upon depletion of Yeats2 or Zzz3. Indeed, GSEA analysis in the double mutant cell lines identified the GO term 'cytoplasmic translation' as the most enriched and downregulated category (Figure S7A and S7B). Similarly, we confirmed that the expression of several RPGs was altered upon inactivation of both SAGA and ATAC but not upon depletion of Tada3, further highlighting the HAT-independent functions of SAGA and ATAC on Pol II transcription (Figure 6C and 6D). Analyses for enrichment of transcription factors bound to significantly downregulated genes in the double mutant cells also revealed an enrichment of Gcn5 (Kat2a), Myc and E2f4-bound genes, comparable to *Yeats2*$^{AID/AID}$ and *Zzz3*$^{AID/AID}$ cell lines (Figure S7C). When comparing the gene expression fold changes in the double mutant line with that of any single mutant, we also observed a significant correlation between the transcriptome of the *Supt7l*$^{-/-}$+*Yeats2*$^{AID/AID}$ with that of the *Supt7l*$^{-/-}$, *Yeats2*$^{AID/AID}$ and *Zzz3*$^{AID/AID}$ cell lines (Figure S7D).

We further analysed the transcriptional effects of the individual HAT modules of SAGA and ATAC in *Tada2a*$^{-/-}$ and *Tada2b*$^{-/-}$ cells, through RT-qPCR quantification of newly synthesized mRNA levels for selected genes that were found downregulated in the SAGA and ATAC core mutant ESCs. Genes downregulated in SAGA mutant ESCs were not affected by the loss of Tada2b and genes

downregulated upon depletion of ATAC core subunits were not affected by the loss of Tada2a (Figure S7E and S7F). These results together further confirm that SAGA and ATAC regulate Pol II transcription in ESCs in a differential manner that is largely independent of their HAT activity.

Discussion

In this study, we demonstrate a key role for the SAGA and ATAC coactivator complexes in the maintenance of ESC self-renewal and growth. The inactivation of the SAGA or ATAC complex significantly affected the expression of distinct gene groups with a varying impact on global Pol II transcription. Importantly, the phenotypes and transcriptional anomalies observed in SAGA and ATAC mutant cells are mostly independent of the activities of their HAT modules, which contain the same Kat2 enzymes (Gcn5 or Pcaf). Transcriptional effects of ATAC on RPGs and other genes involved in translation and of SAGA on genes related to ESC self-renewal most likely explain the phenotypes observed upon inactivation of these complexes. Therefore, each complex makes use of its specific activities to regulate different sets of genes, eventually leading to self-renewal defects of mouse ESCs.

SAGA and ATAC have distinct HAT-independent functions needed for Pol II transcription in mouse ESCs

The distinct HAT-independent functions of SAGA and ATAC for Pol II transcription in mouse ESCs were demonstrated through three major observations: i) disruption of the HAT module of SAGA or ATAC did not reproduce any of the effects seen upon inactivation of core subunits of these complexes; ii) newly synthesized RNA analysis in ATAC and SAGA mutant cell lines revealed that SAGA and ATAC affect transcription of distinct sets of genes, although they share the same HAT enzymes; iii) CRISPR-Cas9-mediated inactivation of the core ATAC subunits, Yeats2 and Zzz3, most likely results in lethality of mouse ESCs, as no homozygous knockout clones could be obtained, while homozygous inactivation of the ATAC-specific HAT subunit Tada2a was viable in mouse ESCs. This further demonstrates that SAGA cannot fully compensate for ATAC loss and therefore acts, at least partially, non-redundantly to ATAC.

Inactivation of the SAGA core subunit Supt7l, resulted in decreased expression of genes related to LIF signalling, a key pathway involved in the maintenance of the pluripotent state of mouse ESCs. These data agree with recent studies suggesting that two other subunits of the core SAGA, Taf5l and Taf6l, are involved in the regulatory networks important for self-renewal maintenance (Seruggia et al., 2019). In addition, our study shows that genes related to cytoplasmic translation such as RPGs were especially enriched in genes downregulated upon depletion of ATAC core subunits. Therefore,

14

our study is in agreement with previous analyses carried out in Drosophila or differentiated human cells suggesting that SAGA and ATAC predominantly regulate different sets of genes and therefore have non-overlapping as well as overlapping roles in the regulation of Pol II transcription (Arede et al., 2020; Gamper et al., 2009; Krebs et al., 2011; Nagy et al., 2009; Pankotai et al., 2005, 2010).

Catalytic-independent functions of coactivator complexes were previously demonstrated through the analysis of catalytic mutants of different chromatin modifying complexes, such as TIP60 and MLL3/4 COMPASS-like complexes (Acharya et al., 2017; Dorighi et al., 2017; Rickels et al., 2017). This suggests that several histone modifying complexes in addition to SAGA and ATAC have important functions besides their histone modifying activities. As earlier studies on $Gcn5^{-/-}$ ESCs demonstrated a requirement of the HAT activities of SAGA and ATAC during differentiation of mouse ESCs (Lin et al., 2007; Wang et al., 2018), the histone modifying activities of ATAC and SAGA appear to have a more critical role during differentiation than for ESC self-renewal. This agrees with the requirement of Gcn5 catalytic activity during mouse embryonic development (Bu et al., 2007; Xu et al., 2000; Yamauchi et al., 2000). Similarly, catalytic inactivation of the histone modifying activities of TIP60 did not impair mouse ESC growth or self-renewal, but resulted in defects during mouse embryonic development (Acharya et al., 2017).

Surprisingly, loss of Supt7l or Supt20h, two subunits of the core module of SAGA, which were reported to be required for the integrity of SAGA structure in *S. cerevisiae* (Grant et al., 1997; Sterner et al., 1999; Wu and Winston, 2002), had very different effects in ESCs. While Supt7l is required for ESC growth and self-renewal as well as for SAGA structural integrity, the loss of Supt20h did not affect ESC growth and had more limited consequences on SAGA assembly. Similar observations were made in *Schizosaccharomyces pombe*, in which *spt7Δ* cells were severely impaired for growth, whereas deletion of *SPT20* showed more modest defects (Helmlinger et al., 2011). Although orthologs of Supt7l were consistently found important for SAGA structure and functions in all tested species, the role of Supt20h orthologs appears more variable between species. Homozygous inactivation of *Supt20h* (also called p38IP) in mice was reported to cause severe gastrulation defects with abnormalities in mesoderm migration (Zohn et al., 2006). Similarly, inactivation of genes encoding catalytic subunits of SAGA, *Gcn5*, *Pcaf* or *Usp22* resulted in embryonic lethality at stages beyond gastrulation (Bu et al., 2007; Koutelou et al., 2019; Xu et al., 2000; Yamauchi et al., 2000). These phenotypes do not argue for a crucial role of SAGA or ATAC in the inner cell mass of the blastocyst, from which ESCs are derived. In agreement, no significant growth defects have been observed in mouse ESCs mutant for genes encoding catalytic subunits of SAGA (Lin et al., 2007; Sussman et al., 2013). However, the role of genes encoding subunits of SAGA or ATAC, which were shown to play a role for ESC growth and self-renewal, i.e. Supt7l, Yeats2, Zzz3 (this study), Taf5l and

Taf6l ((Seruggia et al., 2019)), has not yet been investigated in mouse development. Thus, it will be crucial to determine whether the inactivation of these genes affects the peri-implantation development of mouse embryos.

Gene-specific and global roles of SAGA and ATAC in Pol II transcription

Our newly synthesized RNA analyses suggest that in addition to their gene specific activities, SAGA and ATAC also have mild effects on global Pol II transcription in mammalian cells, although to different extents, as ATAC seems to influence the expression of more genes than SAGA. Nevertheless, many genes are likely to be co-regulated by SAGA and ATAC, with each complex having a predominant effect on their specific targets and pathways. Studies in *S. cerevisiae* suggested a major role for SAGA in Pol II transcription of a vast majority of actively transcribed genes (Baptista et al., 2017; Donczew et al., 2020). However, the transcriptional effects of SAGA appear less pronounced in mouse ESCs than in *S. cerevisiae*. These differences between yeast and mouse ESCs could indicate a more tightly controlled Pol II transcription process with several redundantly acting mechanisms in mammalian cells, which could make this cellular system less sensitive to inactivation of SAGA.

ATAC functions are essential for mouse ESC survival and for expression of RPGs

The reduced expression of RPGs seen upon rapid depletion (24 hours of auxin treatment) of ATAC core subunits, suggest that ATAC directly regulates these genes as well as other genes related to translation. In good agreement, recent studies in human cells found that Yeats2 and Zzz3 bind to the promoters of RPGs and regulate their expression in leukemia and lung cancer cells (Arede et al., 2020; Mi et al., 2017, 2018). Decreased expression of RPGs upon Yeats2 depletion were linked to defects in cancer cell growth and survival (Mi et al., 2017). In mouse ESCs, we could not evidence any obvious cell cycle defect following 24 hours Yeats2 or Zzz3 depletion (Figure S2G), whereas cell cycle and growth defects were detectable upon more extended depletion (Figure S2D-F and S2H). Thus, the reduced expression of translation-related genes precedes defects in cell cycle, suggesting that the observed growth defects and cell cycle abnormalities might be direct consequences of impaired assembly and function of ribosomes. Such striking effects on translation-related genes likely explain the lethality of *Yeats2* or *Zzz3* null cells, as most genes associated with ribosome biogenesis are essential for cell growth, proliferation and survival (Bertomeu et al., 2018).

SAGA stabilizes the naïve pluripotency network while ATAC is required to maintain the whole pluripotency network

A recent study found that the core SAGA subunits Taf5l and Taf6l maintain the self-renewal of mouse ESCs, mainly through acetylation and subsequent expression of SAGA target genes (Seruggia et al.,

2019). Although we could confirm that SAGA core subunits are required to maintain the ESC state, our analyses revealed that the SAGA HAT activity did not affect ESC self-renewal. Serrugia et al. also suggested that SAGA would especially regulate the expression of Oct4- and Myc-dependent genes in mouse ESCs as assessed by transcriptomic profiling and correlation analysis of binding sites for Taf5l and pluripotency factors (Seruggia et al., 2019). On the contrary, we did not find any obvious link between SAGA and Myc-dependent gene expression, an apparent discrepancy that could be explained by differences in medium compositions and assessments based on different methodologies.

Genes downregulated in our *Supt7l⁻/⁻* cells were enriched for genes responsive to LIF signalling and bound by the pluripotency factors Oct4 (*Pou5f1*), Sox2 and Nanog, suggesting a direct role of SAGA in the pluripotency network by functioning as a coactivator for these factors. Our observations that *Supt7l⁻/⁻* cells in FCS+LIF medium expressed reduced levels of naive pluripotency factors such as Esrrb, Nanog Klf4 and Tfcp2l1, but not of core pluripotency regulators (Oct4, Sox2), indicates that SAGA plays an important role in stabilizing the naive pluripotency network, in line with the increased sporadic differentiation and the reduced self-renewal efficiency observed in these cells (Martello and Smith, 2014; Navarro, 2018).

In contrast, we found that genes downregulated upon ATAC inactivation were enriched for Myc-bound as well as E2f4-bound genes, two transcription factors important for cell cycle progression, in agreement with a previously reported interaction between Yeats2 and E2f4 (Chappell and Dalton, 2013; Chen et al., 2009; Fagnocchi and Zippo, 2017; Hsu et al., 2019; Matsumura et al., 2003). Inactivation of *E2f4* or the combined inactivation of *c-Myc* and *n-Myc* in mouse ESCs affected cell growth and self-renewal, similarly to our observations in ATAC mutant cell lines (Smith and Dalton, 2010; Varlakhanova et al., 2010). Interestingly, the mRNA levels of all tested pluripotency transcription factors, including the core pluripotency factors Oct4 (*Pou5f1*) and Sox2, were decreased upon depletion of Yeats2 in ESCs cultured in FCS+LIF medium. This suggests that ATAC is globally required for proper expression of most pluripotency factors, which may explain the self-renewal defects in ATAC mutant cells, whereas the effects of ATAC on mouse ESC proliferation and growth might be partially through E2f4- and Myc-targeted genes.

In summary, we generated a large series of mutant ESC lines for SAGA and ATAC subunits, allowing a comprehensive and comparative analyses of these two coactivator complexes in embryonic stem cells. Our study allowed the identification of important, but differential roles for SAGA and ATAC in mouse ESC growth and self-renewal. Our findings therefore suggest that SAGA and ATAC affect the self-renewing capacities of ESCs through HAT-independent functions within the

pluripotency network. These results pave the way to determine the ATAC- and SAGA-specific activities used in mouse ESCs to control their specific gene expression programs.

Methods

Cell culture

Mouse ES E14 cells were cultured on plates coated with 0.1% gelatine solution in 1x PBS (Dutcher, Cat# P06-20410) using DMEM medium supplemented with 15% foetal calf serum ES-tested (ThermoFisher Scientific, Cat# 10270-106), 2 mM L-glutamine (ThermoFisher Scientific, Cat# 25030-024), 0.1% β-mercaptoethanol (ThermoFisher Scientific, Cat# 31350-010), 100 UI/ml penicillin and 100 μg/ml streptomycin (ThermoFisher Scientific, Cat# 15140-122), 0.1 mM non-essential amino acids (ThermoFisher Scientific, Cat# 11140-035) and 1,500 U/ml leukaemia inhibitory factor (home-made). For medium described as FCS+LIF+2i, 3 μM CHIR99021 (axon medchem, Cat# 1386) and 1 μM PD0325901 (axon medchem, Cat# 1408) were added freshly to the medium. Cells were grown at 37°C with 5% CO_2 levels. Cells were passaged every second day. To induce the auxin-inducible degradation (AID), cells were treated with 500 μM Indole-3-acetic acid (Sigma-Aldrich, Cat# I3750).

Drosophila Schneider S2 cells (CRL-1963, ATCC) were grown in Schneider's Drosophila medium (ThermoFisher Scientific, Cat# 21720-024) containing 10% FCS (heat inactivated) (Sigma Aldrich, Cat# F7524) and 0.5% penicillin and streptomycin at 27°C.

Schizosaccharomyces pombe cells were grown in autoclaved YES medium (yeast extract, adenine, histidine, uracil, leucine, lysine, 3% glucose) at 32°C.

Saccharomyces cerevisiae cells were grown in autoclaved YPD medium (yeast extract, bactopeptone, 2% glucose) at 30°C.

Plasmid construction

All homologous recombination (HR) templates and plasmids expressing 1 or 2 gRNAs and co-expressing high-fidelity Cas9 (Kleinstiver et al., 2016) fused to EGFP (Cas9-HF-EGFP) were generated by Golden Gate cloning (Engler et al., 2009). For the HR templates, silent mutations were introduced by PCR to prevent Cas9-HF-mediated cleavage of the HR template or the knockin allele. The sequences of the gRNAs for the different constructs are indicated in Table 1.

The plasmid containing the mouse *Supt7l* coding sequence (CDS) for the generation of *Supt7l^tg* cell lines was constructed as follows. The CDS of Supt7l was amplified by PCR from a cDNA bank of mouse embryos (day 9-12) and inserted together with the PGK promoter into a pcDNA3.1 hygro vector (Invitrogen) by replacing the CMV promoter.

All plasmids were verified by sequencing.

Generation of stable cell lines

19

For the generation of *Supt7l^tg* cell lines and the *Tir1-BirA* stable cell line, the linearized plasmids containing the coding sequences were transfected into either two independent *Supt7l^-/-* cell lines or wildtype ES E14 cells, respectively, using Lipofectamine2000 (ThermoFisher Scientific, Cat# 11668019) following manufacturer's instruction. Antibiotic selection was started 48 hours post-transfection (250 µg/ml hygromycin (Sigma Aldrich, Cat# H0654) or 400 µg/ml geneticin (ThermoFisher Scientific, Cat# 11811031)). Selection medium was exchanged every second day for a week. For *Supt7l^tg* cell lines, the polyclonal population was used for subsequent experiments. For *Tir1-BirA* stable cell lines, monoclonal cell lines were established by colony picking.

Generation of knockout and auxin-inducible degron (AID) cell lines

Mouse ESCs were transfected with the plasmid constructs at a confluency of 70-80% using Lipofectamine2000 (ThermoFisher Scientific, Cat# 11668019) following manufacturer's instruction. For knock-in (AID and HA-tag) cell lines, donor plasmids were linearized using unique cutter restriction enzymes before transfection and transfected together with the Cas9-containing transient plasmid in a *Tir1-BirA* expressing cell line. Two to three days after transfection, cells were selected for expression of the fluorescent tags (for knockout cell lines: fusion protein of Cas9 with fluorescent protein, for knock-in cell lines: fusion protein of protein of interest with fluorescent protein) by fluorescence activated cell sorting (FACS). Three to five 96-well plates were seeded with one fluorescent cell per well using the BD FACSAria™ II (BD Biosciences) instrument.

Clonal assays

For clonal assay analyses, 1500 to 3000 cells, which had been adapted to the respective media through at least three passages, were plated in wells of 6-well plates. Medium was changed every other day. On the sixth day, colonies were washed twice with 1x PBS before fixation with 4% Paraformaldehyde (Electron Microscopy Sciences, Cat# 15710, 16% PFA solution) for 30 minutes followed by two washes with 1x PBS.

To assess the alkaline phosphatase activity of mouse ESC colonies, Alkaline Phosphatase Kit (Vector Laboratories, Cat# SK-5100) was used following manufacture's instruction. Colonies were stained with AP for 5-10 minutes. For clonal assay analyses in FCS+LIF+2i medium, colony areas were measured automatically using ImageJ software. For clonal assay analyses in FCS+LIF medium, an additional staining with crystal violet was performed after AP staining to assess the total number of colonies enabling normalization between replicates. Colonies were stained with 0.1% crystal violet solution for at least 30 minutes and counted manually using the ImageJ interface. The number of AP positive colonies was also counted manually using the ImageJ interface, while the number of AP negative colonies were deduced by subtracting the number of AP+ colonies from the total number of colonies.

We considered colonies as AP+ colonies if they either stained entirely red or if they possessed a center of red cells surrounded by unstained cells.

Cell cycle analysis

For cell cycle analyses, cells were harvested, washed with 1x PBS and permeabilized with 70% of ice-cold ethanol. Cells were stored at 4°C for up to a week prior to analysis. For propidium iodide staining, permeabilized cells were centrifuged, washed with 1x PBS prior to incubation with 75 µg/ml RNAse A and 15 µg/ml propidium iodide (Sigma-Aldrich, Cat# P4170) for at least 30 minutes at room temperature. Samples were filtered and 10,000 to 20,000 cells were analysed using a BD FACTSCelesta™ (BD Biosciences) instrument to determine cell cycle profiles. Data were analysed using FlowJo™ 10.2. software with manual assignment of the cell cycle phases.

Metabolic labelling

Metabolic labelling of newly-transcribed RNA was adapted from previously described protocols (Rabani et al., 2011; Rädle et al., 2013; Schwalb et al., 2016). In brief, the nucleoside analogue 4-thiouridine (4sU) (Glentham Life Sciences, Cat# GN6085 or abcam, Cat# ab143718) was added to the cell culture medium at a final concentration of 500 µM for a 25 minutes pulse. After the labelling period, the medium containing 4sU was removed, the cells were washed with ice cold 1x PBS and immediately lysed using TRI® Reagent (Sigma). Total RNA was extracted following TRI® Reagent (Molecular Research Center Inc., Cat# TR 188) manufacturer's instruction.

To label *Drosophila* S2 cells, medium containing 4sU at a final concentration of 500 µM was added to the cells during 15 minutes under aluminium foil at room temperature. 4sU-containing medium was removed and 1xPBS was added to collect the cells using a cell scratcher. Cells were centrifuged, flash frozen in aliquots and stored at -80°C. For total RNA extraction, S2 cells were defrozen, lysed using TRI® Reagent (Molecular Research Center Inc., Cat# TR 188) and total RNA was isolated following manufacturer's instruction.

Yeast cultures were grown to an OD_{600} of 0.8. 4-thiouracil (Sigma-Aldrich, Cat# 440736) was freshly dissolved in DMSO and added to the cultures at a final concentration of 1 mM. Labelling was performed for 6 minutes. After this time period, yeast cells were pelleted, washed with ice-cold 1x PBS and aliquoted before being flash frozen and stored at -80°C. For total RNA extraction, the RiboPure™ RNA Purification Kit (ThermoFisher Scientific, Cat# AM1926) was used following manufacturer's instruction.

To remove any potential genomic DNA contamination from the total RNA extracts, the TURBO DNA-*free*™ Kit (Thermo Scientific, Cat# AM1907) was used following manufacturer's instructions for rigorous DNase treatment. In brief, TURBO DNase Buffer and TURBO DNase were added to the RNA samples and incubated at 37°C for 30 minutes. DNase Inactivation Reagent was added to the samples

and incubated for 5 minutes at room temperature before centrifugation at 10,000 xg for 2 minutes. The supernatant containing the RNA was transferred into a new tube and stored at -80°C until further use.

Newly synthesized RNA purification

The purification of newly synthesized RNA was based on previously described protocols (Rabani et al., 2011; Rädle et al., 2013; Schwalb et al., 2016). Labelled, total RNA of spike-in cells (*D. melanogaster, S. cerevisiae* or *S. pombe*) was added to labelled, total RNA from mouse ESCs in a ratio 1:5 to 1:10 prior to newly synthesized RNA purification to a final amount of 200-250 µg of total RNA. The RNA was precipitated and resuspended in 130 µl and sonicated using the following program on a Covaris E220 instrument: 1 % duty factor, 100 W, 200 cycles per burst, 80 seconds. Fragment size ranged from 10 kb to 200 bp. For purification, the fragmented total RNA was incubated for 10 minutes at 60°C and immediately chilled on ice for 2 minutes to open secondary RNA structures. The 4sU-labelled RNA was thiol-specific biotinylated by addition of 200 µg EZ-link HPDP-biotin (ThermoFisher Scientific, Cat# 21341), biotinylation buffer (10 mM Hepes-KOH pH 7.5 and 1 mM EDTA) and 20% DMSO (Sigma-Aldrich, Cat# D8418) to prevent precipitation of HPDP-biotin. Biotinylation was carried out for 3 hours at 24°C in the dark and with gentle agitations. After incubation, excess of biotin was removed by adding an equal volume of chloroform and centrifugation at 16,000 xg for 5 minutes at 4°C. RNA was precipitated from the aqueous phase by adding 0.1 volumes of 5 M NaCl and an equal volume of 100% isopropanol followed by centrifugation at 16,000 xg for 30 minutes at 4°C. After washing with 75% ethanol the RNA pellet was resuspended in 100 µl of RNase-free water and denatured for 10 minutes at 65°C followed by immediate chilling on ice for 5 minutes. The samples were incubated with 100 µl of streptavidin-coated µMACS magnetic beads (Miltenyi Biotec, Cat# 130-074-101) for 90 minutes at 24°C under gentle agitations. The µMACS columns (Miltenyi Biotec, Cat# 130-074-101) were placed on a MACS MultiStand (Miltenyi Biotec) and equilibrated with washing buffer (100 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween20) before applying the samples twice to the columns. The columns were then washed one time with 600 µl, 700 µl, 800 µl, 900 µl and 1 ml washing buffer before eluting the newly synthesized RNA with two washes of 100 µl 0.1M DTT. The isolated newly synthesized RNA was recovered using RNeasy MinElute Cleanup Kit (Qiagen, Cat# 74204) following manufacturer's instruction.

Library preparation and sequencing

Total RNA-seq

Total RNA-seq libraries were generated from 1 µg of total RNA using TruSeq Stranded Total RNA LT Sample Prep Kit with Ribo-Zero Gold (Illumina, San Diego, CA, Cat# RS-122-2301/RS-122-2302)

according to the Illumina protocol with the following modifications. Cytoplasmic and mitochondrial ribosomal RNA (rRNA) was removed using Ribo-Zero Gold rRNA (Yeast) (Illumina, Cat# MRZY1324). Following purification, the depleted RNA was fragmented using divalent cations at 94°C for 2 minutes. While, double stranded cDNA synthesis and adapter ligation were performed according to manufacturer instructions, the number of PCR cycles for library amplification was reduced to 10 cycles. After purification using AMPure XP beads (Beckman-Coulter, Villepinte, France, Cat# A63882), the final cDNA libraries were checked for quality and quantified using capillary electrophoresis. The libraries were subsequently sequenced with 1x 50 base pairs on a HiSeq4000 machine (Illumina).

4sU-seq

4sU-seq libraries were generated from 15 to 50 ng of purified, newly synthesized RNA using TruSeq Stranded Total RNA LT Sample Prep Kit with Ribo-Zero Gold (Illumina, San Diego, CA) according to the Illumina protocol with the following modifications. 4sU-labelled RNA was cleaned up using 1.8X RNAClean XP beads and fragmented using divalent cations at 94°C for 1 minutes without depletion of rRNA. While, double stranded cDNA synthesis and adapter ligation were performed according to manufacturer instructions, the number of PCR cycles for library amplification was reduced to 10 cycles. After purification using SPRIselect beads (Beckman-Coulter, Villepinte, France, Cat# B23319), the final cDNA libraries were checked for quality and quantified using capillary electrophoresis. The libraries were subsequently sequenced with 1x 50 base pairs on a HiSeq4000 machine (Illumina).

Sequence analysis total RNA-seq and 4sU-seq

Reads were preprocessed using cutadapt 1.10 (Martin, 1981) in order to remove adaptors and low-quality sequences and reads shorter than 40 bp were removed for further analysis. Remaining reads were mapped to *M. musculus*, *D. melanogaster* and *S. cerevisiae* rRNA sequences for samples VQFR1-12 or *M. musculus* and *S. pombe* rRNA sequences for samples VQFR13-18 and VQFR25-30 using bowtie 2.2.8 (Langmead and Salzberg, 2012) and reads mapping to rRNA sequences were removed for further analysis. For samples VQFR1-12, remaining reads were aligned to a hybrid genome composed of mm10, BDGP6 and sacCer3 assemblies of *M. musculus*, *D. melanogaster* and *S. cerevisiae* genome respectively with STAR 2.5.3a (Dobin et al., 2013). For samples VQFR13-18 and VQFR25-30, the hybrid genome was composed of mm10 and ASM294v2 assemblies of *M. musculus* and *S. pombe* genome respectively. Gene quantification was performed with htseq-count 0.6.1p1 (Anders et al., 2015), using "union" mode and Ensembl 93 annotations for all organisms except for *Schizosaccharomyces pombe* where Ensembl Fungi 41 annotations were used. For 4sU-seq data, "type" option was set to "gene" in order to take also into account reads aligned onto introns. Differential gene expression analysis was performed using DESeq2 1.16.1 (Love et al., 2014) Bioconductor R package on *M. musculus* counts

normalized with size factors computed by the median-of-ratios method proposed by Anders and Huber (Anders and Huber, 2010), on *Drosophila melanogaster* counts for samples VQFR1-12 or on *S. pombe* counts for samples VQFR13-18 and VQFR25-30 (using the following options: cooksCutoff=TRUE, independentFiltering=TRUE, alpha=0.05). P-values were adjusted for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). For subsequent data analyses and visualization, genes of the Y-chromosomes were excluded and only protein-coding genes were considered. Further, a threshold of 100 reads was used to define expressed genes and only genes shared between all 4sU-seq datasets were analyzed. This resulted in the analysis of 8208 protein-coding genes.

RT-qPCR

Reverse Transcription (RT) was performed with 1 - 2 µg total RNA and using 3.2 µg random hexamer primers (ThermoFisher Scientific, Cat# SO142) and Transcriptor Reverse Transcriptase (Roche, Cat# 03531287001) following manufacturer's instruction. In brief, the RNA was preincubated with the random hexamer primers for 10 minutes at 65°C before adding 1x Transcriptor RT Reaction Buffer (Roche, Cat# 03531325001), 20 U RNase Inhibitor (Promega, Cat# N2515), 1 mM deoxynucleotide-Mix (ThermoFisher Scientific, Cat# R0192) and 10 U Transcriptor Reverse Transcriptase. The samples were reverse transcribed using a Mastercycler gradient machine (Eppendorf) with incubation for 10 minutes at 25°C to allow efficient primer annealing followed by 30 minutes at 55°C for reverse transcription and 5 minutes at 85°C to inactivate the Transcriptor Reverse Transcriptase. For qPCR, the cDNA samples were amplified using LightCycler® 480 SYBR® Green 2x PCR Master Mix I (Roche, Cat# 04887352001) and 0.6 µM of forward and reverse primer respectively. The primer pairs used for qPCR are listed in Table 2. The qPCR was conducted using a LightCycler® 480 (Roche) with following program: 1 cycle of 5 minutes at 95°C for pre-denaturation, 45 amplification cycles with 10 seconds at 95°C for denaturation, 20 seconds at 65°C for primer annealing and 20 seconds at 72°C for extension. Melting curves were determined between 65°C and 97°C followed by 1 cycle of cooling for 30 seconds at 40°C. The obtained threshold-values were used to calculate the relative gene expression using the 2-ΔΔCT method and considering the individual primer pair efficiencies (Pfaffl, 2001).

Whole cell protein extraction

For whole cell protein extracts, cells were washed with 1x PBS and harvested using trypsin 0.25% EDTA. After centrifugation at 2,000 xg for 3 minutes, the protein pellets were washed twice with ice-cold 1x PBS to remove any remaining FCS. Proteins were then extracted from the collected cells using 1 volume of whole cell extract buffer (50 mM Tris HCl pH 7.9, 25% glycerol, 0.2 mM EDTA, 0.5 mM DTT, 5 mM $MgCl_2$, 600 mM KCl, 0.5% NP40 and 1x protein inhibitor cocktail) and incubated for 10 minutes. The

salt concentration was neutralized by addition 3 volumes of IP0 buffer (25 mM Tris HCl pH 7.9, 5% glycerol, 5 mM $MgCl_2$, 0.1% NP40, 1 mM DTT and 1x protein inhibitor cocktail) and incubation for 10 minutes. After centrifugation at 12,000x g for 10 minutes at 4°C, the supernatant containing the proteins was collected. The protein concentrations of the extracts were determined using the Coomassie Protein Assay Dye Reagent Concentrate (Bio-Rad, Cat# 5000006) and Synergy HTX Multi-Mode Reader (BioTek).

Acidic extraction of histone proteins

Cells were harvested by trypsinization and washed twice with ice-cold 1x PBS. Pellets were resuspended in 5-10 volumes of lysis buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 10 mM N-ethylmaleimide, 5 mM sodium butyrate and 1x protein inhibitor cocktail) and 0.2 M hydrochloric acid was added before incubation on ice for 30 minutes. Extracts were centrifuged for 10 minutes at 11,000 x g at 4°C and the supernatant stored at -80°C prior to Western blotting.

Western blot analysis

Proteins were separated using 8% to 15% of SDS-PAGE gels prior to blotting onto nitrocellulose membranes. Membranes were blocked in 3% non-fat dry milk for at least 30 minutes at room temperature. The membranes were then incubated overnight with primary antibodies (antibodies used shown in Table 3) in 0.3% non-fat dry milk at 4°C with one exception being Streptavidin-HRP, which was used in 1% BSA. After washing with 1x PBS containing 0.1% Tween20, if required, the membranes were incubated with secondary goat-anti-rabbit or -mouse antibodies conjugated to HRP for 50 minutes at room temperature followed by further three washes. The membranes were developed using the Pierce™ ECL Western Blotting Substrate (ThermoFisher Scientific, Cat# 32109) and the ChemiDocTM Touch Imaging System (Bio-Rad).

Nuclear extraction

To enrich extracts for nuclear proteins, cells were harvested and washed twice with 1x PBS. Cell pellet was resuspended in hypotonic buffer (10 mM Tris-HCl pH 8.0, 1.5 mM $MgCl_2$, 10 mM KCl and 1x protein inhibitor cocktail) and dounced 10 to 20 times using a B dounce homogenizer to isolate the nuclei. After centrifugation at 10,000 xg for 10 minutes at 4°C, supernatant was removed and pellet resuspended in high salt buffer (20 mM Tris-HCl pH 8.0, 25% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 450 mM NaCl, 0.1% NP40 and 1x protein inhibitor cocktail). Suspension was homogenized by douncing as described before prior to centrifugation at 10,000x g for 10 minutes at 4°C. The supernatant was kept as nuclear extract and stored at -80°C.

Immunoprecipitation

Protein-A or Protein-G Sepharose beads were washed three times with filtered 1x PBS and two times with IP100 buffer (25 mM Tris-HCl 7.5, 5 mM $MgCl_2$, 10% glycerol, 0,1% NP40, 100 mM KCl, 2 mM DTT and 1x protein inhibitor cocktail). Nuclear extracts were pre-cleared with 1/5 of 50% bead slurry for 2 hours at 4°C with overhead agitation. For antibody binding, the 50% bead slurry was incubated with 1/10 volume of the respective antibody ascites for 2 hours at 4°C with overhead agitation. After incubation, beads were washed three times with IP500 buffer (25 mM Tris-HCl 7.5, 5 mM $MgCl_2$, 10% glycerol, 0,1% NP40, 500 mM KCl, 2 mM DTT and 1x protein inhibitor cocktail) and twice with IP100 buffer before addition of 1/5 volume of the 50% antibody-bead slurry to the pre-cleared nuclear extracts. Nuclear extracts were incubated with beads overnight at 4°C with overhead agitation. After incubation, resins were washed three times with IP500 buffer and twice with IP100 buffer.

For anti-Taf10 and anti-Taf12 IPs, complexes were eluted from the beads using two subsequent 0.1 M glycine pH 2.8 elutions at room temperature and with agitation. Importantly, as Taf10 and Taf12 are shared between the SAGA and TFIID complexes and to increase the purification efficiency for SAGA in anti-Taf10 and anti-Taf12 IPs, nuclear extracts were depleted for TFIID prior to anti-Taf10 or anti-Taf12 IP by overnight incubation with beads coated with antibodies targeting the TFIID-specific subunit Taf7. Supernatants depleted for Taf7-containing TFIID were subsequently used for anti-Taf10 IPs. For anti-Atac2 IPs, complexes were eluted from the beads using two subsequent steps of elutions with peptide PI264 at a concentration of 2 mg/ml pH 7.5 for 1 hour each at 4°C with overhead agitation.

Mass spectrometry

Liquid digestion Protein mixtures were precipitated with TCA (Sigma Aldrich, Cat# T0699) overnight at 4°C. Samples were then centrifuged at 14,000 xg for 30 minutes at 4°C. Pellet were washed twice with 1 mL cold acetone and centrifuged at 14,000 xg for 10 minutes at 4°C. Washed pellet were then urea-denatured with 8 M urea (Sigma Aldrich, Cat# U0631) in Tris-HCl 0.1 mM, reduced with 5 mM TCEP for 30 minutes, and then alkylated with 10 mM iodoacetamide (Sigma Aldrich, Cat# I1149) for 30 minutes in the dark. Both reduction and alkylation were performed at room temperature and under agitation (850 rpm). Double digestion was performed with endoproteinase Lys-C (Wako, Cat# 125-05061) at a ratio 1/100 (enzyme/proteins) in 8 M urea for 4 hours, followed by an overnight modified trypsin (Promega, Cat# V5111) digestion at a ratio 1/100 (enzyme/proteins) in 2 M urea. Both Lys-C and Trypsin digestions were performed at 37°C. Peptide mixtures were then desalted on C18 spin-column and dried on Speed-Vacuum before LC-MS/MS analysis.

LC-MS/MS Analysis Samples were analyzed using an Ultimate 3000 nano-RSLC (ThermoFisher Scientific, San Jose California) coupled in line with a LTQ-Orbitrap ELITE mass spectrometer via a nano-electrospray ionization source (ThermoFisher Scientific, San Jose California). Peptide mixtures were

loaded on a C18 Acclaim PepMap100 trap-column (75 μm ID x 2 cm, 3 μm, 100Å, ThermoFisher Scientific) for 3.5 minutes at 5 μL/min with 2% ACN (Sigma Aldrich, Cat# 1207802), 0.1% formic acid (Sigma Aldrich, Cat# 94318) in water and then separated on a C18 Accucore nano-column (75 μm ID x 50 cm, 2.6 μm, 150Å, ThermoFisher Scientific) with a 90 minutes linear gradient from 5% to 35% buffer B (A: 0.1% FA in water/ B: 99% ACN, 0.1% FA in water), then a 20 minutes linear gradient from 35% to 80% buffer B, followed with 5 min at 99% B and 5 minutes of regeneration at 5% B. The total duration was set to 120 minutes at a flow rate of 200 nL/min. The oven temperature was kept constant at 38°C. The mass spectrometer was operated in positive ionization mode, in data-dependent mode with survey scans from m/z 350-1500 acquired in the Orbitrap at a resolution of 120,000 at m/z 400. The 20 most intense peaks (TOP20) from survey scans were selected for further fragmentation in the Linear Ion Trap with an isolation window of 2.0 Da and were fragmented by CID with normalized collision energy of 35%. Unassigned and single charged states were rejected. The Ion Target Value for the survey scans (in the Orbitrap) and the MS2 mode (in the Linear Ion Trap) were set to 1E6 and 5E3 respectively and the maximum injection time was set to 100 ms for both scan modes. Dynamic exclusion was used. Exclusion duration was set to 20 s, repeat count was set to 1 and exclusion mass width was ± 10 ppm.

Data Analysis Proteins were identified by database searching using SequestHT (ThermoFisher Scientific) with Proteome Discoverer 2.4 software (PD2.4, ThermoFisher Scientific) on *Mus musculus* database (Swissprot, non-reviewed, release 2019_08_07, 55121 entries). Precursor and fragment mass tolerances were set at 7 ppm and 0.6 Da respectively, and up to 2 missed cleavages were allowed. Oxidation (M) was set as variable modification, and Carbamidomethylation (C) as fixed modification. Peptides were filtered with a false discovery rate (FDR) at 1%, rank 1 and proteins were identified with 1 unique peptide. For the Label-Free Quantification, the protein abundancies were calculated from the average of the peptide abundancies using the TOP N (where N = 3, the 3 most intense peptides for each protein), and only the unique peptide were used for the quantification.

Data Visualization Normalized spectral abundance factors (NSAF) were calculated for each protein as follows. To obtain spectral abundance factors (SAF), spectral counts identifying a protein were divided by the protein length represented by the number of amino acids. To calculate NSAF values, the SAF values of each protein were divided by the sum of SAF values of all SAGA subunits. The NSAF values were subsequently normalized to the NSAF values of the bait protein Taf10. For visualization, log2 changes were calculated for each mutant cell line relative to wildtype cells.

Data availability

The RNA sequencing datasets are available on the Gene Expression Omnibus (GEO) under the accession number GSE154796.

Author contributions

V.F., L.T. and D.D. designed experiments, B.R.S.M. designed CRISPR-Cas9 plasmids, V.F. conducted experiments, M.S. helped generating mutant cell lines, D.P. performed bioinformatics analyses, V.F. carried out graphical representation of data, V.F., L.T. and D.D. wrote manuscript.

Declaration of Interests

The authors declare no competing interests.

Legends to Figures

Figure 1: The SAGA coactivator complex is required to maintain self-renewal of mouse embryonic stem cells. (**A**) Schematic representation of the HAT module incorporation within SAGA or ATAC through the specific subunits Tada2a and Tada2b, respectively. (**B**) Numbers of clones screened, percentage of heterozygous clones (+/-) and numbers of homozygous clones (-/-) obtained. n.d., not determined. All further experiments were performed using at least two independent homozygous mutant lines. (**C**) Western blot analysis of two independent *Supt7l*$^{-/-}$ and *Supt7l*^{tg} cell lines compared to WT cells. γTubulin serves as loading control. *= unspecific bands. (**D**) Representative images of clonal assays of SAGA mutant cells cultured in FCS+LIF medium and stained by crystal violet. (**E**) Mass spectrometry analyses of SAGA complexes purified from *Supt7l*$^{-/-}$, *Supt20h*$^{-/-}$ and *Supt7l*^{tg} and WT cells. Left: purification scheme. Right: heatmap showing for different SAGA subunits the log2 fold change of signal intensity between mutant and WT cells. *= bait protein. (**F**) Quantification of alkaline phosphatase (AP) staining of clonal assays of SAGA mutant cells grown in FCS+LIF medium. Numbers of AP+ and AP- colonies were normalized to the total number of colonies as assessed by crystal violet staining. (**G**) Total mRNA levels of pluripotency factors were quantified by RT-qPCR in WT, *Supt20h*$^{-/-}$ and *Supt7l*$^{-/-}$cells. mRNA levels were normalized to two RNA polymerase III genes (*Rpph1* and *Rn7sk*) and to WT cells. ***, *p* = 0.00093. For (**F** and **G**), the statistical test performed is Wilcoxon rank sum test with Bonferroni correction for multiple testing. Error bars show mean ± standard deviation (SD) of at least 4 biological replicates, using at least two independent clones. Only statistically significant (<0.05) results are indicated.

Figure 2: The ATAC coactivator complex is required for survival and self-renewal of mouse embryonic stem cells. (**A**) Table showing numbers of clones screened, percentages of heterozygous (+/-) and numbers of homozygous clones (-/-). (**B** and **C**) Western blot analyses using Streptavidin-HRP of ESC lines with auxin-inducible degron (AID) fusion to Yeats2 and Zzz3. Three clones with fusion to the N-terminus of Yeats2 (*Yeats2*$^{AID/AID}$) (**B**), two clones with fusion to the N-terminus (clones #1 and #2) or the C-terminus (clones #3 and #4) of Zzz3 (*Zzz3*$^{AID/AID}$) (**C**) were treated with auxin (IAA) for 24h (+) or in the absence of IAA (-). Asterisks indicate unspecific bands. (**D**) ATAC was purified from nuclear extracts of WT, *Yeats2*$^{AID/AID}$ and *Zzz3*$^{AID/AID}$ cells treated (IAA) or not (DMSO) with auxin, by using anti-Atac2 antibody and peptide elution. Eluted complexes were analyzed by western blotting using the indicated antibodies. Beads incubated with no antibody (AB) are shown as control. (**E**) Clonal assays of ATAC mutant cell lines compared to WT cells in FCS+LIF medium treated with DMSO or IAA and stained with crystal violet. (**F**) Quantification of AP staining in clonal assays as shown in (**E**). Numbers of AP+ and AP- colonies were normalized to the total number of colonies as assessed by crystal violet staining. **, *p* = 0.0084; ***, *p* = 6.5 x10^{-5}. (**G**) mRNA levels of pluripotency factors in *Yeats2*$^{AID/AID}$ mutant and

WT cells upon 24h to 72h of IAA treatment were quantified by RT-qPCR. mRNA levels were normalized to that of *Rpph1* and *Rn7sk* and to respective WT cells. Only WT cells treated for 24h with IAA are shown, for simplification. *, *p* = 0.02857. For (**F** and **G**), error bars show mean ± SD of at least 4 biological replicates. At least two independent clones were analyzed per cell line. Statistical test performed is Wilcoxon rank sum test with Bonferroni correction for multiple testing in (F) and two-sided Wilcoxon-Mann-Whitney test for (G). Only statistically significant (<0.05) results are indicated.

Figure 3: Newly synthesized RNA analysis reveals that SAGA and ATAC affect different sets of genes. (**A**) UCSC genome browser view showing newly synthesized RNA sequencing coverage of *Sf3b3* and *Cog4* in *Yeats2^{AID/AID}*, *Zzz3^{AID/AID}* and *Supt7l^{-/-}* cell lines compared to wildtype (WT) cells. The upper panel shows total RNA-seq coverage in WT cells. Blue arrows indicate transcription direction. (**B**) Density plot representations of log2 fold changes of newly synthesized RNA levels in *Yeats2^{AID/AID}*, *Zzz3^{AID/AID}* and *Supt7l^{-/-}* cell lines relative to WT cells against the mean of normalized reads. For each line, two independent clones were treated for 24h with IAA. *Sf3b3* and *Cog4* shown in (**A**) are highlighted. An adjusted *p*-value of 0.05 and absolute log2 fold change of 0.5 were used as thresholds for significantly affected genes. Numbers of significantly up- and downregulated genes are indicated. (**C**) Correlation analyses of log2 fold changes of newly synthesized RNA between *Yeats2^{AID/AID}*, *Zzz3^{AID/AID}* and *Supt7l^{-/-}* lines. (**D**) Venn diagrams comparing overlaps of significantly downregulated genes between *Yeats2^{AID/AID}*, *Zzz3^{AID/AID}* and *Supt7l^{-/-}* cells.

Figure 4. ATAC is required for the expression of translation-related genes. (**A**) Gene set enrichment analyses (GSEA) for enrichment of gene ontology (GO) biological processes based on log2 fold changes in newly synthesized RNA levels from *Yeats2^{AID/AID}*, *Zzz3^{AID/AID}* and *Supt7l^{-/-}* cells relative to WT cells. Colored bars represent statistically significant terms (FDR < 0.05), while non-significant terms are represented in grey. (**B**) Volcano plots representation of differential expression between *Yeats2^{AID/AID}*, *Zzz3^{AID/AID}*, *Supt7l^{-/-}* and WT cells. Genes were considered as significantly affected with an adjusted *p*-value of < 0.05 and an absolute log2 fold change greater than 0.5. Numbers of significantly mis-regulated genes are indicated. Ribosome protein genes (RPGs) are highlighted by blue dots. (**C**) Heatmap showing log2 fold changes (log2 FC) observed for all RPGs in *Yeats2^{AID/AID}*, *Zzz3^{AID/AID}* and *Supt7l^{-/-}* cell lines.

Figure 5: Loss of the shared acetyltransferase activity of SAGA and ATAC does not affect growth or self-renewal of mouse embryonic stem cells. (**A**) Numbers of clone screened, percentages of heterozygous (+/-) and numbers of homozygous (-/-) clones. (**B**) Western blot analyses of three independent homozygous AID-Tada3 (*Tada3^{AID/AID}*) treated with or without within 24 hours of auxin (IAA) for 24 hours were revealed using Strepatividin-HRP. Asterisks indicate unspecific bands. (**C**) ATAC complex was immuno-purified from nuclear extracts of WT and *Tada2a^{-/-}* cells, by using an anti-Atac2

antibody. Eluted complexes were analyzed by western blot with the indicated antibodies. Beads incubated with WT nuclear extracts in the absence of antibody (no AB) served as control. Stars indicate unspecific bands. (**D**) Western blot analyses of histone H3 lysine 9 acetylation (H3K9ac) levels in *Tada3*[AID/AID] treated for 24 hours with IAA, *Tada2a*[-/-] and *Tada2b*[-/-] cells. TATA-box binding protein (Tbp) serves as loading control. (**E**) SAGA complexes were immuno-purified (IP) from *Tada2b*[-/-] and WT cells, as described in Figure 1E. Results of MS analyses of eluted complexes are displayed as log2 fold change between mutant and WT cells. Asterisk indicates bait protein. (**F**) Clonal assays of HAT mutant cell lines cultured in FCS+LIF medium and stained with crystal violet. *Tada3*[AID/AID] and corresponding WT cells were treated with DMSO or IAA. (**G**) Quantification of alkaline phosphatase (AP) staining of clonal assays as shown in (**F**). Numbers of AP+ and AP- colonies were normalized to the total number of colonies as assessed by crystal violet staining. Error bars show mean ± SD of 8 biological replicates. At least two independent clones were analyzed per cell line. Statistical test performed is Wilcoxon rank sum test with Bonferroni correction for multiple testing. No statistically significant differences were detected.

Figure 6: SAGA and ATAC influence RNA polymerase II transcription independently of their HAT activities. (**A**) UCSC genome browser view of newly synthesized RNA-seq at the *Sf3b3* and *Cog4* locus in *Supt7l*[-/-] + *Yeats2*[AID/AID] double mutant, *Tada3*[AID/AID] and WT cells. The upper line shows total RNA-seq coverage in WT cells. (**B**) Density plots showing log2 fold change of newly synthesized RNA levels in *Supt7l*[-/-] + *Yeats2*[AID/AID] double mutant and *Tada3*[AID/AID] cells relative to WT cells against the mean of normalized reads in WT cells. Data were obtained from two independent clones treated with IAA for 24 hours. *Sf3b3* and *Cog4* genes shown in (**B**) are highlighted. Numbers of significantly up- and downregulated genes are indicated. Genes were considered as significantly affected with an adjusted *p*-value of < 0.05 and an absolute log2 fold change above 0.5. A threshold of 100 reads was set to define expressed genes. (**C**) Volcano plot representation of differential expression analyses of newly synthesized RNA data from *Supt7l*[-/-] + *Yeats2*[AID/AID] double mutant and *Tada3*[AID/AID] cell lines relative to WT cells. Numbers on top left and right indicate significantly down- and upregulated genes, respectively. Genes were considered as significantly affected with an adjusted *p*-value of < 0.05 and log2 fold change of < -0.5 or > 0.5 as indicated by the dashed lines. Ribosome protein genes (RPGs) are highlighted as blue dots. (**D**) Heatmap showing the log2 fold changes (log2 FC) observed for RPG mRNA levels in *Supt7l*[-/-] + *Yeats2*[AID/AID] double mutant and *Tada3*[AID/AID] relative to WT cells.

References

Acharya, D., Hainer, S.J., Yoon, Y., Wang, F., Bach, I., Rivera-Pérez, J.A., and Fazzio, T.G. (2017). KAT-Independent Gene Regulation by Tip60 Promotes ESC Self-Renewal but Not Pluripotency. Cell Reports *19*, 671–679.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Arede, L., Foerner, E., Wind, S., Kulkarni, R., Domingues, A.F., Kleinwaechter, S., Gupta, S., Scheer, E., Tora, L., and Pina, C. (2020). Unique roles of ATAC and SAGA - KAT2A complexes in normal and malignant hematopoiesis. Biorxiv 2020.05.14.096057.

Balasubramanian, R., Pray-Grant, M.G., Selleck, W., Grant, P.A., and Tan, S. (2002). Role of the Ada2 and Ada3 Transcriptional Coactivators in Histone Acetylation. Journal of Biological Chemistry *277*, 7989–7995.

Baptista, T., Grünberg, S., Minoungou, N., Koster, M.J.E., Timmers, H.T.M., Hahn, S., Devys, D., and Tora, L. (2017). SAGA Is a General Cofactor for RNA Polymerase II Transcription. Mol Cell *68*, 130-143.e5.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Royal Statistical Soc Ser B Methodol *57*, 289–300.

Bertomeu, T., Coulombe-Huntington, J., Chatr-aryamontri, A., Bourdages, K.G., Coyaud, E., Raught, B., Xia, Y., and Tyers, M. (2018). A High-Resolution Genome-Wide CRISPR/Cas9 Viability Screen Reveals Structural Features and Contextual Diversity of the Human Cell-Essential Proteome. Mol Cell Biol *38*, e00302-17.

Bonnet, J., Wang, C.-Y., Baptista, T., Vincent, S.D., Hsiao, W.-C., Stierle, M., Kao, C.-F., Tora, L., and Devys, D. (2014). The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes & Development *28*, 1999–2012.

Bu, P., Evrard, Y.A., Lozano, G., and Dent, S.Y. (2007). Loss of Gcn5 Acetyltransferase Activity Leads to Neural Tube Closure Defects and Exencephaly in Mouse Embryos. Molecular and Cellular Biology *27*, 3405–3416.

Chappell, J., and Dalton, S. (2013). Roles for MYC in the Establishment and Maintenance of Pluripotency. Csh Perspect Med *3*, a014381.

Chen, H.-Z., Tsai, S.-Y., and Leone, G. (2009). Emerging roles of E2Fs in cancer: an exit from cell cycle control. Nat Rev Cancer *9*, 785–797.

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.

Donczew, R., Warfield, L., Pacheco, D., Erijman, A., and Hahn, S. (2020). Two roles for the yeast transcription coactivator SAGA and a set of genes redundantly regulated by TFIID and SAGA. Elife *9*, e50109.

Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N.V., Scruggs, B.S., Nady, N., Still, C.D., Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Molecular Cell *66*, 568-576.e4.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. Plos One *4*, e5553.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Fagnocchi, L., and Zippo, A. (2017). Multiple Roles of MYC in Integrating Regulatory Networks of Pluripotent Stem Cells. Frontiers Cell Dev Biology *5*, 7.

Festuccia, N., Gonzalez, I., and Navarro, P. (2017). The Epigenetic Paradox of Pluripotent ES Cells. Journal of Molecular Biology *429*, 1476–1503.

Fournier, M., Orpinell, M., Grauffel, C., Scheer, E., Garnier, J.-M., Ye, T., Chavant, V., Joint, M., Esashi, F., Dejaegere, A., et al. (2016). KAT2A/KAT2B-targeted acetylome reveals a role for PLK4 acetylation in preventing centrosome amplification. Nat Commun *7*, 13227.

Gamper, A.M., Kim, J., and Roeder, R.G. (2009). The STAGA Subunit ADA2b Is an Important Regulator of Human GCN5 Catalysis. Molecular and Cellular Biology *29*, 266–280.

Grant, P., Duggan, L., Côté, J., Roberts, S., Brownell, J., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C., Winston, F., et al. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes & Development *11*, 1640–1650.

Guelman, S., Kozuka, K., Mao, Y., Pham, V., Solloway, M.J., Wang, J., Wu, J., Lill, J.R., and Zha, J. (2009). The Double-Histone-Acetyltransferase Complex ATAC Is Essential for Mammalian Development. Molecular and Cellular Biology *29*, 1176–1188.

Han, Y., Luo, J., Ranish, J., and Hahn, S. (2014). Architecture of the Saccharomyces cerevisiae SAGA transcription coactivator complex. The EMBO Journal *33*, 2534–2546.

Helmlinger, D., and Tora, L. (2017). Sharing the SAGA. Trends in Biochemical Sciences *42*, 850–861.

Helmlinger, D., Marguerat, S., Villén, J., Swaney, D.L., Gygi, S.P., Bähler, J., and Winston, F. (2011). Tra1 has specific regulatory roles, rather than global functions, within the SAGA co-activator complex. The EMBO Journal *30*, 2843–2852.

Helmlinger, D., Papai, G., Devys, D., and Tora, L. (2020). What do the structures of GCN5-containing complexes teach us about their function? Biochimica Et Biophysica Acta Bba - Gene Regul Mech 194614.

Henry, K.W., Wyce, A., Lo, W.-S., Duggan, L.J., Emre, N.C., Kao, C.-F., Pillus, L., Shilatifard, A., Osley, M., and Berger, S.L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes & Development *17*, 2648–2663.

Hsu, J., Arand, J., Chaikovsky, A., Mooney, N.A., Demeter, J., Brison, C.M., Oliverio, R., Vogel, H., Rubin, S.M., Jackson, P.K., et al. (2019). E2F4 regulates transcriptional activation in mouse embryonic stem cells independently of the RB family. Nat Commun *10*, 2939.

Jin, Q., Yu, L., Wang, L., Zhang, Z., Kasper, L.H., Lee, J., Wang, C., Brindle, P.K., Dent, S.Y., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. The EMBO Journal *30*, 249–262.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Koutelou, E., Wang, L., Schibler, A., Chao, H.-P., Kuang, X., Lin, K., Lu, Y., Shen, J., Jeter, C.R., Salinger, A., et al. (2019). Usp22 controls multiple signaling pathways that are essential for vasculature formation in the mouse placenta. Development *146*, dev.174037.

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

Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC Histone Acetyl Transferase Complexes Regulate Distinct Sets of Genes and ATAC Defines a Class of p300-Independent Enhancers. Molecular Cell *44*, 410–423.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359.

Lee, K.K., Sardiu, M.E., Swanson, S.K., Gilmore, J.M., Torok, M., Grant, P.A., Florens, L., Workman, J.L., and Washburn, M.P. (2011). Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. Mol Syst Biol *7*, 503.

Li, B., Carey, M., and Workman, J.L. (2007). The Role of Chromatin during Transcription. Cell *128*, 707–719.

Li, X., Li, L., Pandey, R., Byun, J.S., Gardner, K., Qin, Z., and Dou, Y. (2012). The Histone Acetyltransferase MOF Is a Key Regulator of the Embryonic Stem Cell Core Transcriptional Network. Cell Stem Cell *11*, 163–178.

Lin, W., Srajer, G., Evrard, Y.A., Phan, H.M., Furuta, Y., and Dent, S.Y. (2007). Developmental potential of Gcn5(-/-) embryonic stem cells in vivo and in vitro. Developmental Dynamics : An Official Publication of the American Association of Anatomists *236*, 1547–1557.

Martello, G., and Smith, A. (2014). The Nature of Embryonic Stem Cells. Cell and Developmental Biology *30*, 647–675.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc National Acad Sci *78*, 7634–7638.

Matsumura, I., Tanaka, H., and Kanakura, Y. (2003). E2F1 and c-Myc in Cell Growth and Death. Cell Cycle *2*, 332–335.

Mi, W., Guan, H., Lyu, J., Zhao, D., Xi, Y., Jiang, S., Andrews, F.H., Wang, X., Gagea, M., Wen, H., et al. (2017). YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. Nature Communications *8*, 1088.

Mi, W., Zhang, Y., Lyu, J., Wang, X., Tong, Q., Peng, D., Xue, Y., Tencer, A.H., Wen, H., Li, W., et al. (2018). The ZZ-type zinc finger of ZZZ3 modulates the ATAC complex-mediated histone acetylation and gene activation. Nat Commun *9*, 3759.

Mohibi, S., Gurumurthy, C.B., Nag, A., Wang, J., Mirza, S., Mian, Y., Quinn, M., Katafiasz, B., Eudy, J., Pandey, S., et al. (2012). Mammalian alteration/deficiency in activation 3 (Ada3) is essential for embryonic development and cell cycle progression. The Journal of Biological Chemistry *287*, 29442–29456.

Nagy, Z., Riss, A., Fujiyama, S., Krebs, A., Orpinell, M., Jansen, P., Cohen, A., Stunnenberg, H.G., Kato, S., and Tora, L. (2009). The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. Cell Mol Life Sci Cmls *67*, 611–628.

Natsume, T., Kiyomitsu, T., Saga, Y., and Kanemaki, M.T. (2016). Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. Cell Reports *15*, 210–218.

Navarro, P. (2018). 2i, or Not 2i: The Soliloquy of Nanog-Negative Mouse Embryonic Stem Cells. Stem Cell Rep *11*, 1–3.

Orpinell, M., Fournier, M., Riss, A., Nagy, Z., Krebs, A.R., Frontini, M., and Tora, L. (2010). The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates. Embo J *29*, 2381–2394.

Pankotai, T., Komonyi, O., Bodai, L., Ujfaludi, Z., Muratoglu, S., Ciurciu, A., Tora, L., Szabad, J., and Boros, I. (2005). The Homologous Drosophila Transcriptional Adaptors ADA2a and ADA2b Are both Required for Normal Development but Have Different Functions. Mol Cell Biol *25*, 8215–8227.

Pankotai, T., Popescu, C., Martín, D., Grau, B., Zsindely, N., Bodai, L., Tora, L., Ferrús, A., and Boros, I. (2010). Genes of the Ecdysone Biosynthesis Pathway Are Regulated by the dATAC Histone Acetyltransferase Complex in Drosophila ▽ . Mol Cell Biol *30*, 4254–4266.

Papai, G., Frechard, A., Kolesnikova, O., Crucifix, C., Schultz, P., and Ben-Shem, A. (2020). Structure of SAGA and mechanism of TBP deposition on gene promoters. Nature *577*, 711–716.

Perez-Garcia, V., Fineberg, E., Wilson, R., Murray, A., Mazzeo, C.I., Tudor, C., Sienerth, A., White, J.K., Tuck, E., Ryder, E.J., et al. (2018). Placentation defects are highly prevalent in embryonic lethal mouse mutants. Nature *555*, 463–468.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res *29*, e45–e45.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nature Biotechnology *29*, 436–442.

Rädle, B., Rutkowski, A.J., Ruzsics, Z., Friedel, C.C., Koszinowski, U.H., and Dölken, L. (2013). Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. Journal of Visualized Experiments : JoVE.

Rickels, R., Herz, H.-M., Sze, C.C., Cao, K., Morgan, M.A., Collings, C.K., Gause, M., Takahashi, Y., Wang, L., Rendleman, E.J., et al. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat Genet *49*, 1647–1653.

Sainsbury, S., Niesser, J., and Cramer, P. (2012). Structure and function of the initially transcribing RNA polymerase II–TFIIB complex. Nature *493*, 437.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science (New York, N.Y.) *352*, 1225–1228.

Scognamiglio, R., Cabezas-Wallscheid, N., Thier, M.C., Altamura, S., Reyes, A., Prendergast, Á.M., Baumgärtner, D., Carnevalli, L.S., Atzberger, A., Haas, S., et al. (2016). Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. Cell *164*, 668–680.

Seruggia, D., Oti, M., Tripathi, P., Canver, M.C., LeBlanc, L., Giammartino, D.C.D., Bullen, M.J., Nefzger, C.M., Sun, Y.B.Y., Farouni, R., et al. (2019). TAF5L and TAF6L Maintain Self-Renewal of Embryonic Stem Cells via the MYC Regulatory Network. Mol Cell.

Smith, K., and Dalton, S. (2010). Myc transcription factors: key regulators behind establishment and maintenance of pluripotency. Regen Med *5*, 947–959.

Spedale, G., Timmers, H.Th., and Pijnappel, W.W.M. (2012). ATAC-king the complexity of SAGA during evolution. Genes & Development *26*, 527–541.

Sterner, D.E., Grant, P.A., Roberts, S.M., Duggan, L.J., Belotserkovskaya, R., Pacella, L.A., Winston, F., Workman, J.L., and Berger, S.L. (1999). Functional Organization of the Yeast SAGA Complex: Distinct Components Involved in Structural Integrity, Nucleosome Acetylation, and TATA-Binding Protein Interaction. Molecular and Cellular Biology *19*, 86–98.

Sussman, R.T., Stanek, T.J., Esteso, P., Gearhart, J.D., Knudsen, K.E., and McMahon, S.B. (2013). The epigenetic modifier ubiquitin-specific protease 22 (USP22) regulates embryonic stem cell differentiation via transcriptional repression of sex-determining region Y-box 2 (SOX2). The Journal of Biological Chemistry *288*, 24234–24246.

Thomas, M.C., and Chiang, C.-M.M. (2006). The general transcription machinery and general cofactors. Critical Reviews in Biochemistry and Molecular Biology *41*, 105–178.

Varlakhanova, N.V., Cotterman, R.F., deVries, W.N., Morgan, J., Donahue, L.R., Murray, S., Knowles, B.B., and Knoepfler, P.S. (2010). myc maintains embryonic stem cell pluripotency and self-renewal. Differentiation *80*, 9–19.

Wang, L., and Dent, S.Y. (2014). Functions of SAGA in development and disease. Epigenomics *6*, 329–339.

Wang, H., Dienemann, C., Stützer, A., Urlaub, H., Cheung, A.C.M., and Cramer, P. (2020). Structure of the transcription coactivator SAGA. Nature *577*, 717–720.

Wang, L., Koutelou, E., Hirsch, C., McCarthy, R., Schibler, A., Lin, K., Lu, Y., Jeter, C., Shen, J., Barton, M.C., et al. (2018). GCN5 Regulates FGF Signaling and Activates Selective MYC Target Genes during Early Embryoid Body Differentiation. Stem Cell Reports *10*, 287–299.

Weake, V.M., and Workman, J.L. (2012). SAGA function in tissue-specific gene expression. Trends in Cell Biology *22*, 177–184.

Wu, P.-Y., and Winston, F. (2002). Analysis of Spt7 function in the Saccharomyces cerevisiae SAGA coactivator complex. Molecular and Cellular Biology *22*, 5367–5379.

Xu, W., Edmondson, D.G., Evrard, Y.A., Wakamiya, M., Behringer, R.R., and Roth, S.Y. (2000). Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. Nat Genet *26*, ng1000_229.

Yamauchi, T., Yamauchi, J., Kuwata, T., Tamura, T., Yamashita, T., Bae, N., Westphal, H., Ozato, K., and Nakatani, Y. (2000). Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. Proceedings of the National Academy of Sciences of the United States of America *97*, 11303–11306.

Yang, X.-J., Ogryzko, V.V., Nishikawa, J., Howard, B.H., and Nakatani, Y. (1996). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. Nature *382*, 319–324.

Young, R.A. (2011). Control of the Embryonic Stem Cell State. Cell *144*, 940–954.

Zohn, I.E., Li, Y., Skolnik, E.Y., Anderson, K.V., Han, J., and Niswander, L. (2006). p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation. Cell *125*, 957–969.

Figure 1 Fischer et al.

A

HAT
Tada3
Tada2a
Yeats2
Zzz3
ATAC

HAT
Tada3
Tada2b
Supt20h
Supt7l
SAGA

B

	Clones screened	% +/-	n°-/-
Supt7l	45	n.d.	5
Supt20h	13	23.1	4

C

	WT	*Supt7l^-/-* #1	*Supt7l^tg* #1	*Supt7l^-/-* #2	*Supt7l^tg* #2	
95 —						*
72 —						◄ Supt7l
						*
						*
55 —						◄ γTubulin

D

WT   *Supt7l^-/-*

Supt20h^-/-   *Supt7l^tg*

E

mES nuclear extract
immunodepleted for TFIID
|
anti-Taf10 IP
|
Elution
|
MS

shared
Taf12
Taf10 *
Taf9
Taf6l
Taf5l
Tada1
Supt3h
Supt7l
Supt20h
Trrap
Kat2b
Tada3
Tada2b
Sgf29

log2
0
-1
-2
-3
-4
NA

WT   *Supt7l^-/-*   *Supt7l^tg*   *Supt20h^-/-*

F

% of AP+/AP- colonies

WT   *Supt7l^-/-*   *Supt7l^tg*   *Supt20h^-/-*

AP+
AP−

G

Relative Expression

Pou5f1   *Sox2*   *Tfcp2l1*   *Nanog*   *Esrrb*   *Klf4*

WT
Supt20h^-/-
Supt7l^-/-

***   ***   ***   ***

Figure 2 Fischer et al.

A

	Clones screened	% +/-	# -/-
Yeats2	264	18.9	0
Zzz3	317	17.9	0

B

Yeats2$^{AID/AID}$

#1 #2 #3

IAA − + − + − +

250 —

150 —

*

◄ AID-Yeats2

C

Zzz3$^{AID/AID}$

#1   #2   #3   #4

IAA − + − + − + − +

250 —

150 —

*

◄ AID-Zzz3

D

Atac2 IP

no AB   WT   *Yeats2$^{AID/AID}$*   *Zzz3$^{AID/AID}$*   WT   *Yeats2$^{AID/AID}$*   *Zzz3$^{AID/AID}$*

55 —

◄ Tada3

36 —

◄ Wdr5
◄ Sgf29

DMSO   IAA

E

WT   *Yeats2$^{AID/AID}$*   *Zzz3$^{AID/AID}$*

DMSO

IAA

F

% of AP+/AP- colonies

AP+
AP−

WT   *Yeats2$^{AID/AID}$*   *Zzz3$^{AID/AID}$*   WT   *Yeats2$^{AID/AID}$*   *Zzz3$^{AID/AID}$*

DMSO   IAA

G

Relative Expression

Pou5f1   *Sox2*   *Tfcp2l1*   *Nanog*   *Esrrb*   *Klf4*

24h WT
24h *Yeats2$^{AID/AID}$*
48h *Yeats2$^{AID/AID}$*
72h *Yeats2$^{AID/AID}$*

Figure 3 Fischer et al.

A

B

C
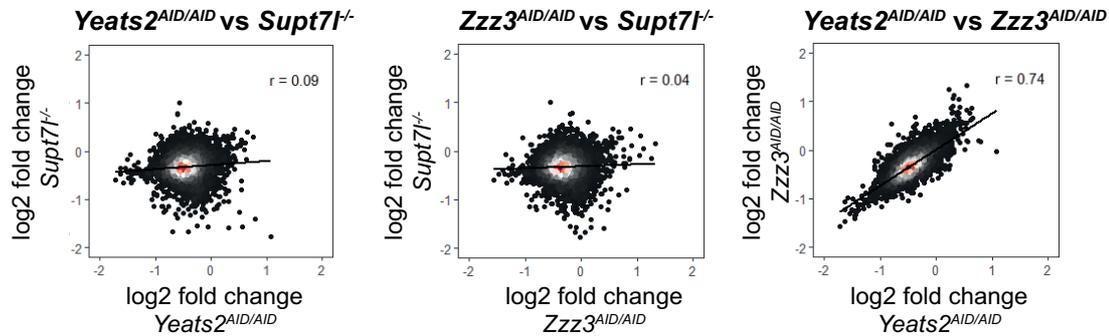
D
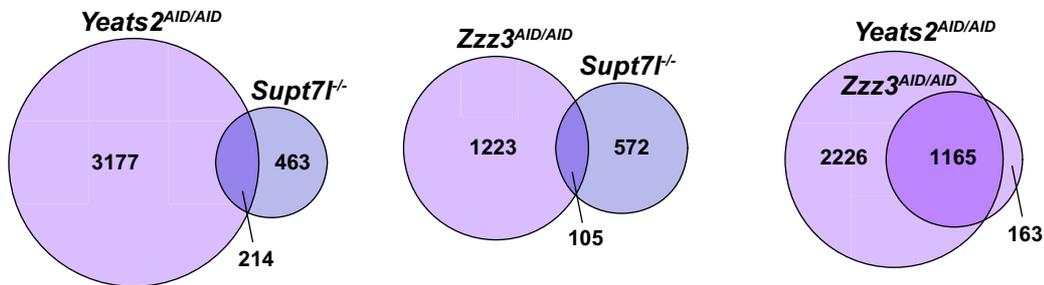
Figure 4 Fischer et al.

Figure 5 Fischer et al.

A

	Clones screened	% +/-	n° -/-	Complex
Tada2a	51	23.5	9	ATAC HAT
Tada2b	60	n.d.	6	SAGA HAT

B

C

Input — Atac2 IP

D

E

Taf10 IP elution

F

G
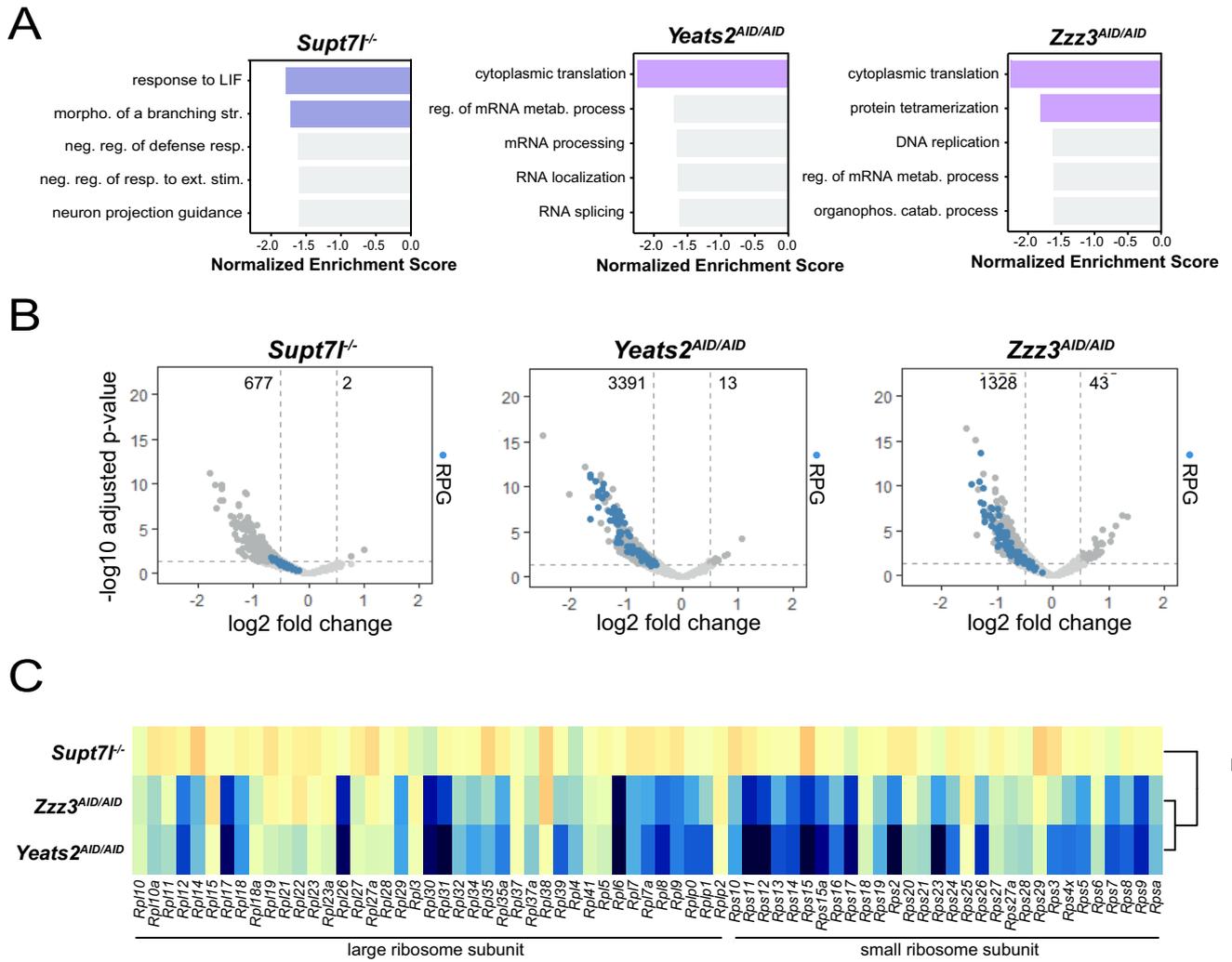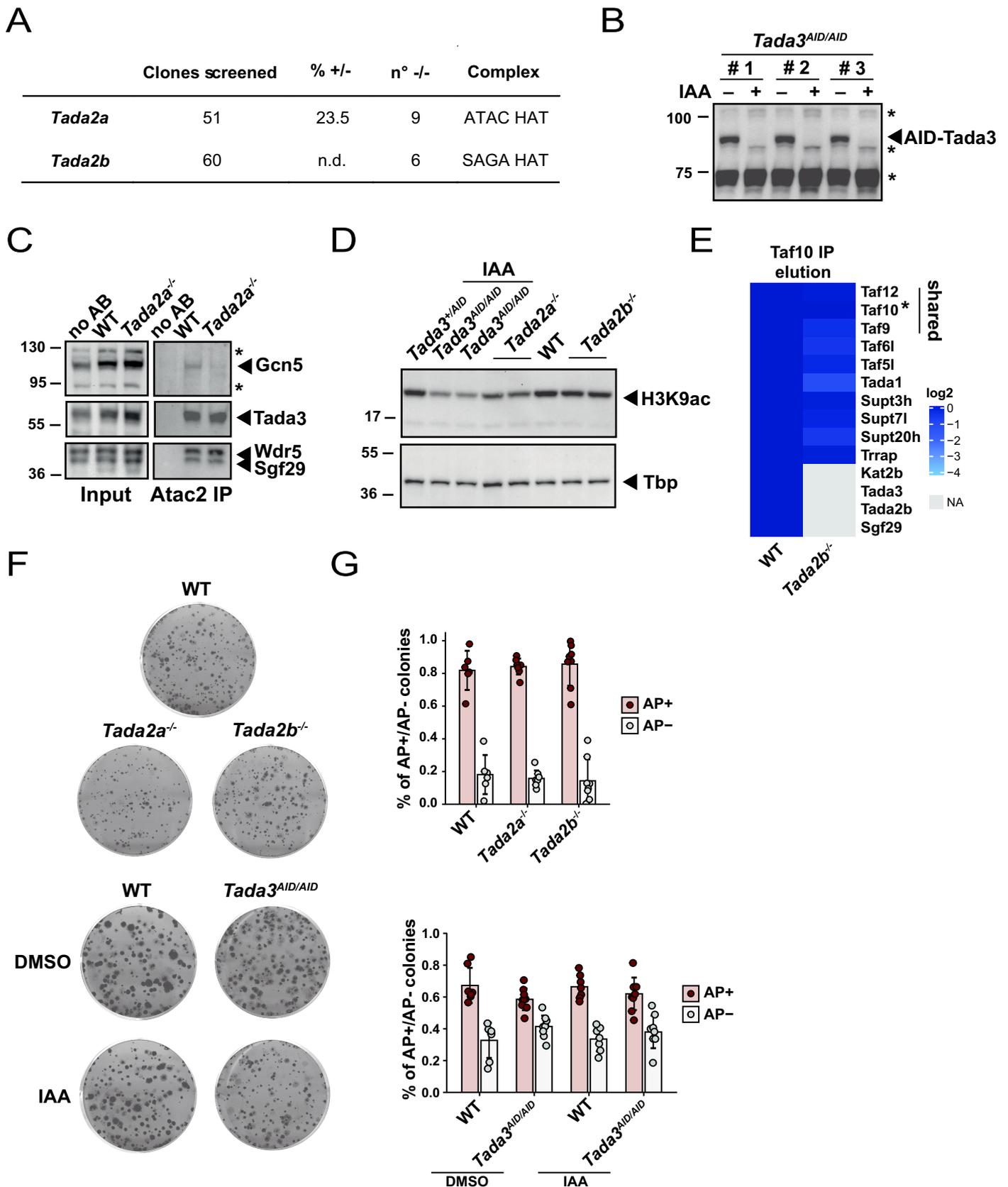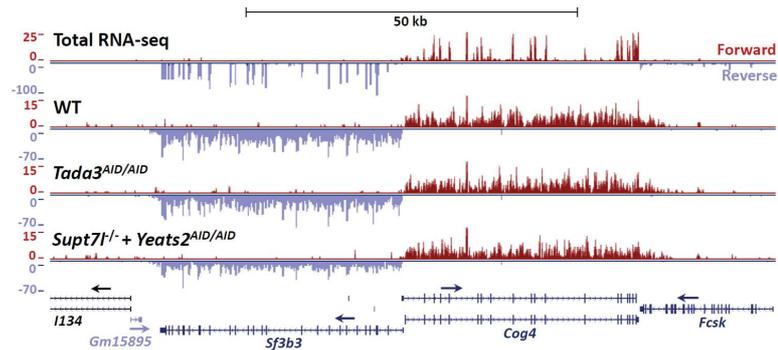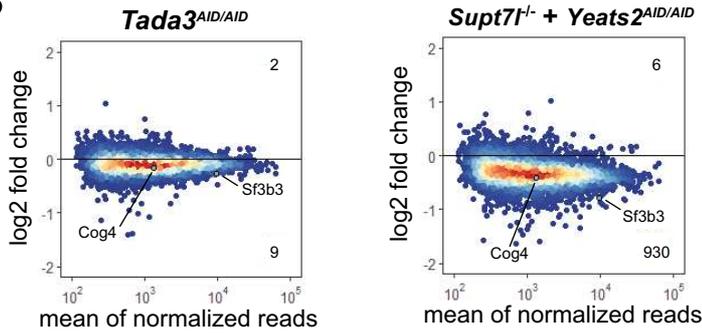
Figure 6 Fischer et al.
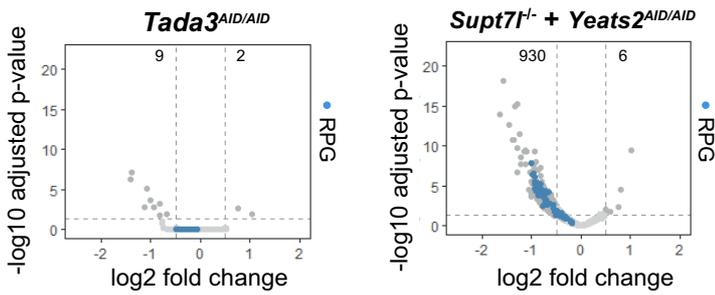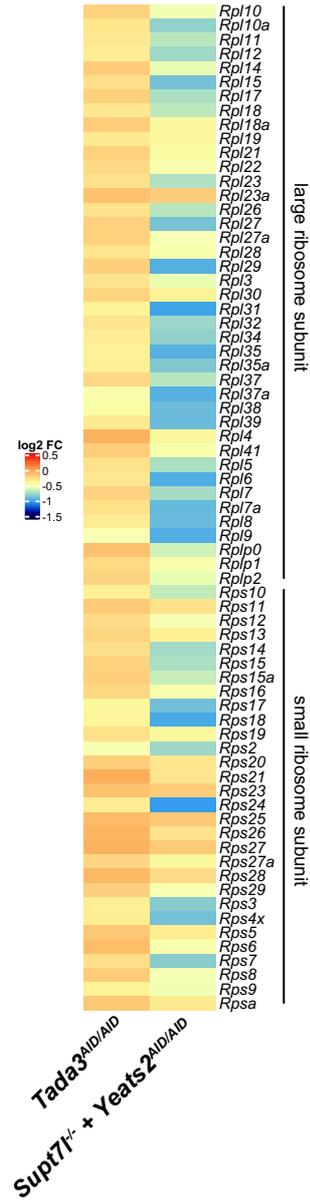
Supplementary Materials (Fischer et al.): contain 3 Supplementary Tables and 7

Supplementary Figures

Supplementary Table 1: CRISPR-Cas9 gRNA table. The following table shows the gRNA sequences used to generate the cell lines presented in this study including the type of modification, the targeted gene, the targeted region of the gene, the sequence of the gRNA, the PAM sequence and the strand of the gRNA. KO, knockout; AID, auxin-inducible degron; E, exon

Type	Target	Region	gRNA sequence	PAM	Strand
KO	Supt20h	5'E4	TCGCTTGCACTCACTCGT	AGG	+
		3'E4	GTAGAGCAGTCCAGTCGG	AGG	-
	Supt7l	5'E3	ACCAGTACGTATTCAGAG	TGG	+
		3'E3	ACCATCTCCCTCGCCCCG	AGG	+
	Tada2b	5'E2	CCTACATAGATGTACCTGAG	CGG	-
		E2	TTATGAGATAGAGTATGACC	AGG	+
	Tada2a	5'E3	GCTACAGGTAGTCTTCCCTG	CGG	-
		3'E3	CTGCTGTGTAGTAGACAGAG	TGG	-
	Yeats2	5'E6	TCACTGAAACAGTATTCAGT	AGG	-
		3'E6	CCGTTACTGCATATTCACAG	TGG	+
	Zzz3	5'E5	GACTAGGTACTTCGTAACTC	AGG	+
		3'E5	AGATATCACTGCATTACATG	GGG	-
AID	Yeats2	E1	TGTTCGCTTGATTCCAGACA	TGG	-
	Tada3	E1	GGAGGCCGGCCTAATCATGC	AGG	-
		E1	ACAAACCTGCATGATTAGGC	CGG	+
	Zzz3	E1	GTGTTACAAGATCAACAGTG	GGG	+
	Zzz3	Last E	TCCCAGCCAACAGATGACAT	GGG	+
HA-tagged	Tada2a	Last E	AGATAGACGTGAACAAAACC	CGG	-
		Last E	AAGGAATGTGAACAGTCAGA	GGG	-

Supplementary Table 2: Table of primers used for RT-qPCR analysis. This table contains the primers used for RT-qPCR analyses show in this study including the gene name and forward and reverse primer sequences. KO, knockout.

Gene		Forward Primer [5' => 3']	Reverse Primer [5' => 3']
KO validation			
Supt20h	targeted exon	TACATCGTGGAAAGTGCTCAG	ACCTCGGGTTCTTTTTCACAT
	untargeted exon	AAGACAAACTTTTGCTTGAGAGC	CCGGTTATAGAGCAGCCTATTG
Supt7l	targeted exon	ATTGTGGCGACTGCTTGATAG	ACCCAGAGAGTGACTTTTACCG
	untargeted exon/5'UTR	GCAGTTCCCACATAAGAAGCA	AGCCGCGTATACCACTCCT
Tada2b	targeted exon	TACATGCTAACGTAGTGCTCCATC	AGTTCGGCTTCGGCAACT
	untargeted exon	CTGTTCAGTCACACCAGCTACT	CACCAGAAGATGCTGAGCAATG
Tada2a	targeted exon	CTCTGCAGGGCGGCTTATC	CAAGGGAGATCAAGCAGCCATC
	untargeted exon	TCCAGCTGGGATCAAGAACAG	TCACTCGAGGATTTGAGTACAAGA
Pluripotency genes			
	Pou5f1 (Oct4)	CTAGCATTGAGAACCGTGTGAG	GATTGGCGATGTGAGTGATCT
	Sox2	GCGGAGTGGAAACTTTTGT	CGGGAAGCGTGTACTTATCCTT
	Tfcp2l1	ACTACAACCAGCACAACTCTGG	CCCATTCTCAGGAGATAGCTG
	Nanog	CTCCAGCAGATGCAAGAACTC	CTTGCACTTCATCCTTTGGTTT
	Esrrb	GAGGACTCCGCCATCAAAT	TAGTGGTAGCCAGAGGCAATGT
	Klf4	GTGGGTTAGCGAGTTGGAAA	GTGCAGCTTGCAGCAGTAAC
House-keeping genes			
	Rpph1	GGGGGAGAGTAGTCTGAATTGG	CGGAGCTTGGAACAGACTCA
	Rn7sk	TTCCCCGAATAGAGGAGGAC	TGGACCTTGAGAGCTTGTTTG

Supplementary Table 3: Table of antibodies used. The following table shows the references for the antibodies used in this study.

Antibodies	Reference (nr.)
Streptavidin protein coupled to HRP	ThermoFisher Scientific (Cat# 21126)
anti-Supt7l antibody	Bethyl Laboratories (Cat# A302-803A)
anti-γTubulin antibody	Sigma-Aldrich (Cat# T6557)
anti-histone H3 lysine 9 acetylation (H3K9ac) antibody	Abcam (Cat# ab4441)
anti-HA tag antibody	Abcam (Cat# ab9110)
secondary goat anti-rabbit antibody	Jackson ImmunoResearch (Cat# 111-035-144)
secondary goat anti-mouse antibody	Jackson ImmunoResearch (Cat# 115-036-071)
anti-Tbp antibody mouse monoclonal	In-house (3TF1-3G3) (Brou et al., 1993)
anti-Taf7 antibody rabbit polyclonal	In-house (3475) (Bardot et al., 2017)
anti-Taf10 antibody mouse monoclonal	In-house (6TA-2B11) (Jacq et al., 1994)
anti-Taf12 antibody mouse monoclonal	In-house (22TA-2A1) (Mengus et al., 1995)
anti-Atac2 antibody rabbit polyclonal	In-house (2734) (Nagy et al., 2010)

Figure S1 Fischer et al.
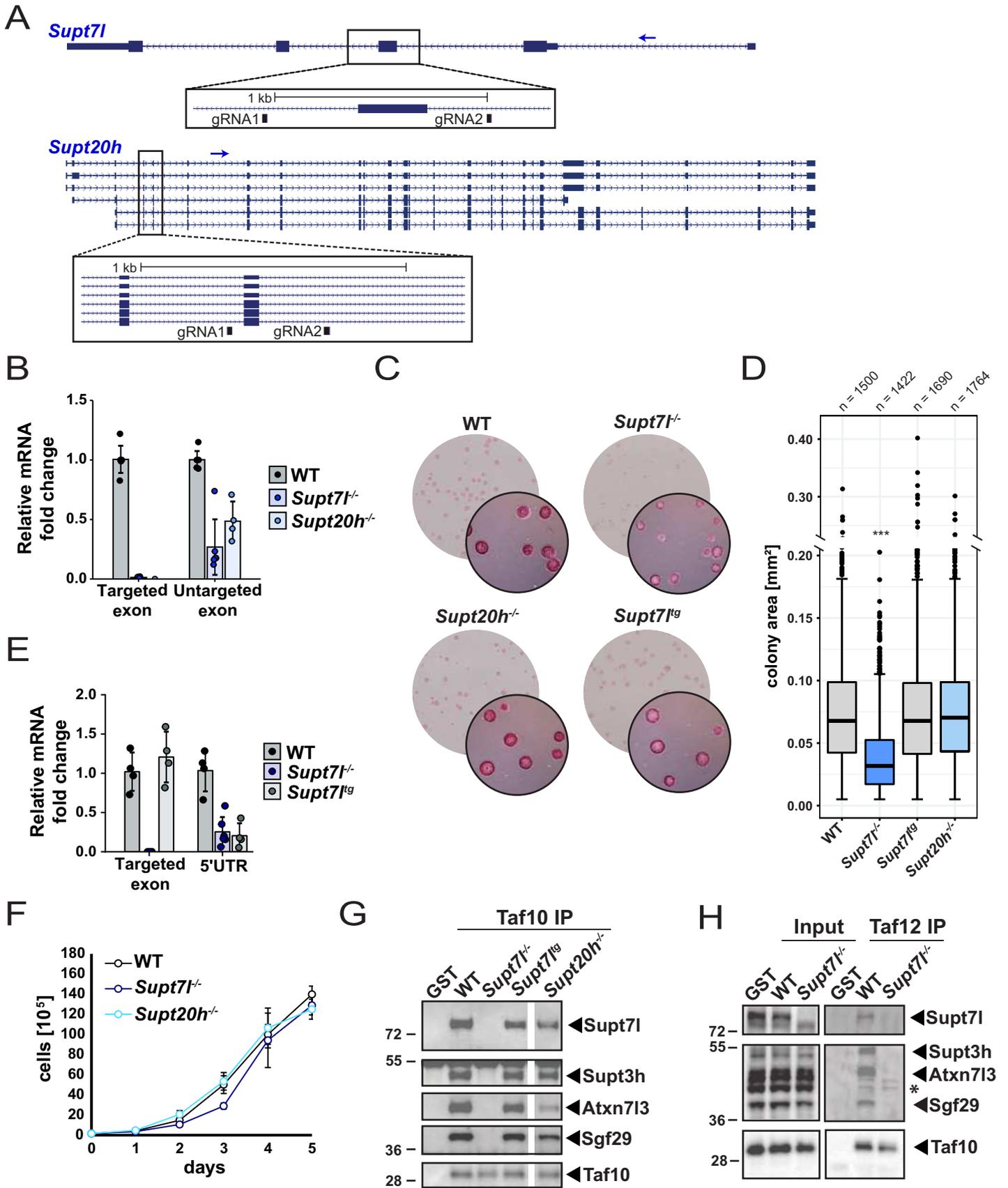
Figure S1, related to Figure 1. Validation of SAGA mutant cell lines. (A) Representation of the *Supt7l* and *Supt20h* loci based on UCSC genome browser views. All RefSeq transcript variants are shown. Blue arrows indicate transcription direction. Inserts highlight sequences targeted by the two gRNAs in introns flanking the first out-of-frame exon shared by all transcript variants. **(B)** RT-qPCR analyses using primers amplifying the deleted exon (left, targeted) or an untargeted exon (right). The untargeted exon primer pair revealed the induction of non-sense mediated decay (NMD). **(C)** Alkaline phosphatase staining on clonal assays of SAGA mutant lines cultured in FCS+LIF+2i medium. Colony morphologies are depicted at higher magnification in foreground images. **(D)** Quantification of the size of colonies analyzed by clonal assays in (**C**). Colony areas were measured with ImageJ, using at least two independent clones (n = 4-9). Statistical test performed is Anova test. ***, $p = < 2 \times 10^{-16}$. Only statistically significant (<0.05) results are indicated. **(E)** Comparison of *Supt7l* mRNA levels in *Supt7l$^{-/-}$* and *Supt7l^{tg}* cell lines. 5'UTR primers represent the same primer pair as the untargeted primer in (C), which is localized within the 5' untranslated region (5'UTR). **(F)** Growth curve analysis of viable cells of *Supt7l$^{-/-}$*, *Supt20h$^{-/-}$* and WT cells grown in FCS+LIF+2i medium. At least two independent clones were analyzed per cell line. Cells were seeded on day 0 and maintained in culture for five days with medium changes every second day. Viable cells were counted every day using trypan blue staining. **(G)** SAGA was purified from WT and *Supt7l$^{-/-}$* cells following the same procedure as in (E top) with the exception that anti-Taf12 antibodies were used. IP against GST (glutathione S-transferase) served as a control for unspecific binding. For (**B** and **E**), RNA polymerase III genes (*Rpph1* and *Rn7sk*) were used for normalization and results were compared to WT cells. Error bars show mean ± SD of at least 4 biological replicates, each the mean of 3 technical RT-qPCR replicates.

A

Yeats2

gRNA1   gRNA2

Zzz3

gRNA1   gRNA2

B

C

D

E

F

G

H
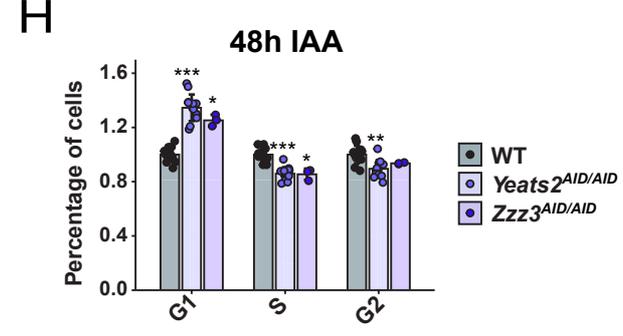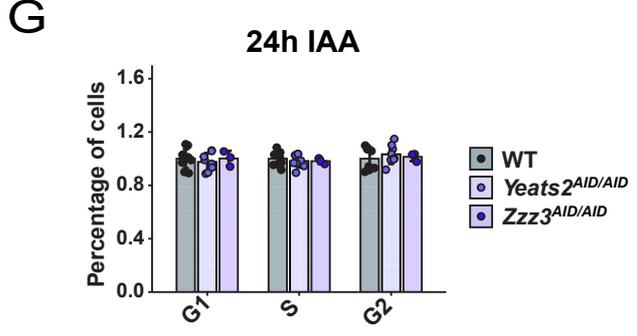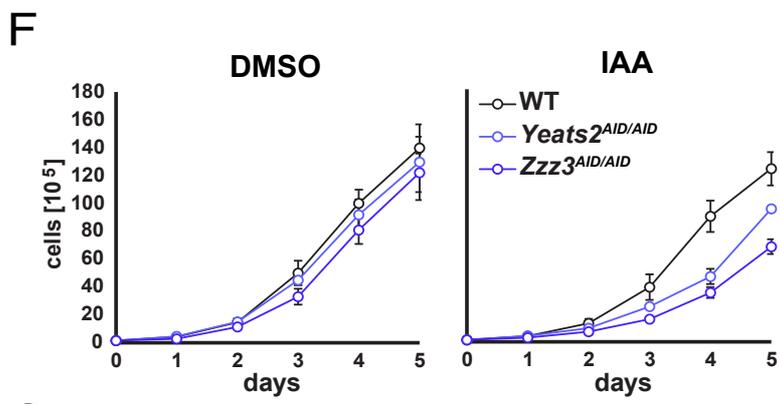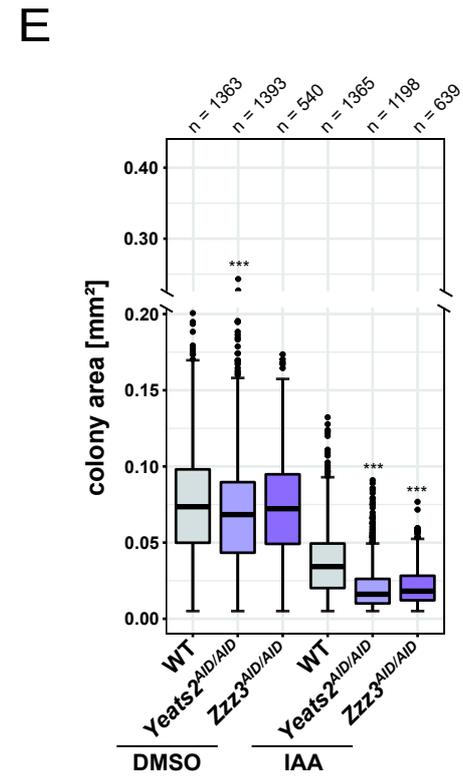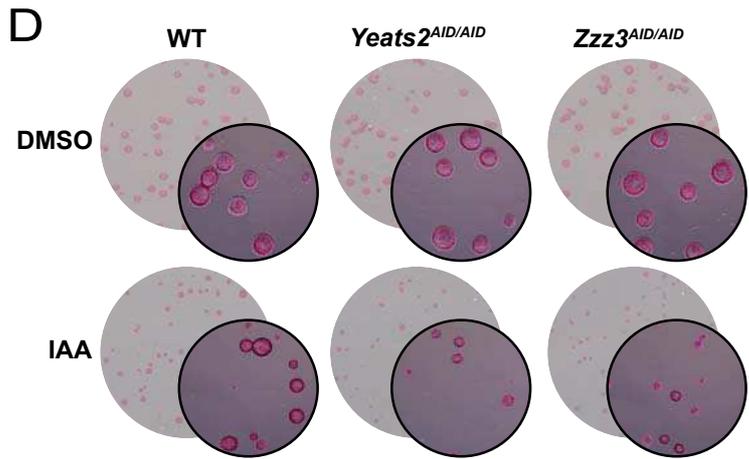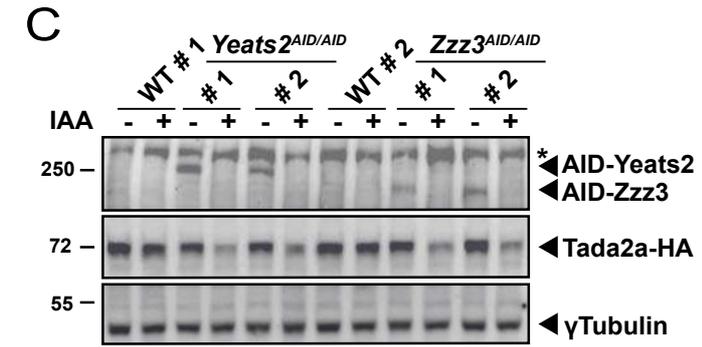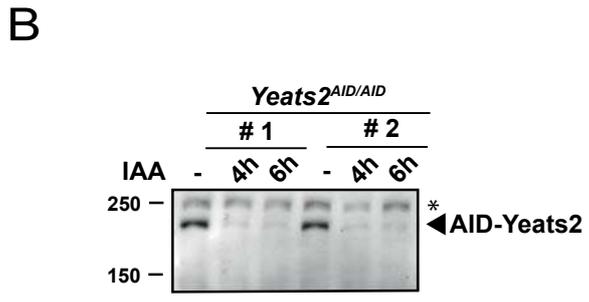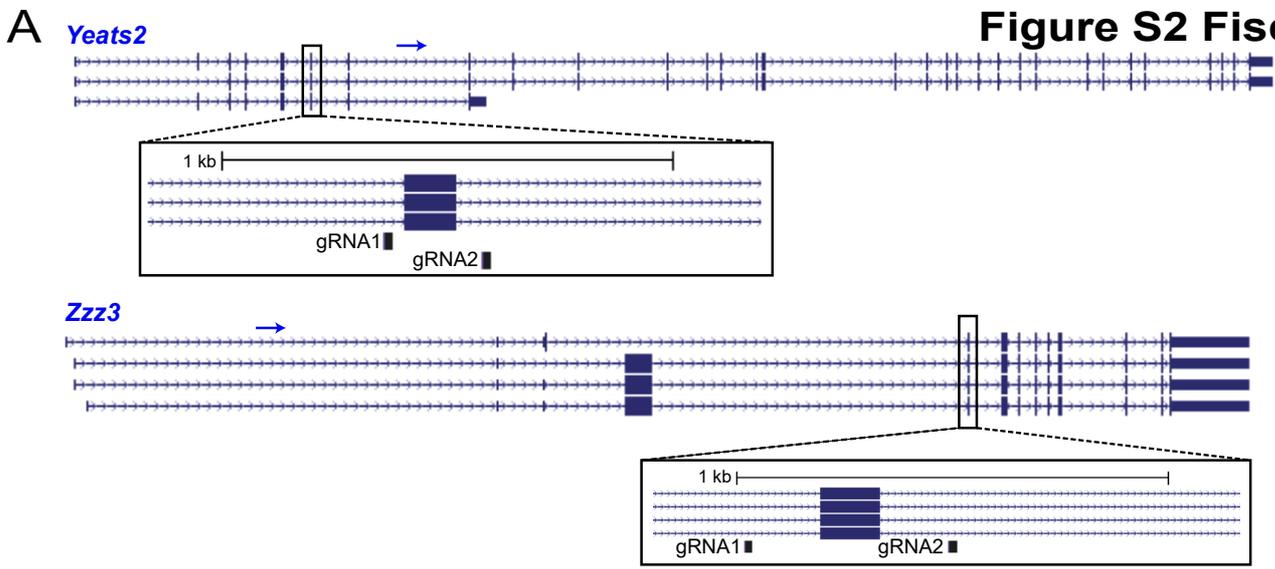
Figure S2, related to Figure 2. Characterization of ATAC mutant cell lines. (**A**) UCSC genome browser views of the *Yeats2* and *Zzz3* loci, showing all RefSeq gene transcript variants. Blue arrows show transcription direction. Insert highlights location of the two gRNAs recruiting Cas9 to the introns flanking the first out-of-frame exon shared by all transcript variants. (**B**) Time course auxin (IAA) treatment of two independent *Yeats2$^{AID/AID}$* cell lines analyzed by western blot using Streptavidin-HRP. Cells were treated with auxin for 4 or 6 hours prior to protein extraction and compared to untreated (-) conditions. Asterisk indicates unspecific band. (**C**) Western blot analyses of whole cell extracts from two independent *Yeats2$^{AID/AID}$* and *Zzz3$^{AID/AID}$* cells in which an HA-tag was fused to the coding sequence of *Tada2a*. Cells were either treated (+) or not (-) with auxin for 24h and western blot were revealed with Streptavidin-HRP or with anti-HA antibody. γTubulin serves as loading control. Asterisk indicates an unspecific band. (**D**) Clonal assays of ATAC mutant cell lines relative to WT cells in FCS+LIF+2i medium treated with either DMSO or IAA and stained with alkaline phosphatase. Images in the foreground represent microscopy images showing colony morphologies. (**E**) Quantification of colony areas of clonal assays as shown in (**D**). Colony areas were measured using ImageJ. At least two independent clones were analyzed per cell line (n = 4-9). ***, p = 0.0004191 for DMSO *Yeats2$^{AID/AID}$* and $p < 2 \times 10^{-16}$ for the remaining conditions. (**F**) Growth curves of viable cells of *Yeats2$^{AID/AID}$* and *Zzz3$^{AID/AID}$* cell lines compared to WT cells in the absence (DMSO) or presence of auxin (IAA). At least two independent clones were analyzed per cell line. Cells were seeded on day 0 and maintained in culture for five days with medium changes every second day. Viable cells were counted every day using trypan blue staining. (**G** and **H**) Cell cycle analyses of *Yeats2$^{AID/AID}$* and *Zzz3$^{AID/AID}$* cell lines were performed using propidium iodide staining after 24h or 48h IAA treatment and compared relative to WT cells. At least two independent clones were analyzed per cell line. Error bars show mean ± SD of at least 4 biological replicates. *, $p < 0.05$; **, p = 0.0024; ***, $p < 9.7 \, 10^{-5}$. For (**E**, **G** and **H**), statistical test performed is Anova test for (**E**) and Wilcoxon rank sum test with Bonferroni correction for multiple testing for (**G** and **H**). Only statistically significant (<0.05) results are indicated.

Figure S3 Fischer et al.
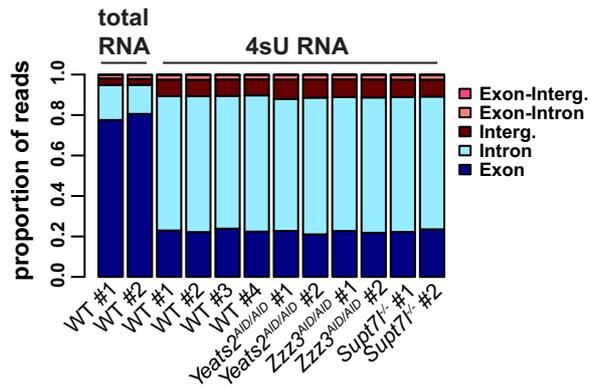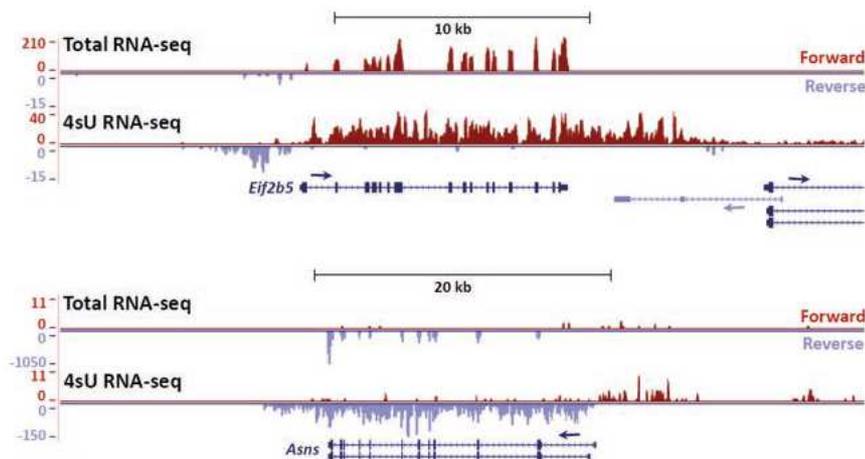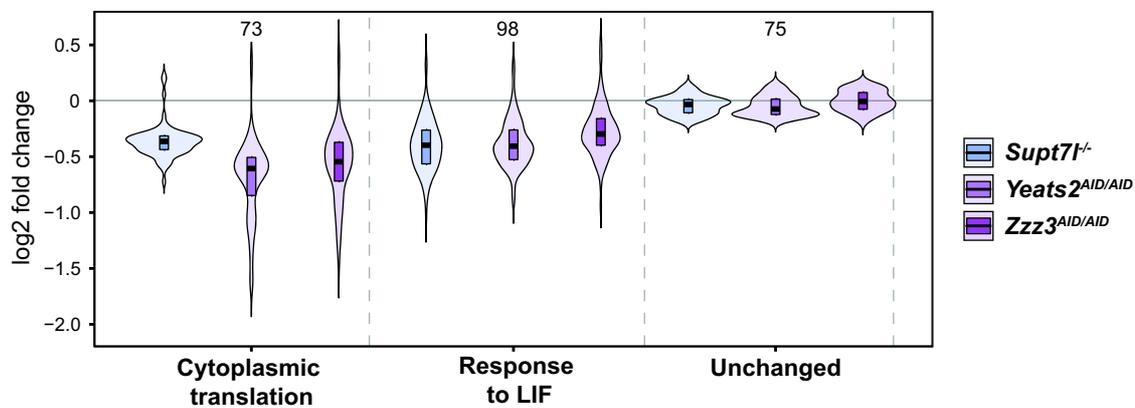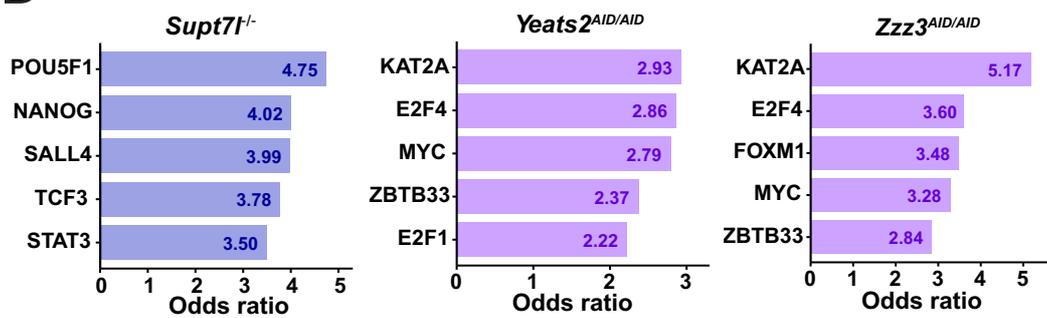
Figure S3, related to Figure 3 and 4. Quantification of newly synthesized RNAs from *Yeats2$^{AID/AID}$*, *Zzz3$^{AID/AID}$* and *Supt7l$^{-/-}$* cell lines. (**A**) Proportion of reads found on genomic elements for total RNA-seq and 4sU-seq experiments shown in Figure 3. Besides reads aligning against exons and introns, reads matching exon-intron junctions, exon-intergenic (exon-interg.) junctions and intergenic (interg.) regions were considered. (**B**) Genome browser views comparing total RNA-seq and 4sU-seq results at two loci (*Eif2b5*, top panel and *Asns*, bottom panel). Blue arrows indicate transcription direction. Forward and reverse strands are shown. (**C**) Violin plots of log2 fold changes comparing the distribution of expression changes of genes belonging to the GO categories 'cytoplasmic translation' (73 genes) and 'response to LIF' (98 genes) to unchanged genes (75 genes). Genes were considered unchanged with an absolute log2 fold change value below 0.2 and an adjusted p-value of > 0.5. (**D**) Transcription factors binding sites from ChEA and Encode ChIP datasets are enriched in the significantly downregulated genes in *Yeats2$^{AID/AID}$*, *Zzz3$^{AID/AID}$* and *Supt7l$^{-/-}$* cell lines as identified by Enrichr. Only the first five transcription factors with the highest odds ratio are shown.
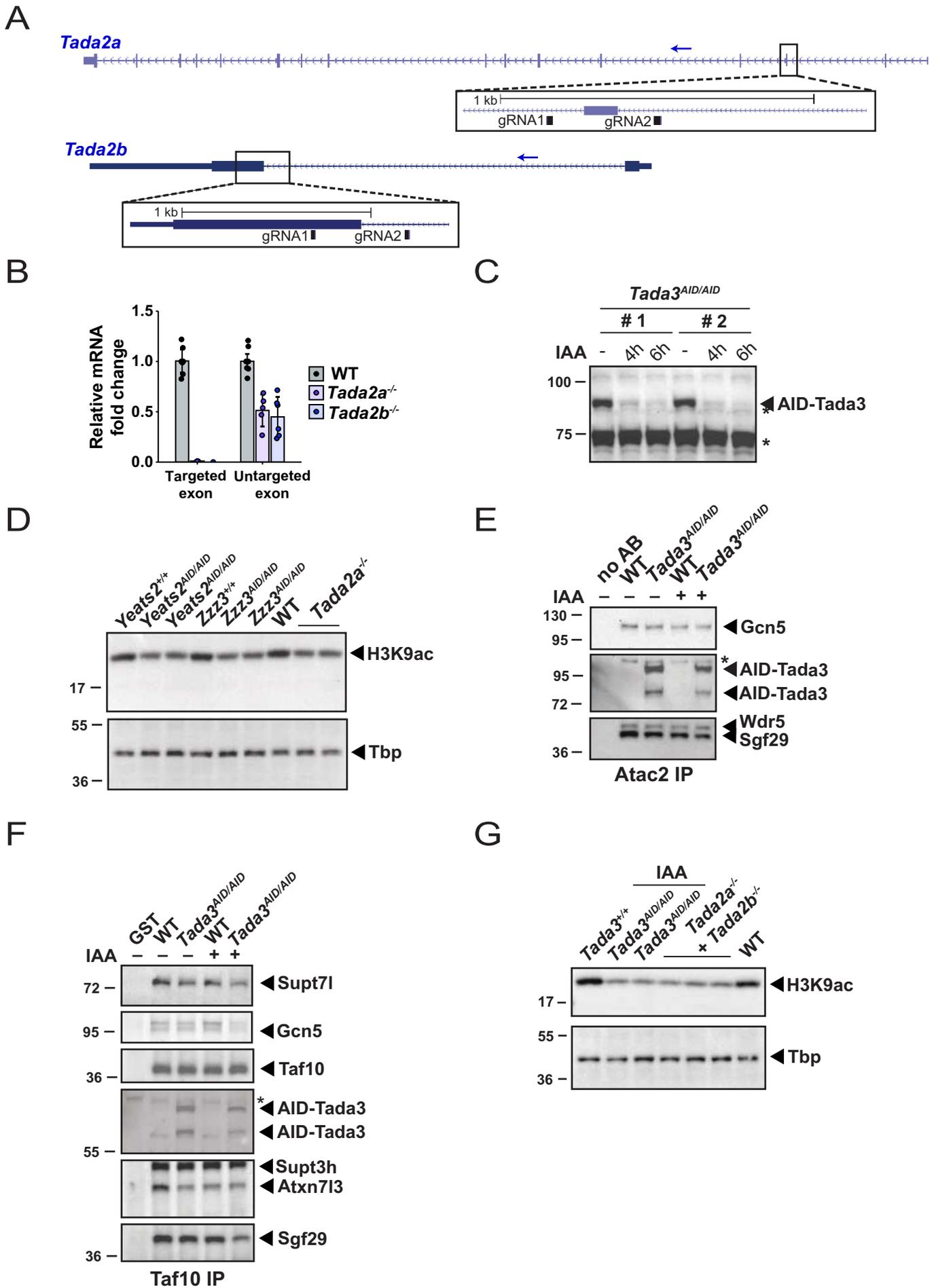
Figure S4, related to Figure 5. Validation of mutant cell lines for subunits of the shared HAT module.
(**A**) UCSC genome browser views of the *Tada2a* and *Tada2b* loci, including all RefSeq gene transcript variants are shown. Blue arrows show transcription direction. Inserts highlight locations of the two gRNAs flanking the first out-of-frame exon of *Tada2a*. For *Tada2b*, the two gRNAs flank the splice acceptor site of the second and last exon. (**B**) mRNA quantification by RT-qPCR in *Tada2a*$^{-/-}$ and *Tada2b*$^{-/-}$ cell lines, using a primer pair located in the targeted exon or primers amplifying an untargeted exon. mRNA levels were normalized to *Rpph1* and *Rn7sk* and results were compared to those in WT cells. Error bars show mean ± SD of at least 5 biological replicates, each the mean of 3 technical RT-qPCR replicates. (**C**) Time course auxin treatment of two independent *Tada3*$^{AID/AID}$ cell lines analyzed by western blot using Streptavidin-HRP. Cells were treated with auxin for 4 or 6 hours and compared to untreated (-) conditions. Asterisks indicate unspecific bands. (**D**) Western blot analyses of histone H3 lysine 9 acetylation (H3K9ac) levels in *Yeats2*$^{AID/AID}$, *Zzz3*$^{AID/AID}$, *Tada2a*$^{-/-}$ and wildtype (WT) cells, treated for 24 hours with IAA. Tbp serves as loading control. (**E** and **F**) ATAC (**E**) and SAGA (**F**) complexes were immuno-purified from WT or *Tada3*$^{AID/AID}$ cells treated with DMSO or IAA for 24 hours, as described in Figure 5C and D, and analyzed by western blots revealed with the indicated antibodies. (**G**) Western blot analyses of H3K9ac levels in *Tada3*$^{AID/AID}$ cell lines treated with IAA or DMSO for 24 hours. Tbp serves as loading control. (**H**) Western blot analyses of H3K9ac levels examined in two independent *Tada3*$^{AID/AID}$ cell lines treated for 24 hours with IAA and three independent *Tada2a*$^{-/-}$ + *Tada2b*$^{-/-}$ double mutant clones compared to WT cells. Tbp serves as a loading control.

Figure S5 Fischer et al.

A

cells [10^5]

WT
Tada2a^{-/-}
Tada2b^{-/-}

days

B

IAA

cells [10^5]

WT
Tada3^{AID/AID}

days

C

WT

Tada2a^{-/-}   *Tada2b^{-/-}*

WT   *Tada3^{AID/AID}*

DMSO

IAA

D

n = 1494   n = 1939   n = 1959   n = 1694   n = 1846   n = 1793   n = 1956

colony area [mm²]

0.40
0.30
0.20
0.15
0.10
0.05
0.00

***   ***   ***   ***

WT   *Tada2a^{-/-}*   *Tada2b^{-/-}*   WT   *Tada3^{AID/AID}*   WT   *Tada3^{AID/AID}*

DMSO   IAA
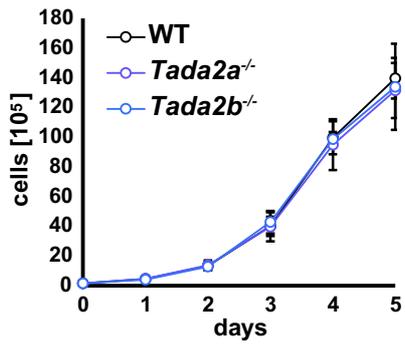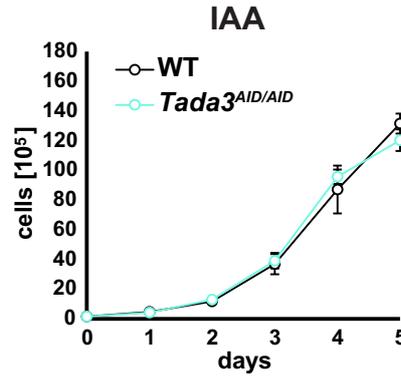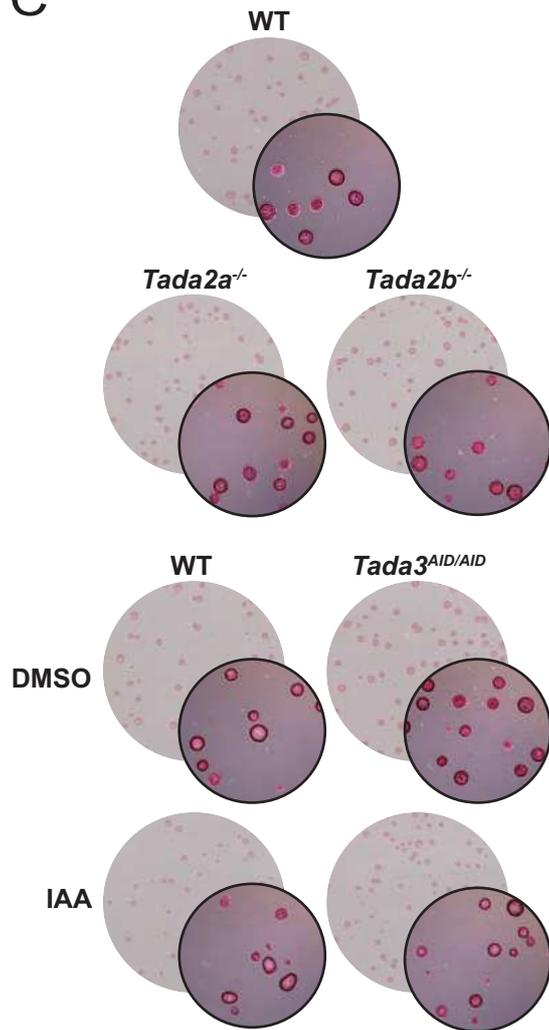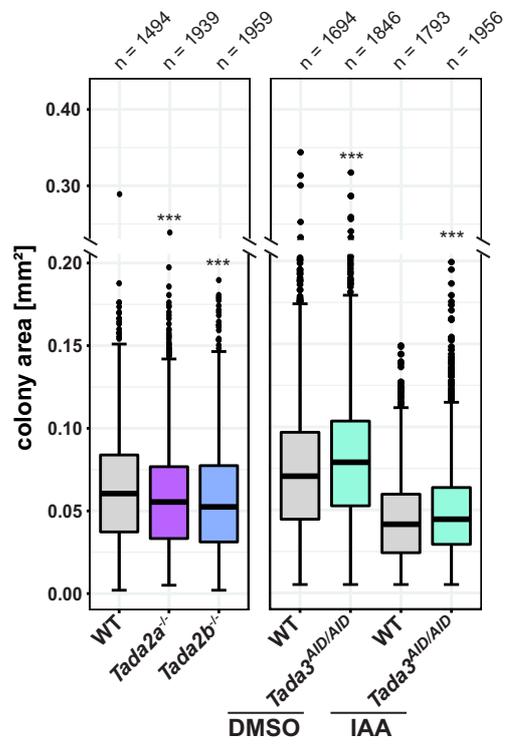
Figure S5, related to Figure 5. Additional analyses of HAT mutant cell lines. (**A** and **B**) Growth curve analyses of *Tada2a*$^{-/-}$, *Tada2b*$^{-/-}$ (**A**) and *Tada3*$^{AID/AID}$ (**B**) cells cultured in FCS+LIF+2i medium compared to WT cells. *Tada3*$^{AID/AID}$ and corresponding WT cells were treated with IAA. At least two independent clones were analyzed per cell line. Cells were seeded on day 0 and maintained in culture for five days with medium changes every second day. Viable cells were counted every day using trypan blue staining. (**C**) Representative images of clonal assay analyses of HAT mutant cell lines cultured in FCS+LIF+2i medium. *Tada3*$^{AID/AID}$ and WT cells were treated with DMSO or IAA and stained for alkaline phosphatase (AP). Images in the foreground represent microscopy images showing colony morphologies. (**D**) Quantification of colony areas in clonal assays of HAT mutant cells as shown in (**C**). Colony areas were measured using ImageJ. At least two independent clones were analyzed per cell line. n = 4-9. Statistical test performed is Anova test. ***, $p < 0.0001795$. Only statistically significant (<0.05) results are indicated.

Figure S6 Fischer et al.

A

$Supt7l^{-/-}$ + Yeats2$^{AID/AID}$ Tada3$^{AID/AID}$

WT   # 1   # 2   WT   # 1   # 2

17 —

◄ H3K9ac
◄ H4K16ac

55 —
36 —

◄ Tbp

B

DMSO

cells [10^5]

—○— WT
—○— $Supt7l^{-/-}$ + Yeats2$^{AID/AID}$

IAA

cells [10^5]

days

C

WT $Supt7l^{-/-}$ + Yeats2$^{AID/AID}$

DMSO

IAA

colony area [mm²]

n = 798   n = 641   n = 773   n = 326

***   ***

WT $Supt7l^{-/-}$ + Yeats2$^{AID/AID}$ WT $Supt7l^{-/-}$ + Yeats2$^{AID/AID}$

DMSO   IAA
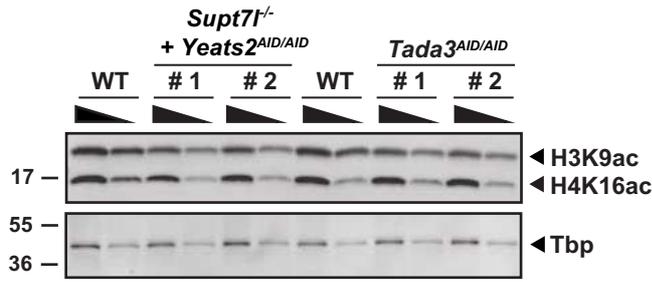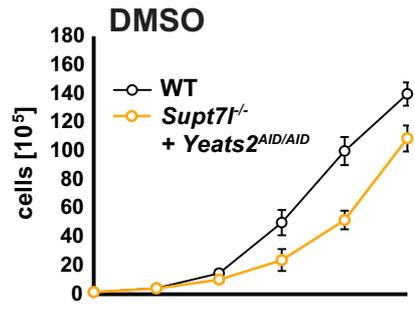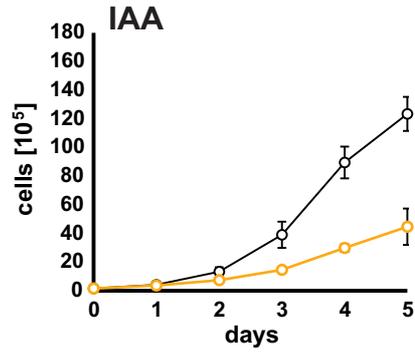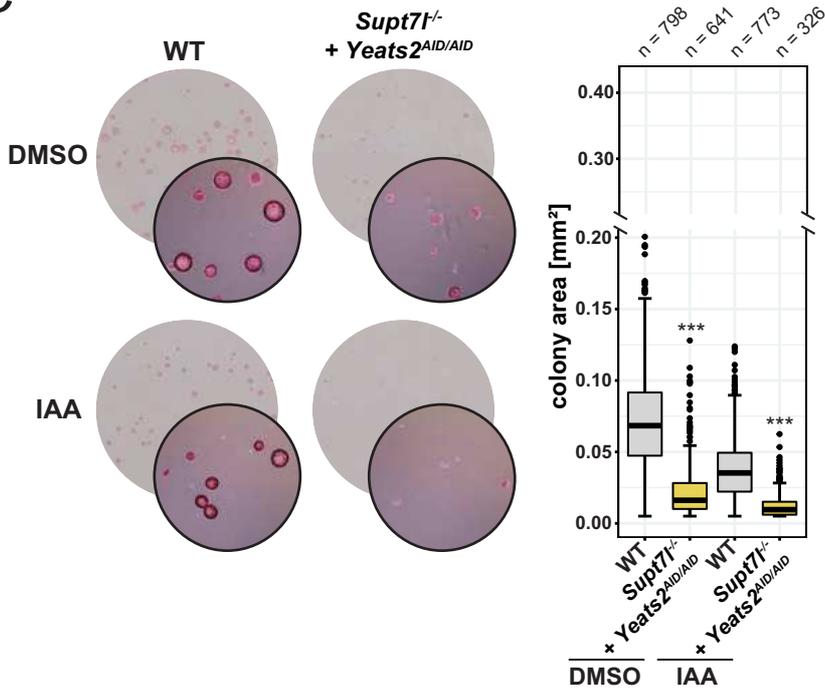
Figure S6, related to Figure 6. Growth analysis of *Supt7l⁻ᐟ⁻ + Yeats2^AID/AID* double mutant cell lines. (A) Western blot analyses of H3K9ac levels examined in two independent *Supt7l⁻ᐟ⁻ + Yeats2^AID/AID* double mutant, *Tada3^AID/AID* and WT cells treated for 48 hours with auxin (IAA). For each sample xx and xx µg of acidic extracts were loaded successively. Tbp serves as loading control. (**B**) Left, clonal assays for *Supt7l⁻ᐟ⁻ + Yeats2*^AID/AID double mutant and WT cells cultured in FCS+LIF+2i medium and treated with DMSO or IAA. Colonies were stained for alkaline phosphatase. Images in the foreground represent microscopy images showing colony morphologies. Right, quantification of colony areas of clonal assays as shown on the left. Colony areas were measured using ImageJ. At least two independent clones were analysed per cell line. n = 2. Statistical test performed is Anova test. ***, p = < 2 x10⁻¹⁶. (**C**) Growth curve analyses of viable cells of *Supt7l⁻ᐟ⁻ + Yeats2*^AID/AID double mutant and WT cells in the absence (DMSO) or presence of IAA. At least two independent clones were analysed per cell line. Cells were seeded on day 0 and maintained in culture for five days with medium changes every second day. Viable cells were counted every day using trypan blue staining

Figure S7 Fischer et al.
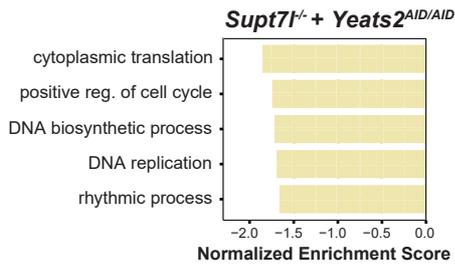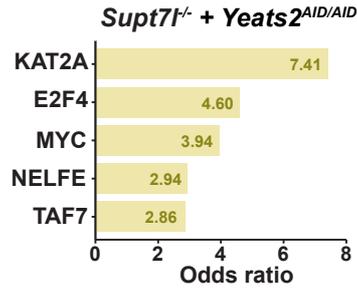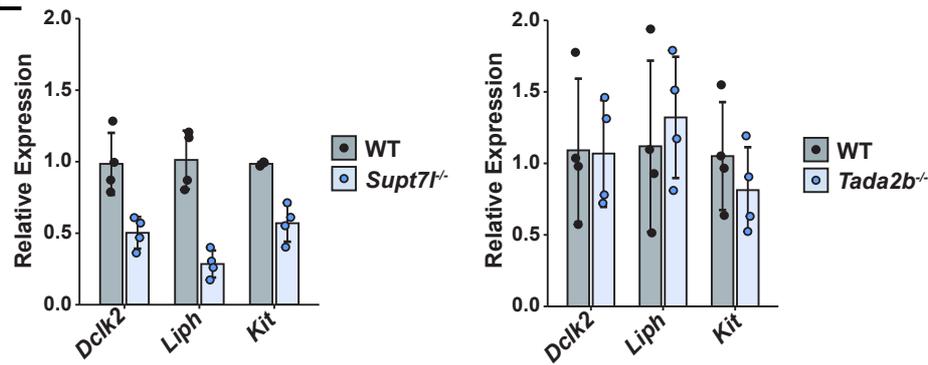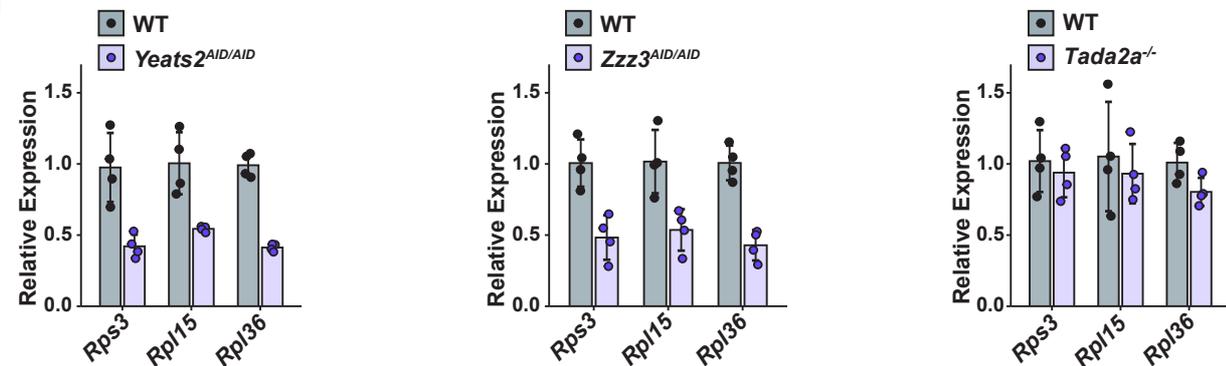
Figure S7, related to Figure 6. Additional analyses of newly synthesized RNA datasets of *Supt7l^{-/-} +* **Yeats2^{AID/AID} and Tada3^{AID/AID} cell lines.** (**A**) Gene set enrichment analysis (GSEA) of GO biological processes based on log2 fold changes observed upon newly synthesized RNA analyses of *Supt7l^{-/-} +* *Yeats2^{AID/AID}* double mutant cells relative to WT cells. Colored bars represent statistically significant terms (FDR < 0.05). Only the first five GO categories with the highest normalized enrichment score are shown. (**B**) Violin plots of log2 fold changes comparing the distribution of expression changes of genes belonging to the GO categories 'cytoplasmic translation' and 'response to LIF' to unchanged genes. Genes were considered unchanged with an absolute log2 fold change < 0.2 and an adjusted *p*-value of > 0.5. Numbers on top of the violin graphs indicate the numbers of genes per category. (**C**) Transcription factors of ChEA and Encode ChIP datasets enriched in the significantly downregulated genes of the *Supt7l^{-/-} + Yeats2*^{AID/AID} double mutant cell lines as identified by the Enrichr database. Only the first five transcription factors with the highest odds ratio are shown. (**D**) Correlation analyses of log2 fold changes of newly synthesized RNA analyses between *Supt7l^{-/-} + Yeats2^{AID/AID}* double mutant cells and *Yeats2^{AID/AID}, Zzz3^{AID/AID}* and *Supt7l^{-/-}* cell lines. (**E**) (**F**).

2. Analysis of the impact of loss of Supt3h on self-renewal capacities of mouse embryonic stem cells and RNA polymerase II transcription.

This result section contains preliminary and unpublished data from experiments I designed and performed on *Supt3h$^{-/-}$* cell lines, which were generated with the help of Matthieu Stierle. The experiments and work were conceived by Didier Devys.

Introduction

Early studies, based on a genetic screen involving the Ty transposon, found a genetic link between subunits of the yeast SAGA complex (Spt3, Spt7, Spt8, Spt20) and TBP (Spt15) (Winston et al., 1984b; Winston et al., 1987; Eisenmann et al., 1989; Roberts & Winston, 1996). Interestingly, mutations in Spt3 were subsequently found to rescue mutations of TBP, suggesting a physical interactions between these two proteins (Eisenmann et al., 1992). Recent cryo-EM structures further highlight the interaction between SAGA subunits and TBP (Papai et al., 2020): The two subunits Spt3 and Spt8, which are part of the core of yeast SAGA, were found to form contacts with TBP on opposite sides. Interestingly, when TBP is bound by SAGA, its DNA-binding domain is inaccessible for DNA, implying that TBP needs to be released from SAGA to bind to gene promoters (Papai et al., 2020). Therefore, yeast SAGA was demonstrated to possess a TBP-loading function through genetic, biochemical and structural studies, which requires its Spt3 and Spt8 subunits.

As inactivation of the HAT module could not explain phenotypes and transcriptional defects observed upon inactivation of the SAGA core subunit Supt7l (see result section 1), we were interested to assess the importance of the putative TBP-loading function of SAGA on mouse ESC physiology and RNA polymerase II (Pol II) transcription. Although, SAGA is highly conserved from yeast to humans, metazoan genomes lack a homolog of the *SPT8* gene of budding yeast (Spedale et al., 2012). Therefore, the putative TBP-loading function of metazoan SAGA only involves its Supt3h subunit (yeast Spt3). To answer **aim d)** and **aim e)** in relation to the importance of the putative TBP-loading function of the mammalian SAGA complex for the self-renewal capacities of mouse embryonic stem cells (ESCs) and for RNA polymerase II transcription, we inactivated *Supt3h* in mouse ESCs.

Results

Following the same strategy as described in result section 1, we deleted the first out-of-frame exon, which is shared by all transcript variants and which does not contain the translation start codon, of *Supt3h* (exon 3) using the CRISPR-Cas9 technology (scheme of the gRNA positions in Figure 51A). We could observe a frequency of heterozygous clones similar to previous constructs (see result section 1, Figure 1B, 2A or Figure 5A) and obtained two clones with homozygous deletion of the targeted

region, which we verified by RT-qPCR (Figure 51B and 51C). Deletion of the out-of-frame exon caused the induction of non-sense mediated mRNA decay (NMD), as indicated by reduced *Supt3h* mRNA levels when assessed by primers against an untargeted exon (Figure 51C).
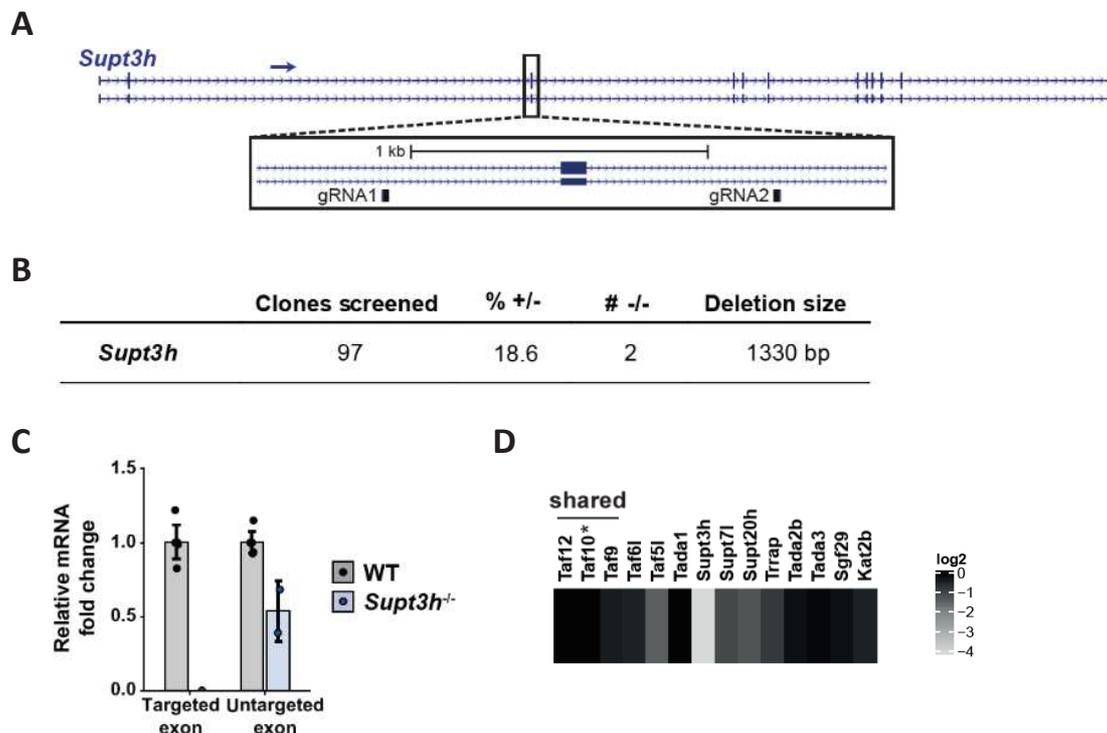
Figure 51: Generation and validation of *Supt3h*[-/-] cell lines. A. Representation of the locus of *Supt3h* based on UCSC genome browser views. All RefSeq gene transcript variants are shown. Small blue arrows show transcription direction. Insert highlights location of the two gRNAs used to recruit Cas9 to the introns surrounding the first out-of-frame exon shared by all transcript variants and excluding exons containing the translation start site. Scale bars reflect 1 kb. **B.** Table showing number of clones screened, percentage of heterozygous (+/-) clones, number of homozygous (-/-) clones and size of the deletion for *Supt3h*. **C.** Verification of *Supt3h*[-/-] cell lines by RT-qPCR analysis using primers against the exon targeted for deletion (targeted exon) and an exon outside of the deletion region (untargeted exon). The untargeted exon primer pair served to examine the induction of non-sense mediated decay (NMD). RNA polymerase III genes (*Rpph1* and *Rn7sk*) were used for normalization and results were compared to wildtype (WT) cells. Error bars show mean ± SD of at least 2 biological replicates, each the mean of 3 technical RT-qPCR replicates. **D.** Immunoprecipitation of the SAGA complex using anti-Taf10 IP followed by mass spectrometry analysis in *Supt3h*[-/-] cell lines. Results of MS analysis are displayed as log2 difference relative to SAGA purified from WT cells. Star indicates bait protein.

We further performed immunopurification (IP) experiments of SAGA in *Supt3h*[-/-] cells coupled to mass spectrometry analysis (MS), to assess the impact of loss of Supt3h on SAGA structure. Therefore, nuclear extracts from *Supt3h*[-/-] and wildtype (WT) cells were depleted for the TFIID complex using antibodies targeting the TFIID-specific subunit Taf7. This depletion of TFIID was necessary to increase the purification efficiency for SAGA by the subsequently IP against Taf10, one of the core subunits of SAGA present also in TFIID. Subsequent MS analyses of the anti-Taf10 IP elutions revealed limited effects of loss of Supt3h on SAGA structure as most SAGA subunits were found except for Supt3h (Figure 51D). These results are in agreement with studies from budding yeast (Lee et al., 2011).
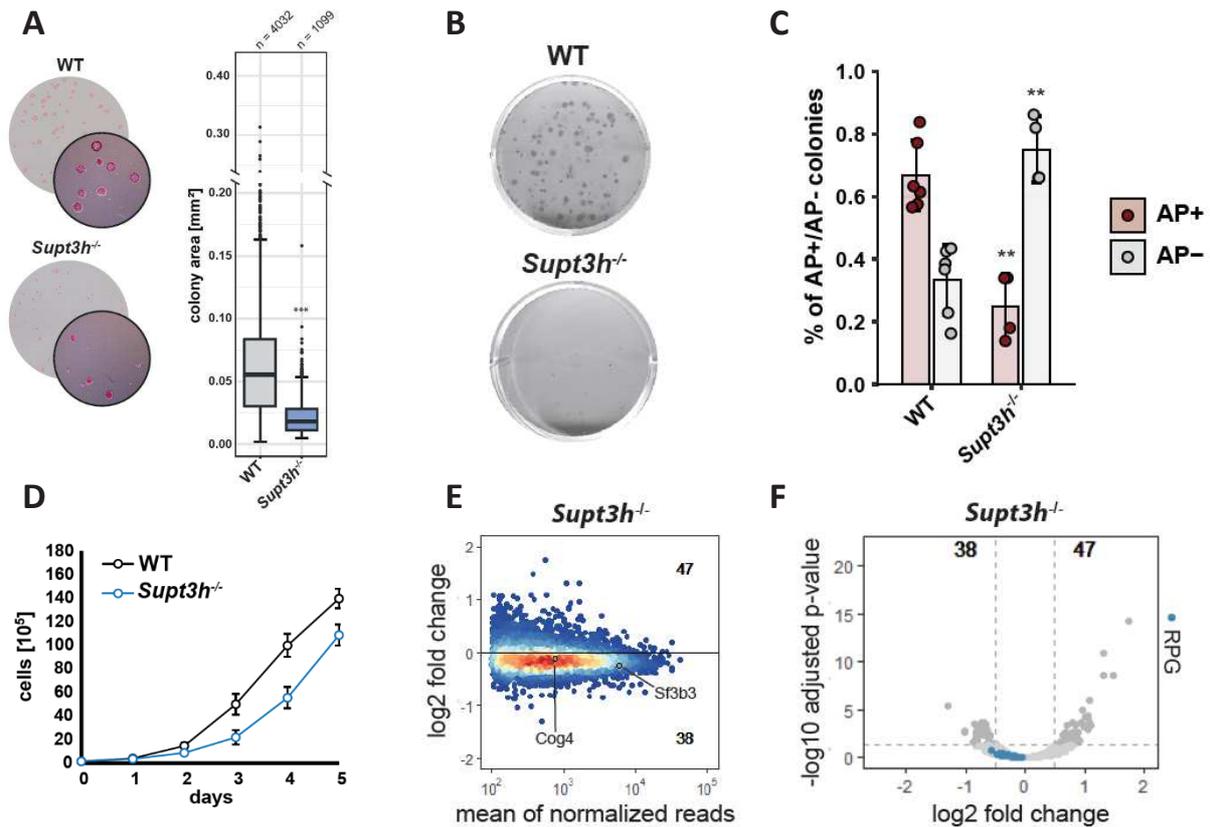
Figure 52: Characterisation of phenotypes and analysis of newly synthesized RNA levels of *Supt3h*−/− cells. A. Left, representative images of clonal assay analyses of *Supt3h*−/− cell lines inFCS + LIF + 2i medium compared to wildtype (WT) cells and stained for alkaline phosphatase levels. Images in the foreground represent microscopy images showing colony morphologies. Right, quantification of colony areas of clonal assays as shown on the left. Colony areas were measured using ImageJ. n = 5-9. Statistical test performed is one-sided Wilcoxon-Mann-Whitney test. ***, $p = < 2.2 \times 10^{-16}$. **B.** Representative images of clonal assay analyses of *Supt3h*−/− cells in medium containing LIF relative to WT cells and stained with crystal violet. **C.** Quantification of alkaline phosphatase (AP) staining of clonal assays as shown in (B). Numbers of AP positive (AP+) and AP negative (AP-) colonies were normalized to total number of colonies as assessed by crystal violet staining. Error bars show mean ± SD of at least 4 biological replicates. Statistical test performed is two-sided Wilcoxon-Mann-Whitney test. **, $p = 0.009524$. **D.** Growth curves analysis of viable cells of *Supt3h*−/− cell lines compared to WT cells in medium containing 2i. Cells were seeded on day 0 and maintained in culture for five days with medium changes every second day. Viable cells were counted every day using trypan blue staining. **E.** Density plot represe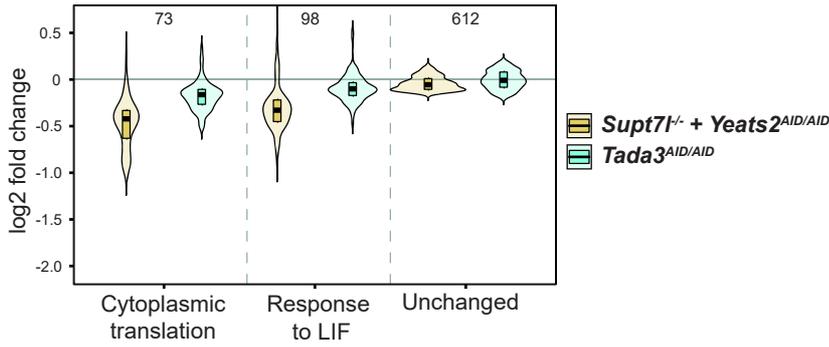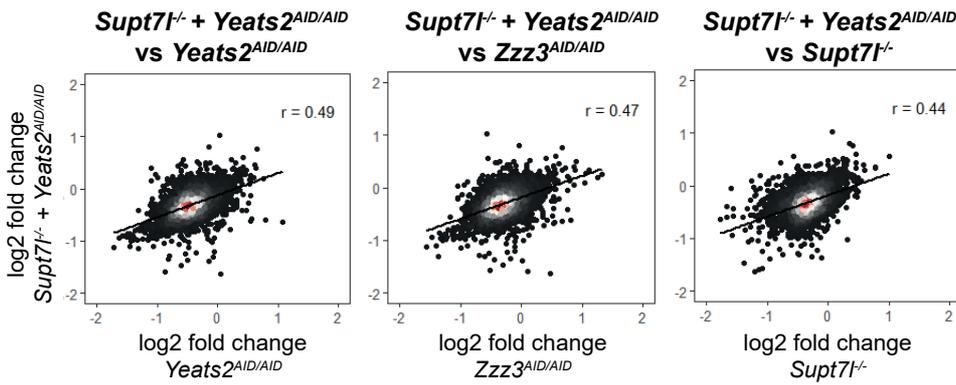ntation of log2 fold change of newly synthesized RNA levels of two independent clones of *Supt3h*−/− cell lines relative to WT cells against the mean of normalized reads. Cell lines were treated for 24 hours with IAA. Localization of the *Sf3b3* and *Cog4* genes are highlighted. Numbers at the top and bottom indicate the numbers of significantly up- and downregulated genes, respectively. Genes were considered as significantly affected with an adjusted p-value of < 0.05 and log2 fold change of < -0.5 or > 0.5. A threshold of 100 reads was set to define expressed genes. **F.** Volcano plot representation of newly synthesized RNA analyses of *Supt3h*−/− cells compared to WT cells showing the adjusted p-values relative to log2 fold changes. Numbers on top left and right indicate significantly down- and upregulated genes, respectively. Genes were considered as significantly affected with an adjusted p-value of < 0.05 and log2 fold change of < -0.5 or > 0.5 as indicated by the dashed lines. Ribosome protein genes (RPG) are highlighted by blue dots.

To reveal the potential impact of loss of Supt3h on mouse ESC growth, we performed clonal assay analyses in medium containing serum and LIF (FCS + LIF medium) or medium additionally containing two potent inhibitors of differentiation (FCS + LIF + 2i medium). For clonal assays, cells were seeded at low density and grown for six days before fixation and staining. Surprisingly, we found that *Supt3h*−/− cells displayed on average significantly smaller colonies in FCS + LIF + 2i medium (Figure

52A) and strongly reduced colony formation in FCS + LIF medium (Figure 52B) compared to WT cells. Growth curve analyses of viable cells in FCS + LIF + 2i medium further suggested that the smaller colony areas observed for *Supt3h*^{-/-} cells in FCS + LIF + 2i medium could be due to reduced proliferation rates (Figure 52D).

Following the strong impact of loss of Supt3h on mouse ESC growth, we were further interested in assessing the self-renewal capacities of *Supt3h*^{-/-} ESCs. Therefore, we performed similar clonal assays in FCS + LIF medium as shown in Figure 52B with additionally staining for alkaline phosphatase (AP) levels. Undifferentiated colonies possess high levels of AP revealed by a red coloration in this assay (AP positive), while differentiating colonies remain unstained (AP negative). AP positive and AP negative colonies were counted and normalized to the total amount of colonies as assessed by crystal violet staining. These analyses indicated a major impact of inactivation of *Supt3h* on mouse ESC self-renewal (Figure 52C). While WT cells displayed on average a ratio of roughly 70% undifferentiated and 30% differentiated colonies, *Supt3h*^{-/-} cells displayed the opposite trend with roughly 30-40% differentiated and 60-70% undifferentiated colonies (Figure 52C).

In general, these phenotypic analyses indicated a major impact of loss of Supt3h on mouse ESC growth and self-renewal. In contrast to inactivation of the gene encoding the SAGA-specific HAT subunit Tada2b (see Figure 5F and 5G as well as Figure S5 in result section 1), phenotypes of *Supt3h*^{-/-} cells resembled phenotypes observed in *Supt7l*^{-/-} cells (see Figure 1D and 1F as well as Figure S1C-S1D and S1F in result section 1), suggesting that the TBP-loading function might be crucial for the role of the SAGA core in mouse ESC proliferation and self-renewal.

Due to the resemblance of the phenotypes between *Supt3h*^{-/-} and *Supt7l*^{-/-} cells, we were interested in assessing if loss of Supt3h would cause similar effects on newly synthesized RNA levels as loss of Supt7l. However, newly synthesized RNA analysis in *Supt3h*^{-/-} cells revealed that the transcriptional defects upon inactivation of the putative TBP-loading function of mammalian SAGA are much less severe than upon inactivation of its core subunit Supt7l (compare Figure 52E and 52F to Figure 3B and Figure 4B of result section 1). Indeed, very few genes were found significantly downregulated (38 genes) and the expression of RPGs was also not affected by the loss of Supt3h (Figure 52E and 52F). Therefore, although inactivation of *Supt3h* recapitulates the self-renewal defects observed upon inactivation of *Supt7l*, it did not recapitulate all transcriptional effects.

Overall, these preliminary findings suggest that Supt3h and the putative TBP-loading function of SAGA is especially required for mouse ESC growth and self-renewal, while RNA polymerase II transcription might additionally be sensitive for other functions of SAGA, such as its deubiquitylation (DUB) module or its transcription factor-interacting subunit Trrap. The mechanistic details of the functions of Supt3h in mouse ESC proliferation and self-renewal remain unclear.

Contributions

<u>Veronique Fischer</u> – Generated cell lines. Designed and performed experiments, analysis and graphical representations of data shown in Figure 51 and Figure 52.

Matthieu Stierle – Helped in generating and validating cell lines.

Bernardo Reina San Martin – designed the CRISPR-Cas9 plasmids.

Didier Devys – Conceived the work and designed experiments.

3. Analysis of the impact of loss of the deubiquitylation module subunit Atxn7l3 on mouse embryonic stem cells.

This result section contains preliminary and unpublished data from experiments I designed and performed on *Atxn7l3⁻/⁻* cell lines, which were generated with the help of Matthieu Stierle. The experiments and work were conceived by Didier Devys.

Introduction

Besides its HAT and TBP-loading functions, which we studied extensively in result section 1 and 2, SAGA also possesses a DUB activity. In contrast to the two other functions, which have a role especially during the step of transcription initiation, the DUB module of SAGA is involved in the removal of H2BK120ub following the transcription elongation machinery.

In mammalian cells, H2BK120ub is catalysed by RNF20 and RNF40, which are thought to associate with the elongating RNA polymerase II (Pol II) through Paf1C thereby modifying histone H2B proteins in proximity to the transcription elongation complex (Kim et al., 2009). H2BK120ub is generally found at all actively transcribed genes along the length of the gene body and was suggested to facilitate nucleosome reassembly along the path of elongating Pol II (Fleming et al., 2008; Bonnet et al., 2014). Recent structural findings in yeast further highlight the importance of monoubiquitylation of H2B in enabling methylation of H3K4 and H3K79 (Worden et al., 2019; Valencia-Sánchez et al., 2019; Hsu et al., 2019b). The importance of removal of H2BK120ub through the DUB of SAGA remains however largely unclear.

In the previous result section we had focused our attention especially on the HAT and TBP-loading functions of SAGA as these two functions were found to be especially required for transcription regulation in budding yeast (Baptista et al., 2017). In contrast, deletion of Ubp8, the catalytic subunit of the DUB module of yeast SAGA and ortholog to mammalian Usp22, did not cause detectable effects on Pol II transcription although levels of H2BK123ub were drastically increased (Baptista et al., 2017). Nevertheless, we were also interested in assessing the importance of the DUB module and consequently removal of H2BK120ub in mammalian cells.

To inactivate the DUB module of SAGA in mouse embryonic stem cells and to respond to **aim e)** in relation to the functions of the DUB module in self-renewal and proliferation of mouse ESC, we decided to inactivate one of the four subunits of the DUB module (Atxn7l3, Atxn7, Usp22 or Eny2). As Eny2 is also present within the TREX2 complex and therefore involved in the process of RNA export, we rapidly excluded this subunit as a target for inactivation (Rodríguez-Navarro et al., 2004; Helmlinger & Tora, 2017). Additionally, as mentioned in the Introduction (see chapter 4.3. Subunit composition of mammalian SAGA), in mammalian cells, a triplication event led to the presence of three seemingly

redundantly acting Atxn7 proteins, the subunit responsible in integrating the DUB module within SAGA (Vermeulen et al., 2010). Based on the challenge imposed by inactivating several genes, we consequently also excluded this subunit.

At the time of these experiments, genetic predictions suggested that besides Usp22 two additional Usp proteins, Usp27x and Usp51, might be incorporated within the SAGA DUB module in a mutually exclusive way to Usp22. Based on this prediction, inactivation of the catalytic subunit of the DUB module would also require the inactivation of three independent genes. Consequently, we ended up with only one remaining possibility, Atxn7l3. Atxn7l3 represents a key subunit of the DUB module which has several important roles within the module by activating the catalytic site of Usp22 and recognizing the substrate through interactions with the acidic patch of histone H2A/H2B dimers (Samara et al., 2010; Köhler et al., 2010; Morgan et al., 2016). A gene duplication event also caused the presence of a paralog, Atxn7l3b, in mammalian cells, which however was reported to not function within SAGA as it mainly localizes to the cytoplasm (Li et al., 2016). Therefore, we targeted only *Atxn7l3* for inactivation by CRISPR-Cas9 in mouse ESCs.

When we initiated the below described experiments, the presence of SAGA-independent DUB modules, mentioned in the Introduction, was not yet demonstrated, although the presence of the two alternative Usp proteins, Usp27x and Usp51, had been genetically predicted and found to be expressed in human cells (Kobayashi et al., 2015; Atanassov et al., 2016). The identified SAGA-independent DUB modules are composed of ATXN7L3, ENY2 and either USP27X or USP51 (Atanassov et al., 2016). As these DUBs do not seem to interact with ATXN7 proteins, they are thought to not be incorporated within SAGA and to rather reflect 'free', SAGA-independent DUB modules (Atanassov et al., 2016). For the results described in this section, it is therefore important to keep in mind that with the inactivation of *Atxn7l3*, we not only target the DUB module of SAGA but also the Usp27x- and Usp51-containing, free DUB modules.

Results

To inactivate *Atxn7l3*, we targeted exons 3 and 4 for deletion using CRISPR-Cas9 following the same strategy as described in the previous results sections (scheme of the gRNA positions in Figure 53A). Loss of these two exons is thought to cause a shift within the ribosome reading frame subsequently generating a scrambled protein downstream of exon 2. We obtained four clones with homozygous deletion of the targeted region on the genomic level, which were further verified by Sanger sequencing (example of genotyping and sequencing shown in Figure 53B and 53C). Subsequent analyses of RNA expression levels of *Atxn7l3* by qPCR in the homozygous mutant cell lines compared to wildtype (WT)
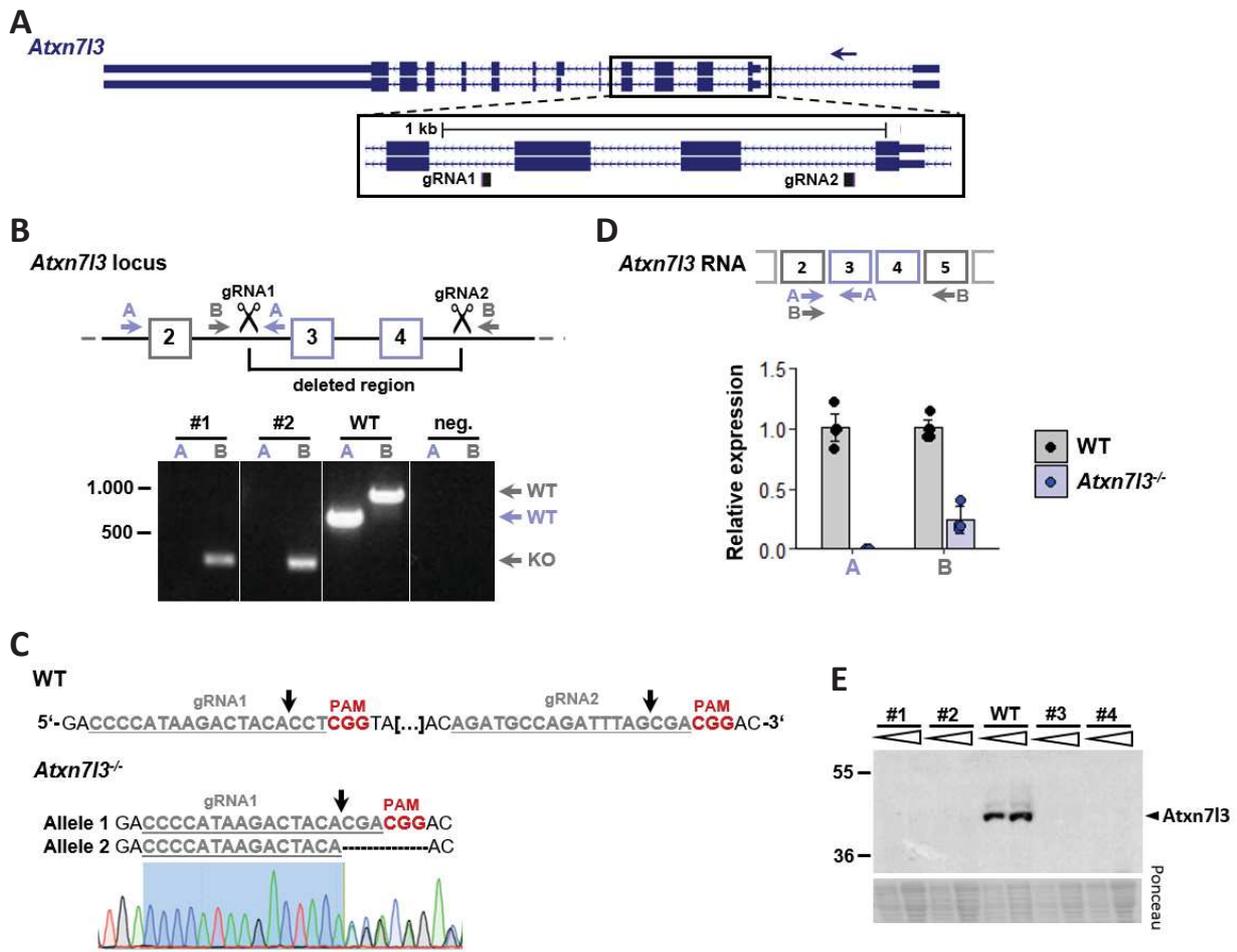
Figure 53: Generation and characterisation of *Atxn7l3*-/- mouse embryonic stem cell lines. A. UCSC genome browser view on the locus of *Atxn7l3* with its two transcript variants. The insert highlights the position of the guide RNAs (gRNAs) localized to the surrounding introns of exons 3 and 4. Arrow indicates the direction of transcription. **B.** Examples of genotyping of *Atxn7l3*-/- clones. Top, scheme displaying the position of the gRNAs (scissors indicate the double strand breaks induced by Cas9) as well as primer pairs (A and B) used for the genotyping PCR. One primer of primer pair A is situated within the deleted region and therefore not amplifying in homozygous mutant clones, while primer pair B is located outside of the deleted region and is consequently shorter in the homozygous mutant clones. Bottom, examples of PCR results are shown for two *Atxn7l3*-/- clones compared to wildtype (WT) and negative (neg.) controls. **C.** Example of sequencing results for one *Atxn7l3*-/- clone compared to WT sequences. Sequencing histogram is show (gRNA1 highlighted in blue) with the respective DNA sequences identified for the two alleles displayed on top. The two alleles were perfectly cut at the expected sites (indicated by arrows) of gRNA1. Similarly, allele 1 shows a perfect cut at the expected site of gRNA2, while allele 2 displays a deletion of the complete gRNA2 including the PAM sequence. gRNA sequences are underlined and in grey and PAM sequences are shown in dark red. PAM, protospacer adjacent motif. **D.** Analysis of *Atxn7l3* RNA levels in *Atxn7l3*-/- clones compared to WT cells. Top, scheme indicating the positions of the primer pairs used. Bottom, expression of *Atxn7l3* as assessed by primer pairs A and B in *Atxn7l3*-/- clones relative to WT cells. Expression was normalized to two RNA polymerase III genes (*Rpph1* and *Rn7sk*). Reduced levels of *Atxn7l3* revealed by primer pair B indicates the induction of non-sense mediated decay (NMD). **E.** Western blot analysis of the four *Atxn7l3*-/- clones compared to WT cells showing the loss of Atxn7l3 protein levels in the mutant cells.

cells, further demonstrated the loss of the targeted exons (primer pair A in Figure 53D). We further found an overall reduction of *Atxn7l3* mRNA levels in the *Atxn7l3*-/- cell lines likely induced by non-sense mediated decay (NMD) suggesting the generation of a premature stop codon upon loss of exons 3 and 4 (primer pair B in Figure 53D). Western blot analysis on whole cell protein extracts further confirmed the loss of Atxn7l3 in the *Atxn7l3*-/- cell lines (Figure 53E).

Additionally, we could observe that loss of Atxn7l3 resulted in an increase of global H2BK120ub levels as assessed by acidic extraction of histone proteins from the mutant cell lines and compared relative to WT cells (Figure 54A). The increased levels of H2BK120ub were comparable to previously reported observations upon depletion of Atxn7l3 in human cells (Bonnet et al., 2014). To assess the impact of inactivation of *Atxn7l3* on mouse ESC physiology, we performed on one hand growth curve analyses to evaluate effects on proliferation (Figure 54B) and on the other hand clonal assay analyses in FCS + LIF + 2i medium followed by alkaline phosphatase (AP) staining to determine effects on colony formation capacities (Figure 54C). Additionally, we assessed the impact of loss of Atxn7l3 on mouse ESC self-renewal by performing clonal assay analyses in FCS + LIF medium, which allows the proliferation of both undifferentiated and differentiating ESCs in contrast to FCS + LIF + 2i medium, which selects for undifferentiated ESCs (Figure 54D. Using AP staining in the FCS + LIF medium clonal assays, we were able to determine the numbers of undifferentiated or mixed colonies (AP positive colonies) or differentiating colonies (AP negative colonies) and calculate relative frequencies by using the total number of colonies as assessed by crystal violet staining.

Overall, we found that growth of *Atxn7l3*$^{-/-}$ cell lines was affected in FCS + LIF + 2i medium with the estimated doubling time in the exponential phase reaching 17 hours compared to 14 hours in WT
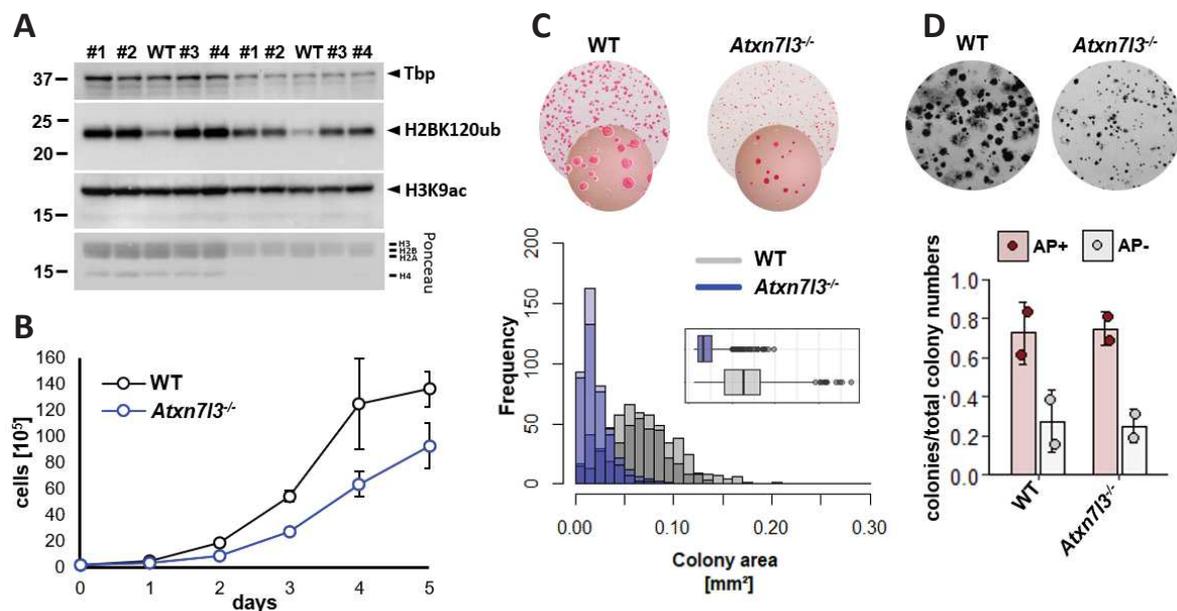
Figure 54: Effects of loss of Atxn7l3 on histone modifications and mouse embryonic stem cells physiology. **A.** Western blot analysis of acidic histone extracts from the four *Atxn7l3*$^{-/-}$ clones (indicated by numbers) compared to wildtype (WT) cells. While H3K9ac levels seem largely unchanged similar to the loading control Tbp, H2BK120ub levels are drastically increased in the *Atxn7l3*$^{-/-}$ cells compared to WT cells. Tbp, TATA-box binding protein. **B.** Growth curve of WT (black) and *Atxn7l3*$^{-/-}$ cells (blue) showing viable cell numbers as assessed by trypan blue staining. Cells were grown in FCS + LIF + 2i medium. n = 2-4. **C.** Clonal assay analysis in FCS + LIF + 2i medium of *Atxn7l3*$^{-/-}$ cells compared to WT cells. Top, example images of alkaline phosphatase (AP) stained colonies with smaller circles showing microscopy images of colonies. Bottom, quantification of colony area distribution with inset showing box plot representation. Colony areas were assessed using ImageJ. **D.** Clonal assay analysis in LIF medium of *Atxn7l3*$^{-/-}$ cells compared to WT cells. Top, example images of colonies stained with crystal violet to assess the total number of colonies. Bottom, quantification of ground state, AP positive (AP+) or primed, AP negative (AP-) colonies normalized by the total number of colonies as assessed by crystal violet staining. n = 2.

cells (Figure 54B). Similarly, loss of Atxn7l3 in mouse ESCs caused the formation of on average clearly smaller colonies in both FCS + LIF medium and medium containing 2i compared to WT cells (Figure 54C and 54D, top). Interestingly, although proliferation was affected, the level of self-renewal of mouse ESCs lacking Atxn7l3, as assessed by the ratio of AP positive colonies relative to the total amount of colonies, was comparable to WT cells (Figure 54D, bottom).

At this point, the functions of ATXN7L3 in SAGA-independent DUB modules was biochemically demonstrated in human cells additionally showing that USP22 is the only DUB enzyme which can interact with ATXN7 and therefore be incorporated into SAGA (Atanassov et al., 2016). Also, around the same time, studies in my host laboratory on mouse lines with either inactivation of *Usp22* or *Atxn7l3* demonstrated that inactivation of *Usp22*, the SAGA-specific DUB enzyme, only mildly affected H2BK120ub and embryonic development in stark contrast to inactivation of *Atxn7l3* (El-Saafin, Wang et al. manuscript in revision, Annexe 2). These findings indeed suggested, that the free Usp27x- and Usp51-containing DUB modules are acting redundantly to the Usp22-containing DUB module incorporated within SAGA, in strong agreement with earlier results (Kobayashi et al., 2015; Atanassov et al., 2016).

Consequently, the preliminary results presented here do not allow proper interpretation of the importance of the DUB module of SAGA on mouse ESC physiology, as loss of Atxn7l3 also affects the activity of SAGA-independent DUBs. It is therefore also difficult to compare results described in this section to findings on the HAT and TBP-interacting activities of SAGA shown in earlier result sections. Nevertheless, our findings indicate an importance of proper H2BK120ub removal for mouse ESC physiology.

Contributions

<u>Veronique Fischer</u> – Generated cell lines. Designed and performed experiments, analysis and graphical representations of data shown in Figure 53 and Figure 54.

Matthieu Stierle – Helped in generating and validating cell lines.

Bernardo Reina San Martin – designed the CRISPR-Cas9 plasmids.

Didier Devys – Conceived the work and designed experiments.

4. Analysis of the impact of acute loss of TRRAP on mRNA synthesis in human cells.

This result section contains preliminary and unpublished experimental data which were generated in collaboration with the Helmlinger research group of the Institut de génétique moléculaire de Montpellier (IGMM), France.

Introduction

The massive TRRAP subunit of the SAGA complex is believed to be the key interaction surface for contacts with transcription factors (TFs) (McMahon et al., 1998; Knutson & Hahn, 2011; Setiaputra et al., 2015; Elías-Villalobos et al., 2019a). It is also part of the TIP60 coactivator complex in which TRRAP was suggested to additionally serve as a structural subunit for complex assembly (Wang et al., 2018b; Elías-Villalobos et al., 2019a; Elías-Villalobos et al., 2019b).

Trrap is essential during mouse embryonic development as its inactivation causes very early embryonic lethality and is required for cell growth and survival of cultured cells (Herceg et al., 2001; Fazzio et al., 2008). Deletion of the gene encoding its counterpart in budding yeast, Tra1, is also detrimental to yeast growth and survival (Helmlinger & Tora, 2017).

These drastic phenotypes suggest a major role of Trrap and its associated complexes in RNA polymerase II (Pol II) transcription and cell physiology. Therefore, and to address **aim f)** in relation to the functional importance of the TF-interacting module of SAGA in Pol II transcriptoin, we were interested in analysing the impact of depletion of TRRAP on newly synthesized RNA levels in mammalian cells.

Results

Our collaborators from the Helmlinger research group in Montpellier had successfully introduced the auxin-inducible degron (AID) tag to the endogenous locus of *Trrap* in human HCT116 colon cancer cells. Using these cell lines, AID-TRRAP can be acutely depleted to near undetectable protein levels following four hours of auxin treatment (Figure 55A). Total RNA extraction followed by deep sequencing (total RNA-seq) from three independent *Trrap*[AID/AID] cell lines treated for five hours with either NaOH, the dissolving agent (+TRRAP), or auxin (-TRRAP) revealed that roughly 800 genes were differentially expressed upon depletion of AID-TRRAP with roughly half being down- or upregulated (Figure 55C). As TRRAP is a subunit of two coactivator complexes which positively stimulate Pol II transcription, the observation of roughly equal amounts of genes being upregulated compared to genes being downregulated upon its depletion was unexpected.
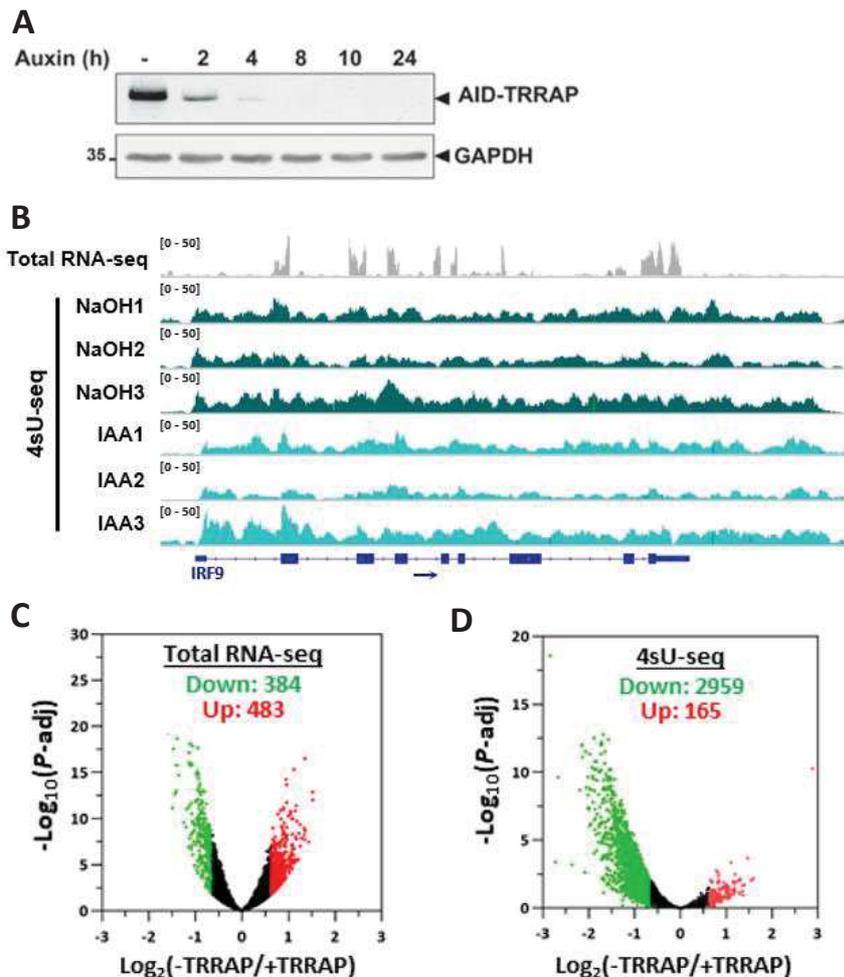
Figure 55: Transcriptional effects upon acute depletion of TRRAP in human cells. A. Western blot analysis of the effects of auxin (IAA) treatment from 2 to 24 hours (h) on *Trrap^{AID/AID}* human colon cancer cell lines compared to treatment with the IAA-dissolving reagent NaOH (-). GAPDH represents the loading control. **B.** Comparison of total RNA-seq profiles of the gene *Irf9* to newly synthesized RNA (4sU-seq) profiles. For 4sU-seq, results from three independent *Trrap^{AID/AID}* cell lines treated either with NaOH or with IAA for five hours before labelling with 4sU are shown. Arrow indicates direction of transcription. **C.** Volcano plot showing total RNA-seq results of three independent *Trrap^{AID/AID}* cell lines. Log10 transformed adjusted *p*-values (*P*-adj) are displayed in respect to log2 fold changes observed between five hours of IAA (-TRRAP) and NaOH (+TRRAP) treatment. **D.** Same as in (C) but showing results of 4sU-seq experiments.

Amongst other reasons, this upregulation of genes on total RNA levels following AID-TRRAP depletion could be indicative of secondary transcriptional effects and prompted us to analyse newly synthesized RNA levels in these cell lines. Newly synthesized RNA analysis enables a direct assessment of effects on Pol II transcription and is largely independent from RNA degradation events, which take an influence on total RNA levels. Additionally, especially within mammalian cells, RNA half-lives vary greatly from few minutes to several hours with an estimated average of roughly six hours (Duffy et al., 2019). Consequently, the analysis of total RNA levels following acute depletion of few hours, as described here for the *Trrap^{AID/AID}* system, might underestimate the effects on transcription of long-lived RNAs and be biased towards more unstable RNAs with short half-lives.

Indeed, newly synthesized RNA purification followed by deep sequencing (4sU-seq) analysis revealed more dramatic effects of depletion of AID-TRRAP on Pol II transcription compared to total RNA-seq (Figure 55C and 55D, comparison of total RNA-seq and 4sU-seq profiles for the unaffected *Irf9* gene in Figure 55B)). Roughly 3000 genes were found significantly downregulated by 4sU-seq in three independent *Trrap^{AID/AID}* cell lines treated for five hours with auxin compared to roughly 400 genes by total RNA-seq. Interestingly, still roughly 200 genes are seemingly upregulated upon depletion of AID-TRRAP as assessed by 4sU-seq.

Overall, these preliminary analyses revealed a potential considerable role of TRRAP-containing complexes on Pol II transcription in human cells, which could be the underlying cause of the requirement of TRRAP for cell survival.

Contributions

Dylane Detilleux – Generated and validated *Trrap$^{AID/AID}$* cell lines and performed western blot and total RNA-seq experiments shown in Figure 55A, 55B and 55C.

Peggy Raynaud – Performed 4sU labelling in *Trrap$^{AID/AID}$* cell lines and bioinformatic analysis as well as graphical representation of data shown in Figure 55B, 55C and 55D.

Veronique Fischer – Performed 4sU RNA purification experiments shown in Figure 55B and 55D.

Didier Devys – Conceived the work and designed experiments.

Dominique Helmlinger – Conceived the work, designed experiments and analysed data.

Discussion & Perspectives

Discussion and Perspectives

1. Distinct functions of the SAGA and ATAC coactivator complexes in mammalian Pol II transcription

Although the SAGA and ATAC coactivator complexes share the same histone acetyltransferase (HAT) module capable of modifying histone H3 tails, we found that SAGA and ATAC have mainly distinct functions within Pol II transcription in mouse embryonic stem cells (ESCs). This was apparent through two main observations: i) CRISPR-Cas9-mediated inactivation of core ATAC subunits, Yeats2 and Zzz3, caused a potential lethality in mouse ESCs as no homozygous KO clones could be obtained. This indirectly suggested that SAGA cannot fully compensate for ATAC loss and therefore might act non-redundantly to ATAC. ii) Newly synthesized RNA analysis in SAGA and ATAC mutant cell lines revealed that SAGA and ATAC affect transcription of distinct genes.

Inactivation of the SAGA core subunit, Supt7l, resulted amongst others in decreased expression of genes related to LIF signalling, a key pathway involved in the maintenance of the pluripotent state of mouse ESCs. The implication of SAGA in the pluripotency network of mouse ESCs is in agreement with recent studies suggesting that Taf5l and Taf6l, two additional core subunits of SAGA, are involved in the regulatory networks important for self-renewal in mouse ESCs (Seruggia et al., 2019).

In contrast, genes downregulated upon depletion of structural ATAC subunits were enriched for genes related to cytoplasmic translation such as ribosomal protein genes (RPGs). Importantly, the effects on RPGs and other translation-related genes were readily detectable on newly synthesized RNA levels upon 24 hours of auxin treatment, when no obvious defects in cell cycle could be detected by propidium iodide staining. However, growth and cell cycle defects were detectable upon prolonged auxin treatment. This suggests that the reduction in expression of translation-related genes proceeds defects in cell cycle and could imply that the observed proliferation and cell cycle abnormalities might be caused by impaired assembly and function of ribosomes. Also, the effects on translation-related genes could explain the lethality observed for constitutive inactivation of *Yeats2* and *Zzz3*, as genes associated with ribosome biogenesis and translation were generally found to be essential (Bertomeu et al., 2018). However, it cannot be excluded that the effects on cell proliferation and growth might be due to changes in gene expression of other important regulators, as transcription of several other genes was affected upon depletion of core ATAC subunits including genes involved in LIF signalling.

On a side note, loss of homozygous mutant clones for *Yeats2* and *Zzz3* could have been partially enhance due to the harsh way (single cell sorting and culturing) in which we generated the cell lines. We cannot exclude that a gentler way of generating KO cell lines, such as for example using colony peaking approaches and avoiding single cell culturing, could have led to the generation of *Yeats2*[-/-] and *Zzz3*[-/-] cell lines.

We do not exclude that SAGA and ATAC might act redundantly at some genes. Indeed, our newly synthesized RNA analysis revealed that roughly 200 genes might be significantly coregulated by SAGA and ATAC. This agrees with earlier genome-wide studies which suggested that some genes might be transcriptionally dependent on or occupied by subunits of both complexes (Pankotai et al., 2010; Krebs et al., 2011; Arede et al., 2020).

Overall, together with earlier studies, our results indicate that SAGA and ATAC significantly affect different sets of genes and therefore might especially have non-overlapping roles in the regulation of Pol II transcription (Pankotai et al., 2005; Gamper et al., 2009; Pankotai et al., 2010; Nagy et al., 2010; Krebs et al., 2011; Arede et al., 2020). However, our newly synthesized RNA analysis also suggested a potential general but mild importance of SAGA and ATAC subunits for Pol II transcription in mammalian cells (more details discussion section 1.2. Do SAGA and ATAC act as general cofactors for Pol II transcription in mammalian cells?).

1.1. Specific functions of ATAC on translation-related genes or does ATAC regulate cell-type specific transcription programs?

It is not clear yet, if RPGs and other translation-related genes are direct targets of ATAC in mouse ESCs, but recent studies in human cells found that Yeats2 and Zzz3 directly bind to gene promoters of RPGs and regulate their expression in human lung and leukaemia cancer cell lines (Mi et al., 2018; Mi et al., 2017; Arede et al., 2020). These findings further suggest that the effects on expression of translation-related genes upon depletion of ATAC subunits is not restricted to mouse ESCs but also found in human cells. However, the importance of ATAC for transcription of translation-related genes in these two highly proliferative cell systems (mouse ESCs and human cancer cells) could simply reflect the high dependency of these cells on protein biogenesis (Bulut-Karslioglu et al., 2018; Pelletier et al., 2018). This raises the question if RPGs and other translation-related genes represent general targets of the ATAC complex also in cell lines which are less proliferative such as terminally differentiated cells.

Several protocols exist allowing the differentiation of mouse ESCs into various terminally differentiated cells such as neurons (Bibel et al., 2007). Following for example the neuronal differentiation protocol, we could assess if depletion of AID-Yeats2 and AID-Zzz3 also affects expression of translation-related genes in neuronal precursor cells and terminally differentiated neurons. Assessing the effects of depletion of ATAC subunits in other cell types, could further allow to reveal if ATAC is generally implicated in cell-type specific transcription programs or important for expression of a fixed set of genes such as translation-related genes. It also remains unclear why some RPGs are more strongly affected upon depletion of ATAC subunits than others.

1.2. Do SAGA and ATAC act as general cofactors for Pol II transcription in mammalian cells?

Recent findings suggest that SAGA affects Pol II transcription of nearly all actively transcribed genes in yeast (Bonnet et al., 2014; Baptista et al., 2017; Donczew et al., 2020). Also, findings in human cells suggest that the shared HAT activity of SAGA and ATAC acts on gene promoters of all actively transcribed genes (Bonnet et al., 2014). Similarly, the DUB activity of SAGA was found to deubiquitylate histone H2B along gene bodies of all actively transcribed genes in human cells. These results strongly suggested a general role of especially SAGA in Pol II transcription in mammalian cells.

In contrast, as described above, we found that SAGA seems to significantly affect transcription at a limited set of genes in mouse ESCs. However, our newly synthesized RNA analysis also revealed a general although mild downward shift of Pol II transcription in *Supt7l*$^{-/-}$ cells. Together with findings of earlier studies suggesting that SAGA activities act on all actively transcribed genes in human cells (Bonnet et al., 2014), our results indicate that, although SAGA might be found at all active genes, Pol II transcription might not be as dependent on SAGA in mammalian cells compared to yeast (Baptista et al., 2017; Donczew et al., 2020). One can imagine that these limited effects of SAGA on Pol II transcription in mammalian cells could be due to differences in i) SAGA composition or in ii) the regulation of transcription between yeast and mammalian cells, which will be non-exhaustively detailed in the following paragraphs.

i) Differences in SAGA composition: In yeast SAGA, two subunits, Spt3 and Spt8, interact with TBP, while in mammalian SAGA, due to the absence of a Spt8 ortholog, the interactions with TBP are mediated solely by Supt3h. The absence of Spt8 could lead to a less stable binding of TBP in mammalian SAGA, as suggested by biochemical evidences upon loss of Spt8 in budding yeast (Wieczorek et al., 1998). In consequence, Pol II transcription in mammalian cells might be more dependent on other TBP loading factors besides SAGA such as the general transcription factor TFIID or potentially ATAC (more details on a potential function of ATAC in TBP deposition in discussion section 3.2. How does ATAC influence Pol II transcription if it is not through its HAT activity?). However, deletion of Spt8 in yeast does not lead to major effects on Pol II transcription in contrast to deletion of Spt3, suggesting that the main determinant of TBP binding to SAGA resides within Spt3 (Baptista et al., 2017). This has further been confirmed by recent cryo-EM structures of the yeast SAGA complex (Papai et al., 2020). These observations suggest however that differences in TBP-binding affinities between yeast and mammalian SAGA might not be the main contributor for the weakened impact of SAGA on Pol II transcription in mammalian cells.

ii) Differences in the regulation of transcription between yeast and mammalian cells: Several differences have been observed distinguishing Pol II transcription in mammalian cells from yeast. For example, enhancer-promoter interactions in mammalian cells, which have been suggested to depend on transcriptional coactivators, are more complex and diverse than in yeast. More specifically, in yeast

UAS (upstream activating sequences), which have functions identical to mammalian enhancers, are generally localized upstream and in proximity to the gene promoter. In contrast, in vertebrates enhancer-promoter contacts must regularly overcome large distances of on average 20 to 50 kb (Furlong & Levine, 2018). Also, gene promoters in mammalian cells are thought to receive input from several enhancers (Osterwalder et al., 2018; Schoenfelder & Fraser, 2019). Based on this, enhancer-promoter contacts in mammalian cells might dependent on a more sophisticated regulatory mechanism compared to UAS-promoter contacts in yeast. This regulation could rely on several additional factors besides coactivators. Also, based on the increased number of enhancers per gene, a functional redundancy among enhancers could prevent major impacts on Pol II transcription (Osterwalder et al., 2018). Additionally to potential redundancies amongst enhancers, mammalian genomes possess a more extended set of transcription factors and coactivator complexes. Examples include cell-type specific transcription factors, such as Nanog in mouse ESCs, the ATAC coactivator, for which no clear ortholog has been described in yeast, and six COMPASS and COMPASS-like coactivator complexes, compared to one COMPASS complex in yeast. This extended set of regulatory factors could lead to a specification of the regulation of transcription programs in mammalian cell. Overall, these and other differences in mammalian cells could indicate a more tightly controlled Pol II transcription process with several redundantly acting mechanisms, which could make these cellular systems less sensitive to inactivation of SAGA.

Our newly synthesized RNA analyses also revealed a mild but general reduction of Pol II transcription upon depletion of structural ATAC subunits, which was slightly more pronounced than the effects observed in $Supt7l^{-/-}$ cells. Therefore, our newly synthesized RNA analyses could indicate that SAGA and ATAC might both have a mild but general impact on global Pol II transcription with however some genes being more dependent on one or the other complex. To validate a potential general impact of these two coactivators for Pol II transcription in mammalian cells, additional experiments are required. For example, analyses of the genome-wide binding profiles of these two coactivator complexes could reveal if they are found at all actively transcribed genes.

We decided to assess the genome-wide binding profiles of SAGA and ATAC by using an alternative approach to ChIP-seq, which is suggested to allow more efficient profiling of highly dynamic factors such as SAGA and ATAC: the CUT&Run or CUT&Tag techniques (Kaya-Okur et al., 2019; Hainer et al., 2019). These methodologies are based on fusion proteins of either MNase or Tn5 transposase to protein A, respectively. The protein A-MNase or protein A-Tn5 transposase can bind to primary antibodies, which recognize the protein of interest, and are subsequently thought to cut the genomic DNA only at sites in proximity of the protein of interest. These methodologies further do not depend on a cross-linking step and allow to assess binding events on native chromatin. Nevertheless, these techniques also rely on a specific primary antibody and need to be tightly controlled to avoid over digestion at open chromatin regions.

To avoid primary antibody-dependent effects and since very specific primary antibodies are also lacking for SAGA and ATAC subunits, we decided to generated endogenously HA-tagged cell lines for two ATAC subunits (Tada2a and Atac2) as well as two SAGA subunits (Taf5l and Tada1) (Annexe 1). Unfortunately, based on different reasons described in Annexe 1, these cell lines could not be used to properly assess the genome-wide binding profiles of SAGA and ATAC, with the potential exception of the HA-tagged Tada2a cells. The tagged Tada2a cell lines could however be used to assess if ATAC binding sites can be analysed using the CUT&Run or CUT&Tag technologies. If successful and informative, additional tagged cell lines targeting other subunits of SAGA and ATAC could be generated. These analyses could subsequently allow to reveal if SAGA and ATAC are recruited to all actively transcribed genes or only to specific genes.

1.3. How important is the interaction with transcription factors for the recruitment of SAGA and ATAC to gene promoters?

1.3.1. Major role of the TF-interacting subunit TRRAP for Pol II transcription in human cells

As shown in result section 3. Analysis of the impact of acute loss of TRRAP on mRNA synthesis in human cells, we further assessed the impact of loss of the TF-interacting module of SAGA, composed of the TRRAP protein, on Pol II-mediated RNA synthesis in mammalian cells. Using a human colon cancer cell line encoding endogenously auxin-inducible degron (AID)-tagged TRRAP, a strong decrease in newly synthesized RNA levels could be detected shortly after induction of TRRAP depletion through auxin treatment. However, TRRAP is not only part of SAGA but also found within the TIP60 complexes. Therefore, the observed effects on Pol II transcription could reflect the loss of function of one or both coactivators. To be able to attribute transcriptional effects observed upon depletion of TRRAP to functions within TIP60 or SAGA, additional experiments upon depletion of other subunits of TIP60 or SAGA would be required. However, findings from earlier studies suggest that the importance of TRRAP for Pol II transcription could be linked to its function within TIP60.

1.3.1.1. TRRAP might be especially important for the function of the TIP60 complex

Phenotypic evidences from mouse KO studies and cultured cells suggest that the essentiality of TRRAP and its major impact on Pol II transcription could mainly originate from its function within the TIP60 complex (Herceg et al., 2001; Fazzio et al., 2008; Hu et al., 2009b). While inactivation of SAGA subunits in mouse embryos generally leads to phenotypes after the gastrulation stage, inactivation of *Trrap* causes very early embryonic defects and further does not enable generation of viable mouse embryonic stem cell lines (Herceg et al., 2001). Indeed, inactivation of *Trrap* displays several

similarities with inactivation of *Tip60*, the catalytic HAT subunit of the TIP60 complex, which also leads to early embryonic defects (Hu et al., 2009b).

Further, evidence from our own experiments in mouse ESCs support the notion that the considerable role of TRRAP for Pol II transcription might not be due to its function within the SAGA coactivator. In budding yeast, the Tra1 subunit (ortholog to TRRAP) is integrated within SAGA mainly through contacts mediated by Spt20 (Elías-Villalobos et al., 2019b). Indeed, in our *Supt20h⁻/⁻* mouse ESC lines we could observe, amongst other effects, the markedly reduced incorporation of TRRAP. Nevertheless, *Supt20h⁻/⁻* did not seem to cause major changes in phenotypes of mouse ESCs in contrast to inactivation of *Supt7l* or *Supt3h*. Therefore, these findings would further support that the considerable impact of loss of TRRAP on Pol II transcription in human cells might not be due to loss of functions within the SAGA complex.

Additional evidence from the fission yeast *S. pombe* further support this notion. Intriguingly, *S. pombe* was found to possess two Tra1 genes (ortholog of *Trrap*). One version encodes for the Tra1 protein, which was subsequently reported to be specifically incorporated within SAGA, while the second protein, Tra2, is exclusively part of the NuA4 complex (the orthologous complex to the mammalian TIP60 complex) (Helmlinger & Tora, 2017; Elías-Villalobos et al., 2019a). In consequence, fission yeast represents an exclusive model system to assess the functional importance of the TF-interacting Tra1-related proteins in either SAGA or NuA4. Interestingly, while deletion of Tra2 (NuA4) is inviable in fission yeast, deletion of Tra1 (SAGA) does not markedly affect viability. This further suggests that the essentiality of Tra1 in budding yeast or TRRAP in mammalian cells might be especially due to their roles within the NuA4 or TIP60 complexes, respectively (Helmlinger & Tora, 2017; Elías-Villalobos et al., 2019a).

1.3.1.2. Does TRRAP impact Pol II transcription through interactions with TFs?

TRRAP was reported in several studies to interact with key transcription factors such as c-MYC and factors of the E2F family such as E2F1 (McMahon et al., 1998; Liu et al., 2003; Lang et al., 2001; Murr et al., 2007). Therefore, a key function of TRRAP is thought to be the recruitment of SAGA and TIP60 to regions bound by these TFs. Intriguingly and in contrast to inactivation of *Trrap*, inactivation of either *Myc* or *E2f1* in mouse embryos were found to cause embryonic lethality either at late stages (around E10.5) or to be completely compatible with embryonic development, respectively (Charron et al., 1992; Davis et al., 1993; Field et al., 1996; Murr et al., 2007).

The stronger requirement for TRRAP during mouse development compared to transcription factors, which are thought to be key mediators of the influence of TRRAP on Pol II transcription, could be explained by several reasons. Foremost, complementary functions of other TFs in recruiting the SAGA or TIP60 coactivators through TRRAP could be responsible for the milder mouse embryo phenotypes

upon inactivation of *Myc* or *E2f1*. Alternatively, an essential, TF-independent function of TRRAP for early mouse embryo development could be its role in serving as assembly platform for the TIP60 complex. Indeed, recent cryo-EM structures of budding yeast NuA4 revealed that Tra1 (ortholog of TRRAP) represents not only a TF-interacting subunit, but is part of the assembly platform required for TIP60 complex formation (Wang et al., 2018b). In contrast, within SAGA Tra1 forms only few contacts with the remaining subunits (Wang et al., 2018b; Wang et al., 2020; Papai et al., 2020). Therefore, inactivation of *Trrap* could lead to a major impact on TIP60 complex assembly, while causing few changes on the integrity of SAGA.

Overall, the strong impact of depletion of TRRAP on Pol II transcription, which we could observe in human cells, could be mainly based on its role for the assembly of the TIP60 complex rather than its function as TF-interacting platform. That the TF-interacting function of TRRAP might not be as crucial for Pol II transcription is supported by observations in fission yeast. Loss of the SAGA-specific subunit Tra1 (ortholog to TRRAP) in fission yeast, does not recapitulate growth phenotypes observed upon loss of core subunits of SAGA such as Spt7 and Ada1 (Helmlinger et al., 2011). This suggests that Tra1, and therefore potential interactions with TFs, are not as crucial as core subunits for the functions of SAGA. Consequently, this might imply that the TF-interacting function of Tra2 or TRRAP might also not be central for NuA4 and mammalian TIP60 functions, respectively.

In summary, the roles of TRRAP within the SAGA and TIP60 coactivator complexes and their dependency on TFs for their recruitment to gene promoters are starting to become apparent but need further evaluation.

1.3.2.   How could SAGA and ATAC be recruited to gene promoters independent of TFs?

As indicated from studies in fission yeast, SAGA might not majorly depend on interactions of its Trrap subunit with TFs for recruitment to gene promoters. TF-independent recognition of actively transcribed regions by SAGA and ATAC could be enable through chromatin reader domains of their subunits, such as for example the Tudor domain of Sgf29, which was reported to recognize H3K4me3 (Vermeulen et al., 2010; Bian et al., 2011). Also, chromatin reader domains in the ATAC subunits, Yeats2 and Zzz3, were recently suggested to be crucial for recruitment of ATAC to gene promoters in human cells (Mi et al., 2017; Mi et al., 2018). This notion of recruitment of SAGA and ATAC through histone modifications such a H3K4me3 would however imply a dependency on earlier acting chromatin modifying factors. Also, as discussed in more details later, several findings from yeast and mouse ESCs indicate that cell growth and Pol II transcription is not majorly impaired upon loss of individual histone tails or modifications (Dai et al., 2008; Hödl & Basler, 2009, 2012; Kim et al., 2012; Rickels et al., 2017; Dorighi et al., 2017; Acharya et al., 2017; Zhang et al., 2020). This could imply that coactivator

complexes do not solely depend on specific histone modifications and their reader domains for recruitment.

Additional findings highlight that SAGA and ATAC are highly dynamic (Bonnet et al., 2014; Vosnakis et al., 2017). For example, subunits of SAGA and ATAC were found to have a fast-moving fraction with a diffusion constant of roughly 8 μm^2/s, which is comparable to results measured for transcription factors, which range from 0.5 to 5 μm^2/s (Hager et al., 2009; Vosnakis et al., 2017). These dynamic properties could enable SAGA and ATAC to scout the nuclear volume similarly to TFs. Consequently, this would imply that SAGA and ATAC could reach their substrates and target regions independently of other factors. However, interaction of SAGA and ATAC with TFs and chromatin modifications could be important to increase the time SAGA and ATAC reside at given gene promoters and therefore the degree to which they influence transcription at these genes (Vosnakis et al., 2017). This might also involve the formation of the recently postulated enhancer-promoter hubs, which might be facilitated by liquid-liquid phase separation (Furlong & Levine, 2018).

2. How do SAGA and ATAC impact self-renewal of mouse ESCs?

Several screens have been performed in mouse ESCs to identify factors involved in the maintenance of their self-renewal and pluripotent capacities (Bilodeau et al., 2009; Das et al., 2014; Cooper & Brockdorff, 2013; Fazzio et al., 2008; Hu et al., 2009a; Kagey et al., 2010). These screens were based on the use of either siRNA or shRNA-mediated depletion of the targeted factors. Unfortunately, ATAC-specific subunits were generally not included in these screens. Also, only few SAGA-specific subunits or subunits of the shared HAT module were tested, with usually variable results between the different siRNAs or shRNAs used within a given study. Overall, these studies did not reveal an importance of SAGA and ATAC for mouse ESC physiology.

Interestingly, however a recent study using a loss-of-function screen based on the CRISPR-Cas9 technology found that several subunits of SAGA are important for self-renewal of mouse ESCs cultured in FCS + LIF medium (Seruggia et al., 2019). Unfortunately, no ATAC-specific subunit was included in this study apart from the HAT-specific ATAC subunit Tada2a, which, in agreement with our data, was not found to be significantly involved in maintaining mouse ESC self-renewal.

Based on these previous findings, we revealed for the first time that structural subunits of the ATAC coactivator complex are important for mouse ESC growth and self-renewal and further confirmed the recently described implication of SAGA in this process. Interestingly, the effects on self-renewal seemed to be largely independent of the shared HAT function of these two complexes (detailed discussion of potential HAT-independent functions of SAGA and ATAC later).

The mechanism by which SAGA and ATAC are implicated in self-renewal of mouse ESCs remains unclear. Indeed, SAGA and ATAC could be implicated directly or indirectly in the pluripotency network

required for mouse ESC self-renewal. For example, SAGA and ATAC could act as coactivators for pluripotency transcription factors and be recruited to specific genes through these factors. This would indicate that SAGA and ATAC are directly implicated in mouse ES self-renewal by transferring the information of the pluripotency factors to the transcription machinery. On the other hand, if SAGA and ATAC are generally found at actively transcribed genes and involved to some extend in global Pol II transcription, they might be required for cell-type specific transcription programs. Thereby, SAGA and ATAC might be important to ensure the expression levels of pluripotency factors in mouse ESCs and thereby be indirectly implicated in the pluripotency network.

2.1. SAGA might be especially important for stabilizing the naïve pluripotency network

Interestingly, downregulated genes in $Supt7l^{-/-}$ cells, as assessed by newly synthesized RNA analysis, showed to be enriched for genes responsive to LIF signalling and bound by the pluripotency TFs Oct4, Sox2 and Nanog, which could suggest a direct role of SAGA in the pluripotency network by functioning as a coactivator for these TFs. Additionally, when growing $Supt7l^{-/-}$ cells in FCS + LIF medium, we found reduced mRNA levels of the naïve-specific pluripotency TFs Nanog, Klf4, Esrrb and Tfcp2l1, while mRNA levels of the core pluripotency TFs Oct4 and Sox2 remained largely unchanged. In general, mouse ESCs grown in FCS + LIF medium have been found to display highly variable levels of the naïve-specific pluripotency factors (Nanog, Esrrb, Klf4 and Tfcp2l1), while Oct4 and Sox2 levels are maintained rather homogeneously (Martello & Smith, 2014; Navarro, 2018). Cells possessing low levels of the naïve-specific pluripotency factors are considered to be sensitive for differentiation cues. Therefore, the specific reduction of mRNA levels for the naïve-specific pluripotency factors Nanog, Klf4, Esrrb and Tfcp2l1 in $Supt7l^{-/-}$ cells, might suggest a destabilization of the naïve pluripotency network and an increased sporadic differentiation, which is in agreement with clonal assays and alkaline phosphatase staining results. Consequently, this could imply that specifically the naïve-specific pluripotency network is dependent on SAGA function.

The potential effect of inactivation of SAGA on the stability of the naïve pluripotency network can be imagined to originate through two possible ways: i) SAGA could represent a coactivator for the naïve-specific pluripotency factors such as Nanog or ii) could be responsible in ensuring the expression of Nanog, Esrrb, Klf4 and Tfcp2l1 thereby indirectly affecting the naïve pluripotency network. These two alternative possibilities could be assessed in our mutant cell lines based on the following observation: Overexpression of Nanog can force differentiating mouse ESCs back into an undifferentiated state, which can be easily monitored through alkaline phosphatase staining in clonal assays (Festuccia et al., 2012). Based on the current model of transcription induction, the restoration of undifferentiated state through overexpression of Nanog would dependent on the functions of coactivator complexes such as potentially SAGA. Overexpression of Nanog in our $Supt7l^{-/-}$ cells could provide us with an indication if SAGA acts as a coactivator for Nanog or if increased differentiation in $Supt7l^{-/-}$

ESCs is simply caused by reduced expression of naïve-specific pluripotency factors such as Nanog. More specifically, if SAGA acts as a crucial coactivator for Nanog, overexpression of Nanog in *Supt7l^-/-* cells would not lead to an increase of alkaline phosphatase positive, undifferentiated colonies compared to wildtype cells. In contrast, if *Supt7l^-/-* cells can reentry into an undifferentiated state by forced expression of Nanog, SAGA might not act as a major coactivator for Nanog, but instead be required to ensure its expression.

Interestingly, a recent study indicates that SAGA would especially transcriptionally activate Oct4- and Myc-dependent genes in mouse ESCs as assessed by correlation analysis between binding of Taf5l and Taf6l, two SAGA-specific core subunits, and pluripotency TFs (Seruggia et al., 2019). In our study, we did not find an obvious link between SAGA and Myc. This discrepancy could be explained by differences in medium conditions used, as Seruggia et al., 2019 analysed mouse ESCs in FCS + LIF medium, while we largely cultured our cell lines in the presence of two potent inhibitors of differentiation (medium containing 2i). In contrast, we found that genes downregulated upon depletion of ATAC subunits were enriched for Myc-bound as well as E2f4-bound genes, two transcription factors important for the regulation of cell cycle progression (more details below) (Matsumura et al., 2003; Chen et al., 2009; Chappell & Dalton, 2013; Fagnocchi & Zippo, 2017; Hsu et al., 2019a).

2.2. Inactivation of ATAC might lead to a general destabilization of the pluripotency network through indirect means

As mentioned above, genes downregulated upon depletion of structural ATAC subunits were found to be enriched for the cell cycle-regulating transcription factors, Myc and E2f4. Interestingly, Yeast2 has been previously reported to interact with E2f4 (Hsu et al., 2019a). Inactivation of E2f4 or combined inactivation of c-Myc and n-Myc in mouse ESCs also leads to an increase in G1-phase and reduced cell growth, similar to what we could observe for our ATAC mutant cell lines. However, inactivation of *E2f4* did not cause defects in the self-renewal capacities of mouse ESCs (Hsu et al., 2019a). In contrast, c-Myc and n-Myc were found to be important to sustain self-renewal when mouse ESCs were cultured in FCS + LIF medium, which mimics the effects we could observe in our ATAC mutant cell lines (Smith et al., 2010; Varlakhanova et al., 2010). Overall, this suggests that the effects of inactivation of ATAC on mouse ESC proliferation and growth might be through E2f4- and Myc-targeted genes. In contrast, the impaired self-renewal of mouse ESCs observed upon depletion of structural ATAC subunits might be explained by Myc functions.

Interestingly, when we analysed mRNA levels of pluripotency factors upon depletion of AID-Yeats2 in mouse ESCs cultured in FCS + LIF medium, all tested pluripotency TFs were affected including the core pluripotency regulators Oct4 (*Pou5f1*) and Sox2. This suggests that ATAC might be generally required for proper expression of the pluripotency factors. The effects observed on self-

renewal in ATAC mutant cell lines might therefore be caused indirectly due to a general loss of expression of pluripotency factors. This indirect effects on the pluripotency network might be linked to the effects of ATAC on translation-related genes, as a recent study suggests that translation might be crucial to maintain the euchromatic state and nascent transcription in mouse ESCs (Bulut-Karslioglu et al., 2018). Indeed, if the translational output is affected in ATAC mutant cell lines, this could cause a reduction of protein levels of the pluripotency TFs and thereby affect self-renewal capacities of mouse ESCs.

Overall, inactivation of SAGA or ATAC seems to affect the pluripotency network in slightly different ways.

3. How important are histone modifications and histone modifying activities of transcriptional coactivators for Pol II transcription?

Using our mutant cell lines, we found that, while inactivation of core subunits of SAGA and ATAC impaired self-renewal capacities of mouse ESCs, inactivation of their shared HAT activity did not. This indicates that SAGA and ATAC possess functions which are important for self-renewal of mouse ESCs and which are largely independent of their shared HAT activity. Together with earlier results on catalytic inactivation of chromatin modifying functions of the TIP60 and MLL3/4 COMPASS-like complexes, this suggests important catalytic-independent functions of transcriptional coactivator complexes (Acharya et al., 2017; Dorighi et al., 2017; Rickels et al., 2017). Further, this raises the question of how important histone modifications are for mouse ESC physiology and Pol II transcription.

As described in the Introduction, several histone modifications are correlated with active chromatin regions, which has raised the possibility that they could be instructive for Pol II transcription. However, based on the catalytic-independent functions of chromatin modifying complexes mentioned above, it has been recently suggested that deposition of histone modifications might instead be a side product of recruiting chromatin modifiers to genetic elements (Pollex & Furlong, 2017).

Assessing the role of histone modifications in mammalian cells is majorly impeded by the presence of several large histone genes clusters. Recent studies have therefore been focusing on histone variants such as histone H3.3, which is encoded by only two genes and incorporated within active chromatin. Interestingly, while complete inactivation of H3.3 is not essential for development and viability in *Drosophila*, it causes early embryonic and cellular lethality in mouse (Sakai et al., 2009; Hödl & Basler, 2009). However, substitutions of either H3.3K4, K27 or K36 with non-modifiable amino acids was found to be viable in mouse ESCs and largely dispensable for Pol II transcription at most genes (Zhang et al., 2020; Gehre et al., 2020). This led to the suggestion that individual histone modifications are not detrimental for most actively transcribed genes and therefore not instructive for Pol II transcription.

More systematic studies on the role of histone modifications for cell viability have been enabled in yeast, where canonical histone genes are encoded by only two loci. In one such study, 486 histone H3 and H4 deletion and substitution mutants were screened (Dai et al., 2008). This revealed that residues within the H3 and H4 N-terminal tail are not essential for yeast growth in contrast to residues that interact directly with the nucleosomal DNA. Also, N-terminal tails of the four histones H2A, H2B, H3 and H4 can be individually deleted in yeast without causing obvious growth defects when cultured under optimal conditions (Morgan et al., 1991; Ling et al., 1996; Kim et al., 2012). These findings suggested that Pol II transcription is not prevented upon loss of individual N-terminal histone tails or histone modifications. However, combined deletion of histone tails was found to cause severe growth defects with, for instance, lethality observed upon combined loss of H2A and H4 tails or H3 and H4 N-terminal tails. These observations indicated that modifications at different histone tail residues might possess functional redundancies under optimal growth conditions. Consequently, this suggests that combined deletion of histone residues with modifications linked to active chromatin might impact global Pol II transcription. This could further imply a certain redundancy among coactivators depositing histone modifications on different histone tails. Growth of the individual histone tail deletion mutants was further found to be affected when the yeast strains were challenged by different stresses such as heat-shock or DNA damage induction (Kim et al., 2012). This suggests that individual histone tails and their modifiable residues might be more crucial when the Pol II transcription program needs to be adapted to environmental changes such as in response to stress or DNA damage.

In general, knowledge on the functional relevance of individual and combinations of histone modifications for Pol II transcription seems to be incomplete, especially in mammalian cells were genetic complexity impedes systematic analyses.

3.1. How does SAGA influence Pol II transcription if it is not through its HAT activity?

SAGA was discovered based on its interaction with subunits of the histone acetyltransferase (HAT) module (Grant et al., 1997; Saleh et al., 1997; Horiuchi et al., 1997). However, our study indicates that the HAT activity of SAGA does not seem to be its most crucial function in regulating Pol II transcription. As described earlier, we found that inactivation of the HAT activity of SAGA does not majorly affect self-renewal capacities of mouse ESCs or newly synthesized RNA levels in contrast to loss of the structural subunit Supt7l. This raises the question what functions of SAGA, beside its HAT activity, could explain its impact on Pol II transcription and mouse ESC self-renewal.

When examining the complex integrity of SAGA in *Supt7l⁻/⁻* cells, we found that Taf10, histone fold partner of Supt7l, and also Taf12, histone fold partner of Tada1, no longer incorporate within SAGA revealing that at least two of the four histone fold pairs of SAGA are missing. This loss of histone fold pairs could affect the overall formation of SAGA, as suggested by findings in yeast. Deletion of Spt7

(ortholog to Supt7l) in budding yeast was found to lead to destabilization of Ada1 (ortholog to Tada1) and Spt20 (ortholog Supt20h) and was therefore suggested to generally affect the complex integrity of SAGA (Wu & Winston, 2002). As highlighted by recent cryo-EM studies, destabilization of the histone octamer-like structure could especially affect the TBP-loading function of SAGA (Papai et al., 2020; Wang et al., 2020). The TBP-interacting subunit, Spt3 (ortholog to Supt3h), is part of the histone octamer-like structure and forms contacts through its C-terminal tail with the other histone fold pairs in yeast SAGA (Papai et al., 2020). Additional findings in budding yeast suggest that deletion of Spt7 (ortholog to Supt7l) affects the histone modifications H3K9ac and H2BK123ub to an extend comparable to deletion of the HAT or DUB enzymatic subunits of SAGA, respectively (Baptista et al., 2017). Overall, this indicates that inactivation of Supt7l in mouse ESCs could especially affect the TBP-interacting module and potentially also SAGA-related HAT and DUB functions.

3.1.1. TBP-loading function of SAGA is required for self-renewal of mouse ESCs

When we inactivated Supt3h, the TBP-interacting subunit, we could indeed observe effects on self-renewal and growth comparable to those observed for $Supt7l^{-/-}$ cells. This suggests that the TBP-loading function of SAGA might be crucial for self-renewal in mouse ESCs. Curiously, when we examined newly synthesized RNA in $Supt3h^{-/-}$ cells, we observed a general reduction of Pol II transcription which was however not as severe as found in $Supt7l^{-/-}$ cells. This consequently suggests that, although inactivation of Supt3h recapitulates the effects on mouse ESC self-renewal observed in $Supt7l^{-/-}$ cells, Pol II transcription does not solely dependent on the TBP-loading function of SAGA in mouse ESCs. These findings are in agreement with results from yeast, where loss of Spt3 (ortholog to Supt3h) leads to less severe effect on Pol II transcription compared to loss of either Spt20 (ortholog to Supt20h) or Spt7 (ortholog to Supt7l) (Baptista et al., 2017). Interestingly, combined deletion of the HAT catalytic subunit Gcn5 and Spt3 in yeast affected growth and Pol II transcription to a comparable extend as observed for inactivation of Spt20 or Spt7 (Sterner et al., 1999; Baptista et al., 2017). This suggests a synergistic effect of the HAT and TBP-interacting functions of SAGA on Pol II transcription. Consequently, these findings in yeast indicate that a combined loss of Tada2b, the SAGA-specific HAT subunit, and Supt3h in mouse ESCs could potentially recapitulate the transcriptional effects observed in $Supt7l^{-/-}$ cells.

A synergistic effect between the TBP-loading and HAT functions of SAGA could be based on a potential requirement of the HAT module subunits and its catalytic activity for the recruitment of SAGA or other downstream factors to gene promoters. Consequently, it could be interesting to assess if the synergistic effect between the two SAGA functions would rely on the enzymatic activity of the HAT or chromatin reader domains found within HAT subunits, such as the Tudor domain of Sgf29 (Vermeulen et al., 2010; Bian et al., 2011). If the synergistic effects would depend especially on chromatin reader domains, this could indicate that the recruitment of SAGA itself would underlie the synergistic effect.

In contrast, if the synergistic effect on Pol II transcription would rely on the enzymatic activity of the HAT, this could suggest that also other downstream factors or secondary effects could be involved in the synergistic impact of combined loss of TBP-loading and HAT functions of SAGA. Interestingly, as mentioned in the Introduction, loss of the catalytic activity of Gcn5, a HAT enzyme shared between SAGA and ATAC, did not fully recapitulate developmental defects observed in *Gcn5*-null mice implying HAT-independent functions (Bu et al., 2007; Spedale et al., 2012). These findings on mouse embryos could indeed suggest that the chromatin reader domains of Gcn5 and its interactions with the Tada2 proteins might have important roles for SAGA function.

3.1.2. Is the DUB activity of SAGA more important in mouse ESCs than its HAT activity?

Interestingly, our findings indicate an importance of H2BK120ub removal for proper mouse ESC growth without impacting on their self-renewal capacities (see result section 3). In contrast to inactivation of subunits of the HAT module of SAGA, inactivation of *Atxn7l3*, encoding a subunit of the DUB module of SAGA, was found to cause smaller colony sizes and to affect cell proliferation of mouse ESCs. However, as mentioned in the Introduction, Atxn7l3 is also part of SAGA-independent DUB modules (Atanassov et al., 2016). Its inactivation therefore might not only affect functions of the DUB module of SAGA but also the free DUB modules. On a side note, very recent cryo-EM structures of yeast SAGA suggest that upon encountering a nucleosome the HAT and DUB modules might dissociate from SAGA thereby also forming SAGA-independent or free versions of these modules (Wang et al., 2020). It remains however unclear if this represents an artefact of *in vitro* reactions or if it is also found *in vivo*.

The effects that we observed on mouse ESC growth upon inactivation of *Atxn7l3* might therefore reflect the combined function of SAGA-dependent and -independent DUB modules. This is suggested by findings in mouse embryos with either inactivation of *Usp22*, the gene encoding the SAGA-specific DUB enzyme, or *Atxn7l3* (El-Saafin, Wang et al. manuscript in revision, Annexe 2). Inactivation of *Atxn7l3* caused phenotypes drastically more severe with lethality by E11.5 than inactivation of *Usp22* with lethality by E14.5 (Lin et al., 2012; Koutelou et al., 2019). Also, H2BK120ub levels were not dramatically increased in *Usp22*$^{-/-}$ embryos in contrast to *Atxn7l3*$^{-/-}$ embryos. These findings imply that the defect observed on mouse ESC proliferation in our *Atxn7l3*$^{-/-}$ cells likely does not only reflect specific functions of the SAGA-dependent DUB module, but the overall impairment of Atxn7l3-dependent DUBs and increased levels of H2Bub. Overall and especially based on the phenotypes of *Usp22*$^{-/-}$ embryos, the DUB function of SAGA might not majorly impact on mouse ESC proliferation or Pol II transcription similarly to its HAT function.

3.1.3. Pol II transcription in mouse ESCs might depend on a combination of TBP-loading function and the histone modifying activities of SAGA

Findings in our *Supt20h*[−/−] cell lines further indicate that the histone modifying activities of SAGA and its TF-interacting subunit, Trrap, might not be most crucial for self-renewal or cell growth of mouse ESCs. Deletion of Spt20 (ortholog to *Supt20h*) in yeast was found to lead to complete or partial loss of the two histone modifying modules, DUB and HAT respectively, and in loss of its TF-interacting subunit, Tra1 (ortholog to Trrap) (Lee et al., 2011). However, the core module of SAGA could still be found upon deletion of *Spt20* although with partially reduced levels (Wu & Winston, 2002; Lee et al., 2011). We found similar effects on SAGA structure in our *Supt20h*[−/−] cells, with reduced levels of Trrap and several subunits of the HAT module with exception of Tada2b, but with an overall intact histone octamer-like structure and Taf5l. We unfortunately could not assess the impact on the DUB module as it was only weakly detected by mass spectrometry. Nevertheless, based on findings in yeast we think that the DUB module of SAGA is likely disturbed upon loss of Supt20h (Lee et al., 2011; Elías-Villalobos et al., 2019b). Overall, this suggests that upon loss of Supt20h the core of SAGA with its TBP-interacting module might still be functional, while its histone modifying and TF-interacting functions might be impaired.

Interestingly, inactivation of *Supt20h* did not cause obvious effects on mouse ESC self-renewal or growth. This observation reveals that in *Supt20h*[−/−] cells the remaining core-like structure of SAGA with its TBP-interacting function can maintain mouse ESC self-renewal and growth comparable to wildtype SAGA function. This finding further supports the notion that the histone modifying activities of SAGA might not be most crucial for SAGA function in mouse ESCs. As *Supt20h*[−/−] also leads to loss of the TF-interacting subunit Trrap without causing a major phenotype in mouse ESCs, SAGA might not be highly dependent on interactions with TFs through Trrap (see also earlier discussion section). Interestingly, the dramatic difference observed on mouse ESC growth between inactivation of *Supt20h* or *Supt7l*, two subunits of the core of SAGA, are comparable to effects observed on cell growth in deletion strains of *Spt20* and *Spt7* in the fission yeast *S. pombe* (Helmlinger et al., 2011).

Nevertheless, as mentioned earlier, inactivation of the TBP-interacting subunit, Supt3h, which is part of the core of SAGA, did not fully recapitulate effects observed on Pol II transcription upon inactivation of *Supt7l*. By immunoprecipitation experiments, we found that loss of Supt3h resulted into loss of itself from the SAGA complex without majorly affecting the incorporation of other subunits. Therefore, in the absence of Supt3h, a nearly complete SAGA can still form. This could suggest an importance for the other functions of SAGA, the histone modifying and TF-interacting functions, which might not be as crucial for survival but important for proper Pol II transcription in mouse ESCs.

Indeed, as mentioned in an earlier discussion section, findings in yeast indicate that the DUB, HAT and TBP-interacting functions of SAGA might act synergistically on Pol II transcription (Sterner et al., 1999; Baptista et al., 2017). As such, combined deletion of Gcn5 and Ubp8, the HAT and DUB enzymes,

or combined deletion of Gcn5 and Spt3, the HAT enzyme and a TBP-interacting subunit, in budding yeast affected Pol II transcription stronger than expected from individual deletion strains (Baptista et al., 2017). The mechanism behind this synergistic effect remains unclear but could be caused for example by a histone crosstalk such as H3 acetylation by the HAT module of SAGA enabling efficient deubiquitylation of H2BK120ub by its DUB module or *vice versa*. The chromatin reader domains within the HAT and DUB modules could facilitate recruitment of SAGA to actively transcribed genes and be involved in the observed synergistic effect (see also earlier discussion section 'TBP-loading function of SAGA is required for self-renewal in mouse ESCs'). Additionally, recruitment of downstream factors recognizing SAGA-dependent histone modifications or TBP could also account for the synergistic effect. More concretely, considering a hypothetical downstream factor (factor H), which would be able to recognize both H3 acetylation and TBP at gene promoters: Upon loss of one of the two functions of SAGA (HAT or TBP-loading), the recruitment of factor H would be impaired but maintained to some extend by the other function, leading to mildly reduced Pol II transcription. However, upon combined loss of the HAT and TBP-loading functions of SAGA, the recruitment of factor H would be majorly affected, causing a seemingly synergistic effect of the HAT and TBP-loading functions of SAGA on Pol II transcription.

Overall, our findings suggest that SAGA might affect self-renewal capacities and Pol II transcription in mouse ESCs especially through its TBP delivery function. However, the histone modifying and TF-interacting functions of SAGA might further contribute in defining the overall functional importance of SAGA for Pol II transcription. Indeed, we hypothesis that the HAT and DUB modules of SAGA as well as its TF-interacting subunit, Trrap, might have important synergistic effects with the TBP loading function of SAGA mediated by its core, as suggested by studies in yeast (Sterner et al., 1999; Baptista et al., 2017). We could test this hypothesis by assessing if combined inactivation of subunits of the DUB and HAT modules together with Supt3h would recapitulate the effects observed on Pol II transcription upon inactivation of the core subunit Supt7l, as discussed earlier.

3.2. How does ATAC influence Pol II transcription if it is not through its HAT activity?

H3K9ac levels at gene promoters have been found to correlated with effects of ATAC subunits on Pol II transcription in human lung cancer cells, suggesting an importance of the HAT module in this cellular context (Mi et al., 2017; Mi et al., 2018). However, our findings indicate that the HAT module of ATAC might not be its most crucial function in Pol II transcription as inactivation of the ATAC-specific HAT subunit Tada2a could not recapitulate phenotypes observed upon depletion of either Yeats2 or Zzz3, two key subunits of ATAC. Besides its HAT module, two other functions have been described within ATAC that might be involved in its function for Pol II transcription: i) Its Atac2 (Kat14) subunit was described to possess histone H4 acetylation capacities, which however were found *in vitro* to be greatly weaker compared to acetyltransferase activities of Gcn5 and are under debate (Suganuma

et al., 2008; Wang et al., 2008; Guelman et al., 2009). ii) The interaction surface formed by dimerization of Yeast2 and Nc2β has been suggested to enable TBP binding, which implies that besides its chromatin modifying functions ATAC could also be involved in the recruitment of TBP to gene promoters (Wang et al., 2008).

Related to i) we could not find major impacts on acetylation levels at lysine 16 of histone H4, a residue which was suggested to be an *in vivo* target of Atac2, in our *Supt7l$^{-/-}$ + Yeats2$^{AID/AID}$* double mutant cell lines. Also, we think that, as loss of the histone H3 acetylation activity of ATAC did not seem to be of major importance for Pol II transcription, inactivation of the putative histone H4 acetylation activity of Atac2 might also cause no major consequences. This is further supported by findings that the catalytic activity of Tip60, also responsible for acetylation of H4, does not cause major effects on Pol II transcription (Acharya et al., 2017).

Regarding ii) it might be that TBP delivery at gene promoters could be an important function of ATAC in Pol II transcription in mouse ESCs. We could address this possibility by performing ChIP-seq against TBP in our *Yeats2$^{AID/AID}$* mutant cell lines following auxin treatment. Evaluation by qPCR with primer sets directed against gene promoters of RPGs, most severely affected on newly synthesized RNA levels, could allow us some first indication. Importantly, however this assay will only allow us to indirectly assess the dependency of TBP recruitment on Yeats2. As such, reduced levels of TBP might be expected at genes that display reduced levels of Pol II transcription and might be caused by other reasons. For example, if absence of ATAC prevents efficient remodelling and opening of chromatin structures at gene promoters, TBP binding to gene promoters might be occluded through nucleosomes and therefore lead to reduced TBP levels independent of delivery by ATAC. More extensive experiments such as time course auxin treatment of *Yeats2$^{AID/AID}$* cells followed by a combination of ATAC-seq, to assess chromatin accessibility, and ChIP-seq against TBP, to assess TBP binding, might reveal effects on chromatin accessibility and TBP binding upon inactivation of ATAC. If TBP binding is reduced without major effects on chromatin accessibility, this could more strongly suggest that ATAC might be required for the delivery of TBP to gene promoters.

3.3. Histone modifications and histone modifying activities might be more important during differentiation and development

Earlier findings indicated that inactivation of Gcn5, the major HAT enzyme of SAGA and ATAC, affects the differentiation potential of mouse ESCs (Lin et al., 2007; Wang et al., 2018a). This suggests a potentially more important role of the histone modifying activities of SAGA and ATAC during differentiation, which is in agreement with the requirement of Gcn5 during mouse embryonic development (Yamauchi et al., 2000; Xu et al., 2000; Bu et al., 2007). Similar findings have been reported for the histone modifying function of the TIP60 complex (Acharya et al., 2017). Catalytic

inactivation of the histone modifying activities of TIP60 did not impair mouse ESC growth or self-renewal but resulted in defects during mouse embryonic development.

This dependency on histone modifying activities during development but not in ESCs could be based on different reasons. Foremost, during differentiation and development, transcription at several genes must be repressed, while induced at others. These processes might be highly dependent on transcription regulating functions such as histone modifying complexes. Indeed, the defects observed in differentiation of mouse ESCs with catalytic inactive Tip60, mentioned earlier, were suggested to be based on the delayed induction of expression of cell type-specific genes (Acharya et al., 2017). The importance of histone modifications upon changes of transcription programs is further highlighted by findings in yeast, as described earlier. Deletion of histone tails, where most modification sites are localized, does not lead to major growth differences when yeast cells are cultured under optimal conditions, but growth defects appear in stress conditions (Kim et al., 2012). On the other hand, transcription in ESCs might be less dependent on histone modifications based on their specific chromatin structure such as a generally open chromatin and few heterochromatic regions.

It could be interesting to assess if histone modifying activities would be important during transitions or changes of transcription programs and if they might be less crucial once the transcription program has been established. This could be assessed by generating inducible cell lines for histone modifying enzymes such as Gcn5 using for example the auxin-inducible degron (AID) system. Important, addition of the AID-tag should not affect the enzymatic functions of Gcn5 in the absence of auxin, which we could unfortunately observe for our $Tada3^{AID/AID}$ cell lines. Subsequently, AID-Gcn5 could be depleted at different time points along the differentiation of mouse ESCs to assess impacts during transitions of transcription programs. For example, a well-established differentiation protocol for mouse ESCs is their differentiation into terminally differentiated neurons (Bibel et al., 2007). This protocol involves several intermediate steps such as formation of embryonic bodies and generation of neuronal precursor cells. It would be interesting to assess if terminally differentiated neurons would display different phenotypes depending on at which time point Gcn5 is depleted during the differentiation process. Would depletion of Gcn5 at the stage of embryonic bodies cause more severe phenotypes in terminally differentiated neurons, than when depleting Gcn5 from the transition of neuronal precursor cells to terminally differentiated neurons? In general, this system could allow to assess if Gcn5 is important for the transition from one cell type to another.

Further, this system would allow to address if the functions of Gcn5 might not be important when cells are cultured under unstressed conditions independent of their cell type. In other words, can the HAT-independent survival of mouse ESCs be reproduced in differentiated cells. For example, if depletion of Gcn5 during cell type transitions leads to defects in neuronal precursors or terminally differentiated neurons, it would be interesting to assess if depletion of Gcn5, once the neuronal precursor cells or terminally differentiated neurons are established and maintained in culture, causes similar

phenotypes. More precisely, if Gcn5 depletion during the differentiation program leads to misshaped terminally differentiated neurons, would depletion of Gcn5 once the terminally differentiated neurons are established also cause the same phenotypes? Would depletion of Gcn5 once the cell type-specific transcription program is defined still lead to phenotypic defects or would it cause no major impacts as observed for mouse ESCs?

Importantly, there are several experimental settings that need to be considered. Differentiation of *Gcn5$^{-/-}$* mouse ESCs through embryonic body formation assays cause defects especially in mesodermal lineages (Wang et al., 2018a). Therefore, phenotypes might be more drastic and consequently easier to assess following a differentiation protocol for mesoderm lineages instead of neuronal lineages. Also, we found that prolonged auxin treatment over several days leads to growth defects in wildtype cells, meaning that the duration of auxin treatment needs to be limited to avoid secondary effects of the AID system. It might further be important to consider the inactivation of *Pcaf*, the gene encoding the second HAT enzyme of SAGA and ATAC, to avoid potential compensatory functions upon depletion of Gcn5. Inactivation of *Pcaf* should not cause major defects in differentiation as *Pcaf$^{/-}$* mice are viable (Yamauchi et al., 2000).

4. Why are some chromatin modifying complexes more essential for mouse ESCs than others?

As summarized in the Introduction, the effects of inactivation of subunits of several chromatin modifying complexes on mouse ESC growth and self-renewal have been studied (Surani et al., 2007; Festuccia et al., 2017a). Overall, factors involved in the formation of heterochromatic regions were found to be less important for the physiology of mouse ESCs, while factors involved in active gene transcription were generally found to be more crucial (Festuccia et al., 2017a). The following discussion section will be focused on chromatin modifying complexes related to euchromatic regions only.

Interestingly, inactivation of subunits of the chromatin remodelling complexes esBAF, INO80 and TIP60 were often found to be incompatible with mouse ESC viability (Bultman et al., 2000; Ho et al., 2009a; Ho et al., 2009b; Kidder et al., 2009; Wang et al., 2014; Hu et al., 2009a; Fazzio et al., 2008; Herceg et al., 2001). Further, depletion or induced deletion of subunits of these complexes were reported to affect self-renewal of mouse ESCs. Subunits of the COMPASS and MSL/NSL complexes, which possess histone modifying functions, were also found to affect mouse ESC self-renewal to some extend without however being essential for cell survival (Ravens et al., 2014; Fang et al., 2016; Glaser et al., 2006; Wang et al., 2016; Bledau et al., 2014; Glaser et al., 2009; Li et al., 2009; Li et al., 2012b). These observations suggest differences between chromatin modifying complexes, with some having crucial functions in mouse ESC survival, while others have important functions in ensuring their self-renewal

capacities. Indeed, in our study we found that inactivation of key subunits of ATAC caused a seemingly non-viable phenotype when homozygous, while inactivation of SAGA subunits did not.

Overall, chromatin modifying complexes and coactivators seem to have varying impacts on mouse ESCs. However, comparison and interpretation of the phenotypes caused by loss of subunits of the different complexes are difficult between studies as several biases might exist. i) Only few subunits were in general assessed per complex, which might lead to differences in severity of the phenotypes depending on how important the targeted subunits are for complex integrity and function. This is highlighted also in our study by observations that subunits of the HAT module of SAGA or ATAC did not cause any impact on self-renewal or growth of mouse ESCs. In this regard, frequently studies have been focusing on subunits containing the histone modifying activity of coactivators, which might have caused an underestimation of effects. ii) Varying backgrounds of mouse ESCs as well as varying culturing conditions were used between studies, which can have a major impact on the phenotypes observed as recently highlighted for Nanog$^{-/-}$ cells (Hastreiter et al., 2018; Navarro, 2018).

For proper comparison, the effects of inactivation of subunits of chromatin modifying complexes on Pol II transcription and phenotypes in mouse ESCs would need to be assessed in a more standardized manner. In combination with the appropriate positive and negative controls, this might allow to answer if defects in cell viability and self-renewal following inactivation of different chromatin modifying complexes originate from comparable effects on Pol II transcription. For example, if subunits crucial for complex integrity of each of the chromatin modifying complexes could be inactivated in the same background of mouse ESCs and assessed under the same culturing conditions, would different chromatin modifying complexes cluster in groups (such as viable vs non-viable, effects on self-renewal vs no effect) and could this grouping allow a functional interpretation of roles in Pol II transcription? More precisely, by further analysing effects on Pol II transcription upon conditional inactivation of subunits of chromatin modifying complexes causing a 'non-viable' phenotype would it be possible to observe the reduction of comparable genes important for cell viability and proliferation, such as in our case translation-related genes upon ATAC inactivation?

Viability upon inactivation of subunits of chromatin modifying complexes could be assessed by a lentivirus-based CRISPR-Cas9-mediated loss-of-function screen with assessment of sgRNA frequencies at several time points (Bertomeu et al., 2018). This would potentially allow to classify subunits of chromatin modifying complexes based on the kinetics of the decrease of sgRNA frequencies over time. For example, loss of some factors could lead to an immediate non-viable phenotype and the loss of sgRNAs targeting these factors, while others might cause lethality gradually which would be reflected by gradual loss of the targeting sgRNAs. The monitoring of sgRNA frequencies over several time points could also allow to estimate effects on growth and proliferation upon inactivation of 'viable' subunits of chromatin modifying complexes.

Impact on self-renewal capacities could be further analysed following the same CRISPR-Cas9-mediated loss-of-function screen by distinguishing undifferentiated pluripotent mouse ESCs from differentiating cells through for example alkaline phosphatase (AP) staining. The AP-stained cell population could be sorted by FACS into AP positive (undifferentiated) and AP negative (differentiating) cells at given time points. Analysing the sgRNA frequencies over time in the two fractions (AP+ vs AP-) would allow to assess if sgRNAs targeting specific subunits of chromatin modifying complexes would be depleted more strongly in the AP+ fraction compared to the AP-fraction, which could suggest defects in maintaining self-renewal upon inactivation of the targeted subunits. Alternatively, to AP staining, a Nanog-GFP fusion cell line could be used for the screen. This would allow to distinguish undifferentiated cells from differentiating cells through high or low levels of Nanog expression, respectively (Chambers et al., 2007).

Subsequent validation experiments on single cell knockouts or conditional cell lines, targeting especially subunits of chromatin modifying complexes crucial for complex integrity, could allow to further study effects on Pol II transcription for several chromatin modifying complexes. This could, as mentioned above, allow to reveal if similar phenotypes are caused by comparable effects on transcription. For instance, do chromatin modifying complexes, which show to be required for self-renewal but which are not essential, show a given signature of genes affected upon their inactivation? Once established, these cell lines together with the established genome editing and screening systems could additionally be used to assess phenotypes upon combinatorial inactivation of subunits of chromatin modifying complexes and be extended to other cell types.

Important considerations for these extensive experiments are improvement of sgRNA libraries for the CRISPR-Cas9 screen, proper positive and negative controls for the screen, preliminary tests of the screening systems using these controls and extensive literature searches and exchanges with experts to identify subunits crucial for complex integrity of the different chromatin modifying complexes. Positive controls could include, beside pluripotency transcription factors, also factors generally involved in transcription such as subunits of Pol II and general transcription factors and might further include factors with known essentiality such as factors involved in ribosome biogenesis, cell cycle or metabolism (Bertomeu et al., 2018).

5.   How frequent are PTMs on subunits of the SAGA and ATAC coactivator?

Our results presented in Annexe 3. 'Examination of post-translational modification states of endogenous SAGA and ATAC subunits by immunoprecipitations followed by mass spectrometry analyses from human cells' indicate that various subunits of the SAGA and ATAC coactivator complexes undergo post-translational modifications (PTMs) at several residues in agreement with previously published reports (Mischerikow et al., 2009; Spedale et al., 2012). Especially the YEATS2

subunit of ATAC was found to potentially contain as much as 18 possible residues which can be phosphorylated. Of these 18 possible sites eight seemed to be reproducibly detected in our IP experiments. Seven of these phosphorylation sites, with the exception being T132phos, were previously reported (Spedale et al., 2012). Quantification estimations revealed that most modifications occurred only in a small subpopulation (less the half) of the respective subunits, which could however sometimes be considerably different between cell types (HeLa vs U2OS cells). Some frequently modified residues identified include for example ZZZ3 S777phos in HeLa cells (~ 65%), YEATS2 S447phos in U2OS cells (~ 60%), SUPT7L S108phos in both HeLa and U2OS cells (~ 80%) or SUPT20H S437phos in U2OS cells (~ 80%). Except for S447phos in YEATS2, these modified residues were not previously reported (Spedale et al., 2012).

5.1. PTMs of SAGA and ATAC subunits do generally not localize within known protein domains and are not evolutionary highly conserved

In general, the modified residues considered to be relatively reproducibly identified in our analysis did not localize within any reported protein domain of the respective subunits (Spedale et al., 2012). Exception being the TADA3 S338phos, which was found to lie within the Ada3 homology domain, a domain of unknown function, and SGF29 K288ac at the very end of the Tudor domain (Spedale et al., 2012). Interestingly, K288ac of SGF29 had also been previously identified, in contrast to S338phos of TADA3 (Spedale et al., 2012). As most PTMs identified here do not seem to localize within protein domains required for interaction with other subunits, this suggests that in the case of SAGA and ATAC PTMs might not affect complex assembly. In line with this, the shared HAT subunits (GCN5, PCAF, TADA3, SGF29) incorporated in either SAGA or ATAC did not reveal any major differences in modified residues between the two complexes. This might indicate that the differential incorporation of the HAT module into SAGA or ATAC might not be mainly directed through PTMs. Additionally, for several residues tested, a rough analysis suggested that no corresponding site is found in evolutionary distant species such as fly (ATAC) or yeast (SAGA). This would imply a rather low evolutionary conservation of these residues. Nevertheless, this could suggest that the identified modifications at these residues could be dependent on signalling pathways and effector proteins exclusive to some species, such as mammals, and not others.

5.2. Do the identified PTMs on SAGA and ATAC subunits have a functional importance?

As mentioned above, we could find differences in modification frequencies between cellular compartments (nucleus vs cytoplasm) and more surprisingly also between cell types (HeLa vs U2OS) for some residues. Although several of these observations seem intriguing and could potentially reflect differences in signalling pathways (especially related to the cell type-specific differences), the

interpretations, particularly related to the frequency estimations, should be considered carefully. Several technical biases can cause over- or underestimation of peptide abundancies which were used for the calculations. For instance, the ionization step of the mass spectrometry analysis could be favoured for peptides containing either modified or unmodified residues leading to higher detection of either one or the other. Also, importantly, the general abundancy of the subunits within the purified IP fractions can affect the detection accuracy of modified peptides. This is demonstrated by the fact that PTMs of ATAC subunits are generally not found within cytoplasmic extracts in contrast to nuclear extracts, which is probably because ATAC is not present within the cytoplasm or less abundant.

To validate the identified PTMs and to assess their importance for the function, localization or assembly of either SAGA or ATAC, additional experiments would be required. For example, substitution studies of the respective amino acids could allow to assess the impact of individual or combined PTMs on the function of SAGA and ATAC. If possible, raising of specific antibodies against the modified versions of the identified peptides could further allow to evaluate the cellular distribution of these modifications in living cells by using for instance the VANIMA method described in Annexe 4.

Most importantly however would be the identification of the kinases, acetyl- or methyltransferases responsible for the respective modifications. This could be partially achieved *in silico* by comparing known binding motifs of the modifying enzymes to the amino acid sequence surrounding the modified residues. The search for candidate 'writer' enzymes could be facilitate if a specific antibody against a given modified residue is available. This antibody could be subsequently used in a CRISPR-Cas9-based screen targeting all or most of the known effector enzymes. An intriguing candidate for initial tests could be the ATAC subunit, YEATS2, which showed a remarkably high count of phosphorylation sites. Also, phosphorylation of some residues of YEATS2, such as S447 and S627, displayed reproducibly different levels between HeLa and U2OS cells among the different IP experiments (as estimated by our calculations) which could potentially reflect functionally relevant differences.

Conclusions

Conclusions

During my doctoral studies, I was able to study the importance of subunits of the two coactivator complexes, SAGA and ATAC, for RNA polymerase II (Pol II) transcription in mammalian cells. This was enabled through the generation of an unprecedented range of mouse ESC lines with inactivation for subunits of the two complexes using CRISPR-Cas9. Due to lethal phenotypes, I also generated mouse ESC lines with endogenously AID (auxin inducible degron)-tagged subunits of ATAC, enabling rapid depletion of these subunits and the consequent analysis of early effects on Pol II transcription.

Using these mutant cell lines, I was able to reveal that inactivation or depletion of core subunits of SAGA and ATAC dramatically affect self-renewal capacities of mouse ESCs. I further found that the importance of SAGA and ATAC for mouse ESC self-renewal does not reside within their shared histone acetyltransferase (HAT) activity. This suggests that SAGA and ATAC bear additional, critical functions, which are largely independent of their HAT modules. For SAGA, I could subsequently reveal that its TBP-loading function plays a role for proper self-renewal of mouse ESCs. I also found that an increase of histone H2B ubiquitylation upon inactivation of a subunit of the deubiquitylation (DUB) module of SAGA, which is also shared with other complexes, affected mouse ESC proliferation but not self-renewal. In contrast, it remains unclear how ATAC impacts the self-renewal capacities of mouse ESCs.

Through newly synthesized RNA analyses, I could further show that SAGA and ATAC significantly regulate mainly non-overlapping gene sets with a potential mild general impact on Pol II transcription in mouse ESCs. Although comparable phenotypes on self-renewal had been observed, depletion of core subunits of ATAC significantly affected genes implicated in cytoplasmic translation, such as ribosome protein genes (RPGs), while inactivation of core subunits of SAGA caused a downregulation of genes implicated in the pluripotency network of mouse ESCs. In general, I could find little overlap between genes significantly regulated by SAGA or ATAC subunits, suggesting that SAGA and ATAC might majorly regulate different groups of genes in mammalian cells.

In a collaborative work, I could further show that acute depletion of the TF-interacting subunit, TRRAP, shared between the SAGA and TIP60 complexes, leads to a major downregulation of RNA synthesis by Pol II in human cells. It is currently under investigation if these effects on Pol II transcription mainly reflect functions of the SAGA or TIP60 complex. Additionally, I found that several subunits of SAGA and ATAC are extensively modified post-translationally in human cells.

Overall, I found an important HAT-independent role of the SAGA and ATAC coactivator complexes in the maintenance of mouse ESC self-renewal and growth. I also revealed that, although inactivation of the two complexes led to a comparable impairment of the self-renewal capacities of mouse ESCs, SAGA and ATAC mainly regulate non-overlapping genes but might mildly affect global Pol II transcription in mammalian cells.

Materials & Methods

Material and Methods

1. Step-by-Step Protocols

1.1. Step-by-Step Protocol 1: Protocol for the generation of CRISPR-Cas9 knockout or auxin-inducible degron knock-in cell lines in mouse embryonic stem cells

1.1.1. Principle

Recent advances in the RNA-programmable CRISPR-Cas9 technology greatly facilitate the genetic engineering of mammalian cells (Doudna & Charpentier, 2014). The CRISPR-Cas9 method is derived from the type II CRISPR-Cas system of bacteria, which is frequently referred to as the adaptive immune response of bacteria against invading viruses (Doudna & Charpentier, 2014). The key components of this gene editing technology are the endonuclease Cas9 and an engineered, interchangeable single-guide RNA (sgRNA), which together form a complex capable of inducting double strand breaks at specific genomic locations (Figure 56).

The sgRNA consists of a hairpin, duplex RNA structure at the 3' end, which enables the recognition by Cas9 (Figure 56) (Doudna & Charpentier, 2014). A roughly 20 nucleotide long sequence at the 5' end of the sgRNA is complementary to the target DNA and designed to recruit Cas9 in proximity to a PAM (protospacer adjacent motif) sequence (Figure 56) (Doudna & Charpentier, 2014). These 20 nucleotides, sometimes referred to as gRNA (guide RNA), can be easily replaced making the CRISPR-Cas9 system a flexible tool for genetic engineering at any possible genomic location containing a PAM sequence.

The PAM sequence consists of the nucleotide sequence 'NGG' where N represents any of the four nucleotides followed by two guanine nucleotides. Cas9 recognizes the PAM sequence and subsequently induces a double strand break three nucleotides upstream of the PAM (Figure 56) (Doudna & Charpentier, 2014).
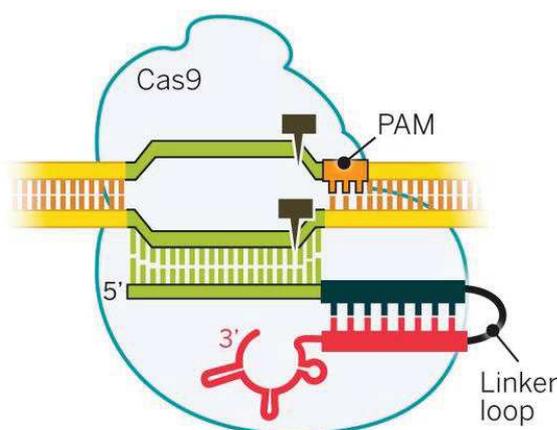
Figure 56: Mechanism of action of CRISPR-Cas9. Cas9 is shown in light blue and associates with the hairpin structure of the sgRNA (in black and red). A roughly 20 nucleotide long segment of the sgRNA additionally binds to its complementary strand within the genomic DNA (represented in green). Upon binding Cas9 induces a double strand break (shown by black pins) three nucleotides upstream of the PAM sequence (shown in orange). PAM, protospacer adjacent motif. Doudna & Carpentier, 2014.

1.1.1.1. Constitutive knockout cell lines using CRISPR-Cas9

To generate constitutive knockout (KO) cell lines for a given gene, we targeted the first out-of-frame exon, shared by all transcript variants and excluding exons containing the translation start site, for deletion by recruiting Cas9 through two sgRNAs to the surrounding introns (gRNA sequences shown in Table 6). Out-of-frame exons represent exons with a nucleotide sequence that is not divisible by three, the length of a codon. The double strand breaks induced by Cas9 within the introns are subsequently repaired by cellular processes frequently causing the loss of the targeted exon. Loss of an out-of-frame exon consequently leads to a shift in the codon-reading frame of the ribosome downstream of the targeted exon resulting in the synthesis of a scrambled protein. Additionally, this frameshift frequently gives rise to a premature stop codon within the RNA molecule causing the induction of NMD (non-sense mediated decay) and therefore reducing the overall levels of the scrambled RNA.

For our KO cell lines, we generally deleted not more than 1.5 kb with an average theoretical deletion size of roughly 600 bp (Table 6). Importantly, we never targeted the exon containing the translation (ATG) start codon to avoid any alternative start codon usage by the ribosome.

Table 6: Table summarizing genes targeted for constitutive knockout by CRISPR-Cas9. For each of the targeted genes the binding region of the gRNAs, the sequence of the gRNA, the respective PAM sequence, the strand location of the gRNA and the theoretical deletion size are shown. gRNAs are generally designed to recruit Cas9 to the surrounding introns of the first out-of-frame exon excluding ATG start codon-containing exons. Ex, exon; +, forward strand; -, reverse strand. Designed by Bernardo Reina San Martin.

Target	Region	gRNA sequence	PAM	Strand	Deletion size
mSupt20h	5'Ex4	TCGCTTGCACTCACTCGT	AGG	+	386
	3'Ex4	GTAGAGCAGTCCAGTCGG	AGG	-	
mSupt7l	5'Ex3	ACCAGTACGTATTCAGAG	TGG	+	1071
	3'Ex3	ACCATCTCCCTCGCCCCG	AGG	+	
mSupt3h	5' Ex3	TCCTGAAGCCTGAATTTGGT	AGG	+	1330
	3' Ex3	GTGATGGGATCTATTCAGTG	TGG	+	
mAtxn7l3	5'Ex3	AGATGCCAGATTTAGCGA	CGG	-	833
	3'Ex4	CCCCATAAGACTACACCT	CGG	-	
mTada2b	5'Ex2	CCTACATAGATGTACCTGAG	CGG	-	514
	Ex2	TTATGAGATAGAGTATGACC	AGG	+	
mTada2a	5'Ex3	GCTACAGGTAGTCTTCCCTG	CGG	-	361
	3'Ex3	CTGCTGTGTAGTAGACAGAG	TGG	-	
mZzz3	5'Ex5	GACTAGGTACTTCGTAACTC	AGG	+	500
	3'Ex5	AGATATCACTGCATTACATG	GGG	-	
mYeats2	5'Ex6	TCACTGAAACAGTATTCAGT	AGG	-	238
	3'Ex6	CCGTTACTGCATATTCACAG	TGG	+	

1.1.1.2. Auxin-inducible degron knock-in cell lines using CRISPR-Cas9

The auxin-inducible degron (AID) system represents an inducible depletion method in which protein levels can be rapidly and reversibly reduced in an efficient manner (Verma et al., 2020). This system is based on the recognition of AID sequences encoded within proteins by the plant-specific F-box protein TIR1 (transport inhibitor response 1) in the presence of the plant-specific hormone IAA (auxin) (Figure 57). Association of TIR1 upon IAA treatment with the SCF (Skp, Cullin, F-box containing) complex subsequently leads to proximity-induced polyubiquitylation and proteasomal degradation of the proteins possessing the AID sequence (Figure 57) (Verma et al., 2020).

A major advantage of this system compared to other inducible depletion methods, such as siRNA-mediated depletion, is that it allows the efficient and acute depletion of the target on protein levels within few hours. Therefore, it enables the determination of loss-of-function consequences largely independent of secondary effects or compensatory mechanisms for a given protein (Verma et al., 2020). A disadvantage of this method is that it requires the ectopic expression of the plant-specific TIR1 protein and the homozygous insertion of the AID sequence to the gene locus of the protein of interest (Verma et al., 2020). Importantly, the SCF machinery is conserved in eukaryotes and is present within the cytoplasm and nucleus allowing for efficient polyubiquitylation and depletion of cytoplasmic and nuclear proteins (Verma et al., 2020).

For our AID cell lines, we first randomly integrated a construct containing *Tir1* fused to one HA-tag and separated by an *Ires* sequence from the *BirA* gene fused to three HA-tags (Figure 58). The IRES (internal ribosome entry site) sequence allows for the generation of two independent proteins from one transcript, as the ribosome can independently initiate translation from the IRES sequence (Carter & Shieh, 2015). As the two gene products encoded within the RNA molecule are synthesized by distinct translation machineries through independent initiation events, the presence of an IRES sequence is
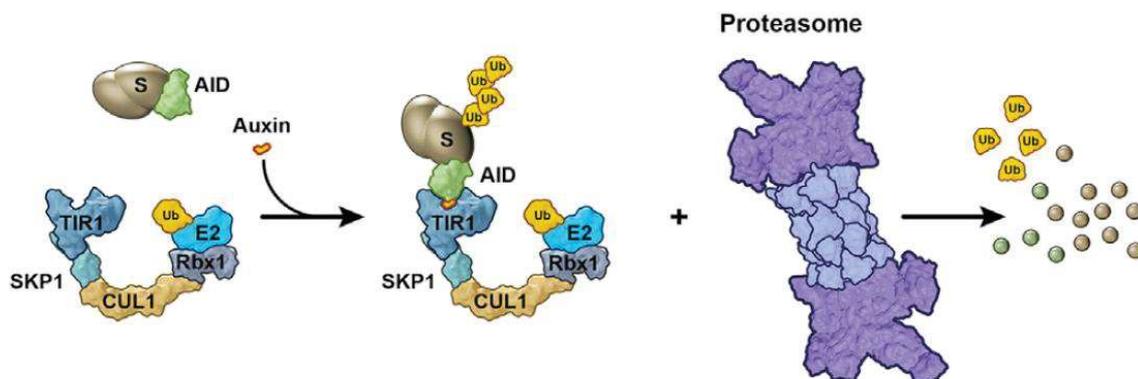
Figure 57: Schematic representation of the auxin-inducible degron system. On the top left, the substrate or target protein (brown, S) is shown containing the auxin-inducible degron (AID) sequence in green. On the bottom left the SCF complex is indicated including the ectopically expressed, plant-specific F-box protein TIR1. SCF subunits SKP1, CUL1, RBX1 and E2 ubiquitin ligases are endogenously present within eukaryotic cells. Ubiquitin (Ub) is shown in yellow. Addition of auxin (shown as orange molecule) leads to the recognition of the AID sequence by TIR1 and the polyubiquitylation of the substrate protein by the SCF complex. The polyubiquitylated protein is subsequently degraded by the proteasome following common cellular processes. From Verma et al., 2020.
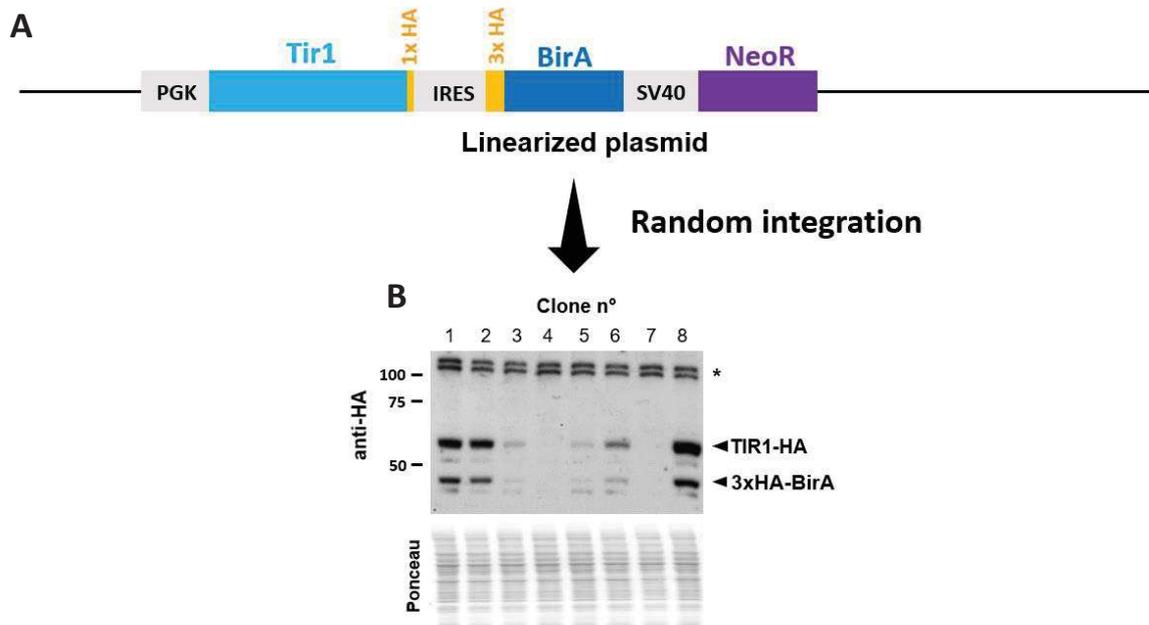
Figure 58: Ectopic expression of TIR1 and BirA in mouse embryonic stem cells. A. Shown is the linearized version of the TIR1 and BirA-encoding plasmid used for ectopic expression. PGK, phosphoglycerate kinase 1 promoter; Tir1, transport inhibitor response 1; HA, hemagglutinin; IRES, internal ribosome entry site; BirA, bifunctional ligase/repressor A; SV40, simian virus 40 promoter; NeoR, Neomycin resistance. **B.** Western blot analysis of the ectopic expression of TIR1 and BirA as assessed by anti-HA antibodies in different mouse embryonic stem cell clones obtained after antibiotic selection. Ponceau staining represents loading control.

generally believed to lead to a non-stochiometric expression of the two products. Downstream of the *Tir1* and *Ires* sequence, our construct further contained the gene encoding for the biotin ligase, BirA.

BirA is of bacterial origin and enables the *in vivo* biotinylation of proteins containing a BioTag sequence, which we included in our knock-in constructs (more details below) (Strübbe et al., 2011). The above described construct is under the control of a PGK (phosphoglycerate kinase 1) promoter reflecting a promoter driving rather low expression compared to viral promoters such as the CMV (cytomegalovirus) promoter. We chose this promoter for two reasons. First, highly expressing promoters such as the CMV promoter are frequently silenced within mouse embryonic stem cells and therefore are only suitable for transient expression (Kawabata et al., 2005). Second, we wanted to perturb our cellular system as little as possible by restricting the overexpression levels of TIR1 and BirA. Importantly, these cell lines were obtained through antibiotic selection with Neomycin (also known as Geneticin or G418) and should therefore be maintained in culture with roughly 300 µg /ml G418-containing medium (details in 1.1.2. Required materials). We also generated cell lines expressing the same construct but with Flag- instead of HA-tags.

For the insertion of the AID sequence to the gene locus of our proteins of interest, we first generally assessed if there was a preference reported in the literature for tagging the respective proteins either N- or C-terminally. If no information was available, the termini shared among the most relevant transcript variants was chosen. For the knock-in of the AID sequence, the transfection of two plasmids designed by Bernardo Reina San Martin was required. One expressing the Cas9 protein with one or two sgRNAs

recruiting Cas9 to PAM sequences close to the start codon (for N-terminus knock-in) or stop codon (for C-terminus knock-in) (Table 7 and Figure 59A). Two sgRNAs were used when the efficiency of the individual sgRNAs were predicted *in silico* to be low and when two PAM sequences were available in relative proximity.

Table 7: Table summarizing genes targeted for knockin of auxin-inducible degron tags by CRISPR-Cas9. For each of the targeted genes the terminus to which the insert was targeted to, the sequence of the gRNA, the respective PAM sequence and the strand location of the gRNAs are indicated. PAM, protospacer adjacent motif; gRNA, guide RNA; +, forward strand; -, reverse strand. Designed by Bernardo Reina San Martin.

Target	Terminus	gRNA sequence	PAM	Strand
mYeats2	N-ter	TGTTCGCTTGATTCCAGACA	TGG	-
mTada3	N-ter	GGAGGCCGGCCTAATCATGC	AGG	-
	N-ter	ACAAACCTGCATGATTAGGC	CGG	+
mZzz3	N-ter	GTGTTACAAGATCAACAGTG	GGG	+
mZzz3	C-ter	TCCCAGCCAACAGATGACAT	GGG	+

The second plasmid, representing the donor plasmid, contained the knock-in sequence which in our case was composed for N-terminus knock-in of mCherry/eGFP-P2A-1xFlag-BioTag-AID and for C-
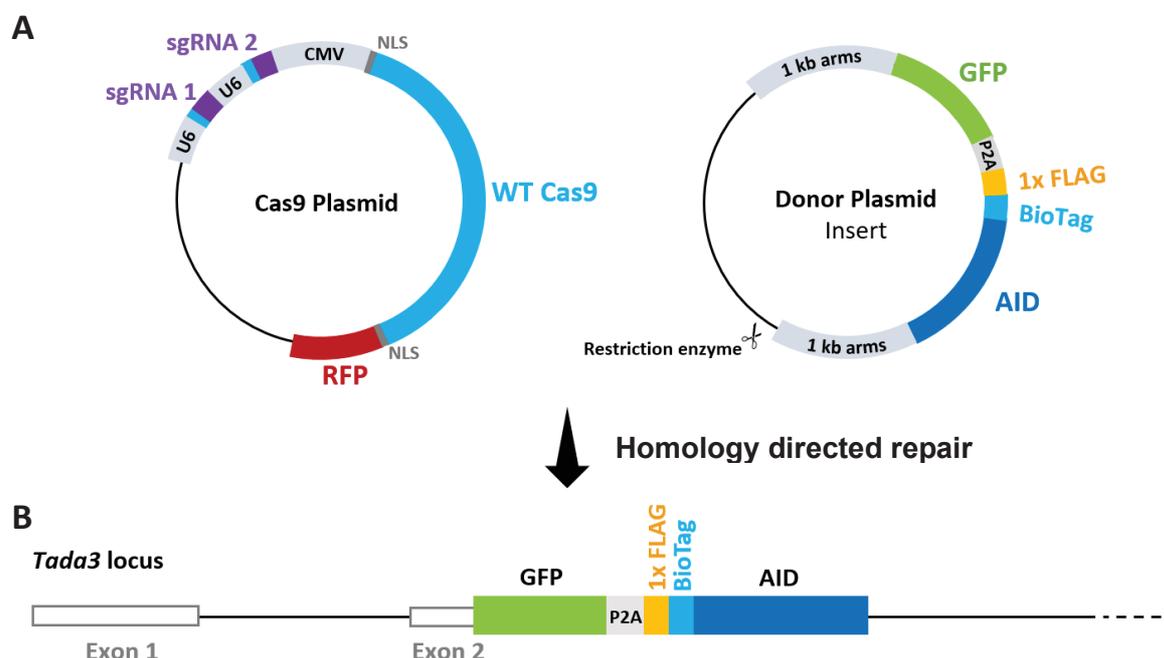
Figure 59: Knock-in of the auxin-inducible degron sequence to the *Tada3* locus using the CRISPR-Cas9 technology. A. The two plasmids required for the CRISPR-Cas9 mediated knock-in are shown. The Cas9 plasmid encodes, in this case, the Cas9 protein fused to RFP (red fluorescent protein, here mCherry) and two sgRNAs. The donor plasmid with the insert containing the AID (auxin-inducible degron) sequence followed by a BioTag, 1x Flag-tag, a P2A sequence and GFP (green fluorescent protein) is shown. The insert is surrounded by roughly 1 kb homologous arms that allow the use of the knock-in construct by the homology directed repair machinery following the Cas9-induced double strand breaks. This plasmid needs to be linearized using a restriction enzyme at an appropriate position before transfection (indicated by scissors). **B.** Shown is the endogenous locus of *Tada3* and the integration of the knock-in insert at the N-terminal end downstream of the start codon in the second exon of *Tada3*. CMV, cytomegalovirus; U6, U6 snRNA promoter; NLS, nuclear localization sequence.

terminus knock-in of AID-BioTag-1xFlag-P2A-mCherry/eGFP (example for N-terminus knock-in at the *Tada3* locus shown in Figure 59).

As the AID tag and the fluorescent protein (either mCherry or eGFP) each represent a roughly 25 kDa big addition to our endogenous proteins of interest, we chose to separate the fluorescent protein from the remaining protein by a P2A sequence thereby reducing the total size of the tag added to the protein by roughly half the size. The P2A sequence, separating the fluorescent protein mCherry or eGFP from the remaining insert, has a similar function as the IRES sequence described above. The presence of the P2A sequence results into the synthesis of two separate polypeptides from one transcript as the ribosome skips the formation of a peptide bond between the glycine and proline amino acids at the C-terminal end of the P2A peptide (Kim et al., 2011). Consequently, this leads to the formation of two independent proteins with one polypeptide ending with the last glycine of the P2A and therefore containing the majority of the P2A peptide, while the other protein starts with the last proline of the P2A peptide. In contrast to the IRES sequence, the P2A sequence is thought to lead to stochiometric amounts of the proteins encoded either up- or downstream of the P2A as the same translation machinery is used for both polypeptides (Kim et al., 2011).

In our constructs, the P2A sequence enables the synthesis of our protein of interest fused to the AID sequence including the 1xFlag tag and the BioTag and the expression of an independent fluorescent protein. The expression of the fluorescent protein upon insertion of the knock-in sequence to the locus of our protein of interest was used for the selection of positive clones by single cell FACS (fluorescence activated cell sorting). We decide for a fluorophore- instead of an antibiotic-based selection of positive clones due to the circumstance that our proteins of interest are generally encoded by lowly expressed genes. The use of antibiotic selection is challenging when the resistance gene is under the control of a lowly expressed gene.

Importantly, we used long homologous arms (at least 800 bp) at each side of the knock-in construct (Figure 59). These homologous arms are complementary to DNA sequences surrounding the start or stop codon of the targeted locus where Cas9-mediated double strand breaks are induced. Through these complementary arms, the cellular homology directed repair pathway subsequently uses the knock-in construct as a template to restore the integrity of the genomic DNA at the double strand breaks.

Importantly, it subsequently turned out that the 1xFlag could not be properly detected by western blot analysis and might require at least three repetitions for proper revelation (3xFlag). Instead, the BioTag showed to work rather well and allowed us to monitor the degradation of our AID-tagged proteins upon auxin treatment. We also found that addition of our knock-in construct to the N-terminal end generally led to reduced mRNA levels, while C-terminal addition did not. The reason for the reduced mRNA levels in N-terminally AID-tagged constructs is unclear.

1.1.2. Required materials

- **Gelatin** Solution 0,1% in PBS (store at 4°C)

- **Medium**: DMEM (4.5g/l glucose) + 2 mM GLUTAMAX-I + 15% FCS ES tested + 0.1% β-mercaptoethanol + 100 UI/ml penicillin + 100 μg/ml streptomycin + 0.1 mM non-essential amino acids + 1500 U/ml leukemia inhibitory factor (LIF) (store at 4°C)

  ⚠ **Important**: for **AID cell lines:** Geneticin/Neomycin G418 is added in a concentration of 300 μg/ml to maintain the Tir1-BirA transgenes; 20 ml G418 [stock 10 mg/ml] for 500 ml medium; resuspend powder in medium without inhibitors and filter before use.

- **Medium without antibiotics**: DMEM (4.5g/l glucose) + 2 mM GLUTAMAX-I + 15% FCS ES tested + 0.1% β-mercaptoethanol + 0.1 mM non-essential amino acids + 1500 U/ml LIF (store at 4°C)

- **Inhibitors**:
  - 3 μM CHIR99021 (axon medchem, Cat# 1386, resuspended in DMSO [stock at 10 mM], store at -20°C in aliquots of 100 μl)
  - 1 μM PD0325901 (axon medchem, Cat# 1408, resuspended in DMSO [stock at 10 mM], store at -20°C in aliquots of 100 μl)

- **0.25% Trypsine**-1mM EDTA INVITROGEN (store at -20°C in aliquots)

- **1x Phosphate buffered saline (PBS) autoclaved** for washing of cells before trypsinization

- **50 μm filters** required for FACS sorting

- **Fetal calf serum 100% (FCS) ES tested** (store at -20°C in aliquots) required for FACS sorting and freezing for subsequent storage of cell lines

- **DMSO** (store at room temperature) required for freezing for subsequent storage of cell lines

- **Opti-MEM** (store at 4°C) required for transfection of ESCs with Lipofectamine2000

- **Lipofectamine2000** (Thermo Fisher, Cat# 11668019) (store at 4°C) required for transfection of ESCs

- **Plasmids** (store at -20°C) to transfect into ESCs. All CRISPR-Cas9 plasmids used for this PhD work were designed by Bernado Reina San Martin at IGBMC.

IMPORTANT: Donor knock-in plasmid for the generation of AID cell lines should be linearized using restriction enzyme-mediated digestion prior to transfection.

- **Phire Tissue Direct PCR kit** (Thermo Scientific, Cat# F170S) required for DNA extraction and genotyping of clones

- **48-well plates** (Thermo Scientific, Cat# 08-772-3D)

1.1.3. Culture conditions

Temperature: 37°C

CO_2 Levels: 5%

1.1.4. General cell culturing protocol

- **Add gelatin** to the required plates

  o Cover plates with gelatin (~10 ml gelatin for 10 cm plates)

  o Incubate des plates for 30 minutes (or longer) with the gelatin

  ⚠ Note: do not keep the plates with gelatin longer than **two days** under the hood

  o Remove gelatin after incubation time

- **Prepare medium**

  o Add **1 µl PD**0325901 and **3 µl CHIR**99021 for **10 ml** of medium

  o Add to the plates (10 ml of medium for 10 cm plate) and keep them in the incubator (hours not days)

- **Collect the mouse ESCs**

  o Remove the medium from the cells

  o Wash cells with **1x PBS** (~5 ml for 10 cm plate)

  o Add **trypsin** (~1 ml for 10 cm plate)

  o Incubate for **2-3 minutes** in incubator

  o Knock the plates and check under the microscope that the cells have detached

  o Add medium to block the trypsin (**no inhibitors** required in the medium)

  o **Centrifuge** at 1000 rpm, ~5 minutes at room temperature

  o Remove medium and resuspend the cells with medium containing inhibitors

  o Add the cells to the prepared plates containing medium

- **Split the cells every second day**!

⚠️ Note: split the cells in a ratio 1:5 - 1:10 if the cells are kept for continuous culturing

⚠️ **Never let the medium (phenol red) get yellow!** (Orange is O.K.)

1.1.5. Transfection protocol (under cell culture hood)

- Prepare 10 cm plates with gelatin (one day before; as described in section 5.1.4)

⚠️ **Important**: Do not forget a plate for the negative control; = no plasmid transfected treated with the transfection reagents only. This is required to set the gate properly for the subsequent FACS to exclude autofluorescence of the cells.

- Prepare 10 cm plates with **medium <u>without</u> antibiotics** (but with the inhibitors; as described in section 5.1.4.)

- Collect the mouse ES as described in section 5.1.4. and resuspend in **medium <u>without</u> antibiotics**

⚠️ Note: to block the trypsin also use medium <u>without</u> antibiotics

- Count the number of cells (e.g. by using the Countess machine and slides) and seed **2 million living cells** (determined e.g. using Trypan blue staining) per 10 cm plate **1-2 hours before** transfection. Transfection should be performed few hours after seeding while the cells are still in suspension and not yet attached to the plate. This is believed to increase the transfection efficiency.

- Transfect **24 µg plasmid for** the generation of **knockout (KO) cell lines** (plasmid containing Cas9 + sgRNAs) or for the generation of **knock-in cell lines 20 µg of Cas9 plasmid** and **30 µg of linearized knock-in donor plasmid** per 10 cm plate. Follow Lipofectamine2000 protocol; for a 10 cm plate:

  - Add 60 µl **Lipofectamine2000** in 1,5 ml **Opti-MEM** and incubate for **5 minutes**

  - Add required amount of **plasmid** in 1,5 ml **Opti-MEM** (or <u>nothing</u> for the <u>negative</u> control) during the 5 minutes incubation of the Lipofectamine2000

  - Add the solution with the plasmid (or <u>nothing</u> for the <u>negative</u> control) in the solution with Lipofectamine2000 and **mix gently** by pipetting; incubate for **20 minutes**

  - **Add** the mixture **droplet by droplet** to the cells in the plates prepared with medium <u>without</u> antibiotics

270

- **Replace the medium** <u>without</u> antibiotics with medium + inhibitors + antibiotics **5-6 hours** after transfection

- Keep the cells in the incubator for **48 hours (for KO) to 72 hours (for knock-in) before FACS sorting.** Change medium if required (e.g. if medium turns yellow)

1.1.6. FACS sorting (48 – 72 hours post-transfection)

- **Prepare 96 well plates** (5 plates per plasmid for KO, 3 plates per plasmid for knock-in) with **50 µl gelatin** per well (the day before) and **100 µl medium** per well (the day of sorting) by using a multichannel pipette

- **Collect the transfected cells** as described in section 5.1.4. with following exceptions:

  o **After centrifugation resuspend** the cells in 1 ml of **1x PBS + 5% FCS ES tested**

  o **Pass** the suspension gently **through** a 50 µm **filter** in a special tube for sorting (both can be asked from the FACS facility)

  ⚠ Note: If suspension very concentrated, dilute by adding 1 - 2 ml of 1x PBS + 5% FCS ES tested (**take the rest** of 1x PBS 5% FCS **with you** for the sorting in case)

  ⚠ **Important**: Collect the cells just before sorting! Don't keep the cells for long time in 1x PBS + 5% FCS ES tested.

- Sort cells by fluorescence (if you have the choice especially for KO cell lines take not to strongly fluorescing cells – this are often false positive – and not weakly fluorescing cells)

  o For KO cell lines, sort by for the uptake of the Cas9 + sgRNA plasmid, e.g. for the eGFP fused to Cas9

  o For knock-in cell lines, if possible (might actually be very few and very weak signal) sort for fluorescent protein of the knock-in donor plasmid, e.g. for mCherry separated from the AID sequence by a P2A signal which should only be expressed in cells that have the knock-in at the gene of interest in at least one allele. Otherwise, sort for the fluorophore fused to Cas9, although this strategy only rarely gave homozygous knock-in clones.

- After FACS finished, put 96-well plates into cell culture **incubator**

1.1.7. Culturing the cells after sorting

- **Add** between **50 µl new medium after 3-4 days** of culture per well of the 96-well plates (**without removing** the old medium from the wells!)

- **4-5 days after sorting select** the wells with **one single ESC colony** by using the microscope

  ⚠ Note: To **find the right plane level** to look for the colonies in the 96-well plates, adjust the microscope so that the **numbers of the wells** (A1, A2, …) can be seen sharply; **mark with a pen** the wells on the **cover and** also on the **bottom** of the plate to find the selected clones especially also when the cover is removed, one can expect up to 20 clones per 96-well plate)

1.1.8. Splitting to 48-well plates

- Prepare **48-well plates** with **200 μl gelatin** per well (the previous day)

- Add **500 μl medium + inhibitors** per well of the **48-well plate** (the same day)

- **Collect the selected clones**

  ○ Remove the medium of the colonies in the 96-well plates

    ⚠ **Important**: Remove medium by adding a **1-20 μl pipette tip** onto the glass Pasteur pipette! **Exchange the tip** of the pipette for each clone!

  ○ Wash the cells with **100 μl 1x PBS** per well

    ⚠ **Important**: Remove the PBS by adding a **1-20 μl pipette tip** onto the glass Pasteur pipette! **Exchange the tip** of the pipette for each clone!

  ○ Add **30 μl trypsin** per well

  ○ Incubate for **2-3 minutes** in the incubator

  ○ Add **100 μl** of **medium + inhibitors** per well to block the trypsin

  ○ Pipet several times (~5 times) the suspension with a 20-200 μl pipette to dissociate the colonies

    ⚠ **Important**: **Exchange the tip** of the pipette for each clone! Avoid making bubbles!

  ○ Transfer all cells directly into the prepared 48-well plates

    ⚠ **Important**: **Mark the wells** where a clone has been added to avoid adding accidentally two clones in one well!

1.1.9. Splitting to 24-well plates

- **4-5 days after** splitting the clones into 48-well plates, **check the confluency** and split

  ⚠ Note: some wells will be empty! One can lose up to 30% of the splitted clones at this step.

- Prepare **24-well plates** with **300 µl** gelatin per well (the day before; enumerate the wells on the cover!)

- Add **600 µl medium + inhibitors** per well of the **24-well plate** (the same day)

- **Collect the clones**

  o Remove the medium of the cells in the 48-well plates

    ⚠ **Important**: Remove medium by adding a **20-200 µl pipette tip** on the glass Pasteur pipette! **Exchange the tip** of the pipette for each clone!

  o Wash the cells with **500 µl 1x PBS** per well

    ⚠ **Important**: Remove PBS by adding a **20-200 µl pipette tip** on the glass Pasteur pipette! **Exchange the tip** of the pipette for each clone!

  o Add **100 µl trypsin** per well

  o Incubate for **2-3 minutes** in the incubator

  o Hit the plates and **check** under the microscope that the cells are detached

  o Add **500 µl medium** to block the trypsin (no inhibitors required)

  o **Transfer** the suspension into Eppendorf tubes labelled with the clone numbers

    ⚠ **Important**: **Exchange the tip** of the pipette for each clone!

  o **Centrifuge** at 2000 rpm, 5 minutes at room temperature (in the lab)

  o Remove the medium and **resuspend** the cells in **300 µl medium + inhibitors**

  o Add **200 µl of the cell suspension** into the prepared and enumerated 24-well plates

    ⚠ **Important**: **Exchange the tip** of the pipette for each clone!

  o The remaining **100 µl** are required for the **extraction of DNA**

1.1.10. DNA Extraction (in the lab)

- **Centrifuge** the 100 µl cell suspension at 5000 rpm, 5 minutes at room temperature

- Wash the cells with **500 µl 1x PBS**

- Utilize the protocol for **Direct PCR kit** (Thermo Fisher) for the DNA extraction:

  o Prepare Master Mix for samples. Per sample:

      20 µl Dilution buffer

      0.5 µl Release reagent

- o **Add 20,5 µl of Master Mix** per sample

- o **Vortex** the samples and centrifuge briefly

- o Incubate at **room temperature for 5 minutes**

- o **Inactivate the enzymes at 100°C** (water bath) for **2 minutes**

- o **Centrifuge** the samples briefly

1.1.11. Screening PCR

1.1.11.1.   Primer design

We generally use two primer pairs to screen the individual KO clones by PCR on genomic DNA. One primer pair is designed such that one of the primers is situated within the deleted region (for example the reverse primer of primer pair A in Figure 60). Upon deletion of the targeted region, this primer can no longer anneal to the DNA, therefore leading to no PCR product in homozygous KO clones compared to wildtype (WT) or heterozygous clones.

As the absence of a PCR product can be caused by technical problems (such as no DNA provided), we use an additional second primer pair which is amplifying from the surrounding regions of the deleted region (primer pair B in Figure 60). This primer pair causes the appearance of a smaller PCR product from KO alleles compared to WT alleles and therefore allows, in contrast to the first primer pair, to distinguish heterozygous KO clones from WT clones: WT clones will display one full length PCR product, while heterozygous KO clones will show one full length and a smaller (full length minus deletion size) PCR product and homozygous KO clones will only present the smaller PCR product.

Importantly, the primers should not be localized too close to the gRNA binding sites as the repair of Cas9-induced double strand breaks can frequently lead to larger than expected deletions, which could cause the loss of primer binding sites if localized in relative proximity. Further, to be able to use the standard PCR program described below, the PCR products should be below 2.5 kb in length, otherwise the number of PCR cycles might need to be adjusted to ensure efficient amplification of the PCR products. Importantly, prior to the genotyping step of the selected clones, the two primer pairs should
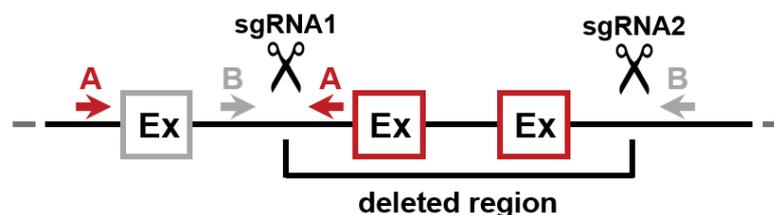
Figure 60: Scheme of primer design for CRISPR-Cas9 knockout screening by PCR. Two primer pairs are generally used with one (A) amplifying from within the deleted region, which leads to the absence of a PCR product when KO was successful. And a second (B), which is situated outside of the targeted region resulting in a smaller PCR product when KO was achieved. Ex, exon. Scissors reflect sgRNAs and therefore Cas9-induced double strand breaks.

be tested on WT genomic DNA to ensure that they result into specific and clearly identifiable PCR products at the expected sizes. Also, for Sanger sequencing only fragments smaller than 1 kb can be properly sequenced. This can be taken into consideration when designing the genotyping primers to enable their subsequent use also for the sequencing validation step of the homozygous KO clones.

The above strategy is also used for the genotyping of AID knock-in clones for which however the design of the primer pair A is slightly different. One primer (for instance the forward primer) of primer pair A is complementary to the knock-in construct, while the second primer (in the example here the reverse primer) is localized within the locus of the gene of interest. This primer pair therefore only leads to a PCR product if the knock-in has occurred in the locus such as in homozygous or heterozygous knock-in clones, while in WT clones no PCR product will be detectable. Unfortunately, this primer pairs cannot be tested for specific amplification prior to the actual genotyping of the clones. However, it is favorable to test for potential unspecific PCR amplification on WT genomic DNA. The second primer pair (primer pair B) is localized, as for the KO screening strategy, at the surrounding sides of the knock-in. Upon integration of the AID sequence, the PCR product while consequently be bigger (full length product plus insert size) compared to amplifications from WT alleles. The PCR products of primer pair B are frequently bigger than 2.5 kb which can lead to weak PCR bands in the knock-in clones. However, combined usage of primer pair A and B should allow for proper identification of the knock-in clones.

Useful software that drastically facilitate primer design are the Primer3Plus website (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi), which can be used to find primer pairs with particular characteristics, such as for instance primer pairs resulting in the amplification of PCR products of a given length, by simply providing the DNA sequence of the region of interest. Primer pairs found by Primer3Plus can be tested *in silico* for their specificity on genomic DNA using the Primer-BLAST tool from NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The Thermo Fisher Multiple Primer Analyzer further allows to verify that selected primers of a primer pair do not form inhibitory secondary structures within themselves or with each other (https://www.thermofisher.com/fr/fr/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html).

1.1.11.2. Master Mix

- Prepare Master Mix for samples. Per sample:

    10 µl of Phire Green Hot Start II PCR Master Mix (from Direct PCR kit)

    7 µl PCR grade water

    1 µl forward primer

1 µl reverse primer

1 µl DNA

- Our **standard PCR program** is the following (might need to be adapted depending on PCR product e.g. melting temperatures of primers, GC content and length of PCR product):

    o 1 cycle of 98.0°C for 5 minutes (initial denaturing – especially because polymerase is a 'Hot start' polymerase)

    o 40 cycles of:

        ▪ 98.0°C for 5 seconds (denaturing)

        ▪ 65.0°C for 5 seconds (annealing)

        ▪ 72.0°C for 45 seconds (elongation)

    o 1 cycle of 72°C for 1 minute (final elongation)

    o Maintain at 4°C (storage)

- Load PCR reactions on **Agarose gel** (1% or 1.5% agarose in TAE buffer depending on sizes of PCR products).

⚠️ **Important**: From Genotyping, **keep** at best **2 wildtype clone, 2 heterozygous clone** and **all homozygous clones** and split them to 6-well plates after 2-3 days (when they are 70-80% confluent

1.1.12. Splitting to 6-well plates (~2-3 days after)

- Prepare **6-well plates** with **2 ml gelatin** per well (day before)

- Add **2 ml medium + inhibitors** per well in the **6-well plate** (same day)

- **Collect the clones**

    o Remove the medium from the cells in the 24-well plates

    ⚠️ **Important**: Remove medium by adding a **20-200 µl pipette tip** on the glass Pasteur pipette! **Exchange the tip** of the pipette for each clone!

    o Wash the cells with **1 ml 1xPBS** per well

    ⚠️ **Important**: Remove medium by adding a **20-200 µl pipette tip** on the glass Pasteur pipette! **Exchange the tip** of the pipette for each clone!

    o Add **150 µl trypsin** per well

    o Incubate for **2-3 minutes** in the incubator

276

o Knock the plates and **check** under the microscope that the cells have detached

o Add **1 ml medium** to block the trypsin (no inhibitor required)

o **Transfer** the suspension in a labelled Eppendorf tube with the number of the clones

⚠ **Important**: **Exchange the tip** of the pipette for each clone!

o **Centrifuge** at 2000 rpm, 5 minutes at room temperature (in the lab)

o Remove the medium and **resuspend** the cells in **500 μl medium + inhibitors**

⚠ **Important**: **Exchange the tip** of the pipette for each clone!

o **Transfer** all **the cells** in the prepared and numerated 6-well plates

⚠ **Important**: **Exchange the tip** of the pipette for each clone!

1.1.13. **Splitting to 10 cm plates** (~2-3 days after)

- Prepare **10 cm plates** with ~**10 ml gelatin** and **6-well plates** with **2 ml gelatin** per well (the previous day)

⚠ Note: 10 cm plates are for freezing and storing the cell lines, the 6-well plate are to detect the proteins per Western blot or for RT-qPCR confirmation if no good antibody is available!

- Add **9 ml medium + inhibitors** per **10 cm plate** and **2 ml** per well of the **6-well plates**

- **Collect the clones**

  o Remove medium from the clones in the 6-well plate

  ⚠ **Important**: Remove medium by adding a **20-200 μl pipette tip** on the glass Pasteur pipette! **Exchange the tip** of the pipette for each clone!

  o Wash the cells with **2 ml 1x PBS** per well

  ⚠ **Important**: Remove medium by adding a **20-200 μl pipette tip** on the glass Pasteur pipette! **Exchange the tip** of the pipette for each clone!

  o Add **200 μl trypsin** per well

  o Incubate for **2-3 minutes** in the incubator

  o Knock the plates and **check** under the microscope that the cells have detached

  o Add **1 ml medium** to block the trypsin (no inhibitors required)

  o **Transfer** into a labelled Eppendorf tube with the number of the clones

  ⚠ **Important: Exchange the tip** of the pipette for each clone!

- o **Centrifuge** at 2000 rpm, 5 minutes at room temperature (in the lab)

- o Remove the medium and **resuspend** the cells in **1 ml medium + inhibitors**

    ⚠️ **Important: Exchange the tip** of the pipette for each clone!

- o Transfer **900 µl of the cells** in the prepared and labelled **10 cm plate** and **100 µl of the cells** in the **6-well plates**

    ⚠️ **Important**: **Exchange the tip** of the pipette for each clone!

1.1.14. Freezing of clones (~2 days later)

- **Collect the cells** of the 10 cm plate as described in section 5.1.4. with following exceptions:

- o **Resuspend** the cells after centrifugation in **6 ml medium + 20% FCS ES tested + 10% DMSO** (do not forget the inhibitors)

    ⚠️ **Important**: **Exchange the tip** of the pipette for each clone!

- o Distribute the resuspended cell suspension in **6 properly labelled freezing tube** (1 ml of suspension per tube)

    ⚠️ **Important**: **Exchange the tip** of the pipette for each clone!

- o Put the tubes into **polystyrene support** and close properly with adhesive band and **store at -80°C for not more than 2-3 days** before putting the freezing tubes in our **liquid nitrogen tanks** (cell lethality upon defreezing is high if tubes are kept to long at **-80°C**).

1.2. Step-by-Step Protocol 2: Protocol for newly synthesized RNA labelling and purification from mammalian cells

1.2.1. Principle

The analysis of RNA levels between wildtype and mutant cell lines is a common way to identify changes in RNA polymerase II (Pol II) transcription. Recently, it has been suggested that total RNA levels can be maintained largely unchanged upon drastic decreases of Pol II transcription in budding yeast (Sun et al., 2012; Bonnet et al., 2014; Baptista et al., 2017; Warfield et al., 2017). This is enabled through an RNA buffering mechanism balancing total RNA levels by regulating RNA decay rates. It is unclear if a similar buffering mechanism exists in mammalian cells. However, mammalian RNA molecules possess a rather long average half-life of roughly 6 hours (Duffy et al., 2019). When analysing total RNA levels, these long RNA half-lives can potentially mask early transcriptional effects especially upon acute depletion of a given factor, such as enabled through the auxin-inducible degron (AID) system.

Several methods were recently developed permitting to directly analyse effects on Pol II transcription and to circumvent any biases potentially introduced through long RNA half-lives or RNA buffering mechanisms (Table 8) (Wissink et al., 2019). One of these methods is the 4-thiouridine (4sU) labelling method and its derivatives in which RNA is labelled for short periods of time with an analogue of uridine containing a thiol group, which subsequently allows the specific purification of the labelled RNAs from the bulk total RNAs (scheme in Figure 61) (Sun et al., 2012; Duffy et al., 2015; Schwalb et al., 2016; Duffy et al., 2019).

Table 8: Table summarizing methods allowing nascent and newly synthesized RNA isolation and analysis and their advantages and disadvantages. PRO, precision nuclear run-on; NTP, nucleoside 5'-triphosphates; caRNA; chromatin-associated RNA; NET, native elongating transcript; Pol II, RNA polymerase II; IP, immunoprecipitation; TT-seq, transient transcriptome sequencing; 4sU, 4-thiouridine; 4tU, 4-thiouracil; EU, 5-ethenyluridine. Based on Wissink et al., 2019.

Method	Short description	Advantages	Disadvantages
Run-on RNA (PRO-seq)	Labelling with biotin-NTP after NTP depletion of isolated nuclei	• nucleotide resolution • streptavidin pull-down	• invasive • extensive protocol
Chromatin-associated RNA (caRNA-seq)	Isolation of chromatin-associated RNA by high salt fractionation	• simplest approach • little invasive • nucleotide resolution	• contamination with matured RNA (e.g. Xist)
Pol II-associated RNA (NET-seq)	Isolation of RNA following Pol II IP from nuclei	• nucleotide resolution • little invasive • CTD modifications • Detection of Pol II pausing	• contamination with small nuclear RNA
Metabolically labelled RNA (e.g. TT-seq)	Labelling with 4sU, 4tU or EU followed by isolation or chemical conversion	• little invasive • streptavidin pull-down • RNA turnover analysis • usable in living organisms • single cell RNA-seq	• no nucleotide resolution
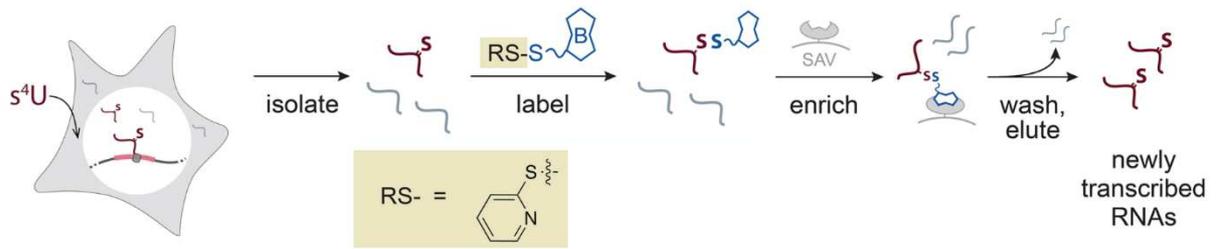
Figure 61: Scheme of the steps involved in the purification of newly synthesized RNAs from mammalian cells following 4-thiouridine labelling. From left to right: Cells are labelled with 4-thiouridine (s⁴U) for a given time. During this duration, s⁴U is incorporated into RNA during the transcription process. Total RNAs are isolated from the cells containing s⁴U-containing, newly synthesized RNAs (purple with S) and unlabelled, pre-existing RNAs (grey). Labelled RNA can be subsequently coupled to biotin (in blue) through a reactive side chain (RS, here shown for HPDP-biotin) which enables the formation of a disulphide bond between the s⁴U-containing RNA and biotin (represented by S-S). Biotinylated RNA is enriched using Streptavidin (SAV)-coated magnetic beads. SAV specifically interacts with biotin and allows the removal of unlabelled RNA through repeated washes. By breaking the disulphide bond, the s⁴U-containing, newly synthesized RNAs can be eluted from the magnetic beads and subjected to subsequent analyses. Adapted from Duffy et al., 2015.

During a short labelling period of generally 15 to 30 minutes, 4sU is added to the culture medium of mammalian cells. Addition to the medium leads to the rapid uptake of 4sU by the cells and subsequent incorporation into RNA synthesized during the duration of the labelling (Figure 61). Restricting the labelling time to a short period allows for the purification of mainly unprocessed, newly synthesized RNAs. After labelling with 4sU, cells are harvested and total RNA is isolated, which contains a small proportion of 4sU-containing, newly synthesized RNAs among unlabelled, pre-existing RNA molecules (Figure 61). Subsequently, 4sU-labelled RNAs can be coupled to biotin by a disulphide bond formed through the reaction of the reactive side chain of HPDP-biotin with the thiol-group of 4sU (Figure 61). Additional steps enable the enrichment of the labelled RNA using streptavidin-coated magnetic beads, which recognize the biotinylated RNA. Repeated wash steps of the 4sU-containing RNA bound to the magnetic beads allow the removal of unwanted contaminations represented by the excess unlabelled RNA. The newly synthesized RNA can be efficiently eluted from the magnetic beads using a reducing agent breaking the disulphide bond connecting the 4sU-containing RNA with the biotin and magnetic beads such as DTT (dithiothreitol). The purified, newly synthesized RNA can subsequently be used for downstream analyses such as RT-qPCR or sequencing analyses.

1.2.2. Required materials

- **4-thiouridine** (Glentham Life Sciences, Cat# GN6085 or abcam, Cat# ab143718) for mammalian and fly cells or **4-thiouracil** (Sigma-Aldrich, Cat# 440736) if working with yeast cells (store in aliquots at -20°C, dissolve in DMSO, stock concentration of 100 mM). *4sU and 4tU are light-sensitive* (wrap in aluminum)
- **Cell culture medium** (store at 4°C)
- **Ice-cold 1x PBS** (store at 4°C)
- **TRI Reagent** (Molecular Research Center Inc., Cat# TR 188) (store at 4°C)

- **Chloroform** (store in chemical board at room temperature)

- **Isopropanol** (store in chemical board at room temperature)

- **75% Ethanol** with RNAse-free water (store at room temperature)

- **DEPC-treated RNAse-free water** (Sigma-Aldrich, Cat# 95284) (store at room temperature)

- **RNAse inhibitor** (Promega, Cat# N2515) (store at -20°C)

- **TURBO DNA-free kit** (Thermo Scientific, Cat# AM1907) (store at -20°C)

- **5 M NaCl** (store at room temperature)

- **Snap Cap tubes** (Covaris, Cat# 520045) (store at room temperature)

- **DMSO molecular grade** (Sigma-Aldrich, Cat# D8418) (store at room temperature)

- **EZ-link HPDP-biotin** (ThermoFisher Cat# 21341) (store at -20°C in aliquots, dissolve in DMSO, stock concentration of 1 mg/ml)

- **10x Biotinylation buffer** (make yourself: 100 mM Tris-HCl, pH 7.5 and 10 mM EDTA, store at room temperature)

- **µMACS columns and magnetic beads** (Miltenyi Biotec, Cat# 130-074-101) (store columns at room temperature, store magnetic beads at 4°C)

- **Washing buffer** (make yourself: 100 mM Tris-HCl pH 7.5, 10 mM EDTA, 1M NaCl, 0.1% Tween20, in RNase-free water, store at room temperature)

- **1 M DTT** (store in aliquots in -20°C)

- **RNeasy MinElute Cleanup kit** (Qiagen, Cat# 74204) (store at room temperature except columns which need to be stored at 4°C)

- **RiboPure RNA Purification kit, yeast** (ThermoFisher, Cat# AM1926)

1.2.3. 4sU-labelling in HeLa cells

Use **3x 10 cm plates of 70-80% confluency** HeLa cells per sample (= should give ~200-300 µg of total RNA)

- **Remove** culture **medium**

- **Add 4-thiouridine (4sU) containing medium** (final concentration of 500 µM for labelling (= 50 µl of stock/10 ml pre-warmed medium), *4sU is __light-sensitive__* (wrap in aluminum)); use **5 ml** of 4sU-medium **per 10 cm plate**

- **Incubate** for **10-20 minutes** in usual culture incubator

- **Remove medium**, wash with **ice cold 1x PBS**

- Add **1 ml Trizol reagent/10 cm plate**

- **Collect** the **lysates** of all three plates in one 15 ml tube and further homogenize by using a syringe with the smallest aperture possible

- **Aliquot** in 1.5 ml Eppendorf tubes for RNA extraction

- For spike-in: Ratio we used for qPCR = HeLa : S2 3:1 plates. S2 cells are labelled in the same way as HeLa cells (culturing see section 1.2.10.)

- Homogenized sample can be **stored at -80°C** for at least one month

1.2.4. 4sU-labelling in mouse ES E14 cells

Use 4x 15 cm plates of trypsinized ESCs (since ESCs are growing in colonies if cultured o/n, we thought that splitting and reseeding them 5-6 hours prior to labelling would allow a better penetrance of 4sU and eliminate potential biases due to the colony formation; ~10 x 10^6 cells per plate; should give ~200-300 µg of total RNA)

- **Remove** culture **medium**

- **Add 4sU-containing medium** (same as for HeLa cells); use **10 ml** of 4sU-medium **per 15 cm plate**

- **Incubate for 10-20 minutes** in incubator

- **Remove medium**, wash with **ice cold 1x PBS**

- **Add 2 ml Trizol reagent/15 cm plate**

- **Collected** all plates in one 15 ml tube and additionally homogenize cells by using a syringe with the smallest aperture possible (for RNA extraction **aliquot in 1.5 ml Eppendorf tubes**)

- Homogenized sample can be **stored at -80°C** for at least one month

1.2.5. Alternative ways for spike-in

The protocol using Trizol to lyse the cells directly in the plate works nicely if the cells (mutants and wildtype (WT)) have the same cell division time. If the growth is affected in the mutants it can result into different numbers of cells per plate especially if cultured o/n. Therefore, we tried following alternative ways for spike-in normalization:

1) After labelling, I trypsinized the cells, centrifuged, washed in 1x ice-cold PBS and counted them after harvesting to be able to mix cell numbers (not Trizol lysates). This requires a very precise cell counting to ensure that indeed the same number of spike-in cells is added

to the same number of test cells. It can easily result in a bias, if counting is not accurately performed.

2) Alternatively, it is possible to extract the total RNA from WT and mutant cell lines and to add in a given ratio total RNA isolated from spike-in cells before purification of the 4sU-labelled RNA (scheme in Figure 62). This is the spike-in method, which we used for all our sequencing data. For our first sequencing experiments, we used a ratio of roughly 2:1 for mouse ESC total RNA versus *D. melanogaster* S2 total RNA. After our first sequencing results, we switched to using total RNA isolated from 4tU-labelled *S. pombe* cells as spike-in with a ratio of 10:1 mouse versus yeast total RNA since it turned out that the S2 cell stocks are contaminated with a kind of *Drosophila*-specific RNA virus. Importantly, for this spike-in alternative we assume that lysis of WT and mutant cells is not majorly different, and that the concentration of total RNA is not greatly affected even upon a potential global decrease in Pol II transcription as total RNA consists to more than 90% of rRNA.
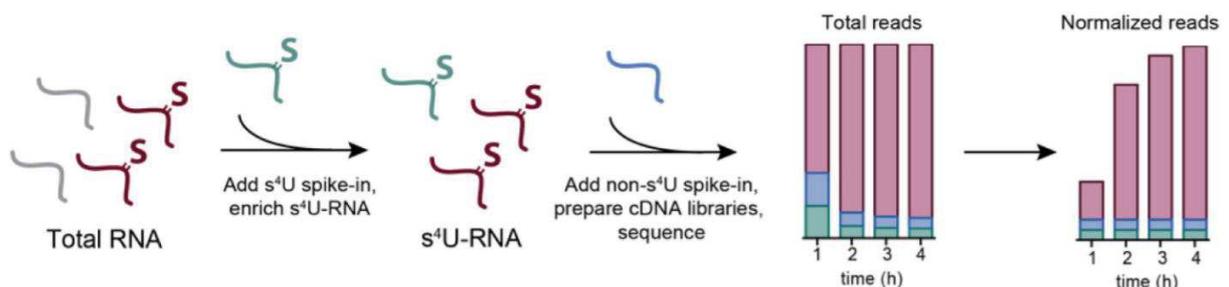
Figure 62: Spike-in possibility using RNA-to-RNA ratios. From left to right: Extracted total RNA containing s^{4}U-labelled RNA (in purple with S) can be spiked in with exogenous s^{4}U-containing RNA from *Drosophila* or *S. pombe*. Additionally, RNA standards (blue) can be added after purification of the newly synthesized, s^{4}U-containing RNAs. After high-throughput sequencing, the total amount of reads can be normalized by the number of reads corresponding to the spike-in. From Duffy et al., 2019.

1.2.6. RNA extraction

Following Trizol extraction protocol

- Incubate for **5 minutes at room temperature** to permit complete dissociation of the nucleoprotein complex

- Add **0.2 mL of chloroform** per 1 mL of Trizol Reagent

- **Shake** tube **vigorously** by hand for 15 seconds

- **Incubate for 2-3 minutes at room temperature**

- **Centrifuge** the sample at 12,000 x g for 15 minutes at 4°C

NOTE: the mixture separates into a lower red phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. RNA remains exclusively in the aqueous phase. The upper aqueous phase is ~50% of the total volume

- **Transfer** the **aqueous phase** of the sample into a new tube by angling the tube at 45° and pipetting the solution out. Avoid drawing any of the interphase or organic layer into the pipette when removing the aqueous phase.

- **Add 0.5 mL of 100% isopropanol** to the aqueous phase, per 1 mL of Trizol Reagent used for homogenization

- **Incubate at room temperature for 10 minutes**

- **Centrifuge** at 12,000 x g for 10 min. at 4°C

- **Remove** the **supernatant** from tube, leaving the RNA pellet

- **Wash** the pellet with **1 mL of 75% ethanol** per 1 mL of Trizol Reagent

- **Centrifuge** at 10,000 x g for 5 minutes at 4°C. **Discard supernatant**, **repeat washing** for better $A_{260/230}$ values

- **Air dry** the RNA pellet for 5 minutes

  NOTE: Do not allow the RNA to dry completely, because the pellet can lose solubility. Partially dissolved RNA samples have an $A_{260/280}$ ratio <1.6

- **Resuspend** the RNA pellet in **RNase-free water**, **add RNase inhibitor**

- **Incubate in heat block** set at 58°C for 15 minutes

- Proceed to downstream application, or store at -80°C

1.2.7. DNase treatment using TURBO DNA-free Kit

Following protocol for rigorous treatment

- Add **0.1 volume of 10X TURBO DNase Buffer** and **2.5-3 μl TURBO DNase** to the RNA, mix gently

- **Incubate at 37°C for 20-30 min**

- Add **0.2 volumes** of **resuspended DNase Inactivation Reagent** (always use at least 2 μl)

- Incubate **5 min. at room temperature**, mixing occasionally, **flick the tube 2-3 times** during the incubation period to redisperse the DNAse Inactivation Reagent

- **Centrifuge** at 10,000 x g for 2 min. and transfer RNA to a **fresh tube**

NOTE: this centrifugation step pellets the DNase Inactivation Reagent. After centrifuging, carefully transfer the supernatant, which contains the RNA, into a fresh tube. Avoid introducing the DNase Inactivation Reagent into solutions that may be used for downstream enzymatic reactions, because it can sequester divalent cations and change the buffer conditions

1.2.8. NanoDrop measurement and Agarose gel

- Use NanoDrop to measure RNA concentration and quality

- Load samples on 1% agarose gel to test RNA quality (ensure to use fresh TAE buffer and freshly made agarose gel, also mix RNA with loading dye only shortly before loading on gel to avoid degradation. Load the samples quickly on gel and start migration (avoid any in gel degradation of RNA).

1.2.9. Fragmentation

- Requires **~250 μg total RNA**

    NOTE: If **final volume exceeds 130 μl**, an additional precipitation step is required

    - Add **1/10 of volume of 5 M NaCl** and mix

    - Add an **equal volume of isopropanol** (under the hood), mix and **centrifuge** at 16,000 x g for at least **30 minutes**, at 4°C

    - Remove supernatant and **add 1 ml of 75% ethanol** to RNA pellet

    - **Centrifuge** at 16,000 x g for 5 minutes, at 4°C

    - **Remove supernatant**, quickspin and remove the remaining supernatant by reversing the tube (do not let the RNA dry)

    - **Resuspend RNA in 130 μl RNase-free water**

- Transfer your **130 μl RNA samples** in **Snap Cap tubes** (add some RNAse inhibitor)

- **Covaris E220 settings**: 1 % duty factor, 100 W, 200 cycles per burst, 80sec

    NOTE: Don't forget to put '**concentrator**' for sonication on Covaris E220

- Use a decapper or simply our thumb to open them to get the fragmented RNA solution

- Fragmented RNA is supposed to be in a **range between 10 kb and above 200 bp (average of >1.5 kb)** and can be stored at -80°C

1.2.10. Newly synthesized RNA extraction

- Add fragmented total RNA in **2 ml Eppendorf tubes** and **heat** it for 10 minutes at 60°C and immediately **chill** it **on ice** for 2 minutes

- Add to RNA:

  o **200 μl HPDP-biotin** (*light-sensitive*, wrap in aluminium)

  o **100 μl of 10x biotinylation buffer** (100mM Tris-HCl, pH 7.5 and 10 mM EDTA, make by yourself: for each 10 ml of buffer pipette 1 ml of 1M Tris-HCl, pH 7.5, 200 μl of 0.5M EDTA and add DEPC-treated water until you complete the volume)

  o **200 μl DMSO** (20% of reaction volume) (without DMSO, flocks of HPDP can appear and might affect purification efficiency)

  o Add 370 μl of DEPC-treated RNase-free water to a **total volume of 1 ml**

  o **Incubate at room temperature** (= 24°C) and protected from light for **3 hours** with gentle agitation (thermomixer, 550 rpm, covered by aluminium)

  o After incubation, add approximately an **equal volume of chloroform** to the tubes and **mix vigorously**

  o **Centrifuge** at 16,000 x g for 5 minutes, at 4°C

  NOTE: This step allows to remove biotin that did not biotinylate the RNA

  o Carefully transfer only upper phase into new 2 ml tubes (don't be greedy, if you transfer *unbound biotin* which is located at the interphase, it *reduces yield of newly synthesized RNA*! Other labs use phase lock tubes for this step)

  o Add **1/10 of volume of 5 M NaCl** and mix

  o Add an **equal volume of isopropanol** (under the hood), mix and **centrifuge** at 16,000 x g for at least **30 minutes**, at 4°C

  o Remove supernatant and **add 1 ml of 75% ethanol** to RNA pellet

  o **Centrifuge** at 16,000 x g for 10 minutes, at 4°C

  o **Remove supernatant**, quickspin and remove the remaining supernatant by reversing the tube (do not let the RNA dry)

  o **Resuspend RNA in 100 μl RNase-free water**

  o **Heat biotinylated RNA** for 10 minutes at 65°C (thermomixer) and then **chill** the samples **on ice** for 5 minutes

- add **100 μl of magnetic streptavidin beads** (μMACS Streptavidin beads and kit, Miltenyi, stored at 4°C) to biotinylated RNA (final volume of 200 μl)

- **incubate** with slight shaking for **90 minutes**, at room temperature (thermomixer, 24°C, 550 rpm, in 2 ml tubes)

- Shortly before the end of the incubation time, **place μMACS columns in the magnetic stand** (one column/sample)

- Add **900 μl of room temperature washing buffer** to columns (for pre-run and equilibration) (make yourself: 100 mM Tris-HCl (pH 7.5), 10 mM EDTA, 1M NaCl, 0.1% Tween20, in RNase-free water, for each 50 ml of buffer pipette 5 ml of 1M Tris-HCl, pH 7.5, 1000 μl of 0.5M EDTA, 10 ml of 5M NaCl, 50 μl of Tween20 and add DEPC-treated water until you complete the volume)

- **Apply beads/RNA mix (200 μl) to the columns**

- **Collect flow-through** in 1.5 ml tubes and **apply** it to the column **again**

- **Wash columns 5 times with 600, 700, 800, 900 and 1000 μl of washing buffer**

- **Elute RNA** with **200 μl of 0.1M DTT (twice with 100 μl with 1 min. pause)** in new 1.5 Eppendorf tubes

- Follow RNeasy MinElute Cleanup Kit protocol to purify the eluted RNA.

<u>NOTE</u>: Following this procedure, the yield of newly synthesized RNA is around 0.1%.

1.2.11. *Drosophila melanogaster* S2 cell culturing

- Culture *Drosophila melanogaster* S2 cells in Schneider medium with 10% inactivated FCS + 0,5% Penicillin and Streptomycin

- Culture them at ~ 27°C (S2 cells don't need special CO_2 levels and can also be grown at room temperature), S2 cells attach to plate until they get confluent, then they also start to grow in suspension

- For labelling, remove medium containing floating cells and add medium containing 4sU (same procedure as for HeLa and mouse ESCs) and keep under aluminium cover at room temperature for 10-20 minutes before harvesting them using a cell scraper.

- Centrifuge at 2,000 xg for 5 min. at 4°C and wash with ice-cold 1x PBS.

- Add to your cells of interesting depending on the spike-in way you chose (either cell-to-cell or total RNA-to-total RNA). Trizol protocol can be used for S2 cells with no problem and no specifications.

1.2.12. *Schizosaccharomyces pombe* culturing

1.2.12.1. Culture medium

S. pombe cells are grown in YES medium with 3% glucose (recipe see Table 9). Medium should be autoclaved before use. Important, add glucose only after autoclavation. Powders can be mixed and stored but should only be dissolved shortly before autoclavation.

1.2.12.2. 4-thiouracil labelling

Precultures (start in the evening and grow o/n at 31°C)

o 4x 10 mL in 50 mL falcon tubes; for each:
- 8.5 mL autoclaved YES medium
- 1.5 mL 20% glucose
- 2-3 colonies of wildtype S. pombe picked from a plate

Main cultures (start in the morning)

o One 2 L falcon
- 850 ml autoclaved YES medium
- 150 mL 20% glucose
- S. pombe preculture at a concentration of [OD_{600} 0.1]

  Important: to measure OD_{600} dilute in YES medium 1/10 before measurement at

Labelling

Labelling is performed when main culture reaches OD_{600} of 0.8. *S. pombe* doubles every 2 ½ hours. OD_{600} 0.8 will be reached after roughly 7-8 hours.

o Freshly dissolve 4-thiouracil (4tU) in DMSO: 64.1 mg 4tU in 250 L of DMSO for 100 mL culture
o Add to cultures, mix and label for 6 minutes
o After incubation time, aliquot cultures into 50 mL falcons and centrifuge at 2000 xg for 2 minutes at 4C
o Wash with ice-cold 1xPBS; keep some cells aside for counting and centrifuge again
o Remove 1xPBS and flash freeze in liquid nitrogen; store at -80 °C
o For RNA extraction use RiboPure RNA Purification kit and follow manufacturer's instruction

Table 9: Recipe for YES medium required for *Schizosaccharomyces pombe* culturing.

YES medium [for 500 mL]	
Yeast extract	2.5 g
Adenine	125 mg
Histidine	125 mg
Uracil	125 mg
Leucine	125 mg
Lysine	125 mg
MilliQ water	425 ml
20% glucose	75 ml

1.3. Step-by-Step Protocol 3: Protocol for ATAC-seq using house-made Tn5 transposase

1.3.1. Principle

ATAC-seq (assay for transposase accessible chromatin followed by high throughput sequencing) represents a method to assess both chromatin accessibility and nucleosome positioning using the modified, hyperactive bacterial Tn5 transposases (Figure 63) (Sun et al., 2019).

In their native environment, DNA transposases, such as the Tn5 transposase, enable the transfer of DNA sequences known as DNA transposons between genomic regions through a 'copy and paste' mechanism (Sun et al., 2019). Instead of DNA transposons, the ATAC-seq method uses the Tn5 transposase to integrate sequencing adaptors into open chromatin regions, such as at linker DNA sequences separating nucleosomes or at nucleosome depleted regions (NDRs) (Figure 63) (Sun et al., 2019).

Compared to other methods enabling the analysis of chromatin accessibility or nucleosome positioning, such as MNase-seq or DNase-seq, ATAC-seq offers several practical advantages such as for instance reduced experimental time (2-3 hours for ATAC-seq compared to 2-3 days for MNase- or DNase-seq) and reduced sample size (commonly 50 thousands cells for ATAC-seq compared to 50 million cells for DNase-seq) (Sun et al., 2019). Importantly, ATAC-seq allows to assess both chromatin accessibility (comparable to DNase-seq) and nucleosome positioning (comparable to MNase-seq).

In general, two distinct sequencing adaptors are used (represented in Figure 63 by red and green sticks), one representing the forward and the other the reverse primer required for the subsequent barcoding and library preparation step. However, the sequencing adaptors bound by the Tn5 transposase are random which leads to a 50% probability that the sequencing adapters at the ends of a given fragment are identical (in contrast to what is shown in Figure 63). This represents a disadvantage of ATAC-seq as fragments containing two identical adaptors on their ends are consequently unusable for the barcoding and library preparation step and are not sequenced.
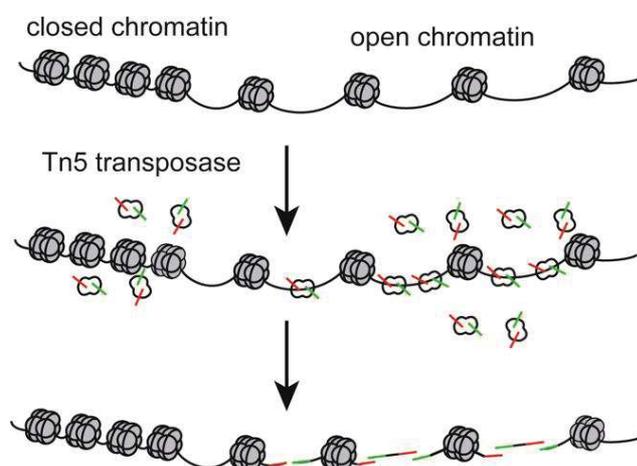
Figure 63: Principle of the ATAC-seq method. Dimers of the Tn5 transposase insert sequencing adaptors (shown as red and green sticks) into accessible chromatin more frequently than at inaccessible chromatin. After DNA purification, the inserted adaptors allow the preparation of barcoded sequencing libraries. Histone proteins are shown as grey spheres and DNA is displayed as black string. From Sun et al., 2019.

1.3.2. Required materials

- House-made Tn5 2M from the protein platform at roughly 7.2 μM (400 ng/μl)

- Mosaic primers, can be order from Sigma

    1. Mosaic end Adapter A: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

    2. Mosaic end Adapter B: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

    3. Mosaic end reverse: [PHO]CTGTCTCTTATACACATCT

- 1x PBS colded to 4⁰C

- Lysis buffer (recipe see 1.3.6 Buffers needed)

- TAPS-DMF buffer (recipe see 1.3.6 Buffers needed)

- NucleoSpin Gel and PCR Clean-up kit

1.3.3. Annealing Mosaic primers

- 10 μl of Mosaic end reverse
  + 10 μl of Mosaic end Adapter A in PCR tube

- 10 μl of Mosaic end reverse
  + 10 μl of Mosaic end Adapter B in PCR tube

- **PCR program**

    - 95⁰C for 2 min

    - Cooling down to 25⁰C with reduced speed **(0.1⁰C/sec ramp rate)**

    - Keep at 4⁰C

1.3.4. Loading of Tn5 transposase

- 50 μl of Tn5 2M in storage buffer from platform
  + 5.2 μl of annealed Mosaic end Adapter A mix
  + 5.2 μl of annealed Mosaic end Adapter B mix

- Incubate at RT mixing platform for 1 hour

- Store at 20⁰C until needed

1.3.5. Experimental protocol for house-made Tn5

- Harvest and count cells (consider roughly 100.000 cells per sample)

- Centrifuge at 2000 xg for 5 min at 4°C

- Wash with 50 µl of cold 1x PBS

- Centrifuge at 2000 xg for 5 min at 4°C

- Resuspend pellet in 50 µl of Lysis buffer

- Centrifuge immediately at 2000 xg for 10 min at 4°C

- Immediately after removing supernatant resuspend pellet in 50 µl of TAPS-DMF-buffer + Tn5 2M (25 nM gives similar signal as Nextera Tn5) (keep pellet on ice while preparing Tn5 mix)

- Incubate reaction at 37°C for 1 hour

- Immediately purify using NucleoSpin Gel and PCR Clean-up kit

- Elute in 15 µl

- Store DNA at -20°C

- *Optional*: Experiments can be verified by qPCR prior to library preparation using the Mosaic primers and 1 µl of tagmented DNA

1.3.6. Buffers needed

Lysis buffer		
Final [c]	Volume for 10 mL	Stock [c]
10 mM Tris HCl ph 7.4	100 µl	1 M
10 mM NaCl	20 µl	5 M
3 mM MgCl2	30 µl	1 M
0.1% NP-40	100 µl	10%
Water	9.75 ml	

TAPS-DMF buffer

Final [c]	Volume for 10 mL	Stock [c]
10 mM TAPS-NaOH pH 8.5	1 ml	100 mM*
5 mM MgCl2	50 µl	1 M
10% DMF	1 ml	100%
Water	7.85 ml	

*100 mM TAPS-NaOH made using 244 mg of TAPS powder + 10 ml of water, pH was adjusted using NaOH

Bibliography

Bibliography

Abranches, E., Guedes, A.M.V., Moravec, M., Maamar, H., Svoboda, P., Raj, A., and Henrique, D. (2014). Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. Development (Cambridge, England) *141*, 2770–2779.

Acharya, D., Hainer, S.J., Yoon, Y., Wang, F., Bach, I., Rivera-Pérez, J.A., and Fazzio, T.G. (2017). KAT-Independent Gene Regulation by Tip60 Promotes ESC Self-Renewal but Not Pluripotency. Cell reports *19*, 671–679.

Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2010). Essential cell biology (New York, NY: Garland Science).

Alfert, A., Moreno, N., and Kerl, K. (2019). The BAF complex in development and disease. Epigenetics & chromatin *12*, 19.

Allard, S., Utley, R.T., Savard, J., Clarke, A., Grant, P., Brandl, C.J., Pillus, L., Workman, J.L., and Côté, J. (1999). NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. The EMBO journal *18*, 5108–5119.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proceedings of the National Academy of Sciences of the United States of America *51*, 786–794.

Andersson, R. (2015). Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. BioEssays news and reviews in molecular, cellular and developmental biology *37*, 314–323.

Andersson, R., Chen, Y., Core, L., Lis, J.T., Sandelin, A., and Jensen, T.H. (2015a). Human Gene Promoters Are Intrinsically Bidirectional. Molecular cell *60*, 346–347.

Andersson, R., Sandelin, A., and Danko, C.G. (2015b). A unified architecture of transcriptional regulatory elements. Trends in genetics TIG *31*, 426–433.

Andrau, J.-C., van de Pasch, L., Lijnzaad, P., Bijma, T., Koerkamp, M.G., van de Peppel, J., Werner, M., and Holstege, F.C.P. (2006). Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. Molecular cell *22*, 179–192.

Andreu-Vieyra, C.V., Chen, R., Agno, J.E., Glaser, S., Anastassiadis, K., Stewart, A.F., and Matzuk, M.M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. PLoS biology *8*.

Ang, Y.-S., Tsai, S.-Y., Lee, D.-F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., Wang, J., Rendl, M., Bernstein, E., Schaniel, C., and Lemischka, I.R. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183–197.

Aoto, T., Saitoh, N., Ichimura, T., Niwa, H., and Nakao, M. (2006). Nuclear and chromatin reorganization in the MHC-Oct3/4 locus at developmental phases of embryonic stem cell differentiation. Developmental biology *298*, 354–367.

Arede, L., Foerner, E., Wind, S., Kulkarni, R., Domingues, A. F., Kleinwaechter, S., Gupta, S., Scheer, E., Tora, L., & Pina, C. (2020). Unique roles of ATAC and SAGA - KAT2A complexes in normal and malignant hematopoiesis.

Armache, K.-J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America *100*, 6964–6968.

Atanassov, B.S., Mohan, R.D., Lan, X., Kuang, X., Lu, Y., Lin, K., McIvor, E., Li, W., Zhang, Y., Florens, L., Byrum, S.D., Mackintosh, S.G., Calhoun-Davis, T., Koutelou, E., Wang, L., Tang, D.G., Tackett, A.J., Washburn, M.P., Workman, J.L., and Dent, S.Y.R. (2016). ATXN7L3 and ENY2 Coordinate Activity of Multiple H2B Deubiquitinases Important for Cellular Proliferation and Tumor Growth. Molecular cell *62*, 558–571.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development *17*, 126–140.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., and Fisher, A.G. (2006). Chromatin signatures of pluripotent cell lines. Nature cell biology *8*, 532–538.

Bajic, V.B., Tan, S.L., Christoffels, A., Schönbach, C., Lipovich, L., Yang, L., Hofmann, O., Kruger, A., Hide, W., Kai, C., Kawai, J., Hume, D.A., Carninci, P., and Hayashizaki, Y. (2006). Mice and men: their promoter properties. PLoS genetics *2*, e54.

Baker, M. (2011). Making sense of chromatin states. Nature methods *8*, 717–722.

Balasubramanian, R., Pray-Grant, M.G., Selleck, W., Grant, P.A., and Tan, S. (2002). Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation. The Journal of biological chemistry *277*, 7989–7995.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell research *21*, 381–395.

Baptista, T., Grünberg, S., Minoungou, N., Koster, M.J.E., Timmers, H.T.M., Hahn, S., Devys, D., and Tora, L. (2017). SAGA Is a General Cofactor for RNA Polymerase II Transcription. Molecular cell *68*, 130-143.e5.

Barlev, N.A., Liu, L., Chehab, N.H., Mansfield, K., Harris, K.G., Halazonetis, T.D., and Berger, S.L. (2001). Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. Molecular cell *8*, 1243–1254.

Bartkowiak, B., and Greenleaf, A.L. (2011). Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? Transcription *2*, 115–119.

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 & development *24*, 2303–2316.

Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. Cell *116*, 699–709.

Bellec, M., Radulescu, O., and Lagha, M. (2018). Remembering the past: Mitotic bookmarking in a developing embryo. Current opinion in systems biology *11*, 41–49.

Belotserkovskaya, R., Sterner, D.E., Deng, M., Sayre, M.H., Lieberman, P.M., and Berger, S.L. (2000). Inhibition of TATA-binding protein function by SAGA subunits Spt3 and Spt8 at Gcn4-activated promoters. Molecular and Cellular Biology *20*, 634–647.

Berger, S.L., Cress, W.D., Cress, A., Triezenberg, S.J., and Guarente, L. (1990). Selective inhibition of activated but not basal transcription by the acidic activation domain of VP16: evidence for transcriptional adaptors. Cell *61*, 1199–1208.

Berger, S.L., Piña, B., Silverman, N., Marcus, G.A., Agapite, J., Regier, J.L., Triezenberg, S.J., and Guarente, L. (1992). Genetic isolation of ADA2: A potential transcriptional adaptor required for function of certain acidic activation domains. Cell *70*, 251–265.

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

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., and Lander, E.S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bertomeu, T., Coulombe-Huntington, J., Chatr-Aryamontri, A., Bourdages, K.G., Coyaud, E., Raught, B., Xia, Y., and Tyers, M. (2018). A High-Resolution Genome-Wide CRISPR/Cas9 Viability Screen Reveals Structural Features and Contextual Diversity of the Human Cell-Essential Proteome. Molecular and Cellular Biology *38*.

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes & development *2*, 754–765.

Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. Nature structural & molecular biology *14*, 1008–1016.

Bian, C., Xu, C., Ruan, J., Lee, K.K., Burke, T.L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B.O., Fleharty, B.E., Paulson, A., Allali-Hassani, A., Zhou, J.-Q., Mer, G., Grant, P.A., Workman, J.L., Zang, J., and Min, J. (2011). Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. The EMBO journal *30*, 2829–2842.

Bibel, M., Richter, J., Lacroix, E., and Barde, Y.-A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nature protocols *2*, 1034–1043.

Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. Nature *493*, 699–702.

Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes & development *23*, 2484–2489.

Birck, C., Poch, O., Romier, C., Ruff, M., Mengus, G., Lavigne, A.C., Davidson, I., and Moras, D. (1998). Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. Cell *94*, 239–249.

Bledau, A.S., Schmidt, K., Neumann, K., Hill, U., Ciotta, G., Gupta, A., Torres, D.C., Fu, J., Kranz, A., Stewart, A.F., and Anastassiadis, K. (2014). The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. Development (Cambridge, England) *141*, 1022–1035.

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., Darzacq, X., and Zweckstetter, M. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. Nature structural & molecular biology *25*, 833–840.

Boeuf, H., Hauss, C., Graeve, F.D., Baran, N., and Kedinger, C. (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. The Journal of cell biology *138*, 1207–1217.

Bondarenko, V.A., Steele, L.M., Ujvári, A., Gaykalova, D.A., Kulaeva, O.I., Polikanov, Y.S., Luse, D.S., and Studitsky, V.M. (2006). Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. Molecular cell *24*, 469–479.

Bonnet, J., Wang, C.-Y., Baptista, T., Vincent, S.D., Hsiao, W.-C., Stierle, M., Kao, C.-F., Tora, L., and Devys, D. (2014). The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes & development *28*, 1999–2012.

Bonnet, J., Wang, Y.-H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W.M.P., Timmers, H.T.M., Tora, L., Devys, D., and Kieffer, B. (2010). The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. EMBO reports *11*, 612–618.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nature cell biology *16*, 516–528.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255–256.

Brand, M., Leurent, C., Mallouh, V., Tora, L., and Schultz, P. (1999). Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. Science (New York, N.Y.) *286*, 2151–2153.

Brouwer, I., and Lenstra, T.L. (2019). Visualizing transcription: key to understanding gene expression dynamics. Current opinion in chemical biology *51*, 122–129.

Brown, C.E., Howe, L., Sousa, K., Alley, S.C., Carrozza, M.J., Tan, S., and Workman, J.L. (2001). Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. Science (New York, N.Y.) *292*, 2333–2337.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843–851.

Bu, P., Evrard, Y.A., Lozano, G., and Dent, S.Y.R. (2007). Loss of Gcn5 acetyltransferase activity leads to neural tube closure defects and exencephaly in mouse embryos. Molecular and Cellular Biology *27*, 3405–3416.

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., and Magnuson, T. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Molecular cell *6*, 1287–1295.

Bulut-Karslioglu, A., Macrae, T.A., Oses-Prieto, J.A., Covarrubias, S., Percharde, M., Ku, G., Diaz, A., McManus, M.T., Burlingame, A.L., and Ramalho-Santos, M. (2018). The Transcriptionally Permissive Chromatin State of Embryonic Stem Cells Is Acutely Tuned to Translational Output. Cell stem cell *22*, 369-383.e8.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Molecular cell *36*, 541–546.

Burke, T.W., and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes & development *11*, 3020–3031.

Bushnell, D.A., and Kornberg, R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. Proceedings of the National Academy of Sciences of the United States of America *100*, 6969–6973.

Cai, Y., Jin, J., Swanson, S.K., Cole, M.D., Choi, S.H., Florens, L., Washburn, M.P., Conaway, J.W., and Conaway, R.C. (2010). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. The Journal of biological chemistry *285*, 4268–4272.

Caliskan, G., Baris, I.C., Ayaydin, F., Dobson, M.J., Senarisoy, M., Boros, I.M., Topcu, Z., and Zencir, S. (2017). Che1/AATF interacts with subunits of the histone acetyltransferase core module of SAGA complexes. PloS one *12*, e0189193.

Callebaut, I., and Mornon, J.-P. (2012). The PWAPA cassette: Intimate association of a PHD-like finger and a winged-helix domain in proteins included in histone-modifying complexes. Biochimie *94*, 2006–2012.

Cao, S., Bendall, H., Hicks, G.G., Nashabi, A., Sakano, H., Shinkai, Y., Gariglio, M., Oltz, E.M., and Ruley, H.E. (2003). The high-mobility-group box protein SSRP1/T160 is essential for cell viability in day 3.5 mouse embryos. Molecular and Cellular Biology *23*, 5301–5307.

Cardozo Gizzi, A.M., Cattoni, D.I., Fiche, J.-B., Espinola, S.M., Gurgo, J., Messina, O., Houbron, C., Ogiyama, Y., Papadopoulos, G.L., Cavalli, G., Lagha, M., and Nollmann, M. (2019). Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. Molecular cell *74*, 212-222.e5.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C., Forrest, A.R.R., Alkema, W.B., Tan, S.L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S.M., Wells, C.A., Orlando, V., Wahlestedt, C., Liu, E.T., Harbers, M., Kawai, J., Bajic, V.B., Hume, D.A., and Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nature Genetics *38*, 626–635.

Carré, C., Ciurciu, A., Komonyi, O., Jacquier, C., Fagegaltier, D., Pidoux, J., Tricoire, H., Tora, L., Boros, I.M., and Antoniewski, C. (2008). The Drosophila NURF remodelling and the ATAC histone acetylase complexes functionally interact and are required for global chromosome organization. EMBO reports *9*, 187–192.

Carter, M., and Shieh, J. (2015). Molecular Cloning and Recombinant DNA Technology. In Guide to Research Techniques in Neuroscience (Elsevier), pp. 219–237.

Chalkley, G.E., and Verrijzer, C.P. (1999). DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. The EMBO journal *18*, 4835–4845.

Chamberlain, S.J., Della Yee, and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem cells (Dayton, Ohio) *26*, 1496–1505.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chapman, R.D., Palancade, B., Lang, A., Bensaude, O., and Eick, D. (2004). The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability. Nucleic acids research *32*, 35–44.

298

Chappell, J., and Dalton, S. (2013). Roles for MYC in the establishment and maintenance of pluripotency. Cold Spring Harbor perspectives in medicine *3*, a014381.

Charron, J., Malynn, B.A., Fisher, P., Stewart, V., Jeannotte, L., Goff, S.P., Robertson, E.J., and Alt, F.W. (1992). Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. Genes & development *6*, 2248–2257.

Chen, F.X., Smith, E.R., and Shilatifard, A. (2018a). Born to run: control of transcription elongation by RNA polymerase II. Nature reviews. Molecular cell biology *19*, 464–478.

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018b). Dynamic interplay between enhancer-promoter topology and gene activity. Nature genetics *50*, 1296–1303.

Chen, H.-Z., Tsai, S.-Y., and Leone, G. (2009). Emerging roles of E2Fs in cancer: an exit from cell cycle control. Nature reviews. Cancer *9*, 785–797.

Chen, J., Zhang, Z., Li, L., Chen, B.-C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., Tjian, R., and Liu, Z. (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell *156*, 1274–1285.

Chen, L., Wei, T., Si, X., Wang, Q., Li, Y., Leng, Y., Deng, A., Chen, J., Wang, G., Zhu, S., and Kang, J. (2013). Lysine acetyltransferase GCN5 potentiates the growth of non-small cell lung cancer via promotion of E2F1, cyclin D1, and cyclin E1 expression. The Journal of biological chemistry *288*, 14510–14521.

Cheung, A.C.M., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249–253.

Chittock, E.C., Latwiel, S., Miller, T.C.R., and Müller, C.W. (2017). Molecular architecture of polycomb repressive complexes. Biochemical Society transactions *45*, 193–205.

Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes & development *15*, 3319–3329.

Ciurciu, A., Komonyi, O., and Boros, I.M. (2008). Loss of ATAC-specific acetylation of histone H4 at Lys12 reduces binding of JIL-1 to chromatin and phosphorylation of histone H3 at Ser10. Journal of cell science *121*, 3366–3372.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nature reviews. Molecular cell biology *18*, 407–422.

Coleman, R.A., and Pugh, B.F. (1995). Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA. The Journal of biological chemistry *270*, 13850–13859.

Cooper, S., and Brockdorff, N. (2013). Genome-wide shRNA screening to identify factors mediating Gata6 repression in mouse embryonic stem cells. Development (Cambridge, England) *140*, 4110–4115.

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. Nature genetics *46*, 1311–1320.

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

Corona, D.F., Eberharter, A., Budde, A., Deuring, R., Ferrari, S., Varga-Weisz, P., Wilm, M., Tamkun, J., and Becker, P.B. (2000). Two histone fold proteins, CHRAC-14 and CHRAC-16, are developmentally regulated subunits of chromatin accessibility complex (CHRAC). The EMBO journal *19*, 3049–3059.

Coronado, D., Godet, M., Bourillot, P.-Y., Tapponnier, Y., Bernat, A., Petit, M., Afanassieff, M., Markossian, S., Malashicheva, A., Iacone, R., Anastassiadis, K., and Savatier, P. (2013). A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. Stem cell research *10*, 118–131.

Craighead, J.L., Chang, W.-h., and Asturias, F.J. (2002). Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. Structure (London, England 1993) *10*, 1117–1125.

Cramer, P. (2004). Structure and function of RNA polymerase II. Advances in protein chemistry *67*, 1–42.

Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45–54.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science (New York, N.Y.) *288*, 640–649.

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

Dai, J., Hyland, E.M., Yuan, D.S., Huang, H., Bader, J.S., and Boeke, J.D. (2008). Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. Cell *134*, 1066–1078.

Daily-Diamond, C.A., Gregg, C.E., and O'Reilly, O.M. (2017). The roles of impact and inertia in the failure of a shoelace knot. Proceedings. Mathematical, physical, and engineering sciences *473*, 20160770.

Daniel, J.A., Torok, M.S., Sun, Z.-W., Schieltz, D., Allis, C.D., Yates, J.R., and Grant, P.A. (2004). Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. The Journal of biological chemistry *279*, 1867–1871.

Danino, Y.M., Even, D., Ideses, D., and Juven-Gershon, T. (2015). The core promoter: At the heart of gene expression. Biochimica et biophysica acta *1849*, 1116–1131.

Das, P.P., Shao, Z., Beyaz, S., Apostolou, E., Pinello, L., Los Angeles, A. de, O'Brien, K., Atsma, J.M., Fujiwara, Y., Nguyen, M., Ljuboja, D., Guo, G., Woo, A., Yuan, G.-C., Onder, T., Daley, G., Hochedlinger, K., Kim, J., and Orkin, S.H. (2014). Distinct and combinatorial functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in mouse embryonic stem cell identity. Molecular cell *53*, 32–48.

Davis, A.C., Wims, M., Spotts, G.D., Hann, S.R., and Bradley, A. (1993). A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. Genes & development *7*, 671–682.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes & development *25*, 1010–1022.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233–1244.

Deng, W., and Roberts, S.G.E. (2007). TFIIB and the regulation of transcription by RNA polymerase II. Chromosoma *116*, 417–429.

Di Stefano, V., Soddu, S., Sacchi, A., and D'Orazi, G. (2005). HIPK2 contributes to PCAF-mediated p53 acetylation and selective transactivation of p21Waf1 after nonapoptotic DNA damage. Oncogene *24*, 5431–5442.

Dias, J., van Nguyen, N., Georgiev, P., Gaub, A., Brettschneider, J., Cusack, S., Kadlec, J., and Akhtar, A. (2014). Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex. Genes & development *28*, 929–942.

Díaz-Santín, L.M., Lukoyanova, N., Aciyan, E., and Cheung, A.C. (2017). Cryo-EM structure of the SAGA and NuA4 coactivator subunit Tra1 at 3.7 angstrom resolution. eLife *6*.

Dodge, J.E., Kang, Y.-K., Beppu, H., Lei, H., and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. Molecular and Cellular Biology *24*, 2478–2486.

Dodonova, S.O., Zhu, F., Dienemann, C., Taipale, J., and Cramer, P. (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. Nature *143*, 1833.

Donczew, R., Warfield, L., Pacheco, D., Erijman, A., and Hahn, S. (2020). Two roles for the yeast transcription coactivator SAGA and a set of genes redundantly regulated by TFIID and SAGA. eLife *9*.

Donovan, B.T., Huynh, A., Ball, D.A., Patel, H.P., Poirier, M.G., Larson, D.R., Ferguson, M.L., and Lenstra, T.L. (2019). Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. The EMBO journal *38*.

Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N.V., Scruggs, B.S., Nady, N., Still, C.D., Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Molecular cell *66*, 568-576.e4.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science (New York, N.Y.) *346*, 1258096.

Doyon, Y., Selleck, W., Lane, W.S., Tan, S., and Côté, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Molecular and Cellular Biology *24*, 1884–1896.

Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Molecular cell *77*, 985-998.e8.

Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. Molecular cell *59*, 858–866.

Duffy, E.E., Schofield, J.A., and Simon, M.D. (2019). Gaining insight into transcriptome-wide RNA population dynamics through the chemistry of 4-thiouridine. Wiley interdisciplinary reviews. RNA *10*, e1513.

Dufourt, J., Trullo, A., Hunter, J., Fernandez, C., Lazaro, J., Dejean, M., Morales, L., Nait-Amer, S., Schulz, K.N., Harrison, M.M., Favard, C., Radulescu, O., and Lagha, M. (2018). Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. Nature communications *9*, 5194.

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. Molecular cell *57*, 674–684.

Eberharter, A., Ferrari, S., Längst, G., Straub, T., Imhof, A., Varga-Weisz, P., Wilm, M., and Becker, P.B. (2001). Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. The EMBO journal *20*, 3781–3788.

Eberharter, A., Sterner, D.E., Schieltz, D., Hassan, A., Yates, J.R., Berger, S.L., and Workman, J.L. (1999). The ADA complex is a distinct histone acetyltransferase complex in Saccharomyces cerevisiae. Molecular and Cellular Biology *19*, 6621–6631.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D.G., Buetow, K.H., Gingeras, T.R., Misteli, T., and Meshorer, E. (2008). Global transcription in pluripotent embryonic stem cells. Cell stem cell *2*, 437–447.

Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. Trends in genetics TIG *24*, 280–288.

Ehara, H., Kujirai, T., Fujino, Y., Shirouzu, M., Kurumizaka, H., and Sekine, S.-I. (2019). Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. Science (New York, N.Y.) *363*, 744–747.

Eisenmann, D.M., Arndt, K.M., Ricupero, S.L., Rooney, J.W., and Winston, F. (1992). SPT3 interacts with TFIID to allow normal transcription in Saccharomyces cerevisiae. Genes & development *6*, 1319–1331.

Eisenmann, D.M., Dollard, C., and Winston, F. (1989). SPT15, the gene encoding the yeast TATA binding factor TFIID, is required for normal transcription initiation in vivo. Cell *58*, 1183–1191.

El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.-R., Park, S., Wang, X., Krebs, J., Tripathi, S., Sakabe, N., Sobreira, D.R., Huang, S.-C., Rao, S.S.P., Pruett, N., Chauss, D., Sadler, E., Lopez, A., Nóbrega, M.A., Aiden, E.L., Asturias, F.J., and Casellas, R. (2019). A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. Cell *178*, 1145-1158.e20.

El-Saafin, F., Wang, F., Ye, T., Stierle, M., Durik, M., Fischer, V., Devys, D., Vincent, S.D., and Tora, L. Histone H2Bub1 deubiquitylation is essential for mouse development, but does not regulate global RNA polymerase II transcription. Manuscript in revision.

Elías-Villalobos, A., Fort, P., and Helmlinger, D. (2019a). New insights into the evolutionary conservation of the sole PIKK pseudokinase Tra1/TRRAP. Biochemical Society transactions *47*, 1597–1608.

Elías-Villalobos, A., Toullec, D., Faux, C., Séveno, M., and Helmlinger, D. (2019b). Chaperone-mediated ordered assembly of the SAGA and NuA4 transcription co-activator complexes in yeast. Nature communications *10*, 5237.

Elling, U., Klasen, C., Eisenberger, T., Anlag, K., and Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. Proceedings of the National Academy of Sciences of the United States of America *103*, 16319–16324.

Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y.-i., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M., and Koseki, H. (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development (Cambridge, England) *135*, 1513–1524.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., and Bernstein, B.E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Evangelista, F.M., Maglott-Roth, A., Stierle, M., Brino, L., Soutoglou, E., and Tora, L. (2018). Transcription and mRNA export machineries SAGA and TREX-2 maintain monoubiquitinated H2B balance required for DNA repair. The Journal of cell biology *217*, 3382–3397.

Evans, M. (2011). Discovering pluripotency: 30 years of mouse embryonic stem cells. Nature reviews. Molecular cell biology *12*, 680–686.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Faast, R., White, J., Cartwright, P., Crocker, L., Sarcevic, B., and Dalton, S. (2004). Cdk6-cyclin D3 activity in murine ES cells is resistant to inhibition by p16(INK4a). Oncogene *23*, 491–502.

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. Molecular cell *11*, 1549–1561.

Fagnocchi, L., and Zippo, A. (2017). Multiple Roles of MYC in Integrating Regulatory Networks of Pluripotent Stem Cells. Frontiers in cell and developmental biology *5*, 7.

Fang, L., Zhang, J., Zhang, H., Yang, X., Jin, X., Zhang, L., Skalnik, D.G., Jin, Y., Zhang, Y., Huang, X., Li, J., and Wong, J. (2016). H3K4 Methyltransferase Set1a Is A Key Oct4 Coactivator Essential for Generation of Oct4 Positive Inner Cell Mass. Stem cells (Dayton, Ohio) *34*, 565–580.

Fant, C.B., Levandowski, C.B., Gupta, K., Maas, Z.L., Moir, J., Rubin, J.D., Sawyer, A., Esbin, M.N., Rimel, J.K., Luyties, O., Marr, M.T., Berger, I., Dowell, R.D., and Taatjes, D.J. (2020). TFIID Enables RNA Polymerase II Promoter-Proximal Pausing. Molecular cell.

Farnung, L., Vos, S.M., and Cramer, P. (2018). Structure of transcribing RNA polymerase II-nucleosome complex. Nature communications *9*, 5432.

Farrants, A.-K.O. (2008). Chromatin remodelling and actin organisation. FEBS letters *582*, 2041–2050.

Farria, A.T., Mustachio, L.M., Akdemir, Z.H.C., and Dent, S.Y.R. (2019). GCN5 HAT inhibition reduces human Burkitt lymphoma cell survival through reduction of MYC target gene expression and impeding BCR signaling pathways. Oncotarget *10*, 5847–5858.

Fassler, J.S., and Winston, F. (1988). Isolation and analysis of a novel class of suppressor of Ty insertion mutations in Saccharomyces cerevisiae. Genetics *118*, 203–212.

Fazzio, T.G., Huff, J.T., and Panning, B. (2008). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell *134*, 162–174.

Ferraro, T., Esposito, E., Mancini, L., Ng, S., Lucas, T., Coppey, M., Dostatni, N., Walczak, A.M., Levine, M., and Lagha, M. (2016). Transcriptional Memory in the Drosophila Embryo. Current biology CB *26*, 212–218.

Festuccia, N., Dubois, A., Vandormael-Pournin, S., Gallego Tejeda, E., Mouren, A., Bessonnard, S., Mueller, F., Proux, C., Cohen-Tannoudji, M., and Navarro, P. (2016). Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. Nature cell biology *18*, 1139–1148.

Festuccia, N., Gonzalez, I., and Navarro, P. (2017a). The Epigenetic Paradox of Pluripotent ES Cells. Journal of molecular biology *429*, 1476–1503.

Festuccia, N., Gonzalez, I., Owens, N., and Navarro, P. (2017b). Mitotic bookmarking in development and stem cells. Development (Cambridge, England) *144*, 3633–3645.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell stem cell *11*, 477–490.

Fica, S.M., and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. Nature structural & molecular biology *24*, 791–799.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., Cook, S.J., Andrews, S., Branco, M.R., and Reik, W. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell stem cell *13*, 351–359.

Field, S.J., Tsai, F.Y., Kuo, F., Zubiaga, A.M., Kaelin, W.G., Livingston, D.M., Orkin, S.H., and Greenberg, M.E. (1996). E2F-1 functions in mice to promote apoptosis and suppress proliferation. Cell *85*, 549–561.

Filipczyk, A., Gkatzis, K., Fu, J., Hoppe, P.S., Lickert, H., Anastassiadis, K., and Schroeder, T. (2013). Biallelic expression of nanog protein in mouse embryonic stem cells. Cell stem cell *13*, 12–13.

Fish, R.N., and Kane, C.M. (2002). Promoting elongation with transcript cleavage stimulatory factors. Biochimica et biophysica acta *1577*, 287–307.

Fishburn, J., Mohibullah, N., and Hahn, S. (2005). Function of a eukaryotic transcription activator during the transcription cycle. Molecular cell *18*, 369–378.

Fleming, A.B., Kao, C.-F., Hillyer, C., Pikaart, M., and Osley, M.A. (2008). H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. Molecular cell *31*, 57–66.

Formosa, T. (2012). The role of FACT in making and breaking nucleosomes. Biochimica et biophysica acta *1819*, 247–255.

Fournier, M., Orpinell, M., Grauffel, C., Scheer, E., Garnier, J.-M., Ye, T., Chavant, V., Joint, M., Esashi, F., Dejaegere, A., Gönczy, P., and Tora, L. (2016). KAT2A/KAT2B-targeted acetylome reveals a role for PLK4 acetylation in preventing centrosome amplification. Nature communications *7*, 13227.

Frontini, M., Soutoglou, E., Argentini, M., Bole-Feysot, C., Jost, B., Scheer, E., and Tora, L. (2005). TAF9b (formerly TAF9L) is a bona fide TAF that has unique and overlapping roles with TAF9. Molecular and Cellular Biology *25*, 4638–4649.

Fujii-Yamamoto, H., Kim, J.M., Arai, K.-i., and Masai, H. (2005). Cell cycle and developmental regulations of replication factors in mouse embryonic stem cells. The Journal of biological chemistry *280*, 12976–12987.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science (New York, N.Y.) *361*, 1341–1345.

Fussner, E., Strauss, M., Djuric, U., Li, R., Ahmed, K., Hart, M., Ellis, J., and Bazett-Jones, D.P. (2012). Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO reports *13*, 992–996.

Gamper, A.M., Kim, J., and Roeder, R.G. (2009). The STAGA subunit ADA2b is an important regulator of human GCN5 catalysis. Molecular and Cellular Biology *29*, 266–280.

Gangloff, Y.G., Sanders, S.L., Romier, C., Kirschner, D., Weil, P.A., Tora, L., and Davidson, I. (2001). Histone folds mediate selective heterodimerization of yeast TAF(II)25 with TFIID components yTAF(II)47 and yTAF(II)65 and with SAGA component ySPT7. Molecular and Cellular Biology *21*, 1841–1853.

Gangloff, Y.G., Werten, S., Romier, C., Carré, L., Poch, O., Moras, D., and Davidson, I. (2000). The human TFIID components TAF(II)135 and TAF(II)20 and the yeast SAGA components ADA1 and TAF(II)68 heterodimerize to form histone-like pairs. Molecular and Cellular Biology *20*, 340–351.

García-Oliver, E., García-Molinero, V., and Rodríguez-Navarro, S. (2012). mRNA export and gene expression: the SAGA-TREX-2 connection. Biochimica et biophysica acta *1819*, 555–565.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. Journal of molecular biology *196*, 261–282.

Garvie, C.W., and Wolberger, C. (2001). Recognition of Specific DNA Sequences. Molecular cell *8*, 937–946.

Gazit, K., Moshonov, S., Elfakess, R., Sharon, M., Mengus, G., Davidson, I., and Dikstein, R. (2009). TAF4/4b x TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes. The Journal of biological chemistry *284*, 26286–26296.

Gehre, M., Bunina, D., Sidoli, S., Lübke, M.J., Diaz, N., Trovato, M., Garcia, B.A., Zaugg, J.B., and Noh, K.-M. (2020). Lysine 4 of histone H3.3 is required for embryonic stem cell differentiation, histone enrichment at regulatory regions and transcription accuracy. Nature genetics *52*, 273–282.

Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R.R., Korbel, J.O., and Furlong, E.E.M. (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nature genetics *51*, 1272–1282.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96–100.

Ghosh, A., Shuman, S., and Lima, C.D. (2008). The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. Molecular cell *32*, 478–490.

Ghosh, S., and Pugh, B.F. (2011). Sequential recruitment of SAGA and TFIID in a genomic response to DNA damage in Saccharomyces cerevisiae. Molecular and Cellular Biology *31*, 190–202.

Gilbert, S. F. (2000). Developmental biology (Sunderland, Mass.: Sinauer Assoc).

Gilchrist, D.A., and Adelman, K. (2012). Coupling polymerase pausing and chromatin landscapes for precise regulation of transcription. Biochimica et biophysica acta *1819*, 700–706.

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 & development *26*, 933–944.

Glaser, S., Lubitz, S., Loveland, K.L., Ohbo, K., Robb, L., Schwenk, F., Seibler, J., Roellig, D., Kranz, A., Anastassiadis, K., and Stewart, A.F. (2009). The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. Epigenetics & chromatin *2*, 5.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.-L., and Stewart, A.F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development (Cambridge, England) *133*, 1423–1432.

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. Molecular and Cellular Biology *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 (New York, N.Y.) *292*, 1876–1882.

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

Grant, P.A., Duggan, L., Côté, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., Berger, S.L., and Workman, J.L. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes & development *11*, 1640–1650.

Grant, P.A., Eberharter, A., John, S., Cook, R.G., Turner, B.M., and Workman, J.L. (1999). Expanded lysine acetylation specificity of Gcn5 in native complexes. The Journal of biological chemistry *274*, 5895–5900.

Grant, P.A., Schieltz, D., Pray-Grant, M.G., Steger, D.J., Reese, J.C., Yates, J.R., and Workman, J.L. (1998a). A Subset of TAFIIs Are Integral Components of the SAGA Complex Required for Nucleosome Acetylation and Transcriptional Stimulation. Cell *94*, 45–53.

Grant, P.A., Schieltz, D., Pray-Grant, M.G., Yates, J.R., and Workman, J.L. (1998b). The ATM-related cofactor Tra1 is a component of the purified SAGA complex. Molecular cell *2*, 863–867.

Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nature reviews. Molecular cell biology *20*, 590–607.

Greenblatt, J. (1997). RNA polymerase II holoenzyme and transcriptional regulation. Current opinion in cell biology *9*, 310–319.

Grünberg, S., Henikoff, S., Hahn, S., and Zentner, G.E. (2016). Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. The EMBO journal *35*, 2435–2446.

Guelman, S., Kozuka, K., Mao, Y., Pham, V., Solloway, M.J., Wang, J., Wu, J., Lill, J.R., and Zha, J. (2009). The double-histone-acetyltransferase complex ATAC is essential for mammalian development. Molecular and Cellular Biology *29*, 1176–1188.

Guelman, S., Suganuma, T., Florens, L., Swanson, S.K., Kiesecker, C.L., Kusch, T., Anderson, S., Yates, J.R., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2006). Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel dAda2A/dGcn5-containing histone acetyltransferase complex in Drosophila. Molecular and Cellular Biology *26*, 871–882.

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, R., Chen, J., Mitchell, D.L., and Johnson, D.G. (2011). GCN5 and E2F1 stimulate nucleotide excision repair by promoting H3K9 acetylation at sites of damage. Nucleic acids research *39*, 1390–1397.

Gupta, A., Guerin-Peyrou, T.G., Sharma, G.G., Park, C., Agarwal, M., Ganju, R.K., Pandita, S., Choi, K., Sukumar, S., Pandita, R.K., Ludwig, T., and Pandita, T.K. (2008). The mammalian ortholog of Drosophila MOF that acetylates histone H4 lysine 16 is essential for embryogenesis and oncogenesis. Molecular and Cellular Biology *28*, 397–409.

Gupta, K., Watson, A.A., Baptista, T., Scheer, E., Chambers, A.L., Koehler, C., Zou, J., Obong-Ebong, I., Kandiah, E., Temblador, A., Round, A., Forest, E., Man, P., Bieniossek, C., Laue, E.D., Lemke, E.A., Rappsilber, J., Robinson, C.V., Devys, D., Tora, L., and Berger, I. (2017). Architecture of TAF11/TAF13/TBP complex suggests novel regulation properties of general transcription factor TFIID. eLife *6*.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nature reviews. Molecular cell biology *19*, 621–637.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., Benedetti, R., Altucci, L., Jansen, J.H., Walter, J., Gut, I.G., Marks, H., and Stunnenberg, H.G. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell stem cell *13*, 360–369.

Hager, G.L., McNally, J.G., and Misteli, T. (2009). Transcription dynamics. Molecular cell *35*, 741–753.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nature structural & molecular biology *11*, 394–403.

Hainer, S.J., Bošković, A., McCannell, K.N., Rando, O.J., and Fazzio, T.G. (2019). Profiling of Pluripotency Factors in Single Cells and Early Embryos. Cell *177*, 1319-1329.e11.

Han, Y., Luo, J., Ranish, J., and Hahn, S. (2014). Architecture of the Saccharomyces cerevisiae SAGA transcription coactivator complex. The EMBO journal *33*, 2534–2546.

Han, Y., Reyes, A.A., Malik, S., and He, Y. (2020). Cryo-EM structure of SWI/SNF complex bound to a nucleosome. Nature *579*, 452–455.

Hancock, R. (2014). The crowded nucleus. International review of cell and molecular biology *307*, 15–26.

Hantsche, M., and Cramer, P. (2017). Conserved RNA polymerase II initiation complex structure. Current opinion in structural biology *47*, 17–22.

Hardy, S., Brand, M., Mittler, G., Yanagisawa, J., Kato, S., Meisterernst, M., and Tora, L. (2002). TATA-binding protein-free TAF-containing complex (TFTC) and p300 are both required for efficient transcriptional activation. The Journal of biological chemistry *277*, 32875–32882.

Hartlepp, K.F., Fernández-Tornero, C., Eberharter, A., Grüne, T., Müller, C.W., and Becker, P.B. (2005). The histone fold subunits of Drosophila CHRAC facilitate nucleosome sliding through dynamic DNA interactions. Molecular and Cellular Biology *25*, 9886–9896.

Hastreiter, S., Skylaki, S., Loeffler, D., Reimann, A., Hilsenbeck, O., Hoppe, P.S., Coutu, D.L., Kokkaliaris, K.D., Schwarzfischer, M., Anastassiadis, K., Theis, F.J., and Schroeder, T. (2018). Inductive and Selective Effects of GSK3 and MEK Inhibition on Nanog Heterogeneity in Embryonic Stem Cells. Stem cell reports *11*, 58–69.

Hayashi, K., Sousa Lopes, S.M.C. de, Tang, F., Lao, K., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell stem cell *3*, 391–401.

He, S., Wu, Z., Tian, Y., Yu, Z., Yu, J., Wang, X., Li, J., Liu, B., and Xu, Y. (2020). Structure of nucleosome-bound human BAF complex. Science (New York, N.Y.) *367*, 875–881.

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.

Helmlinger, D., Marguerat, S., Villén, J., Swaney, D.L., Gygi, S.P., Bähler, J., and Winston, F. (2011). Tra1 has specific regulatory roles, rather than global functions, within the SAGA co-activator complex. The EMBO journal *30*, 2843–2852.

Helmlinger, D., and Tora, L. (2017). Sharing the SAGA. Trends in biochemical sciences *42*, 850–861.

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. Molecular cell *52*, 517–528.

Henry, K.W., Wyce, A., Lo, W.-S., Duggan, L.J., Emre, N.C.T., Kao, C.-F., Pillus, L., Shilatifard, A., Osley, M.A., and Berger, S.L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes & development *17*, 2648–2663.

Herceg, Z., Hulla, W., Gell, D., Cuenin, C., Lleonart, M., Jackson, S., and Wang, Z.Q. (2001). Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression. Nature genetics *29*, 206–211.

Herz, H.-M., Mohan, M., Garruss, A.S., Liang, K., Takahashi, Y.-H., Mickey, K., Voets, O., Verrijzer, C.P., and Shilatifard, A. (2012). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes & development *26*, 2604–2620.

Herzel, L., Straube, K., and Neugebauer, K.M. (2018). Long-read sequencing of nascent RNA reveals coupling among RNA processing events. Genome research *28*, 1008–1019.

Hirsch, C.L., Coban Akdemir, Z., Wang, L., Jayakumaran, G., Trcka, D., Weiss, A., Hernandez, J.J., Pan, Q., Han, H., Xu, X., Xia, Z., Salinger, A.P., Wilson, M., Vizeacoumar, F., Datti, A., Li, W., Cooney, A.J., Barton, M.C., Blencowe, B.J., Wrana, J.L., and Dent, S.Y.R. (2015). Myc and SAGA rewire an alternative splicing network during early somatic cell reprogramming. Genes & development *29*, 803–816.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009a). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proceedings of the National Academy of Sciences of the United States of America *106*, 5187–5191.

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009b). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proceedings of the National Academy of Sciences of the United States of America *106*, 5181–5186.

Hodges, C., Kirkland, J.G., and Crabtree, G.R. (2016). The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. Cold Spring Harbor perspectives in medicine *6*.

Hödl, M., and Basler, K. (2009). Transcription in the absence of histone H3.3. Current biology CB *19*, 1221–1226.

Hödl, M., and Basler, K. (2012). Transcription in the absence of histone H3.2 and H3K4 methylation. Current biology CB *22*, 2253–2257.

Horiuchi, J., Silverman, N., Piña, B., Marcus, G.A., and Guarente, L. (1997). ADA1, a novel component of the ADA/GCN5 complex, has broader effects than GCN5, ADA2, or ADA3. Molecular and Cellular Biology *17*, 3220–3228.

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

Hsu, J., Arand, J., Chaikovsky, A., Mooney, N.A., Demeter, J., Brison, C.M., Oliverio, R., Vogel, H., Rubin, S.M., Jackson, P.K., and Sage, J. (2019a). E2F4 regulates transcriptional activation in mouse embryonic stem cells independently of the RB family. Nature communications *10*, 2939.

Hsu, P.L., Shi, H., Leonen, C., Kang, J., Chatterjee, C., and Zheng, N. (2019b). Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS. Molecular cell *76*, 712-723.e4.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009a). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes & development *23*, 837–848.

Hu, Y., Fisher, J.B., Koprowski, S., McAllister, D., Kim, M.-S., and Lough, J. (2009b). Homozygous disruption of the Tip60 gene causes early embryonic lethality. Developmental dynamics an official publication of the American Association of Anatomists *238*, 2912–2921.

Hughes, A.L., and Rando, O.J. (2014). Mechanisms underlying nucleosome positioning in vivo. Annual review of biophysics *43*, 41–63.

Huisinga, K.L., and Pugh, B.F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Molecular cell *13*, 573–585.

Ingvarsdottir, K., Krogan, N.J., Emre, N.C.T., Wyce, A., Thompson, N.J., Emili, A., Hughes, T.R., Greenblatt, J.F., and Berger, S.L. (2005). H2B ubiquitin protease Ubp8 and Sgf11 constitute a discrete functional module within the Saccharomyces cerevisiae SAGA complex. Molecular and Cellular Biology *25*, 1162–1172.

Janssen, A., Colmenares, S.U., and Karpen, G.H. (2018). Heterochromatin: Guardian of the Genome. Annual review of cell and developmental biology *34*, 265–288.

Jeon, C., Yoon, H., and Agarwal, K. (1994). The transcription factor TFIIS zinc ribbon dipeptide Asp-Glu is critical for stimulation of elongation and RNA cleavage by RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America *91*, 9106–9110.

Jeronimo, C., Langelier, M.-F., Bataille, A.R., Pascal, J.M., Pugh, B.F., and Robert, F. (2016). Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. Molecular cell *64*, 455–466.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & development *21*, 1519–1529.

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. Nature genetics *41*, 941–945.

Jin, Q., Yu, L.-R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.-E., Wang, C., Brindle, P.K., Dent, S.Y.R., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. The EMBO journal *30*, 249–262.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews. Genetics *13*, 484–492.

Joti, Y., Hikima, T., Nishino, Y., Kamada, F., Hihara, S., Takata, H., Ishikawa, T., and Maeshima, K. (2012). Chromosomes without a 30-nm chromatin fiber. Nucleus (Austin, Tex.) *3*, 404–410.

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

Kadauke, S., and Blobel, G.A. (2013). Mitotic bookmarking by transcription factors. Epigenetics & chromatin *6*, 6.

Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Science advances *1*, e1500447.

Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter. Wiley interdisciplinary reviews. Developmental biology *1*, 40–51.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., Taatjes, D.J., Dekker, J., and Young, R.A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kalkan, T., and Smith, A. (2014). Mapping the route from naive pluripotency to lineage specification. Philosophical transactions of the Royal Society of London. Series B, Biological sciences *369*.

Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS biology *7*, e1000149.

Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R.G., Meisterernst, M., and Burley, S.K. (2001). Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. Cell *106*, 71–81.

Kamata, K., Hatanaka, A., Goswami, G., Shinmyozu, K., Nakayama, J.-I., Urano, T., Hatashita, M., Uchida, H., and Oki, M. (2013). C-terminus of the Sgf73 subunit of SAGA and SLIK is important for retention in the larger complex and for heterochromatin boundary function. Genes to cells devoted to molecular & cellular mechanisms *18*, 823–837.

Kamieniarz-Gdula, K., and Proudfoot, N.J. (2019). Transcriptional Control by Premature Termination: A Forgotten Mechanism. Trends in genetics TIG *35*, 553–564.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., and Lorincz, M.C. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell stem cell *8*, 676–687.

Kawabata, K., Sakurai, F., Yamaguchi, T., Hayakawa, T., and Mizuguchi, H. (2005). Efficient gene transfer into mouse embryonic stem cells with adenovirus vectors. Molecular therapy the journal of the American Society of Gene Therapy *12*, 547–554.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nature communications *10*, 1930.

Kedinger, C., Gniazdowski, M., Mandel, J.L., Gissinger, F., and Chambon, P. (1970). Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. Biochemical and biophysical research communications *38*, 165–171.

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. Molecular cell *60*, 294–306.

Kettenberger, H., Armache, K.-J., and Cramer, P. (2003). Architecture of the RNA Polymerase II-TFIIS Complex and Implications for mRNA Cleavage. Cell *114*, 347–357.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.-H.A., Marr, M.T., and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes & development *25*, 2502–2512.

Kidder, B.L., Palmer, S., and Knott, J.G. (2009). SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. Stem cells (Dayton, Ohio) *27*, 317–328.

Kikuchi, H., Takami, Y., and Nakayama, T. (2005). GCN5: a supervisor in all-inclusive control of vertebrate cell cycle progression through transcription regulation of various cell cycle-related genes. Gene *347*, 83–97.

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.-A., Hsu, J.-Y., Smith, M.M., and Allis, C.D. (2012). Mutagenesis of pairwise combinations of histone amino-terminal tails reveals functional redundancy in budding yeast. Proceedings of the National Academy of Sciences of the United States of America *109*, 5779–5784.

Kim, J.H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K.Y., Kim, M.-K., Shin, B.A., and Choi, S.-Y. (2011). High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PloS one *6*, e18556.

Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell *77*, 599–608.

Knuesel, M.T., Meyer, K.D., Bernecky, C., and Taatjes, D.J. (2009). The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. Genes & development *23*, 439–451.

Knutson, B.A., and Hahn, S. (2011). Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. Molecular and Cellular Biology *31*, 818–831.

Kobayashi, T., Iwamoto, Y., Takashima, K., Isomura, A., Kosodo, Y., Kawakami, K., Nishioka, T., Kaibuchi, K., and Kageyama, R. (2015). Deubiquitinating enzymes regulate Hes1 stability and neuronal differentiation. The FEBS journal *282*, 2411–2423.

Köhler, A., Schneider, M., Cabal, G.G., Nehrbass, U., and Hurt, E. (2008). Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. Nature cell biology *10*, 707–715.

Köhler, A., Zimmerman, E., Schneider, M., Hurt, E., and Zheng, N. (2010). Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. Cell *141*, 606–617.

Kolesnikova, O., Ben-Shem, A., Luo, J., Ranish, J., Schultz, P., and Papai, G. (2018). Molecular structure of promoter-bound yeast TFIID. Nature communications *9*, 4666.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & development *14*, 2452–2460.

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. Trends in biochemical sciences *30*, 235–239.

Koutelou, E., Hirsch, C.L., and Dent, S.Y.R. (2010). Multiple faces of the SAGA complex. Current opinion in cell biology *22*, 374–382.

Koutelou, E., Wang, L., Schibler, A.C., Chao, H.-P., Kuang, X., Lin, K., Lu, Y., Shen, J., Jeter, C.R., Salinger, A., Wilson, M., Chen, Y.C., Atanassov, B.S., Tang, D.G., and Dent, S.Y.R. (2019). USP22 controls multiple signaling pathways that are essential for vasculature formation in the mouse placenta. Development (Cambridge, England) *146*.

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

Krebs, A.R., Demmers, J., Karmodiya, K., Chang, N.-C., Chang, A.C., and Tora, L. (2010). ATAC and Mediator coactivators form a stable complex and regulate a set of non-coding RNA genes. EMBO reports *11*, 541–547.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Molecular cell *67*, 411-422.e4.

Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Molecular cell *44*, 410–423.

Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. Nature reviews. Molecular cell biology *12*, 283–294.

Kujirai, T., Ehara, H., Fujino, Y., Shirouzu, M., Sekine, S.-I., and Kurumizaka, H. (2018). Structural basis of the nucleosome transition during RNA polymerase II passage. Science (New York, N.Y.) *362*, 595–598.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development (Cambridge, England) *134*, 2895–2902.

Kuo, M.H., Brownell, J.E., Sobel, R.E., Ranalli, T.A., Cook, R.G., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature *383*, 269–272.

Kusch, T., Guelman, S., Abmayr, S.M., and Workman, J.L. (2003). Two Drosophila Ada2 homologues function in different multiprotein complexes. Molecular and Cellular Biology *23*, 3305–3319.

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

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., La Bastide, M. de, Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., Jong, P. de, Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Lang, G., Bonnet, J., Umlauf, D., Karmodiya, K., Koffler, J., Stierle, M., Devys, D., and Tora, L. (2011). The tightly controlled deubiquitination activity of the human SAGA complex differentially modifies distinct gene regulatory elements. Molecular and Cellular Biology *31*, 3734–3744.

Lang, S.E., McMahon, S.B., Cole, M.D., and Hearing, P. (2001). E2F transcriptional activation requires TRRAP and GCN5 cofactors. The Journal of biological chemistry *276*, 32627–32634.

Lanner, F., and Rossant, J. (2010). The role of FGF/Erk signaling in pluripotent cells. Development (Cambridge, England) *137*, 3351–3360.

Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. Nature *565*, 251–254.

Layer, J.H., and Weil, P.A. (2013). Direct TFIIA-TFIID protein contacts drive budding yeast ribosomal protein gene transcription. The Journal of biological chemistry *288*, 23273–23294.

Lee, J.-E., Wang, C., Xu, S., Cho, Y.-W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W., and Ge, K. (2013). H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. eLife *2*, e01503.

Lee, K.K., Sardiu, M.E., Swanson, S.K., Gilmore, J.M., Torok, M., Grant, P.A., Florens, L., Workman, J.L., and Washburn, M.P. (2011). Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. Molecular systems biology *7*, 503.

Lee, T.I., Causton, H.C., Holstege, F.C., Shen, W.C., Hannett, N., Jennings, E.G., Winston, F., Green, M.R., and Young, R.A. (2000). Redundant roles for the TFIID and SAGA complexes in global transcription. Nature *405*, 701–704.

Leeb, M., and Wutz, A. (2007). Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. The Journal of cell biology *178*, 219–229.

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current biology CB *13*, 1192–1200.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., Surani, M.A., and Hajkova, P. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nature structural & molecular biology *20*, 311–316.

Lelli, K.M., Slattery, M., and Mann, R.S. (2012). Disentangling the many layers of eukaryotic transcriptional regulation. Annual review of genetics *46*, 43–68.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature reviews. Genetics *13*, 233–245.

Lenstra, T.L., Benschop, J.J., Kim, T., Schulze, J.M., Brabers, N.A.C.H., Margaritis, T., van de Pasch, L.A.L., van Heesch, S.A.A.C., Brok, M.O., Groot Koerkamp, M.J.A., Ko, C.W., van Leenen, D., Sameith, K., van Hooff, S.R., Lijnzaad, P., Kemmeren, P., Hentrich, T., Kobor, M.S., Buratowski, S., and Holstege, F.C.P. (2011). The specificity and topology of chromatin interaction pathways in yeast. Molecular cell *42*, 536–549.

Lenstra, T.L., and Holstege, F.C.P. (2012). The discrepancy between chromatin factor location and effect. Nucleus (Austin, Tex.) *3*, 213–219.

Lenstra, T.L., Rodriguez, J., Chen, H., and Larson, D.R. (2016). Transcription Dynamics in Living Cells. Annual review of biophysics *45*, 25–47.

Levens, D., Baranello, L., and Kouzine, F. (2016). Controlling gene expression by DNA mechanics: emerging insights and challenges. Biophysical reviews *8*, 259–268.

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

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13–25.

Levo, M., and Segal, E. (2014). In pursuit of design principles of regulatory sequences. Nature reviews. Genetics *15*, 453–468.

Lewis, B.A., and Reinberg, D. (2006). Promoter activation when the ChIPs are down. Nature structural & molecular biology *13*, 96–97.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., Sim, H.S., Peh, S.Q., Mulawadi, F.H., Ong, C.T., Orlov, Y.L., Hong, S., Zhang, Z., Landt, S., Raha, D., Euskirchen, G., Wei, C.-L., Ge, W., Wang, H., Davis, C., Fisher-Aylor, K.I., Mortazavi, A., Gerstein, M., Gingeras, T., Wold, B., Sun, Y., Fullwood, M.J., Cheung, E., Liu, E., Sung, W.-K., Snyder, M., and Ruan, Y. (2012a). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84–98.

Li, W., Atanassov, B.S., Lan, X., Mohan, R.D., Swanson, S.K., Farria, A.T., Florens, L., Washburn, M.P., Workman, J.L., and Dent, S.Y.R. (2016). Cytoplasmic ATXN7L3B Interferes with Nuclear Functions of the SAGA Deubiquitinase Module. Molecular and Cellular Biology *36*, 2855–2866.

Li, X., Li, L., Pandey, R., Byun, J.S., Gardner, K., Qin, Z., and Dou, Y. (2012b). The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network. Cell stem cell *11*, 163–178.

Li, X., Wu, L., Corsa, C.A.S., Kunkel, S., and Dou, Y. (2009). Two mammalian MOF complexes regulate transcription activation by distinct mechanisms. Molecular cell *36*, 290–301.

Li, Y., Zhao, D., Chen, Z., and Li, H. (2017). YEATS domain: Linking histone crotonylation to gene regulation. Transcription *8*, 9–14.

Li, Z., Wang, W., Hu, H., Wang, M., Yi, H., & Lu, J. (2020). Is Oculudentavis a bird or even archosaur?

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional Interaction of Yeast Pre-mRNA 3′ End Processing Factors with RNA Polymerase II. Molecular cell *9*, 1101–1111.

Lim, B., Heist, T., Levine, M., and Fukaya, T. (2018). Visualization of Transvection in Living Drosophila Embryos. Molecular cell *70*, 287-296.e6.

Lin, W., Srajer, G., Evrard, Y.A., Phan, H.M., Furuta, Y., and Dent, S.Y.R. (2007). Developmental potential of Gcn5(-/-) embryonic stem cells in vivo and in vitro. Developmental dynamics an official publication of the American Association of Anatomists *236*, 1547–1557.

Lin, Z., Yang, H., Kong, Q., Li, J., Lee, S.-M., Gao, B., Dong, H., Wei, J., Song, J., Zhang, D.D., and Fang, D. (2012). USP22 antagonizes p53 transcriptional activation by deubiquitinating Sirt1 to suppress cell apoptosis and is required for mouse embryonic development. Molecular cell *46*, 484–494.

Ling, X., Harkness, T.A., Schultz, M.C., Fisher-Adams, G., and Grunstein, M. (1996). Yeast histone H3 and H4 amino termini are important for nucleosome assembly in vivo and in vitro: redundant and position-independent functions in assembly but not in gene regulation. Genes & development *10*, 686–699.

Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. Genes & development *14*, 792–803.

Liu, D., Ishima, R., Tong, K.I., Bagby, S., Kokubo, T., Muhandiram, D.R., Kay, L.E., Nakatani, Y., and Ikura, M. (1998). Solution Structure of a TBP–TAFII230 Complex. Cell *94*, 573–583.

Liu, G., Zheng, X., Guan, H., Cao, Y., Qu, H., Kang, J., Ren, X., Lei, J., Dong, M.-Q., Li, X., and Li, H. (2019). Architecture of Saccharomyces cerevisiae SAGA complex. Cell discovery *5*, 25.

Liu, L., Scolnick, D.M., Trievel, R.C., Zhang, H.B., Marmorstein, R., Halazonetis, T.D., and Berger, S.L. (1999). p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage. Molecular and Cellular Biology *19*, 1202–1209.

Liu, X., Tesfai, J., Evrard, Y.A., Dent, S.Y.R., and Martinez, E. (2003). c-Myc transformation domain recruits the human STAGA complex and requires TRRAP and GCN5 acetylase activity for transcription activation. The Journal of biological chemistry *278*, 20405–20412.

Liu, Y., Pelham-Webb, B., Di Giammartino, D.C., Li, J., Kim, D., Kita, K., Saiz, N., Garg, V., Doane, A., Giannakakou, P., Hadjantonakis, A.-K., Elemento, O., and Apostolou, E. (2017). Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. Cell reports *19*, 1283–1293.

Liu, Y., Zhou, K., Zhang, N., Wei, H., Tan, Y.Z., Zhang, Z., Carragher, B., Potter, C.S., D'Arcy, S., and Luger, K. (2020). FACT caught in the act of manipulating the nucleosome. Nature *577*, 426–431.

Loh, K.M., and Lim, B. (2011). A precarious balance: pluripotency factors as lineage specifiers. Cell stem cell *8*, 363–369.

Long, M., Sun, X., Shi, W., Yanru, A., Leung, S.T.C., Ding, D., Cheema, M.S., MacPherson, N., Nelson, C.J., Ausio, J., Yan, Y., and Ishibashi, T. (2019). A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. Nucleic acids research *47*, 8399–8409.

Lorberbaum, D.S., and Barolo, S. (2013). Gene regulation: when analog beats digital. Current biology CB *23*, R1054-6.

Louder, R.K., He, Y., López-Blanco, J.R., Fang, J., Chacón, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature *531*, 604–609.

Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., and Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. Nature *558*, 318–323.

Marcus, G.A., Horiuchi, J., Silverman, N., and Guarente, L. (1996). ADA5/SPT20 links the ADA and SPT genes, which are involved in yeast transcription. Molecular and Cellular Biology *16*, 3197–3205.

Marcus, G.A., Silverman, N., Berger, S.L., Horiuchi, J., and Guarente, L. (1994). Functional similarity and physical association between GCN5 and ADA2: putative transcriptional adaptors. The EMBO journal *13*, 4807–4815.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590–604.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. The EMBO journal *32*, 2561–2574.

Martello, G., and Smith, A. (2014). The nature of embryonic stem cells. Annual review of cell and developmental biology *30*, 647–675.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell stem cell *11*, 491–504.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 7634–7638.

Martinez, E., Kundu, T.K., Fu, J., and Roeder, R.G. (1998). A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID. The Journal of biological chemistry *273*, 23781–23785.

Martínez-Balbás, M.A., Bauer, U.M., Nielsen, S.J., Brehm, A., and Kouzarides, T. (2000). Regulation of E2F1 activity by acetylation. The EMBO journal *19*, 662–671.

Marzio, G., Wagener, C., Gutierrez, M.I., Cartwright, P., Helin, K., and Giacca, M. (2000). E2F family members are differentially regulated by reversible acetylation. The Journal of biological chemistry *275*, 10887–10892.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927–931.

Matsumura, I., Tanaka, H., and Kanakura, Y. (2003). E2F1 and c-Myc in cell growth and death. Cell cycle (Georgetown, Tex.) *2*, 333–338.

Maurice, T., Duclot, F., Meunier, J., Naert, G., Givalois, L., Meffre, J., Célérier, A., Jacquet, C., Copois, V., Mechti, N., Ozato, K., and Gongora, C. (2008). Altered memory capacities and response to stress in p300/CBP-associated factor (PCAF) histone acetylase knockout mice. Neuropsychopharmacology official publication of the American College of Neuropsychopharmacology *33*, 1584–1602.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nature structural & molecular biology *17*, 1272–1278.

McCormick, M.A., Mason, A.G., Guyenet, S.J., Dang, W., Garza, R.M., Ting, M.K., Moller, R.M., Berger, S.L., Kaeberlein, M., Pillus, L., La Spada, A.R., and Kennedy, B.K. (2014). The SAGA histone deubiquitinase module controls yeast replicative lifespan via Sir2 interaction. Cell reports *8*, 477–486.

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.

McGinty, R.K., and Tan, S. (2014). Histone, Nucleosome, and Chromatin Structure. In Fundamentals of Chromatin, J.L Workman, and S.M Abmayr, eds. (New York, NY: Springer New York), pp. 1–28.

McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. Nature reviews. Molecular cell biology *17*, 16–29.

McMahon, S.B., van Buskirk, H.A., Dugan, K.A., Copeland, T.D., and Cole, M.D. (1998). The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. Cell *94*, 363–374.

Meeks, J.J., and Shilatifard, A. (2017). Multiple Roles for the MLL/COMPASS Family in the Epigenetic Regulation of Gene Expression and in Cancer. Annual Review of Cancer Biology *1*, 425–446.

Meers, M.P., Janssens, D.H., and Henikoff, S. (2019). Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. Molecular cell *75*, 562-575.e5.

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

Mermelstein, F., Yeung, K., Cao, J., Inostroza, J.A., Erdjument-Bromage, H., Eagelson, K., Landsman, D., Levitt, P., Tempst, P., and Reinberg, D. (1996). Requirement of a corepressor for Dr1-mediated repression of transcription. Genes & development *10*, 1033–1048.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nature reviews. Molecular cell biology *7*, 540–546.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Developmental cell *10*, 105–116.

Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., Braekeleer, E. de, Braekeleer, M. de, Delabesse, E., Oliveira, M.P. de, Cavé, H., Clappier, E., van Dongen, J.J.M., Balgobind, B.V., van den Heuvel-Eibrink, M.M., Beverloo, H.B., Panzer-Grümayer, R., Teigler-Schlegel, A., Harbott, J., Kjeldsen, E., Schnittger, S., Koehl, U., Gruhn, B., Heidenreich, O., Chan, L.C., Yip, S.F., Krzywinski, M., Eckert, C., Möricke, A., Schrappe, M., Alonso, C.N., Schäfer, B.W., Krauter, J., Lee, D.A., Zur Stadt, U., Te Kronnie, G., Sutton, R., Izraeli, S., Trakhtenbrot, L., Lo Nigro, L., Tsaur, G., Fechina, L., Szczepanski, T., Strehl, S., Ilencikova, D., Molkentin, M., Burmeister, T., Dingermann, T., Klingebiel, T., and Marschalek, R. (2009). New insights to the MLL recombinome of acute leukemias. Leukemia *23*, 1490–1499.

Mi, W., Guan, H., Lyu, J., Zhao, D., Xi, Y., Jiang, S., Andrews, F.H., Wang, X., Gagea, M., Wen, H., Tora, L., Dent, S.Y.R., Kutateladze, T.G., Li, W., Li, H., and Shi, X. (2017). YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. Nature communications *8*, 1088.

Mi, W., Zhang, Y., Lyu, J., Wang, X., Tong, Q., Peng, D., Xue, Y., Tencer, A.H., Wen, H., Li, W., Kutateladze, T.G., and Shi, X. (2018). The ZZ-type zinc finger of ZZZ3 modulates the ATAC complex-mediated histone acetylation and gene activation. Nature communications *9*, 3759.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S., and Bernstein, B.E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Mischerikow, N., Spedale, G., Altelaar, A.F.M., Timmers, H.T.M., Pijnappel, W.W.M.P., and Heck, A.J.R. (2009). In-depth profiling of post-translational modifications on the related transcription factor complexes TFIID and SAGA. Journal of proteome research *8*, 5020–5030.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.

Mohibi, S., Gurumurthy, C.B., Nag, A., Wang, J., Mirza, S., Mian, Y., Quinn, M., Katafiasz, B., Eudy, J., Pandey, S., Guda, C., Naramura, M., Band, H., and Band, V. (2012). Mammalian alteration/deficiency in activation 3 (Ada3) is essential for embryonic development and cell cycle progression. The Journal of biological chemistry *287*, 29442–29456.

Montgomery, N.D., Della Yee, Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. Current biology CB *15*, 942–947.

Morales Torres, C., Laugesen, A., and Helin, K. (2013). Utx is required for proper induction of ectoderm and mesoderm during differentiation of embryonic stem cells. PloS one *8*, e60020.

Morgan, B.A., Mittman, B.A., and Smith, M.M. (1991). The highly conserved N-terminal domains of histones H3 and H4 are required for normal cell cycle progression. Molecular and Cellular Biology *11*, 4111–4120.

Morgan, M.T., Haj-Yahya, M., Ringel, A.E., Bandi, P., Brik, A., and Wolberger, C. (2016). Structural basis for histone H2B deubiquitination by the SAGA DUB module. Science (New York, N.Y.) *351*, 725–728.

Moris, N., Edri, S., Seyres, D., Kulkarni, R., Domingues, A.F., Balayo, T., Frontini, M., and Pina, C. (2018). Histone Acetyltransferase KAT2A Stabilizes Pluripotency with Control of Transcriptional Heterogeneity. Stem cells (Dayton, Ohio) *36*, 1828–1838.

Morris, S.A., Baek, S., Sung, M.-H., John, S., Wiench, M., Johnson, T.A., Schiltz, R.L., and Hager, G.L. (2014). Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. Nature structural & molecular biology *21*, 73–81.

Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A., and Berezney, R. (1996). A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. Proceedings of the National Academy of Sciences of the United States of America *93*, 8253–8257.

Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., and Washburn, M.P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Molecular cell *34*, 168–178.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Molecular cell *51*, 606–617.

Moyle-Heyrman, G., Viswanathan, R., Widom, J., and Auble, D.T. (2012). Two-step mechanism for modifier of transcription 1 (Mot1) enzyme-catalyzed displacement of TATA-binding protein (TBP) from DNA. The Journal of biological chemistry *287*, 9002–9012.

Mühlbacher, W., Sainsbury, S., Hemann, M., Hantsche, M., Neyer, S., Herzog, F., and Cramer, P. (2014). Conserved architecture of the core RNA polymerase II initiation complex. Nature communications *5*, 4310.

Mujtaba, S., Zeng, L., and Zhou, M.-M. (2007). Structure and acetyl-lysine recognition of the bromodomain. Oncogene *26*, 5521–5527.

Munchel, S.E., Shultzaberger, R.K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Molecular biology of the cell *22*, 2787–2795.

Murakami, K., Tsai, K.-L., Kalisman, N., Bushnell, D.A., Asturias, F.J., and Kornberg, R.D. (2015). Structure of an RNA polymerase II preinitiation complex. Proceedings of the National Academy of Sciences of the United States of America *112*, 13543–13548.

Muratoglu, S., Georgieva, S., Pápai, G., Scheer, E., Enünlü, I., Komonyi, O., Cserpán, I., Lebedeva, L., Nabirochkina, E., Udvardy, A., Tora, L., and Boros, I. (2003). Two different Drosophila ADA2 homologues are present in distinct GCN5 histone acetyltransferase-containing complexes. Molecular and Cellular Biology *23*, 306–321.

Murr, R., Vaissière, T., Sawan, C., Shukla, V., and Herceg, Z. (2007). Orchestration of chromatin-based processes: mind the TRRAP. Oncogene *26*, 5358–5372.

Myer, V.E., and Young, R.A. (1998). RNA polymerase II holoenzymes and subcomplexes. The Journal of biological chemistry *273*, 27757–27760.

Myers, L.C., Gustafsson, C.M., Bushnell, D.A., Lui, M., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (1998). The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. Genes & development *12*, 45–54.

Nagy, Z., Riss, A., Fujiyama, S., Krebs, A., Orpinell, M., Jansen, P., Cohen, A., Stunnenberg, H.G., Kato, S., and Tora, L. (2010). The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. Cellular and molecular life sciences CMLS *67*, 611–628.

Nagy, Z., Riss, A., Romier, C., Le Guezennec, X., Dongre, A.R., Orpinell, M., Han, J., Stunnenberg, H., and Tora, L. (2009). The human SPT20-containing SAGA complex plays a direct role in the regulation of endoplasmic reticulum stress-induced genes. Molecular and Cellular Biology *29*, 1649–1660.

Nagy, Z., and Tora, L. (2007). Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. Oncogene *26*, 5341–5357.

Navarro, P. (2018). 2i, or Not 2i: The Soliloquy of Nanog-Negative Mouse Embryonic Stem Cells. Stem cell reports *11*, 1–3.

Navarro, P., Festuccia, N., Colby, D., Gagliardi, A., Mullin, N.P., Zhang, W., Karwacki-Neisius, V., Osorno, R., Kelly, D., Robertson, M., and Chambers, I. (2012). OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. The EMBO journal *31*, 4547–4562.

Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochimica et biophysica acta *1809*, 34–45.

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 (New York, N.Y.) *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. Molecular cell *11*, 709–719.

Nichols, J., Chambers, I., Taga, T., and Smith, A. (2001). Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development (Cambridge, England) *128*, 2333–2339.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379–391.

Niwa, H. (2007). How is pluripotency determined and maintained? Development (Cambridge, England) *134*, 635–646.

Noe Gonzalez, M., Conaway, R.C., and Conaway, J.W. (2018). Frozen in Transcription: Cryo-EM Structures of Pol II Transcribing through a Nucleosome. Molecular cell *72*, 802–804.

Nogales, E., Louder, R.K., and He, Y. (2017). Structural Insights into the Eukaryotic Transcription Initiation Machinery. Annual review of biophysics *46*, 59–83.

Nonet, M., Sweetser, D., and Young, R.A. (1987). Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell *50*, 909–915.

Nozaki, T., Yachie, N., Ogawa, R., Kratz, A., Saito, R., and Tomita, M. (2011). Tight associations between transcription promoter type and epigenetic variation in histone positioning and modification. BMC genomics *12*, 416.

Oesterreich, F.C., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell *165*, 372–381.

Ogryzko, V.V., Kotani, T., Zhang, X., Schiltz, R.L., Howard, T., Yang, X.J., Howard, B.H., Qin, J., and Nakatani, Y. (1998). Histone-like TAFs within the PCAF histone acetylase complex. Cell *94*, 35–44.

Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., and Hamada, H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell *60*, 461–472.

Olins, D.E., and Olins, A.L. (2003). Chromatin history: our view from the bridge. Nature reviews. Molecular cell biology *4*, 809–814.

Orekhova, A.S., and Rubtsov, P.M. (2013). Bidirectional promoters in the transcription of mammalian genomes. Biochemistry. Biokhimiia *78*, 335–341.

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.

Orpinell, M., Fournier, M., Riss, A., Nagy, Z., Krebs, A.R., Frontini, M., and Tora, L. (2010). The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates. The EMBO journal *29*, 2381–2394.

O'Shea-Greenfield, A., and Smale, S.T. (1992). Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. The Journal of biological chemistry *267*, 1391–1402.

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., Kato, M., Garvin, T.H., Pham, Q.T., Harrington, A.N., Akiyama, J.A., Afzal, V., Lopez-Rios, J., Dickel, D.E., Visel, A., and Pennacchio, L.A. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science (New York, N.Y.) *357*.

Pankotai, T., Komonyi, O., Bodai, L., Ujfaludi, Z., Muratoglu, S., Ciurciu, A., Tora, L., Szabad, J., and Boros, I. (2005). The homologous Drosophila transcriptional adaptors ADA2a and ADA2b are both required for normal development but have different functions. Molecular and Cellular Biology *25*, 8215–8227.

Pankotai, T., Popescu, C., Martín, D., Grau, B., Zsindely, N., Bodai, L., Tora, L., Ferrús, A., and Boros, I. (2010). Genes of the ecdysone biosynthesis pathway are regulated by the dATAC histone acetyltransferase complex in Drosophila. Molecular and Cellular Biology *30*, 4254–4266.

Papai, G., Frechard, A., Kolesnikova, O., Crucifix, C., Schultz, P., and Ben-Shem, A. (2020). Structure of SAGA and mechanism of TBP deposition on gene promoters. Nature.

Papatsenko, D., and Levine, M. (2007). A rationale for the enhanceosome and other evolutionarily constrained enhancers. Current biology CB *17*, R955-7.

Parry, T.J., Theisen, J.W.M., Hsu, J.-Y., Wang, Y.-L., Corcoran, D.L., Eustice, M., Ohler, U., and Kadonaga, J.T. (2010). The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. Genes & development *24*, 2013–2018.

Patel, A.B., Louder, R.K., Greber, B.J., Grünberg, S., Luo, J., Fang, J., Liu, Y., Ranish, J., Hahn, S., and Nogales, E. (2018). Structure of human TFIID and mechanism of TBP loading onto promoter DNA. Science (New York, N.Y.) *362*.

Patel, A.B., Moore, C.M., Greber, B.J., Luo, J., Zukin, S.A., Ranish, J., and Nogales, E. (2019). Architecture of the chromatin remodeler RSC and insights into its nucleosome engagement. eLife *8*.

Patel, J.H., Du, Y., Ard, P.G., Phillips, C., Carella, B., Chen, C.-J., Rakowski, C., Chatterjee, C., Lieberman, P.M., Lane, W.S., Blobel, G.A., and McMahon, S.B. (2004). The c-MYC oncoprotein is a substrate of the acetyltransferases hGCN5/PCAF and TIP60. Molecular and Cellular Biology *24*, 10826–10834.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J.-C., Ferrier, P., and Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. The EMBO journal *30*, 4198–4210.

Pelletier, J., Thomas, G., and Volarević, S. (2018). Ribosome biogenesis in cancer: new players and therapeutic avenues. Nature reviews. Cancer *18*, 51–63.

Pereira, L.A., Klejman, M.P., and Timmers, H.T.M. (2003). Roles for BTAF1 and Mot1p in dynamics of TATA-binding protein and regulation of RNA polymerase II transcription. Gene *315*, 1–13.

Perez-Garcia, V., Fineberg, E., Wilson, R., Murray, A., Mazzeo, C.I., Tudor, C., Sienerth, A., White, J.K., Tuck, E., Ryder, E.J., Gleeson, D., Siragher, E., Wardle-Jones, H., Staudt, N., Wali, N., Collins, J., Geyer, S., Busch-Nentwich, E.M., Galli, A., Smith, J.C., Robertson, E., Adams, D.J., Weninger, W.J., Mohun, T., and Hemberger, M. (2018). Placentation defects are highly prevalent in embryonic lethal mouse mutants. Nature *555*, 463–468.

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M., and Jenuwein, T. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323–337.

Petrenko, N., Jin, Y., Wong, K.H., and Struhl, K. (2016). Mediator Undergoes a Compositional Change during Transcriptional Activation. Molecular cell *64*, 443–454.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes & development *20*, 2922–2936.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., Bland, M.J., Wagstaff, W., Dalton, S., McDevitt, T.C., Sen, R., Dekker, J., Taylor, J., and Corces, V.G. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

Pijnappel, W.W.M.P., and Timmers, H.T.M. (2008). Dubbing SAGA unveils new epigenetic crosstalk. Molecular cell *29*, 152–154.

Piña, B., Berger, S., Marcus, G.A., Silverman, N., Agapite, J., and Guarente, L. (1993). ADA3: a gene, identified by resistance to GAL4-VP16, with properties similar to and different from those of ADA2. Molecular and Cellular Biology *13*, 5981–5989.

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.

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., Villa, E., and Cramer, P. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature *518*, 376–380.

Pollex, T., and Furlong, E.E.M. (2017). Correlation Does Not Imply Causation: Histone Methyltransferases, but Not Histone Methylation, SET the Stage for Enhancer Activation. Molecular cell *66*, 439–441.

Ponjavic, J., Lenhard, B., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., and Sandelin, A. (2006). Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. Genome biology *7*, R78.

Porrua, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. Trends in genetics TIG *32*, 508–522.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nature reviews. Molecular cell biology *16*, 190–202.

Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. Critical reviews in biochemistry and molecular biology *48*, 575–608.

Pradhan, S.K., Su, T., Yen, L., Jacquet, K., Huang, C., Côté, J., Kurdistani, S.K., and Carey, M.F. (2016). EP400 Deposits H3.3 into Promoters and Enhancers during Gene Activation. Molecular cell *61*, 27–38.

Pray-Grant, M.G., Schieltz, D., McMahon, S.J., Wood, J.M., Kennedy, E.L., Cook, R.G., Workman, J.L., Yates, J.R., and Grant, P.A. (2002). The novel SLIK histone acetyltransferase complex functions in the yeast retrograde response pathway. Molecular and Cellular Biology *22*, 8774–8786.

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 (New York, N.Y.) *322*, 1851–1854.

Price, D.H. (2010). Regulation of RNA polymerase II elongation by c-Myc. Cell *141*, 399–400.

Price, D.H. (2018). Transient pausing by RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America *115*, 4810–4812.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501–512.

Pugh, B.F. (2000). Control of gene expression through regulation of the TATA-binding protein. Gene *255*, 1–14.

Putnam, C.D., and Tainer, J.A. (2005). Protein mimicry of DNA and pathway regulation. DNA repair *4*, 1410–1420.

Qian, C., Zhang, Q., Li, S., Zeng, L., Walsh, M.J., and Zhou, M.-M. (2005). Structure and chromosomal DNA binding of the SWIRM domain. Nature structural & molecular biology *12*, 1078–1085.

Qiao, L., Zhang, Q., Zhang, W., and Chen, J.J. (2018). The lysine acetyltransferase GCN5 contributes to human papillomavirus oncoprotein E7-induced cell proliferation via up-regulating E2F1. Journal of cellular and molecular medicine *22*, 5333–5345.

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. The EMBO journal *31*, 3494–3505.

Rach, E.A., Winter, D.R., Benjamin, A.M., Corcoran, D.L., Ni, T., Zhu, J., and Ohler, U. (2011). Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. PLoS genetics *7*, e1001274.

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.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS biology *4*, e309.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Ramachandran, S., and Henikoff, S. (2016). Transcriptional Regulators Compete with Nucleosomes Post-replication. Cell *165*, 580–592.

Rando, O.J., and Winston, F. (2012). Chromatin and transcription in yeast. Genetics *190*, 351–387.

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., Huang, X., Shamim, M.S., Shin, J., Turner, D., Ye, Z., Omer, A.D., Robinson, J.T., Schlick, T., Bernstein, B.E., Casellas, R., Lander, E.S., and Aiden, E.L. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-320.e24.

Rathjen, P.D., Nichols, J., Toth, S., Edwards, D.R., Heath, J.K., and Smith, A.G. (1990). Developmentally programmed induction of differentiation inhibiting activity and the control of stem cell populations. Genes & development *4*, 2308–2318.

Ravens, S., Fournier, M., Ye, T., Stierle, M., Dembele, D., Chavant, V., and Tora, L. (2014). Mof-associated complexes have overlapping and unique roles in regulating pluripotency in embryonic stem cells and during differentiation. eLife *3*.

Ray-Gallet, D., Woolfe, A., Vassias, I., Pellentz, C., Lacoste, N., Puri, A., Schultz, D.C., Pchelintsev, N.A., Adams, P.D., Jansen, L.E.T., and Almouzni, G. (2011). Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. Molecular cell *44*, 928–941.

Reeves, W.M., and Hahn, S. (2005). Targets of the Gal4 transcription activator in functional transcription complexes. Molecular and Cellular Biology *25*, 9092–9102.

Reimer, K.A., Mimoso, C., Adelman, K., and Neugebauer, K.M. (2020). Rapid and Efficient Co-Transcriptional Splicing Enhances Mammalian Gene Expression. bioRxiv *3*.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.

Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell *160*, 1145–1158.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. Genes & development *23*, 1247–1269.

Rickels, R., Herz, H.-M., Sze, C.C., Cao, K., Morgan, M.A., Collings, C.K., Gause, M., Takahashi, Y.-H., Wang, L., Rendleman, E.J., Marshall, S.A., Krueger, A., Bartom, E.T., Piunti, A., Smith, E.R., Abshiru, N.A., Kelleher, N.L., Dorsett, D., and Shilatifard, A. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nature genetics *49*, 1647–1653.

Riss, A., Scheer, E., Joint, M., Trowitzsch, S., Berger, I., and Tora, L. (2015). Subunits of ADA-two-A-containing (ATAC) or Spt-Ada-Gcn5-acetyltrasferase (SAGA) Coactivator Complexes Enhance the Acetyltransferase Activity of GCN5. The Journal of biological chemistry *290*, 28997–29009.

Roberts, S.M., and Winston, F. (1996). SPT20/ADA5 encodes a novel protein functionally related to the TATA-binding protein and important for transcription in Saccharomyces cerevisiae. Molecular and Cellular Biology *16*, 3206–3213.

Rodríguez-Navarro, S., Fischer, T., Luo, M.-J., Antúnez, O., Brettschneider, S., Lechner, J., Pérez-Ortín, J.E., Reed, R., and Hurt, E. (2004). Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. Cell *116*, 75–86.

Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234–237.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nature reviews. Molecular cell biology *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.

Sakai, A., Schwartz, B.E., Goldstein, S., and Ahmad, K. (2009). Transcriptional and developmental functions of the H3.3 histone variant in Drosophila. Current biology CB *19*, 1816–1820.

Saleh, A., Lang, V., Cook, R., and Brandl, C.J. (1997). Identification of native complexes containing the yeast coactivator/repressor proteins NGG1/ADA3 and ADA2. The Journal of biological chemistry *272*, 5571–5578.

Saleh, A., Schieltz, D., Ting, N., McMahon, S.B., Litchfield, D.W., Yates, J.R., Lees-Miller, S.P., Cole, M.D., and Brandl, C.J. (1998). Tra1p is a component of the yeast Ada.Spt transcriptional regulatory complexes. The Journal of biological chemistry *273*, 26559–26565.

Samara, N.L., Datta, A.B., Berndsen, C.E., Zhang, X., Yao, T., Cohen, R.E., and Wolberger, C. (2010). Structural insights into the assembly and function of the SAGA deubiquitinating module. Science (New York, N.Y.) *328*, 1025–1029.

Samara, N.L., Ringel, A.E., and Wolberger, C. (2012). A role for intersubunit interactions in maintaining SAGA deubiquitinating module structure and activity. Structure (London, England 1993) *20*, 1414–1424.

Sanders, S.L., Jennings, J., Canutescu, A., Link, A.J., and Weil, P.A. (2002). Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. Molecular and Cellular Biology *22*, 4723–4738.

Sapountzi, V., and Côté, J. (2011). MYST-family histone acetyltransferases: beyond chromatin. Cellular and molecular life sciences CMLS *68*, 1147–1156.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nature medicine *10*, 55–63.

Savatier, P., Huang, S., Szekely, L., Wiman, K.G., and Samarut, J. (1994). Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. Oncogene *9*, 809–818.

Savatier, P., Lapillonne, H., van Grunsven, L.A., Rudkin, B.B., and Samarut, J. (1996). Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. Oncogene *12*, 309–322.

Schluesche, P., Stelzer, G., Piaia, E., Lamb, D.C., and Meisterernst, M. (2007). NC2 mobilizes TBP on core promoter TATA boxes. Nature structural & molecular biology *14*, 1196–1201.

Schmidl, C., Klug, M., Boeld, T.J., Andreesen, R., Hoffmann, P., Edinger, M., and Rehli, M. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome research *19*, 1165–1174.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nature reviews. Genetics *20*, 437–455.

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.-M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., Dimitrova, E., Dimond, A., Edelman, L.B., Elderkin, S., Tabbada, K., Darbo, E., Andrews, S., Herman, B., Higgs, A., LeProust, E., Osborne, C.S., Mitchell, J.A., Luscombe, N.M., and Fraser, P. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome research *25*, 582–597.

Schöler, H.R., Dressler, G.R., Balling, R., Rohdewohld, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. The EMBO journal *9*, 2185–2195.

Schor, I.E., Degner, J.F., Harnett, D., Cannavò, E., Casale, F.P., Shim, H., Garfield, D.A., Birney, E., Stephens, M., Stegle, O., and Furlong, E.E.M. (2017). Promoter shape varies across populations and affects promoter evolution and expression noise. Nature genetics *49*, 550–558.

Schram, A.W., Baas, R., Jansen, P.W.T.C., Riss, A., Tora, L., Vermeulen, M., and Timmers, H.T.M. (2013). A dual role for SAGA-associated factor 29 (SGF29) in ER stress survival by coordination of both histone H3 acetylation and histone H3 lysine-4 trimethylation. PloS one *8*, e70035.

Schröder, M., and Kaufman, R.J. (2005). ER stress and the unfolded protein response. Mutation research *569*, 29–63.

Schüller, R., Forné, 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. Molecular cell *61*, 305–314.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science (New York, N.Y.) *352*, 1225–1228.

Scognamiglio, R., Cabezas-Wallscheid, N., Thier, M.C., Altamura, S., Reyes, A., Prendergast, Á.M., Baumgärtner, D., Carnevalli, L.S., Atzberger, A., Haas, S., Paleske, L. von, Boroviak, T., Wörsdörfer, P., Essers, M.A.G., Kloz, U., Eisenman, R.N., Edenhofer, F., Bertone, P., Huber, W., van der Hoeven, F., Smith, A., and Trumpp, A. (2016). Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. Cell *164*, 668–680.

Sela, D., Chen, L., Martin-Brown, S., Washburn, M.P., Florens, L., Conaway, J.W., and Conaway, R.C. (2012). Endoplasmic reticulum stress-responsive transcription factor ATF6α directs recruitment of the Mediator of RNA polymerase II transcription and multiple histone acetyltransferase complexes. The Journal of biological chemistry *287*, 23035–23045.

Selleck, W., Howley, R., Fang, Q., Podolny, V., Fried, M.G., Buratowski, S., and Tan, S. (2001). A histone fold TAF octamer within the yeast TFIID transcriptional coactivator. Nature structural biology *8*, 695–700.

Seruggia, D., Oti, M., Tripathi, P., Canver, M.C., LeBlanc, L., Di Giammartino, D.C., Bullen, M.J., Nefzger, C.M., Sun, Y.B.Y., Farouni, R., Polo, J.M., Pinello, L., Apostolou, E., Kim, J., Orkin, S.H., and Das, P.P. (2019). TAF5L and TAF6L Maintain Self-Renewal of Embryonic Stem Cells via the MYC Regulatory Network. Molecular cell *74*, 1148-1163.e7.

Setiaputra, D., Ross, J.D., Lu, S., Cheng, D.T., Dong, M.-Q., and Yip, C.K. (2015). Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. The Journal of biological chemistry *290*, 10057–10070.

Shao, H., Revach, M., Moshonov, S., Tzuman, Y., Gazit, K., Albeck, S., Unger, T., and Dikstein, R. (2005). Core promoter binding by histone-like TAF complexes. Molecular and Cellular Biology *25*, 206–219.

Sharov, G., Voltz, K., Durand, A., Kolesnikova, O., Papai, G., Myasnikov, A.G., Dejaegere, A., Ben Shem, A., and Schultz, P. (2017). Structure of the transcription activator target Tra1 within the chromatin modifying complex SAGA. Nature communications *8*, 1556.

Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Molecular cell *32*, 491–502.

Shi, S., Lin, J., Cai, Y., Yu, J., Hong, H., Ji, K., Downey, J.S., Lu, X., Chen, R., Han, J., and Han, A. (2014). Dimeric structure of p300/CBP associated factor. BMC structural biology *14*, 2.

Shimosato, D., Shiki, M., and Niwa, H. (2007). Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells. BMC developmental biology *7*, 80.

Shivaswamy, S., and Iyer, V.R. (2008). Stress-dependent dynamics of global chromatin remodeling in yeast: dual role for SWI/SNF in the heat shock stress response. Molecular and Cellular Biology *28*, 2221–2234.

Shy, B.R., Wu, C.-I., Khramtsova, G.F., Zhang, J.Y., Olopade, O.I., Goss, K.H., and Merrill, B.J. (2013). Regulation of Tcf7l1 DNA binding and protein stability as principal mechanisms of Wnt/β-catenin signaling. Cell reports *4*, 1–9.

Sigurdsson, S., Dirac-Svejstrup, A.B., and Svejstrup, J.Q. (2010). Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. Molecular cell *38*, 202–210.

Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. Current opinion in cell biology *21*, 344–351.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.

Silva, J., and Smith, A. (2008). Capturing pluripotency. Cell *132*, 532–536.

Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. Nature structural & molecular biology *16*, 1128–1133.

Skene, P.J., and Henikoff, S. (2013). Histone variants in pluripotency and disease. Development (Cambridge, England) *140*, 2513–2524.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordân, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. Trends in biochemical sciences *39*, 381–399.

Smith, A. (2013). Nanog heterogeneity: tilting at windmills? Cell stem cell *13*, 6–7.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688–690.

Smith, E.R., Cayrou, C., Huang, R., Lane, W.S., Côté, J., and Lucchesi, J.C. (2005). A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. Molecular and Cellular Biology *25*, 9175–9188.

Smith, K.N., Singh, A.M., and Dalton, S. (2010). Myc represses primitive endoderm differentiation in pluripotent stem cells. Cell stem cell *7*, 343–354.

Soffers, J.H.M., Li, X., Saraf, A., Seidel, C.W., Florens, L., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2019). Characterization of a metazoan ADA acetyltransferase complex. Nucleic acids research *47*, 3383–3394.

Somesh, B.P., Reid, J., Liu, W.-F., Søgaard, T.M.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2005). Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. Cell *121*, 913–923.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nature reviews. Molecular cell biology *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 (New York, N.Y.) *331*, 1451–1454.

Spedale, G., Mischerikow, N., Heck, A.J.R., Timmers, H.T.M., and Pijnappel, W.W.M.P. (2010). Identification of Pep4p as the protease responsible for formation of the SAGA-related SLIK protein complex. The Journal of biological chemistry *285*, 22793–22799.

Spedale, G., Timmers, H.T.M., and Pijnappel, W.W.M.P. (2012). ATAC-king the complexity of SAGA during evolution. Genes & development *26*, 527–541.

Spilianakis, C.G., and Flavell, R.A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nature immunology *5*, 1017–1027.

Stead, E., White, J., Faast, R., Conn, S., Goldstone, S., Rathjen, J., Dhingra, U., Rathjen, P., Walker, D., and Dalton, S. (2002). Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. Oncogene *21*, 8320–8333.

Sterner, D.E., Belotserkovskaya, R., and Berger, S.L. (2002). SALSA, a variant of yeast SAGA, contains truncated Spt7, which correlates with activated transcription. Proceedings of the National Academy of Sciences of the United States of America *99*, 11622–11627.

Sterner, D.E., Grant, P.A., Roberts, S.M., Duggan, L.J., Belotserkovskaya, R., Pacella, L.A., Winston, F., Workman, J.L., and Berger, S.L. (1999). Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. Molecular and Cellular Biology *19*, 86–98.

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., Houtsmuller, A.B., and Marteijn, J.A. (2018). Live-cell analysis of endogenous GFP-RPB1

uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. Proceedings of the National Academy of Sciences of the United States of America *115*, E4368-E4376.

Stewart-Ornstein, J., Nelson, C., DeRisi, J., Weissman, J.S., and El-Samad, H. (2013). Msn2 coordinates a stoichiometric gene expression program. Current biology CB *23*, 2336–2345.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nature cell biology *9*, 1428–1435.

Strübbe, G., Popp, C., Schmidt, A., Pauli, A., Ringrose, L., Beisel, C., and Paro, R. (2011). Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners. Proceedings of the National Academy of Sciences of the United States of America *108*, 5572–5577.

Suganuma, T., Gutiérrez, J.L., Li, B., Florens, L., Swanson, S.K., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2008). ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. Nature structural & molecular biology *15*, 364–372.

Suganuma, T., Mushegian, A., Swanson, S.K., Abmayr, S.M., Florens, L., Washburn, M.P., and Workman, J.L. (2010). The ATAC acetyltransferase complex coordinates MAP kinases to regulate JNK target genes. Cell *142*, 726–736.

Suganuma, T., Swanson, S.K., Florens, L., Washburn, M.P., and Workman, J.L. (2016). Moco biosynthesis and the ATAC acetyltransferase engage translation initiation by inhibiting latent PKR activity. Journal of molecular cell biology *8*, 44–50.

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. Molecular cell *61*, 297–304.

Sun, J., Paduch, M., Kim, S.-A., Kramer, R.M., Barrios, A.F., Lu, V., Luke, J., Usatyuk, S., Kossiakoff, A.A., and Tan, S. (2018). Structural basis for activation of SAGA histone acetyltransferase Gcn5 by partner subunit Ada2. Proceedings of the National Academy of Sciences of the United States of America *115*, 10010–10015.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome research *22*, 1350–1359.

Sun, Y., Miao, N., and Sun, T. (2019). Detect accessible chromatin using ATAC-sequencing, from principle to applications. Hereditas *156*, 29.

Surani, M.A., Hayashi, K., and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. Cell *128*, 747–762.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., and Shinkai, Y. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes & development *16*, 1779–1791.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes & development *19*, 815–826.

Takagi, Y., and Kornberg, R.D. (2006). Mediator as a general transcription factor. The Journal of biological chemistry *281*, 80–89.

Talbert, P.B., and Henikoff, S. (2017). Histone variants on the move: substrates for chromatin dynamics. Nature reviews. Molecular cell biology *18*, 115–126.

Teves, S.S., An, L., Hansen, A.S., Xie, L., Darzacq, X., and Tjian, R. (2016). A dynamic mode of mitotic bookmarking by transcription factors. eLife *5*.

Thomas, M.C., and Chiang, C.-M. (2006). The general transcription machinery and general cofactors. Critical reviews in biochemistry and molecular biology *41*, 105–178.

Thomas, T., Dixon, M.P., Kueh, A.J., and Voss, A.K. (2008). Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture. Molecular and Cellular Biology *28*, 5093–5105.

Thomson, M., Liu, S.J., Zou, L.-N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875–889.

Trott, J., and Martinez Arias, A. (2013). Single cell lineage analysis of mouse embryonic stem cells at the exit from pluripotency. Biology open *2*, 1049–1056.

Tsai, K.-L., Sato, S., Tomomori-Sato, C., Conaway, R.C., Conaway, J.W., and Asturias, F.J. (2013). A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. Nature structural & molecular biology *20*, 611–619.

Tsunaka, Y., Fujiwara, Y., Oyama, T., Hirose, S., and Morikawa, K. (2016). Integrated molecular mechanism directing nucleosome reorganization by human FACT. Genes & development *30*, 673–686.

Valencia-Sánchez, M.I., Ioannes, P. de, Wang, M., Vasilyev, N., Chen, R., Nudler, E., Armache, J.-P., and Armache, K.-J. (2019). Structural Basis of Dot1L Stimulation by Histone H2B Lysine 120 Ubiquitination. Molecular cell *74*, 1010-1019.e6.

Vamos, E.E., and Boros, I.M. (2012). The C-terminal domains of ADA2 proteins determine selective incorporation into GCN5-containing complexes that target histone H3 or H4 for acetylation. FEBS letters *586*, 3279–3286.

van der Stoop, P., Boutsma, E.A., Hulsman, D., Noback, S., Heimerikx, M., Kerkhoven, R.M., Voncken, J.W., Wessels, L.F.A., and van Lohuizen, M. (2008). Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells. PloS one *3*, e2235.

van Emmerik, C.L., and van Ingen, H. (2019). Unspinning chromatin: Revealing the dynamic nucleosome landscape by NMR. Progress in nuclear magnetic resonance spectroscopy *110*, 1–19.

van Welsem, T., Korthout, T., Ekkebus, R., Morais, D., Molenaar, T.M., van Harten, K., Poramba-Liyanage, D.W., Sun, S.M., Lenstra, T.L., Srivas, R., Ideker, T., Holstege, F.C.P., van Attikum, H., El Oualid, F., Ovaa, H., Stulemeijer, I.J.E., Vlaming, H., and van Leeuwen, F. (2018). Dot1 promotes H2B ubiquitination by a methyltransferase-independent mechanism. Nucleic acids research *46*, 11251–11261.

van Werven, F.J., van Bakel, H., van Teeffelen, H.A.A.M., Altelaar, A.F.M., Koerkamp, M.G., Heck, A.J.R., Holstege, F.C.P., and Timmers, H.T.M. (2008). Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome. Genes & development *22*, 2359–2369.

Varlakhanova, N.V., Cotterman, R.F., deVries, W.N., Morgan, J., Donahue, L.R., Murray, S., Knowles, B.B., and Knoepfler, P.S. (2010). myc maintains embryonic stem cell pluripotency and self-renewal. Differentiation; research in biological diversity *80*, 9–19.

Venters, B.J., Wachi, S., Mavrich, T.N., Andersen, B.E., Jena, P., Sinnamon, A.J., Jain, P., Rolleri, N.S., Jiang, C., Hemeryck-Walsh, C., and Pugh, B.F. (2011). A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. Molecular cell *41*, 480–492.

Verma, R., Mohl, D., and Deshaies, R.J. (2020). Harnessing the Power of Proteolysis for Targeted Protein Inactivation. Molecular cell.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., and Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967–980.

Verrijzer, C.P., Yokomori, K., Chen, J.L., and Tjian, R. (1994). Drosophila TAFII150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. Science (New York, N.Y.) *264*, 933–941.

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.e15.

Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Pol II-DSIF-NELF. Nature *560*, 601–606.

Vosnakis, N., Koch, M., Scheer, E., Kessler, P., Mély, Y., Didier, P., and Tora, L. (2017). Coactivators and general transcription factors have two distinct dynamic populations dependent on transcription. The EMBO journal *36*, 2710–2725.

Wagner, F.R., Dienemann, C., Wang, H., Stützer, A., Tegunov, D., Urlaub, H., and Cramer, P. (2020). Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. Nature *579*, 448–451.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Walter, M., Teissandier, A., Pérez-Palacios, R., and Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. eLife *5*.

Wang, C., Lee, J.-E., Lai, B., Macfarlan, T.S., Xu, S., Zhuang, L., Liu, C., Peng, W., and Ge, K. (2016). Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proceedings of the National Academy of Sciences of the United States of America *113*, 11871–11876.

Wang, H., Dienemann, C., Stützer, A., Urlaub, H., Cheung, A.C.M., and Cramer, P. (2020). Structure of the transcription coactivator SAGA. Nature.

Wang, L., and Dent, S.Y.R. (2014). Functions of SAGA in development and disease. Epigenomics *6*, 329–339.

Wang, L., Du, Y., Ward, J.M., Shimbo, T., Lackford, B., Zheng, X., Miao, Y.-l., Zhou, B., Han, L., Fargo, D.C., Jothi, R., Williams, C.J., Wade, P.A., and Hu, G. (2014). INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. Cell stem cell *14*, 575–591.

Wang, L., Koutelou, E., Hirsch, C., McCarthy, R., Schibler, A., Lin, K., Lu, Y., Jeter, C., Shen, J., Barton, M.C., and Dent, S.Y.R. (2018a). GCN5 Regulates FGF Signaling and Activates Selective MYC Target Genes during Early Embryoid Body Differentiation. Stem cell reports *10*, 287–299.

Wang, X., Ahmad, S., Zhang, Z., Côté, J., and Cai, G. (2018b). Architecture of the Saccharomyces cerevisiae NuA4/TIP60 complex. Nature communications *9*, 1147.

Wang, Y.-L., Faiola, F., Xu, M., Pan, S., and Martinez, E. (2008). Human ATAC Is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. The Journal of biological chemistry *283*, 33808–33815.

Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., and Hahn, S. (2017). Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. Molecular cell *68*, 118-129.e5.

Warfield, L., Ranish, J.A., and Hahn, S. (2004). Positive and negative functions of the SAGA complex mediated through interaction of Spt8 with TBP and the N-terminal domain of TFIIA. Genes & development *18*, 1022–1034.

Warrier, S., Nuwayhid, S., Sabatino, J.A., Sugrue, K.F., and Zohn, I.E. (2017). Supt20 is required for development of the axial skeleton. Developmental biology *421*, 245–257.

Weake, V.M., Lee, K.K., Guelman, S., Lin, C.-H., Seidel, C., Abmayr, S.M., and Workman, J.L. (2008). SAGA-mediated H2B deubiquitination controls the development of neuronal connectivity in the Drosophila visual system. The EMBO journal *27*, 394–405.

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Molecular cell *53*, 819–830.

Welstead, G.G., Creyghton, M.P., Bilodeau, S., Cheng, A.W., Markoulaki, S., Young, R.A., and Jaenisch, R. (2012). X-linked H3K27me3 demethylase Utx is required for embryonic development in a sex-specific manner. Proceedings of the National Academy of Sciences of the United States of America *109*, 13004–13009.

Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. Nature reviews. Microbiology *9*, 85–98.

West, M.L., and Corden, J.L. (1995). Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. Genetics *140*, 1223–1233.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. Science (New York, N.Y.) *303*, 1014–1016.

White, J., Stead, E., Faast, R., Conn, S., Cartwright, P., and Dalton, S. (2005). Developmental activation of the Rb-E2F pathway and establishment of cell cycle-regulated cyclin-dependent kinase activity during embryonic stem cell differentiation. Molecular biology of the cell *16*, 2018–2027.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wieczorek, E., Brand, M., Jacq, X., and Tora, L. (1998). Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II. Nature *393*, 187–191.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684–687.

Winston, F., Chaleff, D.T., Valent, B., and Fink, G.R. (1984a). Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics *107*, 179–197.

Winston, F., Dollard, C., Malone, E.A., Clare, J., Kapakos, J.G., Farabaugh, P., and Minehart, P.L. (1987). Three genes are required for trans-activation of Ty transcription in yeast. Genetics *115*, 649–656.

Winston, F., Durbin, K.J., and Fink, G.R. (1984b). The SPT3 gene is required for normal transcription of Ty elements in S. cerevisiae. Cell *39*, 675–682.

Wissink, E.M., Vihervaara, A., Tippens, N.D., and Lis, J.T. (2019). Nascent RNA analyses: tracking transcription and its regulation. Nature reviews. Genetics *20*, 705–723.

Wong, J.M., and Bateman, E. (1994). TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. Nucleic acids research *22*, 1890–1896.

Worden, E.J., Hoffmann, N.A., Hicks, C.W., and Wolberger, C. (2019). Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. Cell *176*, 1490-1501.e12.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochemical Society transactions *38*, 1027–1032.

Wu, P.-Y.J., Ruhlmann, C., Winston, F., and Schultz, P. (2004). Molecular architecture of the S. cerevisiae SAGA complex. Molecular cell *15*, 199–208.

Wu, P.-Y.J., and Winston, F. (2002). Analysis of Spt7 function in the Saccharomyces cerevisiae SAGA coactivator complex. Molecular and Cellular Biology *22*, 5367–5379.

Wu, Y., Ginther, C., Kim, J., Mosher, N., Chung, S., Slamon, D., and Vadgama, J.V. (2012). Expression of Wnt3 activates Wnt/β-catenin pathway and promotes EMT-like phenotype in trastuzumab-resistant HER2-overexpressing breast cancer cells. Molecular cancer research MCR *10*, 1597–1606.

Xing, L., O'Connor, J.K., Schmitz, L., Chiappe, L.M., McKellar, R.C., Yi, Q., and Li, G. (2020). Hummingbird-sized dinosaur from the Cretaceous period of Myanmar. Nature *579*, 245–249.

Xu, C., and Min, J. (2011). Structure and function of WD40 domain proteins. Protein & cell *2*, 202–214.

Xu, W., Edmondson, D.G., Evrard, Y.A., Wakamiya, M., Behringer, R.R., and Roth, S.Y. (2000). Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. Nature genetics *26*, 229–232.

Xu, Y., Bernecky, C., Lee, C.-T., Maier, K.C., Schwalb, B., Tegunov, D., Plitzko, J.M., Urlaub, H., and Cramer, P. (2017). Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. Nature communications *8*, 15741.

Yamaguchi, Y., Narita, T., Inukai, N., Wada, T., and Handa, H. (2001). SPT genes: key players in the regulation of transcription, chromatin structure and other cellular processes. Journal of biochemistry *129*, 185–191.

Yamauchi, T., Yamauchi, J., Kuwata, T., Tamura, T., Yamashita, T., Bae, N., Westphal, H., Ozato, K., and Nakatani, Y. (2000). Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. Proceedings of the National Academy of Sciences of the United States of America *97*, 11303–11306.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene *389*, 52–65.

Ye, S., Li, P., Tong, C., and Ying, Q.-L. (2013). Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. The EMBO journal *32*, 2548–2560.

Ye, Y., Wu, H., Chen, K., Clapier, C.R., Verma, N., Zhang, W., Deng, H., Cairns, B.R., Gao, N., and Chen, Z. (2019). Structure of the RSC complex bound to the nucleosome. Science (New York, N.Y.) *366*, 838–843.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281–292.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. Nature *378*, 505–508.

Yuan, P., Han, J., Guo, G., Orlov, Y.L., Huss, M., Loh, Y.-H., Yaw, L.-P., Robson, P., Lim, B., and Ng, H.-H. (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes & development *23*, 2507–2520.

Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. Nature *408*, 225–229.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556–559.

Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., and Greenleaf, A.L. (1988). The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. Proceedings of the National Academy of Sciences of the United States of America *85*, 3698–3702.

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. Nature genetics *39*, 1512–1516.

Zentner, G.E., Kasinathan, S., Xin, B., Rohs, R., and Henikoff, S. (2015). ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. Nature communications *6*, 8733.

Zhang, T., Zhang, Z., Dong, Q., Xiong, J., and Zhu, B. (2020). Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. Genome biology *21*, 45.

Zhang, X.-Y., Varthi, M., Sykes, S.M., Phillips, C., Warzecha, C., Zhu, W., Wyce, A., Thorne, A.W., Berger, S.L., and McMahon, S.B. (2008). The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression. Molecular cell *29*, 102–111.

Zhang, Y., Mi, W., Xue, Y., Shi, X., and Kutateladze, T.G. (2019). The ZZ domain as a new epigenetic reader and a degradation signal sensor. Critical reviews in biochemistry and molecular biology *54*, 1–10.

Zhao, D., Guan, H., Zhao, S., Mi, W., Wen, H., Li, Y., Zhao, Y., Allis, C.D., Shi, X., and Li, H. (2016). YEATS2 is a selective histone crotonylation reader. Cell research *26*, 629–632.

Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., Suzuki, E., Le Guezennec, X., Stunnenberg, H.G., Krasnov, A., Georgieva, S.G., Schüle, R., Takeyama, K.-I., Kato, S., Tora, L., and Devys, D. (2008). A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. Molecular cell *29*, 92–101.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., Cramer, P., and Taipale, J. (2018). The interaction landscape between transcription factors and the nucleosome. Nature *562*, 76–81.

Zhu, X., Wirén, M., Sinha, I., Rasmussen, N.N., Linder, T., Holmberg, S., Ekwall, K., and Gustafsson, C.M. (2006). Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. Molecular cell *22*, 169–178.

Zohn, I.E., Li, Y., Skolnik, E.Y., Anderson, K.V., Han, J., and Niswander, L. (2006). p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation. Cell *125*, 957–969.

Zsindely, N., Pankotai, T., Ujfaludi, Z., Lakatos, D., Komonyi, O., Bodai, L., Tora, L., and Boros, I.M. (2009). The loss of histone H3 lysine 9 acetylation due to dSAGA-specific dAda2b mutation influences the expression of only a small subset of genes. Nucleic acids research *37*, 6665–6680.

Annexes

Annexes

1. Generation of cell lines with endogenously tagged subunits of the SAGA and ATAC coactivator complex

The results described in this section represent unpublished data on the generation and validation of individual cell lines possessing 3x HA-tags and a BioTag on the C-terminal ends of either Tada2a (ATAC), Atac2 (ATAC), Tada1 (SAGA) or Taf6l (SAGA).

These subunits were tagged endogenously using CRISPR-Cas9 and homologous recombination as used for the generation of the auxin-inducible degron (AID) cell lines (Table 10 shows gRNAs used). These cell lines were constructed to enable the analysis of genome-wide binding profiles of SAGA and ATAC subunits by the CUT&Run or CUT&Tag methodologies, which require highly specific antibodies (Hainer et al., 2019; Kaya-Okur et al., 2019). Further, these cell lines would enable efficient immunoprecipitations of the SAGA and ATAC complexes and would consequently allow to assess more specifically the effects on SAGA and ATAC complex integrity upon inactivation of their subunits.

Table 10: Table summarizing genes targeted for the generation of tagged cell lines by CRISPR-Cas9. For each of the targeted genes the terminus to which the insert was targeted to, the sequence of the gRNA, the respective PAM sequence and the strand location of the gRNAs are indicated. PAM, protospacer adjacent motif; gRNA, guide RNA; +, forward strand; -, reverse strand. Designed by Bernardo Reina San Martin.

Target	Terminus	gRNA sequence	PAM	Strand
mTaf6l	C-ter	TCGCTGTACCTGCCGCTGTG	AGG	-
		GAGGCGCCGCAACCTCACAG	CGG	+
mTada1	C-ter	AGGGCATACACAGTATGTGT	AGG	-
		GTTGTAGCTCTTCGTGATTG	GGG	-
mTada2a	C-ter	AGATAGACGTGAACAAAACC	CGG	-
		AAGGAATGTGAACAGTCAGA	GGG	-
mAtac2	C-ter	TTACACTCCGTACTCTCCAG	CGG	-
		CCTGAGGCTCCGACGCTGAG	GGG	+

We could successfully obtain homozygous knock-in clones for all four constructs. Western blot analysis using antibodies directed against the HA-tag, revealed a band at the expected size of roughly 70 kDa for two individual *Tada2a*-tag cell lines (Figure 64A). In contrast, in two independent *Atac2*-tag cell lines, several bands at various sizes could be detected beside the expected protein at roughly 120 kDa compared to wildtype untagged cell lines (Figure 64A).

In two independent *Taf6l*-tag clones a protein at the expected size of roughly 90 kDa could be found. Unexpectedly, clone #1 additionally showed a dominant band at roughly 35 kDa (Figure 64B). The cause for this product is not clear but could reflect a transcript variant or truncated protein. Of the two *Tada1*-tag clones only clone #1 showed a faint band at the expected size of roughly 40 kDa, while no band could be detected for clone #2 (Figure 64B).

To ensure that we tagged all possible transcript variants of the four genes, we performed RT-qPCR analysis using two primer pairs with the following design. Primer pair 1 is directed against an exon shared by all transcript variants, while for primer pair 2, the forward primer is localized within the C-terminal exon and the reverse primer aligns to the DNA sequence of the inserted tag. Primer pair 1 allowed us to assess the overall level of expression of all transcript variants, while primer pair 2 revealed the levels of expression of tagged transcripts. By normalising to the results of primer pair 1, we could
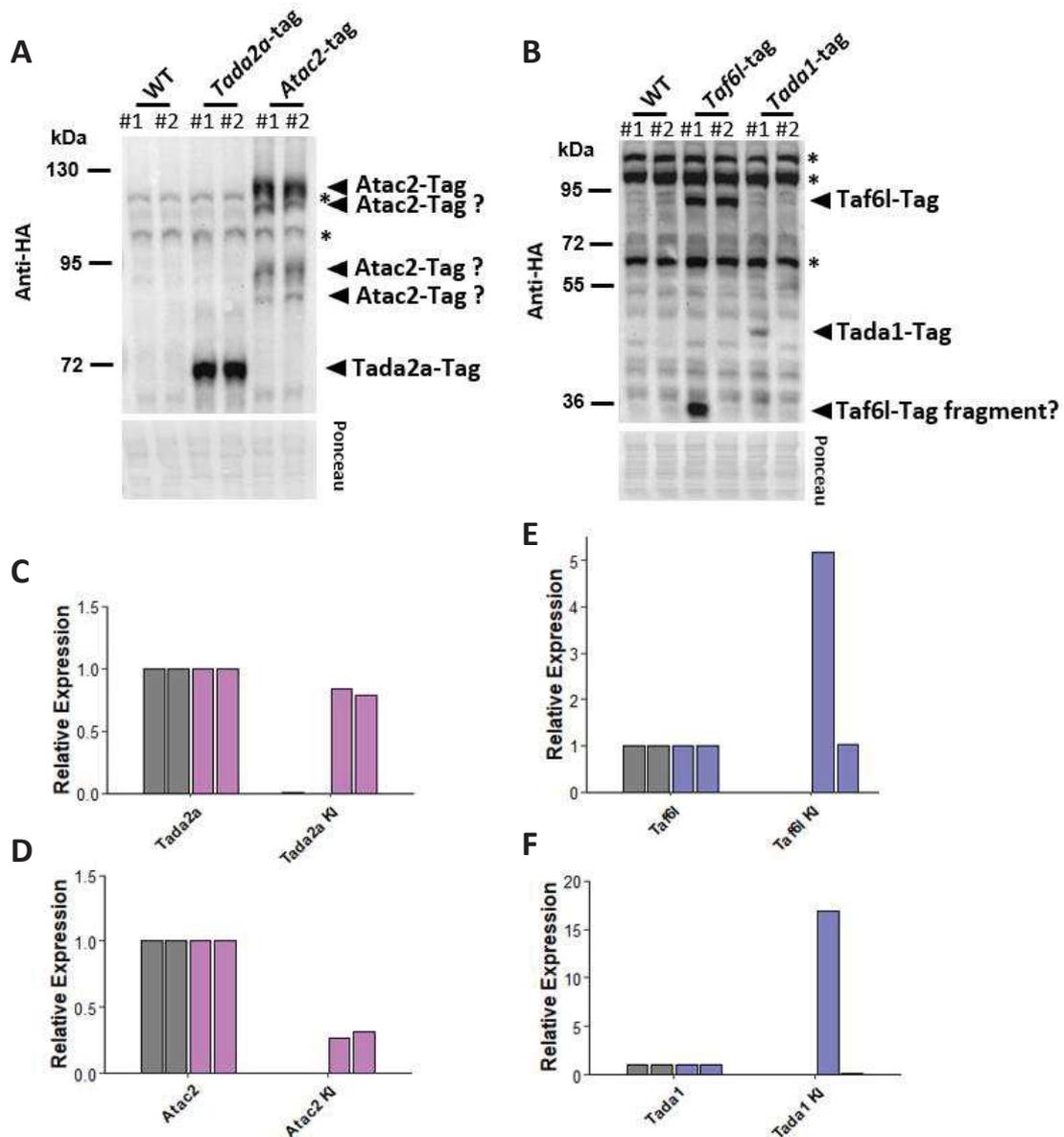
Figure 64: Generation of cell lines with endogenously tagged SAGA and ATAC subunits. A. and **B.** Western blot analysis of cell lines with either endogenously 3xHA and BioTag tagged Tada2a or Atac2 (ATAC subunits in A) and Taf6l or Tada1 (SAGA subunits in B) compared to wildtype (WT) cells. Ponceau staining serves as loading control. Stars indicate unspecific bands. **C., D., E.** and **F.** RT-qPCR analysis of cell lines with tagged SAGA and ATAC subunits shown in (A) and (B) compared to WT cells (in dark grey). Primer pair 1 (named Tada2a, Atac2, Taf6l and Tada1) amplifies from within an exon shared by all transcript variants. Primer pair 2 (named Tada2a KI, Atac2 KI, Taf6l KI and Tada1 KI) amplifies from within the knock-in (KI) tags. Results were normalized to two RNA polymerase III genes (*Rn7sk* and *Rpph1*) and values of primer pair 1 were set to 1. Results for Tada2a-tag cell lines in purple (C), Atac2-tag cell lines in purple (D), for Taf6l-tag cell lines in blue (E) and for Tada-tag cell lines in blue (F).

consequently deduce if the expression level of the tagged transcripts, as assessed by primer pair 2, was equivalent, meaning that all transcript variants were successfully tagged.

This analysis revealed that for the *Tada2a*-tag cell lines, both clones showed levels of the tagged transcripts (primer pair 2: Tada2a KI) very similar to the overall expression levels (primer pair 1: Tada2a) (Figure 64C). In contrast, this analysis revealed that only roughly 25% of all transcripts are tagged in both *Atac2*-tag clones (Figure 64D). For *Taf6l*-tag clone #2, all transcript variants were successfully tagged, while surprisingly clone #1 showed roughly 5-fold higher expression levels of the tagged transcripts (Figure 64E). This increase in tagged transcripts for *Taf6l*-tag clone #1 could suggest a stabilization of these transcripts and could also be at the basis of the truncated protein detected by Western blot analysis (Figure 64B). Astonishingly, clone #1 of the *Tada1*-tag cell lines also displayed a massive increase (roughly 17-fold) of the tagged transcripts over the general expression, while clone #2 only possessed background levels of tagged transcripts (Figure 64F).

In summary, we could validate the two *Tada2a*-tag cell lines and clone #2 of the *Taf6l*-tag cell lines on protein and transcript level. In contrast, in the *Atac2*-tag cell lines only a small portion of transcript variants seemed to be tagged and one of the two clones (clone #2) of the *Tada1*-tag cell lines turned out to be rather comparable to untagged cell lines.

Additionally, to ensure that the addition of the tags to the proteins of interest did not cause any effects on growth of mouse ESCs, we performed clonal assay analysis in medium containing two potent
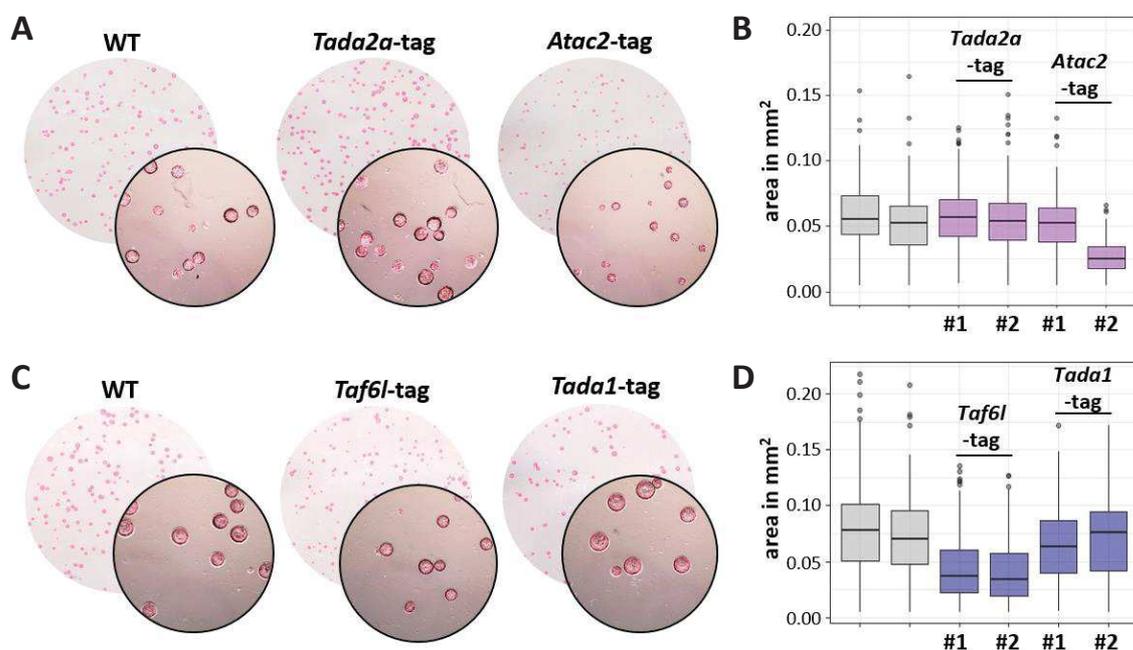

Figure 65: Clonal assay analysis of endogenously tagged SAGA and ATAC subunits. A. and **C.** Representative images of clonal assay analysis of tagged cell lines in medium containing two potent inhibitors of differentiation (2i medium) stained for alkaline phosphatase. Cell lines with endogenously tagged ATAC subunits, Tada2a and Atac2, in (A) and SAGA subunits, Taf6l and Tada1, in (C). **B.** and **D.** Quantification of colony areas of clonal assays shown in (A) and (C), respectively. Colony areas for two independent wildtype cell lines shown in grey. Colony areas were measured using ImageJ.

inhibitors of differentiation (FCS + LIF + 2i medium). These analyses revealed that, while tagging Tada2a and Tada1 did not result in major differences to wildtype cells (Figure 65A to 65D), *Taf6l*-tag cell lines displayed on average smaller colonies (Figure 65C and 65D). As smaller colony sizes had been observed in *Supt7l*^{-/-} cell lines, this suggested that C-terminally tagged Taf6l might not integrate within SAGA. Since Taf6l represents a core subunit of SAGA which forms a dimer with Taf9, its absence from SAGA could lead to a disruption of the complex assembly. Results for the *Atac2*-tag cell lines were very variable, with clone #1 showing no major effect on colony areas, while clone #2 displayed dramatically smaller colonies (Figure 65A and 65B).

Consequently, we aimed to assess the incorporation of the tagged proteins within the respective complexes. For this analysis, the *Atac2*-tag cell lines were excluded as the RT-qPCR analysis suggested that only a small fraction of all transcript variants was actually tagged, and clonal assay analysis showed variable results between the two *Atac2*-tag clones.

We performed immunoprecipitation (IP) experiments against Tada3 and Atac2 in the *Tada2a*-tag cell lines compared to untagged wildtype cells to verify the correct incorporation of Tada2a-tag within ATAC. IP against Tada3 efficiently depleted Tada3 and also Sgf29, Atac2 and Tada2a-tag but not Wdr5 (Figure 66A, compare input (IN) to supernatant (SN). As Wdr5 is also present within other complexes, such as the COMPASS(-like) complexes, it was not astonishing to not reach depletion for this protein. Atac2 IP caused the depletion of Atac2 and partially for Tada2a-tag but not majorly for Tada3, Wdr5 or Sgf29 (Figure 66A). Very surprisingly, Atac2 was suddenly detectable in the *Tada2a*-tag nuclear extracts in contrast to WT nuclear extracts. Also, we found that Tada2a-tag coelutes with the other ATAC subunits tested. Overall, this suggests that the tagged Tada2a successfully incorporates into ATAC with no seemingly major effect on complex assembly. However, it could be that C-terminal tagging of Tada2a could result in the stabilization of Atac2 protein levels and potentially into an overall stabilization of ATAC.

To assess the impact of endogenously tagged Taf6l or Tada1 on SAGA assembly, we performed IP experiments against Taf10 in *Taf6l*-tag and *Tada1*-tag cell lines compared to WT cells. This revealed that SAGA assembly seems affected in the *Taf6l*-tag cell lines as Supt7l (subunit of the core module), Atxn7l3 (subunit of the DUB module) and Sgf29 (subunit of the HAT module) could not be detected in Taf10 IP elutions of *Taf6l*-tag cell lines (Figure 66B). Interestingly, Supt7l levels were also found reduced in the input nuclear extracts in the *Taf6l*-tag cells suggesting that protein stability of Supt7l is affected in the *Taf6l*-tag cells (Figure 66B). Tagging Tada1 did not seem to affect SAGA assembly majorly, although reduced levels were detected for Atxn7l3 in the *Tada1*-tag clone #1 cell line compared to WT cells (Figure 66B).
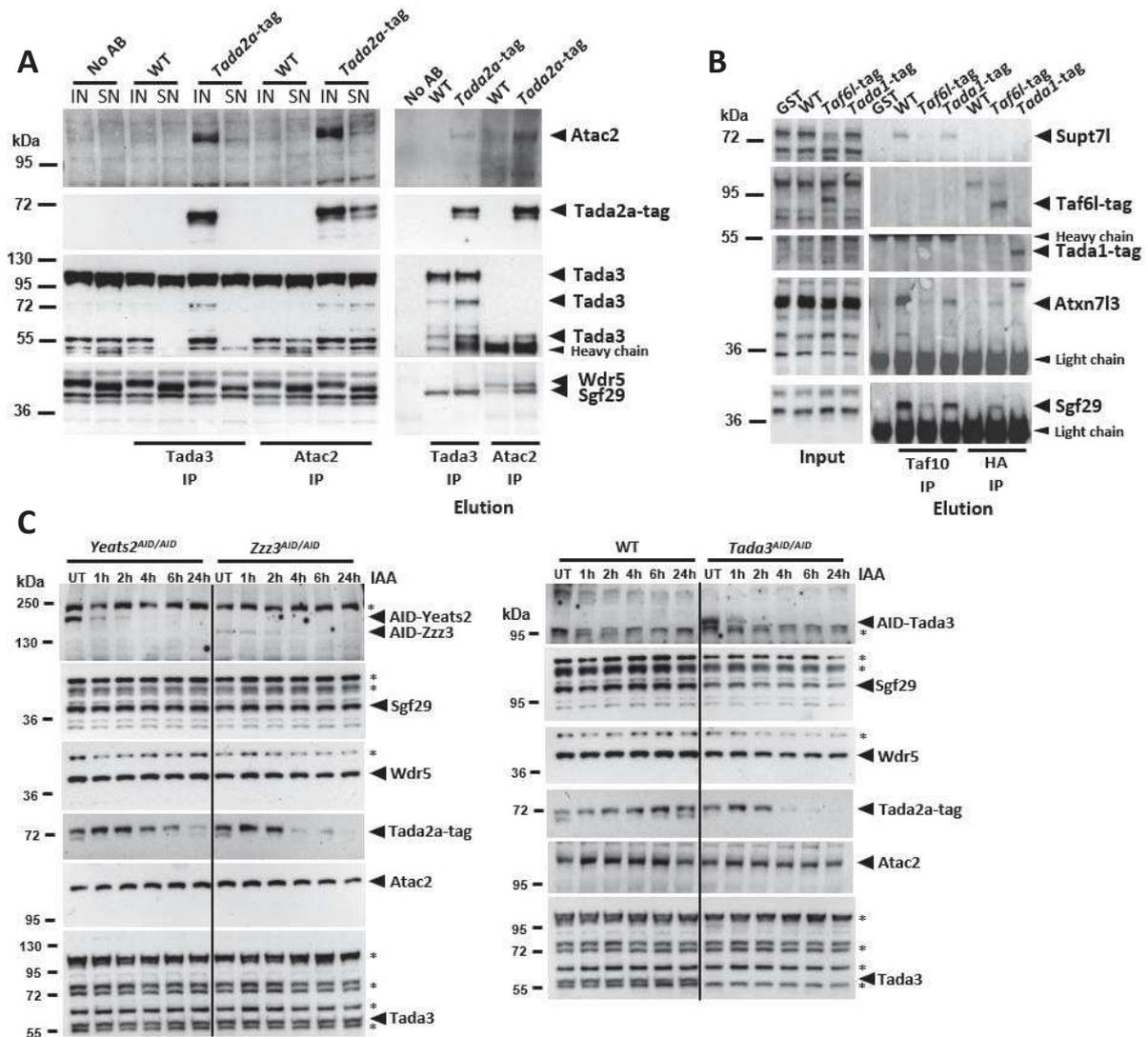
Figure 66: Effects of endogenously tagged SAGA and ATAC subunits on complex integrity. A. Immunoprecipitation (IP) of the ATAC complex from nuclear extracts from wildtype (WT) and *Tada2a*-tag cell lines by using either anti-Tada3 or anti-Atac2 antibodies. IP without antibody (AB) represents the negative control. IN, input; SN, supernatant. **B.** IP against Taf10 to purify the SAGA complex or against the 3xHA-tag using nuclear extracts from *Taf6l*- and *Tada1*-tag cell lines compared to WT cells. IP against GST represents the negative control. Input on the left and elution on the right. **C.** Western blot analysis of whole cell extracts from *Yeats2[AID/AID]*, *Zzz3[AID/AID]*, *Tada3[AID/AID]* and WT cells on a *Tada2a*-tag background untreated (UT) or treated with auxin (IAA) for 1 to 24 hours.

Interestingly, when performing IP experiments against the HA-tag of the insert, we found that Taf6l-tag could interact weakly with Atxn7l3 and Sgf29 but not detectably with Supt7l (Figure 66B). This suggest that Taf6l can partially interact with SAGA subunits, but overall SAGA assembly seems majorly affected as revealed by the anti-Taf10 IP. Anti-HA IP in *Tada1*-tag clone #1 cells did not result into detectably co-purification of the other tested SAGA subunits. The reason for this is unclear but could suggest that Tada1-tag does not successfully incorporate within SAGA.

We consequently constructed *Yeats2[AID/AID]*, *Zzz3[AID/AID]* and *Tada3[AID/AID]* cell lines using a *Tada2a*-tag cell line as a background. These cell lines were constructed to enable subsequent IP experiments

against the HA tag of *Tada2a*-tag for efficient purification of ATAC upon auxin-induced depletion of *Yeats2$^{AID/AID}$*, *Zzz3$^{AID/AID}$* and *Tada3$^{AID/AID}$*.

Unfortunately, it turned out that auxin treatment of *Yeats2$^{AID/AID}$*, *Zzz3$^{AID/AID}$* and *Tada3$^{AID/AID}$* cell lines resulted in an unexpected destabilization of Tada2a-tag compared to WT cells (Figure 66C). This was observed on whole cell extracts with a gradual decrease upon prolonged auxin treatment. Also, this destabilization seemed to be specific to Tada2a protein levels as no major changes could be observed for Sgf29, Wdr5, Atac2 and Tada3 (Figure 66C). This degradation of Tada2a-tag upon auxin treatment specifically in *Yeats2$^{AID/AID}$*, *Zzz3$^{AID/AID}$* and *Tada3$^{AID/AID}$* cell lines could suggest that depletion of these subunits affects the assembly of the HAT module of the ATAC complex. In the case of *Yeats2$^{AID/AID}$* and *Zzz3$^{AID/AID}$* this could indicate a general destabilization of the ATAC complex (see also result section 1).

Overall, the results of this section suggest that unfortunately tagging of endogenous Taf6l and Tada1 leads to an exclusion of the tagged proteins from the SAGA complex and, in the case of tagged Taf6l, in a general destabilization of SAGA. Additionally, only a small portion of *Atac2* transcripts were successfully tagged in the *Atac2*-tag cell lines, which would complicate downstream analyses, leading to the exclusion of the cell lines from further analyses. In contrast, C-terminally-tagged Tada2a was successfully incorporated into the ATAC complex, but unfortunately was destabilized upon depletion of AID-Yeats2, AID-Zzz3 and AID-Tada3. Nevertheless, the *Tada2a*-tag cell lines can be subsequently used in CUT&Run and CUT&Tag experiments to set up the experimental procedure of these experiments and to assess the localization of ATAC at genomic elements in wildtype conditions.

Contributions

Veronique Fischer – Generated cell lines. Designed and performed experiments, analysis and graphical representations of data shown in Figure 64 to 66.

Matthieu Stierle – Helped in generating and validating cell lines.

Bernardo Reina San Martin – designed the CRISPR-Cas9 plasmids.

Didier Devys – Conceived the work.

2. Histone H2Bub1 deubiquitylation is essential for mouse development, but does not regulate global RNA polymerase II transcription

Farrah El-Saafin[1,2,3,4,*,∞] , Fang Wang[1,2,3,4,*], Tao Ye[1,2,3,4,5], Matthieu Stierle[1,2,3,4], Matej Durik[1,2,3,4], Veronique Fischer[1,2,3,4], Didier Devys[1,2,3,4], Stéphane D. Vincent[1,2,3,4], and László Tora[1,2,3,4,#]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France;
[2]Centre National de la Recherche Scientifique (CNRS), UMR7104, 67404 Illkirch, France;
[3]Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, 67404 Illkirch, France;
[4]Université de Strasbourg, 67404 Illkirch, France;
[5]Plateforme GenomEast, infrastructure France Génomique; 67404 Illkirch, France.
*These authors contributed equally to this work
∞Present address: Olivia Newton-John Cancer Research Institute, Melbourne, Victoria, Australia

#Corresponding author: László Tora; Development and stem cells Department; Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), UMR 7104 CNRS, INSERM U1258, Université de Strasbourg (Unistra), 1, rue Laurent Fries, 67404 ILLKIRCH Cedex, FRANCE; Tel +33 3 88 65 34 44, Fax: +33 3 88 65 32 01; e-mail: laszlo@igbmc.fr

Abstract

Coactivator complexes dynamically deposit post-translational modifications (PTMs) on histones, or remove them, to regulate chromatin accessibility and/or to create/erase docking surfaces for proteins that recognize histone PTMs. SAGA (Spt-Ada-Gcn5 Acetyltransferase) is an evolutionary conserved multisubunit co-activator complex with modular organization. The deubiquitylation module (DUB) of mammalian SAGA complex is composed of the ubiquitin-specific protease 22 (USP22) and three adaptor proteins, ATXN7, ATXN7L3 and ENY2, which are all needed for the full activity of the? USP22 enzyme to remove monoubiquitin (ub1) from histone H2B. Two additional USP22-related ubiquitin hydrolases (called USP27X or USP51) have been described to form alternative DUBs with ATXN7L3 and ENY2, which can also deubiquitylate H2Bub1. Here we report that USP22 and ATXN7L3 are essential for normal embryonic development of mice, however their requirements are not identical during this process, as *Atxn7l3* null mutants show developmental delay already at embryonic day (E) 8.5, while *Usp22*$^{-/-}$ mutant embryos are normal at this stage, but die at E14.5. Global histone H2Bub1 levels were only slightly affected in *Usp22* null embryos, in contrast H2Bub1 levels were strongly increased in *Atxn7l3* null mutants and derived cellular systems. Our transcriptomic analyses carried out from wildtype and *Atxn7l3* null mutant mouse embryonic stem cells (mESCs) or primary embryonic fibroblasts (MEFs) suggest that the ATXN7L3-related DUB activity regulates only a subset of genes in both cellular systems, but the gene sets and the extent of their deregulation are different in mESCs and MEFs. Interestingly, the strong genome-wide H2Bub1 increases observed in the *Atxn7l3*$^{-/-}$ mESCs, or *Atxn7l3*$^{-/-}$ MEFs, do not correlate with the modest genome-wide RNA polymerase II occupancy changes observed in the two knock-out cellular systems. Thus, histone H2Bub1 deubiquitylation does not directly regulate global RNA polymerase II transcription.

3. Examination of post-translational modification states of endogenous SAGA and ATAC subunits by immunoprecipitations followed by mass spectrometry analysis from human cells.

This section contains preliminary and unpublished results of downstream analyses which I conducted on data obtained from immunoprecipitation (IP) experiments followed by mass spectrometry (MS) analysis performed by Elisabeth Scheer and Luc Negroni. The work and experiments were conceived by László Tora.

Introduction

Post-translational modifications (PTMs), such as acetylation, methylation and phosphorylation, were reported to occur at subunits of the SAGA and ATAC coactivator complexes (Mischerikow et al., 2009; Spedale et al., 2012). For example, the subunits Spt7 and Sgf73 of yeast SAGA were found to be acetylated at several lysine residues (Mischerikow et al., 2009). Intriguingly, complex-specific differences in PTM levels for proteins shared between complexes were also reported as, for instance, for the Taf5 subunit shared between yeast SAGA and yeast TFIID. Yeast Taf5 was reported to be significantly more phosphorylated when incorporated in SAGA than when part of TFIID (Mischerikow et al., 2009). PTMs could therefore be implicated in regulating the incorporation of subunits within specific complexes by affecting, for instance, their capacity to interact with other subunits.

Further, PTMs on several proteins involved in signal transduction pathways, such as for example phosphorylation of p53 or Stat3, are commonly known to enable the regulation of their localization from the cytoplasm to the nucleus. Consequently, the various PTMs identified for several subunits of the SAGA and also ATAC coactivator complexes were proposed to be mediated by downstream effectors of cellular signalling pathways (Spedale et al., 2012).

Based on this pre-existing knowledge, we wanted to assess if i) specific PTM patterns could be found on subunits of SAGA and ATAC depending on their localization in the cytoplasm or in the nucleus and if ii) the subunits of the histone acetyltransferase (HAT) module shared between these two complexes would display complex-specific PTM marks. To address these questions, cytoplasmic (CE) and nuclear (NE) extracts were prepared from HeLa or U2OS cells and subsequently used for IP experiments using antibodies directed against various subunits of SAGA or ATAC. The purified complexes were analysed by MS to i) assess the general occurrence of PTMs, such as phosphorylation, methylation and acetylation, on SAGA and ATAC subunits, to ii) identify PTMs reproducibly found among the different IP experiments in the cytoplasm and nucleus and to iii) estimate the relative frequencies at which the identified residues are modified.
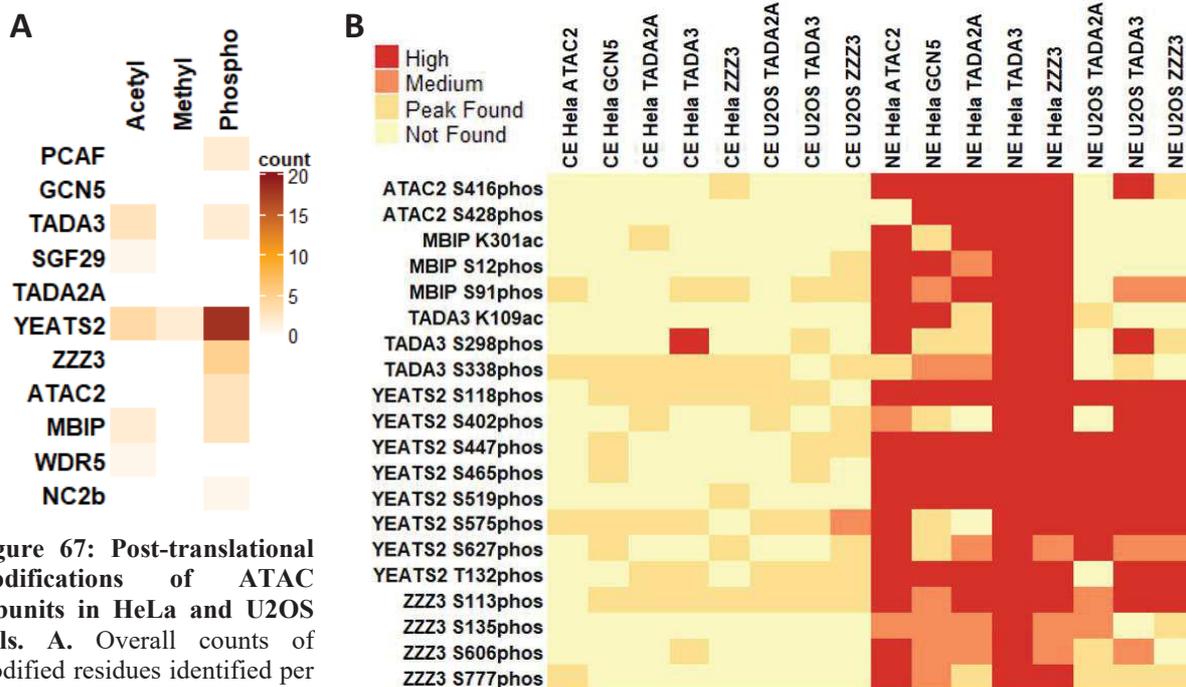
Figure 67: Post-translational modifications of ATAC subunits in HeLa and U2OS cells. A. Overall counts of modified residues identified per subunit of the ATAC complex. All modifications that were identified in at least one immunoprecipitation (IP) experiment are considered. Acetyl, acetylation; Methyl, methylation; Phospho, phosphorylation. **B.** Heatmap of detection accuracy for modified residues showing at least 'Medium' detection accuracy in four IP experiments of ATAC purified from nuclear extracts (NE). Overall, barely any modification is detected within cytoplasmic extracts (CE). **C.** Estimation of frequency of modified peptides for ATAC subunits in HeLa (left) or U2OS (right) nuclear extracts (NE). Each point reflects one IP experiment. Highlighted in orange are residues that show different frequencies between the two cell types, while light blue highlights indicate residues that share similar levels of modifications between HeLa and U2OS cells.
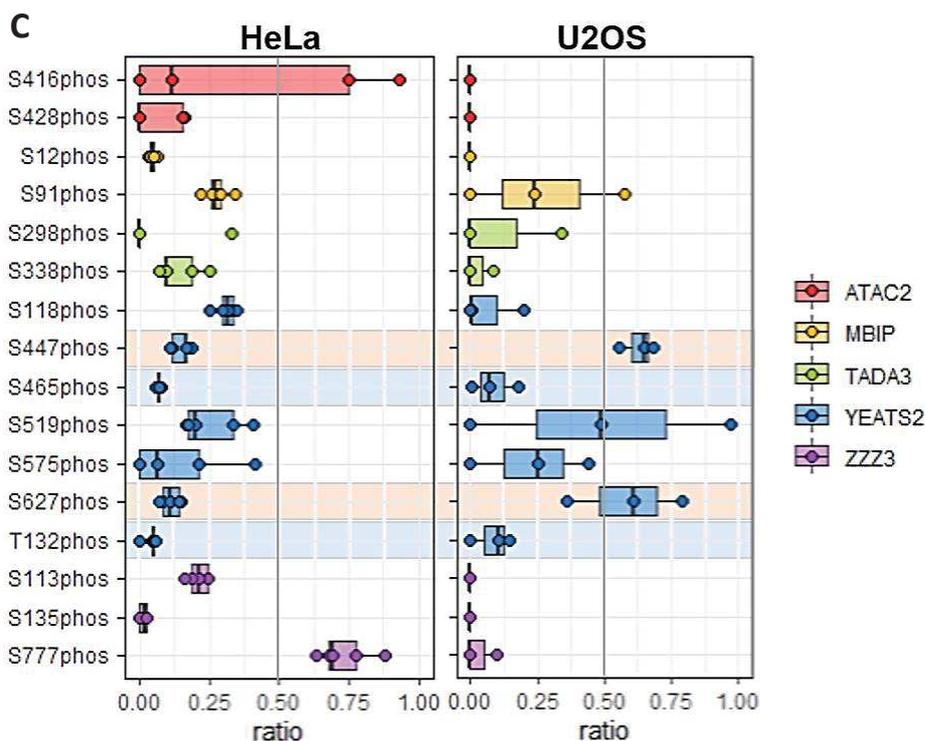
Results

Among the performed IP-MS experiments, acetylation, methylation and phosphorylation events were identified for several subunits of the human ATAC complex, with methylation however only occurring on the YEATS2 subunit (Figure 67A). Indeed, phosphorylation seemed to be the most frequently occurring modification with seven out of eleven ATAC subunits showing at least one

phosphorylated residue. Most strikingly, YEATS2 displayed 18 potential phosphorylation sites (Figure 67A). Importantly, the counting of modified residues represented in Figure 67A considers every modification identified in at least one of the IP experiments.

To identify PTMs which are reproducibly found among the different IP experiments and to be able to analyse potential differences in PTMs between cytoplasmic (CE) or nuclear (NE) extracts, we visualised the confidence of MS identification of the different peptides containing modified residues for each purification experiments (Figure 67B shows modified residues, which were identified to 'Medium' accuracy in at least half of the NE IP experiments; full list in Figure 68A). For 'High' confidence detection a threshold of 1% FDR (False discovery rate) is used, while 'Medium' accuracy is identified at 5% FDR.

This analysis revealed that PTMs on ATAC subunits are mostly detected within the nucleus, while only in one IP experiment TADA3 was found phosphorylated with 'High' confidence in the cytoplasm. This difference in detection of PTMs between CE and NE might be partially due to the fact that ATAC subunits can be generally less well detected when purified from CE (Figure 68B). Interestingly, detection differences for several PTMs of ATAC subunits could be found between nuclear extracts of HeLa or U2OS cell lines such as for the subunits ATAC2, MBIP, TADA3 and ZZZ3 (Figure 67B and Figure 68A).

We were further interested in estimating the relative abundancy of the modified residues compared to its unmodified counterpart. Therefore, we restricted the analysis to ATAC purifications from nuclear extracts and residues shown in Figure 67B, which were found modified rather reproducibly among NE purifications. We subsequently only considered modified peptides for which an unmodified counterpart had also been detected in the respective IP experiments. Additionally, we thresholded for peptides with an abundancy of at least 100.000, which roughly represents the abundancy value separating the confidence categories 'Medium' and 'Peak Found'. This threshold was set to reduce potential biases introduced by technical limitations of detection. Subsequently, we calculated the total quantity of each peptide per IP experiment by summing the abundancy of the modified and unmodified peptides. To roughly estimate the frequency of modification occurring at specific residues of ATAC subunits, the abundancies of the modified peptides were divided by the abundancies of the total peptides for each individual NE IP experiment. Figure 67C shows the calculated ratios for the modified residues, for which the unmodified counterpart could be detected, for either HeLa or U2OS cells separately.

Among the five ATAC purifications from HeLa NE, several phosphorylation marks were reproducibly detected in all of them, such as for example roughly 25% of Serine 91 of MBIP displaying phosphorylation (S91phos) or about 15% of Serine 113 of ZZZ3 being phosphorylated (S113phos) (Figure 67C, left panel). Other examples include S118phos of YEATS2 with a frequency of roughly 35% and S777phos of ZZZ3 occurring at roughly 65% of detected peptides.
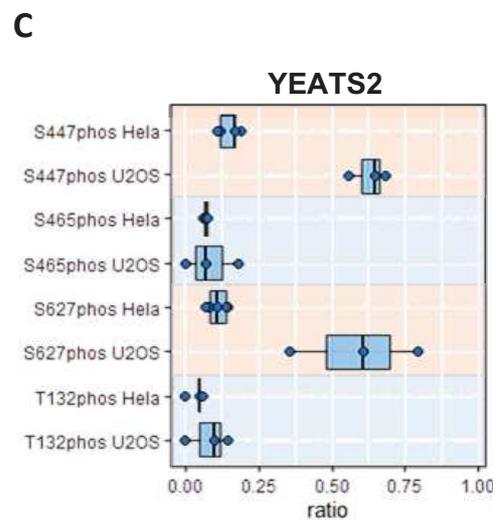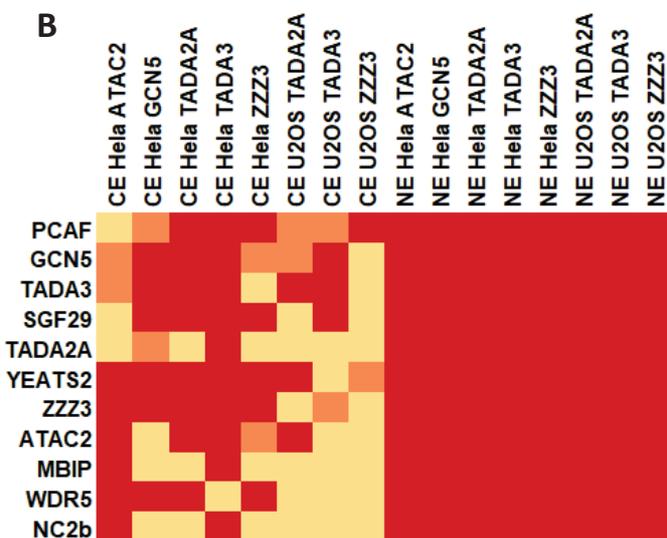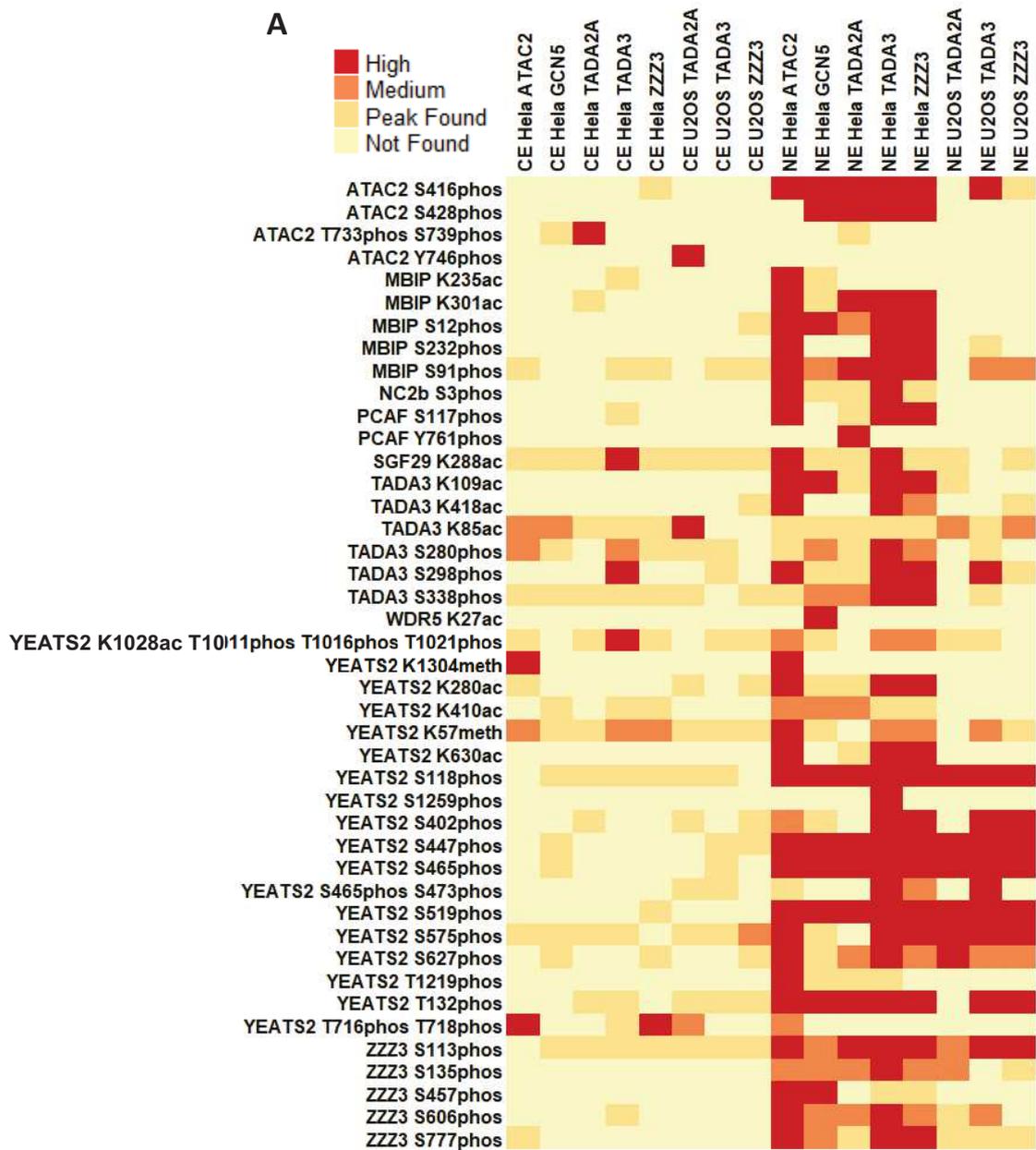
340

Figure legend see next page

Figure 68: *(Figure see previous page)* **Extended data on post-translational modifications of ATAC subunits in HeLa and U2OS cells. A.** Heatmap of detection accuracy for all modified residues identified for ATAC purified from either cytoplasmatic (CE) or nuclear extracts (NE). Overall, few modifications are detected within purifications from CEs. **B.** Heatmap of general detection accuracy for ATAC subunits in CE and NE from HeLa or U2OS cells. ATAC subunits seem to be less well detected within CE purifications. Colour code same as in (A). **C.** Comparison of frequency estimations for modified residues of the ATAC subunit YEATS2 between HeLa and U2OS nuclear extracts. Highlighted in orange are residues that show different frequencies between the two cell types, while light blue highlights indicate residues that share a similar degree of modification between HeLa and U2OS cells.

Modifications among ATAC purifications from U2OS cells were found to frequently not pass the threshold setting, as indicated by values of 0 for the calculated ratios, and to be more variable among the three IP experiments (Figure 67C, right panel). Interestingly, some modifications were found to display different ratios between the two cell types, while others behaved similarly (highlighted in Figure 67C). For example, S465phos and T132phos of YEATS2 showed to be present at roughly 7% of all peptides in both HeLa and U2OS cells, while the residues S447 and S627 of YEATS2 seemed to be more frequently phosphorylated in U2OS cells than in HeLa cells (side-by-side comparison in Figure 68C). It is important to keep in mind that the abundancy of the modified or unmodified peptides can be biased for instance due to differences during the ionization processes of the MS. These biases could lead to an over- or underestimation of the frequency of modification of a given residue (see also discussion section).

We performed the same workflow for IP experiments of SAGA from extracts of HeLa and U2OS cells (Figure 69, 70, 71 and 72). In contrast to the YEATS2 subunit of ATAC showing 18 phosphorylated residues (Figure 67A), the most abundantly modified subunits of SAGA were either TAF9B with six phosphorylation sites or ATXN7 with six acetylation sites (Figure 69A). Comparison between detection accuracy for the different PTMs within CE or NE revealed that several modifications were enriched for nuclear SAGA, such as for instance ATXN7 S711phos, while others were found in both extracts such as, TAF9 T161phos (Figure 69B shows residues, which were found modified with at least 'Medium' detection confidence in half of the NE IP experiments; full list in Figure 71 and continuation in Figure 72A). Intriguingly, in contrast to ATAC, SAGA subunits were generally detected with 'High' accuracy also from cytoplasmic extracts (Figure 72B).

Estimation of the frequencies of the modified residues of SAGA subunits in cytoplasmatic and nuclear extracts of HeLa and U2OS cells showed interesting overlaps and differences (Figure 69C and Figure 70). When comparing modified residues of SAGA subunits purified from either CE or NE of HeLa cells (Figure 69C), several PTMs showed a tendency of being more frequent in nuclear compared to cytoplasmic SAGA. Although not consisted throughout all nine IP experiments, examples include ATXN7 S711phos or S86phos, ATXN7L1 S131phos or SUPT20H S381phos (Figure 69C). One residue that seemed to differ more reproducibly between CE and NE was TAF6L S501phos, which appears more frequent within nuclear SAGA. On the other hand, SUPT20H S437phos occurred at roughly 30%
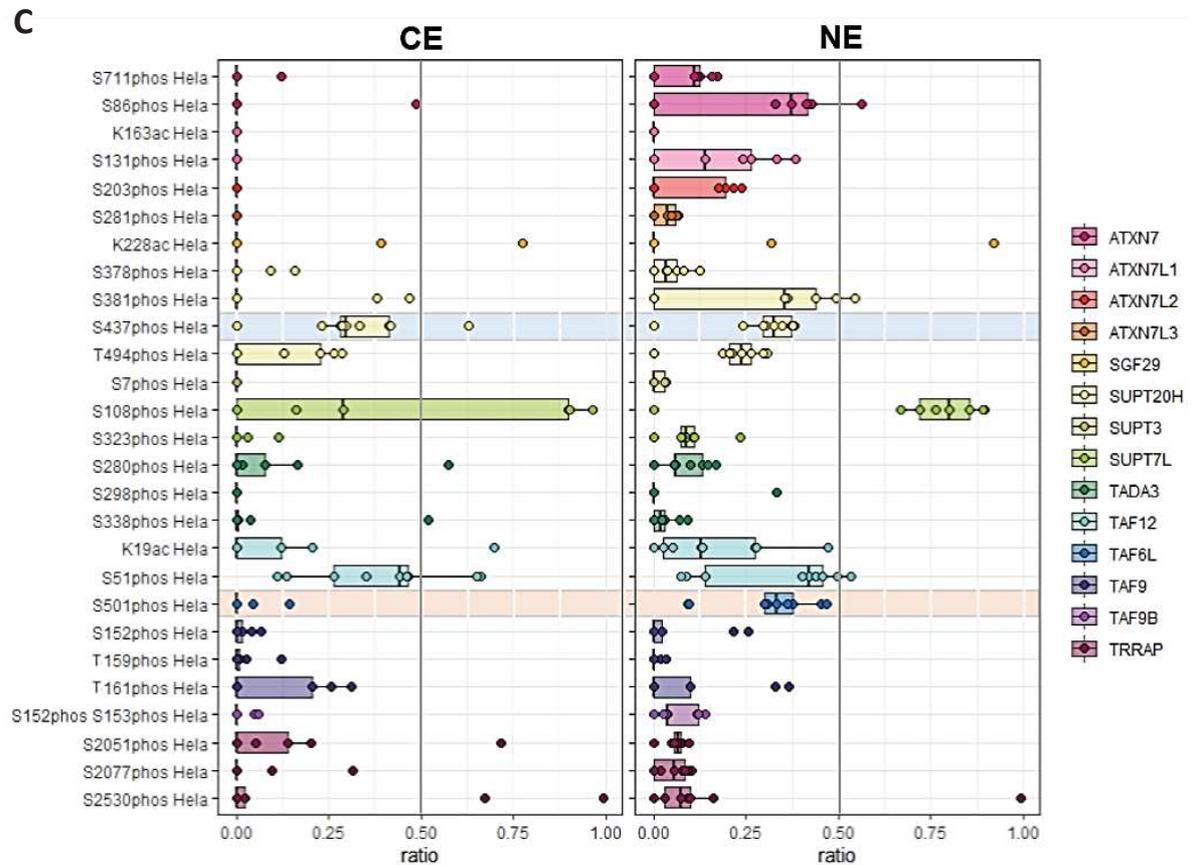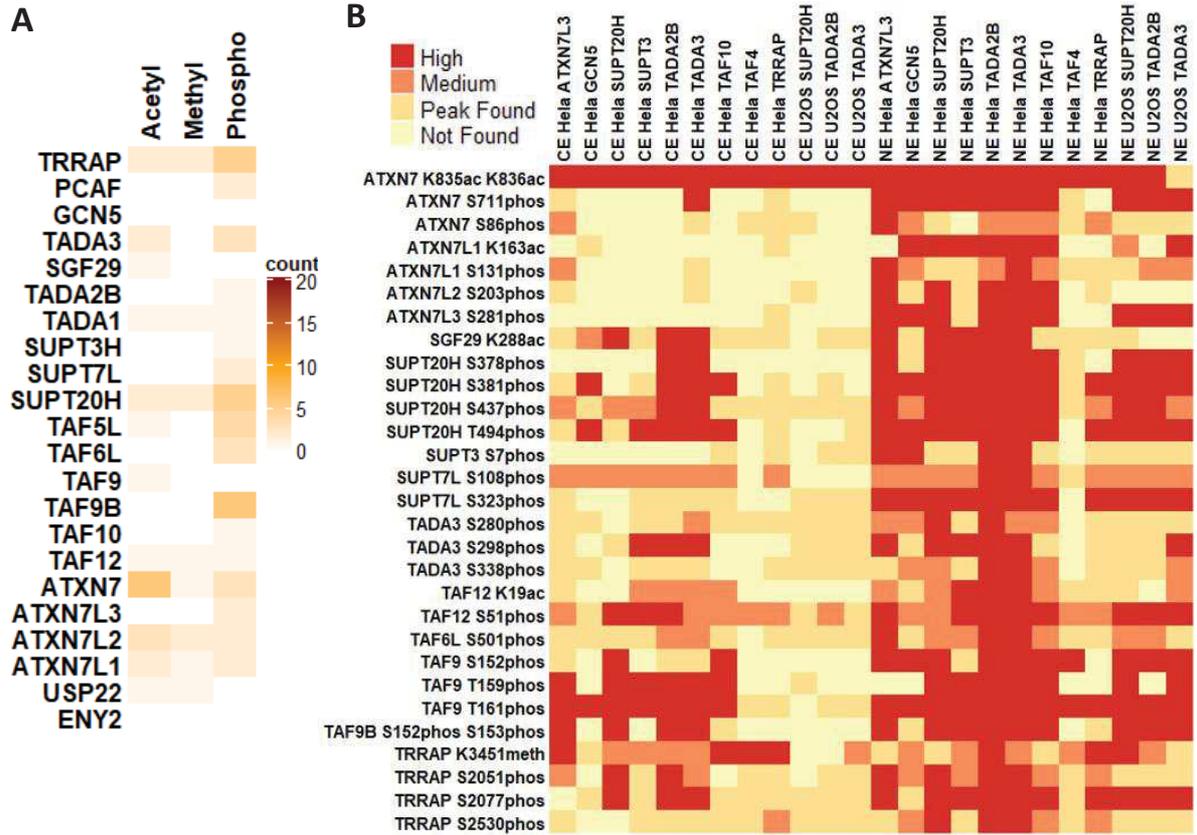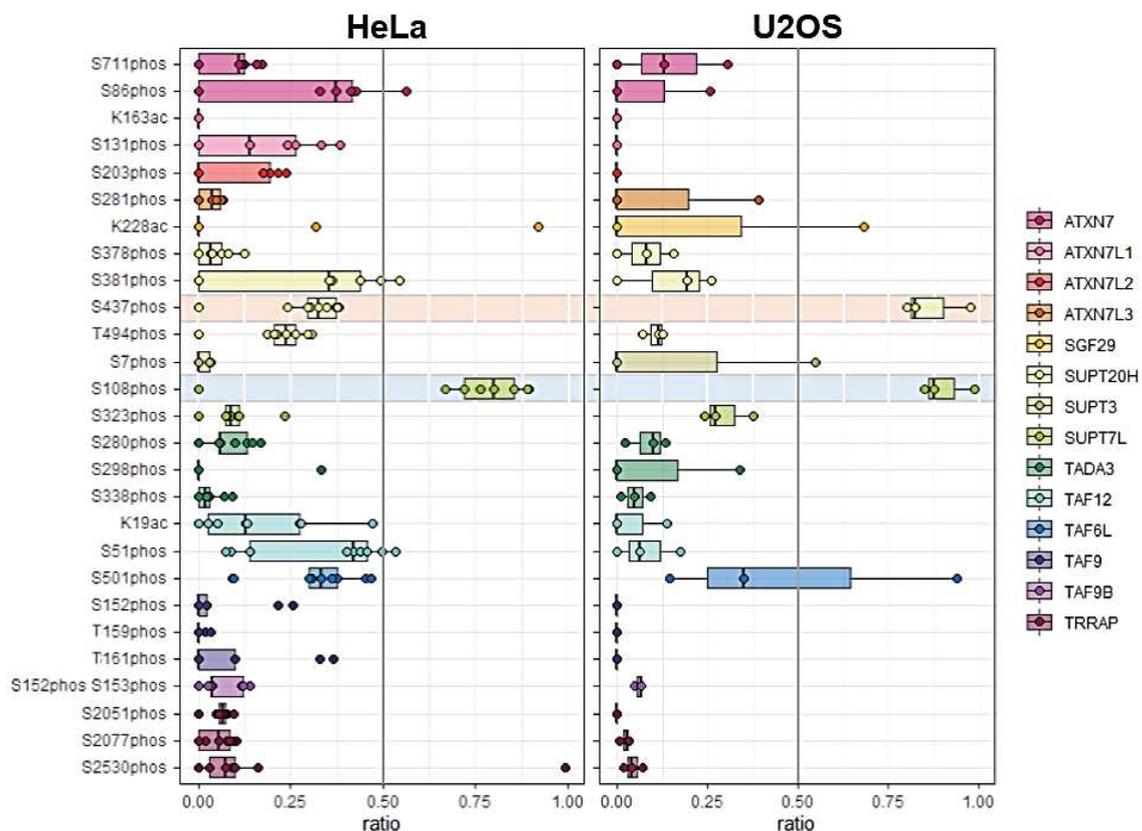
Figure legend see next page

Figure 69: *(Figure see previous page)* **Post-translational modifications of SAGA subunits in HeLa and U2OS cells. A.** Overall counts of modified residues identified per subunit of the SAGA complex. All modifications that are found in at least one immunoprecipitation (IP) experiment are considered. **B.** Heatmap of detection accuracy of modified residues showing at least 'Medium' detection in six IP experiments of SAGA purified from nuclear extracts (NE). CE, cytoplasmic extract. **C.** Estimation of frequency of modified peptides relative to total amount of peptides in CE (left) or NE (right) of HeLa cells. Highlighted in orange is a residue of TAF6L, that shows different frequencies between the two cellular compartments, while light blue highlights indicate a residue of SUPT20H, that shares similar levels of phosphorylation between CE and NE.

in both CE and NE. SUPT7L S108phos was the most frequently and quite reproducibly occurring modification within nuclear SAGA with roughly 80% of peptides being modified in seven out of nine IP experiments.

In general, modifications among SAGA purifications from cytoplasmic extracts frequently did not pass the threshold setting, as indicated by values of 0 for the calculated ratios, and were more variable among the different IP experiments (Figure 69C, right panel). Interestingly, ATXN7 K835ac and K836ac were detected with 'High' accuracy in both cytoplasmic and nuclear extracts (Figure 69B). Unfortunately, no unmodified counterpart of the peptide containing the two modifications sites (ATXN7 K835 and K836) was detected, which consequently led to the exclusion of these residues from the downstream frequency estimation. This highlights a drawback of our quantification estimation, which requires the presence of an unmodified counterpart peptide, consequently excluding in theory modifications which would occur at 100% frequency.

Figure 70: Comparison between estimated frequencies of modified residues between HeLa (left) or U2OS (right) nuclear extracts. Highlighted in orange are a residue of SUPT20H, which shows different frequencies between the two cell types, while light blue highlights indicate a residue of SUPT7L, that shares similar levels of modifications between HeLa and U2OS.
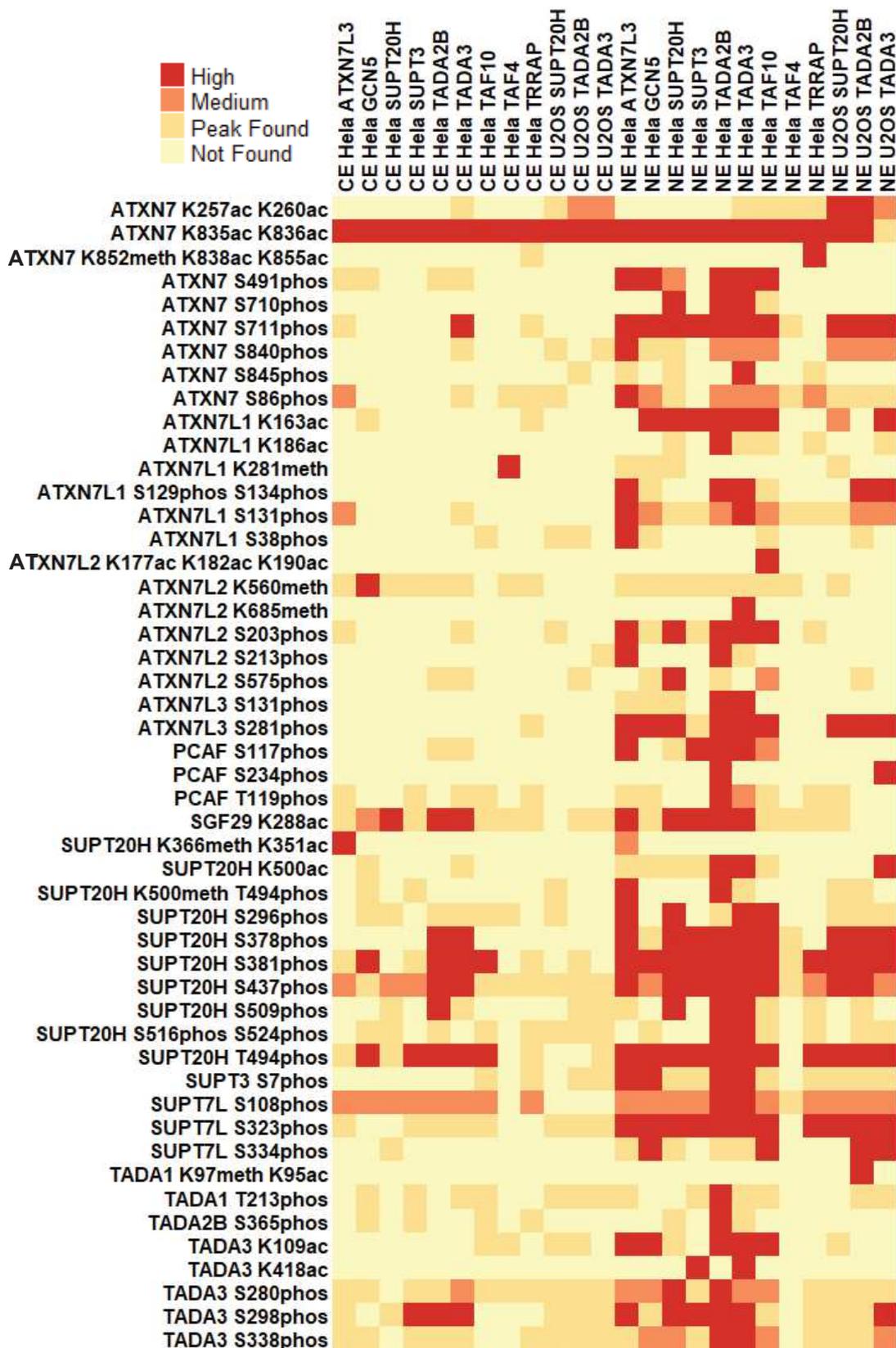
344

Figure 71: Extended data on post-translational modifications of SAGA subunits in HeLa and U2OS cells. First part of heatmap of detection accuracy of all modified residues from IP experiments of SAGA purified from either cytoplasmatic (CE) or nuclear extracts (NE). Continuation in Figure 72, next page.
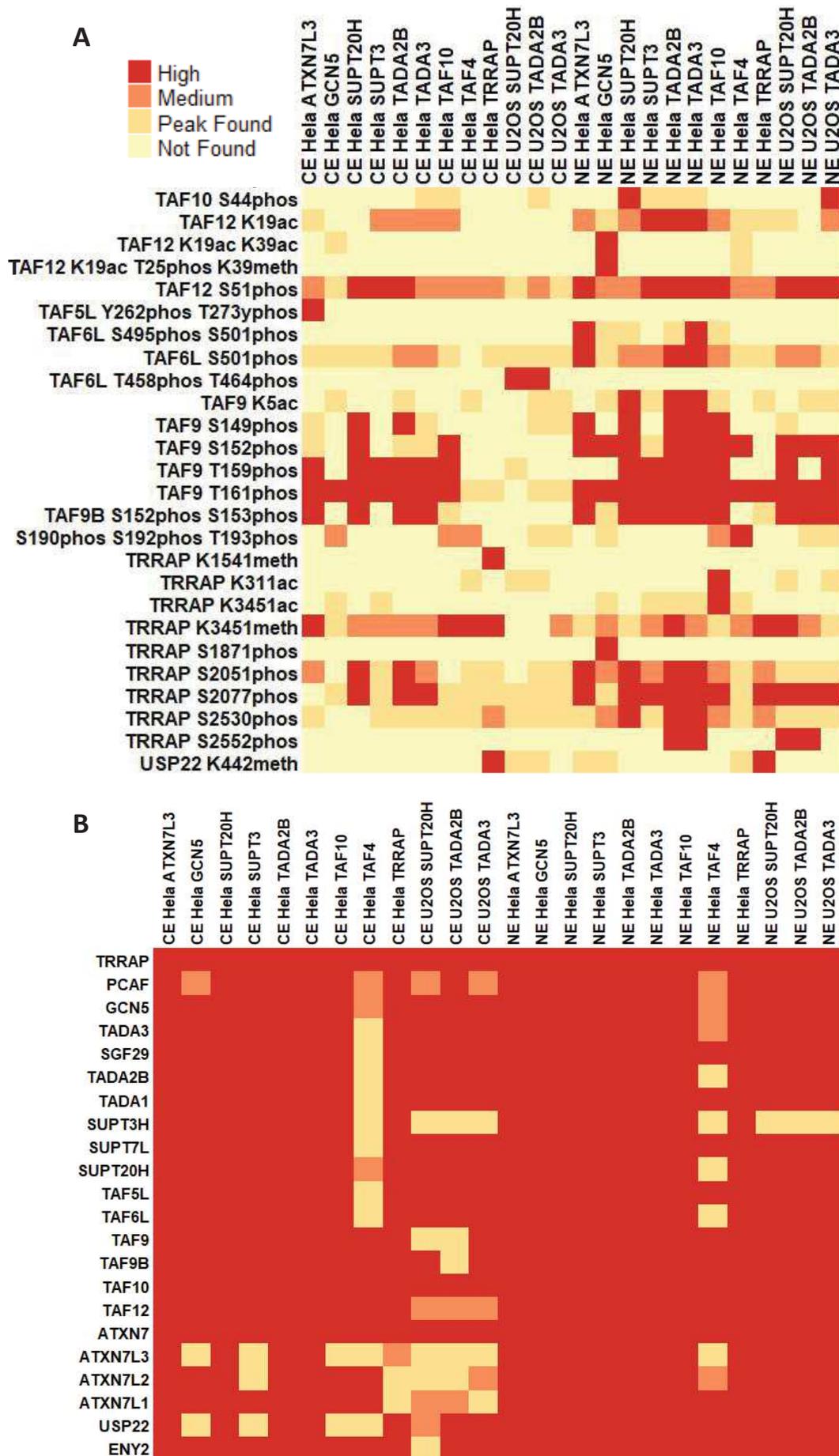
A

B

Figure legend see next page

346

Figure 72: *(Figure see previous page)* **Extended data on post-translational modifications of SAGA subunits in HeLa and U2OS cells. A.** Continuation of heatmap of detection accuracy shown in Figure 71 of all modified residues from IP experiments of SAGA purified from either cytoplasmatic (CE) or nuclear extracts (NE). **B.** Heatmap of general detection accuracy of SAGA subunits in CE and NE from HeLa or U2OS cells. SAGA subunits are also accurately detected within CE purifications. Colour code same as in (A).

Comparison between PTMs occurring in SAGA subunits purified from NE of HeLa or U2OS cells indicated several similar tendencies. In general, however, the results of the SAGA purifications from U2OS NE displayed quite some variability between the three IP experiments (Figure 70). One obvious difference between HeLa and U2OS cells was found for the levels of SUPT20H S437phos, which had a frequency of roughly 30% in HeLa cells while reaching roughly 80% in U2OS cells. In contrast, SUPT7L S108phos seemed to be modified to a similar degree in both HeLa and U2OS cells with roughly 80% vs 88% (Figure 70).

Of the four subunits (PCAF, GCN5, TADA3 and SGF29) shared between SAGA and ATAC only TADA3 seemed to show reproducible phosphorylation marks (Figure 67B and 67C and Figure 69B and 69C). TADA3 S280phos was more accurately detected within SAGA purifications (at least 'Medium' in six out of twelve NE IPs) than ATAC purifications (at least 'Medium' in three out of eight NE IPs) (Figure 68A and Figure 69B). It therefore past the threshold settings for the frequency estimations for SAGA but not for ATAC. Similarly, SGF29 K288ac was found in only 2 out of 8 NE purifications of ATAC with 'High' accuracy, while 5 out of 12 NE purifications of SAGA showed 'High' accuracy for SGF29 K288ac. Overall, however shared subunits of SAGA and ATAC displayed modifications at similar residues. The differences in detection accuracy observed for TADA3 S280phos and SGF29 K288ac might be related to differences in amounts of total TADA3 and SGF29 purified in the respective IP experiments for SAGA or ATAC and could therefore reflect technical limitations.

We consequently were also interested in assessing if the modifications, which were found to be rather reproducibly modified, would occur at residues conserved throughout evolution. Analysis of a few key residues such as YEATS2 T132, S465, S447, S627, SUPT20H S437 or SUPT7L S108 revealed that they were generally conserved amongst human, rat, mouse and bovine homologues, however they were not found within more evolutionary distant species such as *Drosophila* for ATAC subunits or yeast for SAGA subunits in this preliminary analysis. This could suggest that modifications at these residues would be linked to pathways or regulatory stimuli exclusive to some species and not others.

In conclusion, our analysis of PTM occurrences on SAGA and ATAC subunits indicate some potential interesting differences in frequencies of modified residues between cellular compartments and unexpectedly also cell types, which however should be considered carefully. Additional experiments would be required to properly establish the importance of the identified PTMs on functions of SAGA and ATAC subunits (more details see discussion section).

Contributions

Elisabeth Scheer – Performed all immunoprecipitation experiments.

Luc Negroni – Preformed mass spectrometry experiments and analysis.

Veronique Fischer – Performed downstream data analysis and graphical representations shown in Figure 67 to 72.

László Tora – Conceived the work and designed experiments.

4. Imaging of native transcription factors and histone phosphorylation at high resolution in live cells.

Sascha Conic,[1,2,3,4] Dominique Desplancq,[5] Alexia Ferrand,[6] Veronique Fischer,[1,2,3,4] Vincent Heyer,[1,2,3,4] Bernardo Reina San Martin,[1,2,3,4] Julien Pontabry,[1,2,3,4,8] Mustapha Oulad-Abdelghani,[1,2,3,4] Kishore Babu N.,[9] Graham D. Wright,[7] Nacho Molina,[1,2,3,4] Etienne Weiss,[5] and László Tora[1,2,3,4,9]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire

[2]Centre National de la Recherche Scientifique, UMR7104

[3]Institut National de la Santé et de la Recherche Médicale, U964

[4]Université de Strasbourg

[5]Institut de Recherche de l'ESBS, UMR 7242, Illkirch, France

[6]Imaging Core Facility, Biozentrum, University of Basel, Basel, Switzerland

[7]Institute of Medical Biology, A*STAR, Singapore, Singapore

[8]Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Institute of Epigenetics and Stem Cells, München, Germany

[9]School of Biological Sciences, Nanyang Technological University, Singapore

Corresponding authors: Etienne Weiss; Institut de Recherche de l'ESBS, UMR 7242, 300 Boulevard Sébastien Brant, 67412 Illkirch Cedex, France; e-mail: etienne.weiss@unistra.fr and László Tora; Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), UMR 7104 CNRS, INSERM U1258, Université de Strasbourg, 1, rue Laurent Fries, 67404 Illkirch Cedex, France; e-mail: laszlo@igbmc.fr

Abstract

Fluorescent labeling of endogenous proteins for live-cell imaging without exogenous expression of tagged proteins or genetic manipulations has not been routinely possible. We describe a simple versatile antibody-based imaging approach (VANIMA) for the precise localization and tracking of endogenous nuclear factors. Our protocol can be implemented in every laboratory allowing the efficient and nonharmful delivery of organic dye-conjugated antibodies, or antibody fragments, into different metazoan cell types. Live-cell imaging permits following the labeled probes bound to their endogenous targets. By using conventional and super-resolution imaging we show dynamic changes in the distribution of several nuclear transcription factors (i.e., RNA polymerase II or TAF10), and specific phosphorylated histones (γH2AX), upon distinct biological stimuli at the nanometer scale. Hence, considering the large panel of available antibodies and the simplicity of their implementation, VANIMA can be used to uncover novel biological information based on the dynamic behavior of transcription factors or posttranslational modifications in the nucleus of single live cells.

This study was published on the 12th of February 2018 in *Journal of Cell Biology*.

Veronique FISCHER

Etude fonctionnelle des complexes coactivateurs ATAC et SAGA dans les cellules souches de souris

UNIVERSITÉ DE STRASBOURG

École Doctorale
**des Sciences de la Vie
et de la Santé**
S T R A S B O U R G

Résumé

Des études récentes de mon laboratoire d'accueil indiquent que le complexe de modification des histones SAGA agit comme un cofacteur général pour la transcription par l'ARN polymérase II dans la levure, contrairement à sa fonction spécifique supposée précédemment. SAGA est évolutivement bien conservée, de la levure aux mammifères, et possède une fonction d'histone acétyltransférase (HAT). Chez les métazoaires, l'activité HAT de SAGA est partagée avec un autre complexe, le complexe ATAC. Des nouvelles approches m'ont permis de montrer qu'ATAC et SAGA ont des rôles cruciaux pour la maintenance de la pluripotence des cellules souches embryonnaires de souris. Des analyses d'ARN nouvellement-synthétises ont démontré que l'inactivation des complexes ATAC et SAGA modifient l'expression de groupes de gènes différents et aboutent à des phénotypes relativement différents dans ces cellules. Enfin, j'ai pu montrer que les anomalies transcriptionnelles et les phénotypes observé ne semble pas liées à l'activité HAT partagée entre ces deux complexes. Par conséquent, nos données indiquent que les complexes ATAC et SAGA ont des rôles importants et indépendants du HAT dans les cellules mammifères.

Mots clés : histone acétyltransférase, ATAC, SAGA, cellules souches embryonnaires de souris, ARN Polymérase II

Summary

Recent studies from my host laboratory indicate that the histone modifying complex SAGA acts as a general cofactor for RNA polymerase II transcription in budding yeast in contrast to its previously assumed specific functions. SAGA is evolutionarily well conserved, from yeast to mammals, and has a histone acetyltransferase (HAT) function. In metazoans, the HAT activity of SAGA is shared with another complex, the ATAC complex. New approaches have allowed me to demonstrate that SAGA and ATAC have crucial roles in maintaining self-renewal of mouse embryonic stem cells. Newly synthesized RNA analyses revealed that inactivation of the SAGA and ATAC complexes influences the expression of different groups of genes and results in relatively distinct phenotypes in these cells. Finally, I was able to show that the transcriptional anomalies and the observed phenotypes do not seem to be linked to the HAT activity shared by these two complexes. Therefore, our data indicate that SAGA and ATAC have important, HAT-independent roles in mammalian cells.

Keywords: histone acetyltransferase, ATAC, SAGA, mouse embryonic stem cells, RNA polymerase II