

*ÉCOLE DOCTORALE 414 – Sciences de la Vie et de la Santé*

**IGBMC – CNRS UMR 7104 – Inserm U 1258 – Université de Strasbourg**

**THÈSE** présentée par :

**Vincent HISLER**

soutenue le : **9 Décembre 2021**

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

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

**Etude du rôle de TFIID  
et de ses complexes partiels  
dans l'initiation de la transcription**

**THÈSE dirigée par :**

**M. Stéphane VINCENT**

Chargé de recherche INSERM, IGBMC, Université de Strasbourg

**RAPPORTEURS :**

**Mme. Mounia LAGHA**

**M. Nicola IOVINO**

Directrice de recherche CNRS, IGMM, Université de Montpellier

Directeur de recherche, Max Planck Institute of Immunobiology and  
Epigenetics, University of Freiburg, Germany

**AUTRES MEMBRES DU JURY :**

**Mme. Sophie JARRIAULT**

Directrice de recherche CNRS, IGBMC, Université de Strasbourg

**INVITÉS : M. Arnaud KREBS**

Directeur de recherche, EMBL, Heidelberg, Germany

# Remerciements

J'écris ces quelques mots à la hâte avant d'imprimer ce qui a été une grande partie de ma vie durant ces 4 à 5 dernières années.

La réalisation d'une thèse n'est pas que passer de longues heures à la paillasse, à s'énerver sur le fait qu'une manip pourtant optimisée ne marche pas, et râler sur l'ordinateur qui n'avance pas aussi vite que l'on voudrait. C'est avant tout une histoire humaine. Passer de longues heures à lire des articles écrits par des êtres que l'on ne connaît pas, vivant dans une autre région, dans un autre pays, parfois déjà mort, et avoir un sentiment d'admiration par rapport aux travaux réalisés. On voudrait leur dire merci parce que sans le travail qu'ils ont réalisé, on n'aurait pas pu avoir l'idée de faire telles ou telles manips.

C'est une histoire humaine, parce qu'on travaille dans une équipe. Un ensemble de personnes avec qui on partage des histoires privées, des histoires scientifiques, des histoires banales. Je remercie toutes l'équipe Tora pour leur gentillesse ainsi que leur conversation qui m'ont permis de mieux comprendre tellement de processus biologiques, qui je l'avoue reste encore obscure pour moi, mais un peu moins qu'avant.

En particulier, je remercie profondément Stéphane de m'avoir proposé de réaliser ma thèse à ses côtés. Ce n'était pas prévu, mais c'était la meilleure décision. Je ne savais pas parler anglais, alors faire ma thèse à Vienne où tout le monde parle allemand... Quelle idée saugrenue... En faisant ma thèse à ses côtés j'ai appris tellement de choses en biologie mais également à la paillasse, en informatique, en R. Il me faudrait tellement plus de temps pour apprendre tout ce qu'il pourrait m'enseigner. Le problème c'est qu'il apprend au fur et à mesure que j'apprends... Du coup, tristesse, je n'arriverai jamais à le dépasser. Je le remercie également pour tout le temps qu'il a passé à m'entraîner en master pour les différents exams, pour le concours de l'école doctorale, ainsi que pour avoir corrigé tous les choses que j'ai écrites. Je le remercie de m'avoir accompagné dans la rédaction de ma thèse qui fût difficile.

Je remercie Laszlo de m'avoir accueilli dans son laboratoire pour que je puisse réaliser ma thèse. Je le remercie aussi, ainsi que Didier, pour tous les critiques et bons conseils qui m'ont permis de garder le bon cap dans ma recherche durant toutes ces années.

Je remercie Paul Bardot, qui dès le début de mon stage master m'a pris sous son aile et m'a enseigné les rudiments de la biologie et du travail de paillasse. Tout le travail que je présente dans cette thèse a été initié par lui.

Je remercie Kenny, ce complet inconnu déjà présent en tant que stagiaire quand j'ai intégré l'équipe. Cette personne que je connaissais à peine et qui est devenue ma femme de labo avec le temps (je peux écrire ce que je veux ici, personne ne lit jamais les remerciements dépassé le premier paragraphe). Grâce à Kenny et au lien d'amitié que nous avons tissé, la thèse a été moins douloureuse que prévue.

Je remercie Eli qui tout le long m'a montré de nouvelles techniques et m'a toujours bien conseillé. Je remercie Véronique de m'avoir enseigné la préparation des ARNs pour le TT-seq. Je remercie Matthieu de m'avoir aidé à générer mes lignées mESC et de m'aider dorénavant à finir mon projet. Et de façon plus générale, je remercie tous les personnes qui au cours de ces années m'ont aidé à résoudre mes problèmes d'expérimentations et m'ont donné de bonnes idées pour faire avancer mon projet.

Je remercie les différents colocataires que j'ai pu avoir durant toutes ces années, Nathalie, Kevin, Aude, Yurou, Marie Q, s'en oublier Marie F, Louise, Aurélien, Benoît, Clément, Jérémie, Raquelle, Emma. Je remercie Aurélien et Louise d'avoir pris le temps de relire quelques parties de ma thèse pour corriger les fautes.



Les amis... C'est ça aussi une thèse. Se faire un nouvel cercle social. Je remercie toutes les personnes que j'ai pu côtoyer et me lier d'amitié durant ces 4 dernières années, avec qui j'ai grimpé, randonné, nagé, rampé, mangé, bu et rigolé... J'aimerais bien écrire tous leur nom et prénom mais ça prendrait trop de place. Et puis ce ne sont pas des prénoms français, alors je ne sais pas trop comment ça s'écrit. Des fois il y a même des lettres que je n'ai pas sur mon clavier alors bon... Des fois il y'en a même qui ont des prénoms de fruits dans une langue alors qui sont issus d'un pays qui ne parle pas cette langue. Bref compliqué tout ça, comme la biologie.

Cercle social indique des histoires d'amour, des déceptions mais pas que. Je remercie Léonie qui m'a soutenu durant ces quelques semaines, alors que j'avais l'impression que je n'arriverai jamais à écrire cette thèse. Qui m'a rappelé plus d'une fois l'importance d'être content du travail que l'on a fourni, aussi petit soit-il. Et si les grandes parties telles qu'Introduction sont si belles, c'est bien grâce à elle.

Je n'ai pas vraiment plus de temps. Maïté attend que je lui donne ma clef USB avec ma thèse dessus pour l'imprimer. Je la remercie elle aussi. Est-ce qu'elle a déjà été remercié dans une thèse ? Je l'espère, elle qui depuis 40 ans imprime et relie des thèses. A-t-elle déjà été citée ? Du coup, Marie-Therese PENEY, je te remercie d'avoir pris de ton temps pour imprimer ma thèse. Ainsi pour tes conseils et tes sourires.

Je remercie les membres des plateformes de culture cell, de mass-spec, de la laverie, d'anticorps, les femmes de ménage et j'en passe. Toutes ces petites mains qui s'affèrent dans l'ombre et sans lesquelles la science avancerait tellement moins vite.

Je remercie aussi mes stagiaires. Je remercie Cynthia qui m'a mis en évidence l'anticorps anti-TAF7 qui m'a permis de valider mes mutants par IF. Je remercie Lynda qui a permis de développer toute la partie SAGA de mon article.

Je remercie également mes vieux amis et ma famille, qui au cours des ces années, ont toujours cherché à savoir comment ma thèse avançait, à ma mère qui voulait savoir ce que j'avais découvert aujourd'hui, à mon père toujours intéressé de savoir comment avançait mon travail, à mes nièces et à mon neveu qui se demandaient quelle maladie j'allai soigner.

Je pense que je pourrais encore remercier beaucoup de monde. Mais il faut vraiment que j'imprime ma thèse maintenant, Maïté m'attend.

Bonne lecture

Vincent Hisler

Le 05/11/2021

*« La vie oscille, comme un pendu, de droite à gauche, de la souffrance à l'ennui »*

Schopenhauer

Evitez de vous focaliser sur le tic et le tac qui dure une fraction de seconde  
Profiter de la seconde de silence.

## Etude du rôle de TFIID et de ses complexes partiels dans l'initiation de la transcription

(Study of the role of TFIID and its partial complexes in transcription initiation)

### Résumé en français

L'ARN polymérase II (ARN pol II) transcrit les gènes codant pour les protéines et des ARNs non codants (Dergai and Hernandez, 2019). Son recrutement au niveau du promoteur nécessite l'assemblage de facteurs généraux de transcription (GTFs) dans un complexe de pré-initiation (PIC) (Chen et al., 2021). TFIID est le premier GTF qui se lie au promoteur. Par la suite, d'autres GTFs s'agrègent autour du TFIID et du promoteur permettant le recrutement de l'ARN pol II ce qui finalise la formation du PIC. TFIID est composé de la protéine de liaison à la boîte TATA (TBP) et de 13 différents facteurs associés à TBP (TAFs) (Patel et al., 2020). TFIID partage certaines sous-unités, comme TAF10, avec le coactivateur transcriptionnel Spt-Ada-Gcn5 Acétyltransférase (SAGA) (Herbst et al., 2021; Papai et al., 2020).

Le modèle actuel d'assemblage de TFIID a été établi grâce à des expériences *in vitro* telles que des co-immunofluorescences, des immunoprécipitations ou encore des analyses de complexes reconstitués par cryo-microscopie et par spectrométrie de masse (Bieniossek et al., 2013a; Gupta et al., 2017; Kamenova et al., 2019; Soutoglou et al., 2005; Trowitzsch et al., 2015). Dans ce modèle, le core-TFIID est d'abord formé, composé de TAF4-5-6-9-12. Le sous-module TAF2-8-10 rejoint le core-TFIID et brise sa symétrie axiale pour former un complexe 8-TAFs. Cette rupture de symétrie est importante pour permettre l'ancrage des dernières sous-unités (TBP, TAF1 et le sous-module TAF7-11-13). Le complexe TFIID alors assemblé et est composé de 3 lobes : A, B et C. Ces lobes présentent une certaine mobilité les uns par rapport aux autres permettant le placement au mieux de TBP sur le promoteur ainsi que le recrutement des autres GTFs en fonction des caractéristiques du promoteur (Chen et al., 2021; Patel et al., 2018).

Certains des TAFs ont des paralogues qui peuvent les remplacer dans TFIID en changeant sa composition finale. Cette variabilité participe à la mise en place d'un programme d'expression génétique spécifique un type cellulaire. C'est le cas par exemple de TAF4B, qui peut remplacer TAF4 dans TFIID, mais ne le compense pas totalement car les embryons *Taf4<sup>-/-</sup>* arrêtent de se développer vers 9.5jpc (Langer et al., 2016). En revanche, TAF4B est important pour l'expression de gènes nécessaire à l'ovogenèse et la spermatogenèse (Falender et al., 2005; Grive et al., 2016). TAF7L, un paralogue de TAF7, est spécifiquement impliqué dans la spermatogenèse et l'adipogenèse (Pointud, 2003; Zhou et al., 2013). *Taf7l2* est un possible autre paralogue de *Taf7* mais ce gène n'a pas encore été étudié.

Un autre niveau de variabilité dans la composition de TFIID repose sur l'absence de certains TAFs. Dans les cellules HeLa et les cellules F9, un complexe TFIID ne contenant pas TAF10 coexiste avec le complexe canonique (Jacq et al., 1994; Mohan II et al., 2003). Bien que tous les TAFs soient requis pour la survie et la différenciation cellulaire dans la plupart des cas, dans certains contextes cellulaires, la perte de TAF7, de TAF8 ou bien de TAF10 n'a pas d'effet (Bardot et al., 2017; Gegonne et al., 2012;

Indra et al., 2005; Voss et al., 2000), suggérant que la présence de complexes TFIID partiels est suffisante dans certaines situations pour la transcription. Ces découvertes soulèvent des questions sur la composition de ces complexes. De plus, leur rôle dans l'initiation de la transcription est encore inconnu. L'objectif de ma thèse est de mettre en évidence la composition et la possible fonction de ces complexes TFIID alternatifs dans l'initiation de la transcription.

Récemment, chez des embryons de souris mutants *T-Cre ; Taf10<sup>ff</sup>*, dans lesquels TAF10 est conditionnellement perdu dans le mésoderme embryonnaire, nous avons montré que la somitogenèse, un processus hautement régulé sur le plan transcriptionnel n'est pas affecté durant une courte fenêtre de temps. Cependant, la perte de TAF10 empêche l'assemblage de TFIID et de SAGA (Bardot et al., 2017). Ainsi, pour étudier le rôle spécifique de TFIID dans ce processus, des embryons mutants *T-Cre ; Taf7<sup>ff</sup>*, ont été produits. Chez ces mutants, nous avons observé que la somitogenèse n'est pas initialement affectée comme chez les mutants *T-Cre ; Taf10<sup>ff</sup>*. De manière inattendue, leurs phénotypes sont néanmoins légèrement différents. Par exemple, les mutants *T-Cre ; Taf7<sup>ff</sup>* se développe plus longtemps que les mutants *T-Cre ; Taf10<sup>ff</sup>*. Les mutants *T-Cre ; Taf7<sup>ff</sup>* présentent un défaut de développement cardiovasculaire et la morphogenèse de leurs membres antérieurs est initiée. Au contraire, les mutants *T-Cre ; Taf10<sup>ff</sup>* ne semblent pas présenter de problème de développement de la zone cardiovasculaire mais la morphogenèse de leurs membres antérieurs n'est pas du tout initiée. Cela pourrait indiquer qu'en l'absence de TAF7 ou TAF10, des complexes alternatifs ou partiels de TFIID sont soit nouvellement formés, soit déjà présents dans les cellules, et pourraient initier d'une manière ou d'une autre la transcription de l'ARN pol II dans ces contextes cellulaires particuliers.

Afin de caractériser les complexes partiels TFIID, j'ai dérivé des lignées de cellules souches embryonnaires de souris (mESC) dans lesquelles la délétion de *Taf7* (lignée cellulaire *RT7*), de *Taf10* (lignée cellulaire *RT10*) ou des deux (lignée cellulaire *RT710*) est induite par l'ajout de 4-hydroxytamoxifène (4-OHT) (Bardot et al., 2017). Après deux jours de traitement, la déplétion de TAF7 et/ou TAF10 est efficace dans toutes les lignées cellulaires. TAF7 et TAF10 sont essentiels aux mESC, ces dernières ne pouvant pas être maintenues en culture en leur absence. Cependant, à partir du jour 4, des différences en termes de survie cellulaire et de croissance des colonies apparaissent entre les différentes lignées cellulaires. Les cellules mutantes *RT710* présentent le phénotype le plus sévère, suivi des cellules mutantes *RT10* et enfin des cellules mutantes *RT7*. Il est intéressant de noter que les cellules mutantes *RT7* ne présentent pas d'augmentation de l'apoptose contrairement aux autres cellules mutantes, même après 6 jours de traitement. Ceci n'est pas dû à un phénomène de sauvetage du phénotype par l'expression ectopique d'un paralogue de *Taf7* car l'analyse par RT-qPCR a montré que ces cellules sauvages ou mutantes n'expriment ni *Taf7l* ni *Taf7l2*.

Une hypothèse qui pourrait expliquer ces différences de phénotypes est que TAF10 est également inclus dans SAGA et il a été montré que TAF10 est requis pour son assemblage (Bardot et al., 2017). Pour étudier cette question, nous avons muté le gène *Supt7l* dans la lignée cellulaire *RT7* par CRISPR-Cas9. SUPT7L est le partenaire spécifique de TAF10 dans SAGA (Herbst et al., 2021; Papai et al., 2020). Comme attendu, la réalisation d'immunoprécipitation suivi d'analyse par mass spectrométrie a montré

que la perte de SUPT7L empêche l'assemblage de SAGA de manière similaire à la perte de TAF10. De plus, il a été récemment publié qu'au sein de mESC *Supt7l*<sup>-/-</sup>, TAF10 ne peut pas être assemblé dans SAGA mais uniquement dans TFIID (Fischer et al., 2021). Par conséquent, la perte SUPT7L est un bon modèle permettant d'étudier l'impact de la perte de SAGA induit par la perte de TAF10 chez les mESC. La perte de SUPT7L n'a pas d'impact sur la survie des cellules ni sur la croissance des colonies. De plus, contrairement à la perte de TAF7 et/ou TAF10, la perte de SUPT7L n'empêche pas les cellules d'être maintenues en culture. Enfin, la perte supplémentaire de TAF7 dans cette lignée cellulaire *RT7* ; *Supt7l*<sup>-/-</sup> présente le même phénotype que les mutants *RT7*. Par conséquent, les différences de phénotype observées entre les mutants *RT7*, *RT10* et *RT710* ne peuvent pas être expliquées par la perte de SAGA induite par la perte de TAF10.

Une autre hypothèse pour expliquer ces différences de phénotypes est que la perte de TAF7 ou TAF10 n'altère pas l'assemblage du TFIID aux mêmes étapes. En particulier, TAF10 joue un rôle important dans la transition entre le core-TFIID et les étapes ultérieures de l'assemblage au cours desquelles TAF7 est recruté (Bieniossek et al., 2013). Il a déjà été montré qu'en absence de TAF7, un complexe TFIID presque complet peut être formé dans lequel seul TAF7 est absent (Gegonne et al., 2012). Ce dernier pourrait initier la transcription. En absence de TAF10, il est possible que seul le core-TFIID puisse être formé et ne puisse pas initier la transcription. Pour répondre à ces questions, différentes expériences d'immunoprécipitations de TAF12 suivies d'analyse par spectrométrie de masse à différents jours de traitement ont été réalisées. J'ai confirmé qu'en l'absence de TAF7, un TFIID dépourvu de TAF7 est formé (*TAF7-less TFIID*), ce qui correspond aux données publiées dans Gegonne et al., 2012. J'ai également montré qu'en l'absence de TAF10 le core-TFIID est formé sur lequel s'assemble au moins TAF1, nommé ci-après 7-TAFs, indiquant que des sous-unités de TFIID peuvent s'assembler au core-TFIID non réarrangé par TAF2-8-10. A noter que la perte quasi-totale de TFIID pour ces complexes partiels se fait à partir de trois jours de traitement, soit juste avant l'apparition des différences de phénotypes entre les différentes lignées cellulaires mutantes. Ensuite, j'ai montré par des immunoprécipitations de TBP suivies d'analyse par spectrométrie de masse que le *TAF7-less TFIID* peut recruter TBP, encore en accord avec l'étude de Gegonne et al., 2012. Ces résultats ont été confirmés par l'analyse de fractions obtenues par chromatographie d'exclusion de taille. TBP joue un rôle important dans l'initiation permettant de modifier la topologie de l'ADN et interagissant avec les autres GTFs (Chen et al., 2021). Nous aimerions par conséquent savoir si TBP peut être chargé sur le promoteur en absence de TAF7 comme il devrait l'être en conditions contrôles. Pour cela, des expériences d'immunoprécipitation de la chromatine via TBP suivie d'un séquençage (*TBP ChIP-seq*) seront réalisées prochainement.

Puisque la perte de TAF10 a un effet plus dramatique sur l'assemblage de TFIID que la perte de TAF7, nous proposons que la perte de TAF10 a plus d'impact sur la transcription que la perte de TAF7. Ceci pourrait expliquer les différences de phénotypes observés entre les mutants. Dans cette optique, nous avons d'abord voulu examiner le niveau global de transcription lorsque le TFIID canonique a presque complètement disparu. L'analyse par immunofluorescence des niveaux de phosphorylation de l'ARN pol II, RPB1<sup>pSer2</sup>, RPB1<sup>pSer5</sup>, indiquant l'état d'activité transcriptionnelle de l'ARN pol II, ainsi que

l'analyse de l'intégration du 5-éthynyl uridine (EU) après une courte exposition, indiquant le niveau de synthèse active des ARNs naissants, ont montré que l'ARN pol II transcrit toujours à un niveau normal alors que la quasi-totalité des complexes TFIID sont absents, que ce soit en présence du *Taf7-less TFIID* ou du 7-TAFs. Pour avoir une vision plus précise du transcriptome naissant dans ces contextes particuliers, nous avons réalisé un séquençage transitoire du transcriptome (*TT-seq* (Schwalb et al., 2016)) qui est en cours d'analyse.

En conclusion, ma thèse vise à mieux comprendre si les complexes TFIID partiels qui diffèrent du complexe TFIID peuvent initier la transcription, et si oui, est-ce important dans la mise en place ou au maintien d'un programme d'expression génique particulier. Mes résultats montrent que le complexe TFIID canonique est requis à la fois pendant le développement embryonnaire et pour le maintien en culture des mESC. Cependant, lorsque l'assemblage de TFIID est affecté, la sévérité du phénotype observé dépend à la fois du contexte cellulaire et des complexes TFIID partiels résultants. Dans le cas des mESC, le complexe *TAF7-less TFIID* semble être plus apte à compenser la perte de TFIID que le core-TFIID.

Il est important de garder à l'esprit que les mESC expriment également TBPL1, un paralogue de TBP impliqué dans la transcription de l'ARN pol II indépendamment de TFIID. TBPL1 initie la transcription principalement à partir de promoteurs ne contenant pas de boîte TATA et avec la collaboration de TFIIA (Martianov et al., 2016 ; Ohbayashi et al., 2003). Cependant, mes données ont montré que TBPL1 n'est pas associé ni avec TFIIA ni avec TFIID dans les mESC. L'importance de TBPL1 dans la transcription de l'ARN pol II reste à étudier plus précisément.

Bien qu'un phénotype cellulaire soit visible après la perte presque totale de TFIID, ceci n'affecte ni le niveau de phosphorylation de l'ARN polymérase II ni le taux de transcription active. Ainsi, dans les premiers instants de la perte de TFIID, la transcription n'est pas complètement abrogée mais pourrait donc être perturbée qu'aux niveaux de certains gènes particuliers et/ou à des étapes particulières de l'initiation de la transcription. L'analyse des résultats de *TBP ChIP-seq*, *TT-seq* permettra de mieux comprendre l'importance de TFIID dans l'initiation de la transcription et comment les complexes TFIID partiels peuvent ou non maintenir cette activité. Des analyses par *CAGE* (*Cap analysis of gene expression* (Shiraki et al., 2003)) permettant de connaître le site de départ de la transcription (*Transcription Start Site*, TSS) à l'échelle de transcriptome entier seront aussi réalisées. Ceci permettra de savoir si, dans ces différentes lignées cellulaires mutantes, la transcription est initiée aux endroits appropriés et par conséquent de savoir si le rôle fondamental de TFIID n'est pas dans l'initiation de la transcription, mais plus spécifiquement dans le fait de l'initier aux bons endroits.

## Summary in English

The RNA polymerase II (RNA pol II) transcribes protein-coding genes and some non-coding RNAs (Dergai and Hernandez, 2019). Its recruitment to the promoter requires the assembly of general transcription factors (GTFs) in a preinitiation complex (PIC) (Chen et al., 2021). TFIID is the first GTF to bind the promoter. Subsequently, other GTFs as well as the RNA pol II are recruited around the TFIID which finalizes the transcription initiation step. TFIID is composed of the TATA-binding protein (TBP) and 13 different TBP-associated factors (TAFs) (Patel et al., 2020), some of which are shared with other complexes, such as TAF10 also present in the Spt-Ada-Gcn5 Acetyltransferase (SAGA) transcriptional coactivator (Herbst et al., 2021; Papai et al., 2020).

The current model of TFIID assembly has been established through *in vitro* experiments such as co-immunofluorescence, immunoprecipitation and cryo-microscopy/mass spectrometry analysis of reconstituted complexes (Bieniossek et al., 2013b; Gupta et al., 2017; Kamenova et al., 2019; Soutoglou et al., 2005; Trowitzsch et al., 2015). In this model, the core-TFIID is first formed, consisting of TAF4-5-6-9-12. The TAF2-8-10 sub-module joins the core-TFIID and breaks its axial symmetry to form an 8TAFs complex. This symmetry breaking is important to allow the anchoring of the last subunits (TBP, TAF1 and the TAF7-11-13 sub-module). The assembled TFIID complex is composed of 3 lobes: A, B and C. These lobes have a certain degree of mobility in relation to each other, allowing for the optimal placement of TBP on the promoter as well as the recruitment of other GTFs according to the characteristics of the promoter (Chen et al., 2021; Patel et al., 2018).

Some of the TAFs have paralogs that can replace them in TFIID by changing its final composition. This variability would participate in the establishment of a cell type specific gene expression program. For instance, this is the case of TAF4B, which can replace TAF4 in TFIID, but does not fully compensate for it because *Taf4*<sup>-/-</sup> embryos stop developing at around E9.5 (Langer et al., 2016). In contrast, TAF4B is important for the expression of genes required for oogenesis and spermatogenesis (Falender et al., 2005; Grive et al., 2016). TAF7L, a paralogue of TAF7, is specifically involved in spermatogenesis and adipogenesis (Pointud, 2003; Zhou et al., 2013) but its integration into TFIID composition has to be confirmed. *Taf7l2* is another possible paralogue of *Taf7* but this gene has not yet been studied.

Another level of variability in TFIID composition is based on the absence of some TAFs. In HeLa cell and F9 cells, a TFIID complex that does not contain TAF10 coexists with the canonical one (Jacq et al., 1994; Mohan II et al., 2003). Although all TAFs are required for cell survival and differentiation in most cases, in some cellular contexts the loss of TAF7, TAF8 or TAF10 has no effect (Bardot et al., 2017; Gegonne et al., 2012; Indra et al., 2005) suggesting that the presence of partial TFIID complexes is sufficient in some situations for transcription. These discoveries raise questions about the composition of these complexes. Moreover, their role in transcription initiation is still unknown. The goal of my PhD

is to highlight the composition and potential function of these partial TFIID complexes in transcription initiation.

Recently, in *T-Cre; Taf10<sup>ff</sup>* mutant mouse embryos, in which TAF10 is conditionally lost in the embryonic mesoderm, we have shown that somitogenesis, a highly transcriptionally regulated process, is not affected during a short time window. However, the loss of TAF10 prevents the assembly of both TFIID and SAGA (Bardot et al., 2017). Thus, to study the specific role of TFIID in this process, I generated *T-Cre; Taf7<sup>ff</sup>* mutant embryos in which the TFIID-specific subunit TAF7 is depleted in a similar manner. In these mutants we observed that somitogenesis is not affected during a time window as for *T-Cre; Taf10<sup>ff</sup>* mutants. Unexpectedly, their phenotypes are slightly different. For instance, the *T-Cre; Taf7<sup>ff</sup>* mutants live longer than the *Taf10* mutants. The *T-Cre; Taf7<sup>ff</sup>* mutants show a defect in cardiovascular development and the morphogenesis of their forelimbs is initiated. In contrast, the *T-Cre; Taf10<sup>ff</sup>* mutants do not appear to have a problem with the development of the cardiovascular zone but the morphogenesis of their forelimbs is not initiated at all.

In order to characterize TFIID partial complexes, I derived mouse embryonic stem cell (mESC) lines in which the deletion of *Taf7* (*RT7* cell line), *Taf10* (*RT10* cell line) or both (*RT710* cell line) is induced by 4-hydroxytamoxifen (4-OHT) (Bardot et al., 2017). After two days of treatment, the depletion of either TAF7 and/or TAF10 is efficient in all the cell lines. TAF7 and TAF10 are essential for mESCs since they cannot be maintained in culture in their absence. From day 4, differences in term of cell survival and colony growth appear between the different cell lines. The *RT710* mutant presents the most severe phenotype, followed by the *RT10* mutant and finally the *RT7* mutant. Interestingly, the *RT7* mutant does not show an increase in apoptosis unlike the other mutants, even after 6 days of treatment. This is not due to phenotype rescue by ectopic expression of a TAF7 paralogue since RT-qPCR analysis showed that neither *Taf7l* nor *Taf7l2* are expressed in these wild type or mutant cells.

One hypothesis that could explain these differences of phenotypes is that TAF10 is also required for SAGA assembly (Bardot et al., 2017). To study this question, we deleted *Supt7l* gene in *RT7* cell line by CRISPR-Cas9. SUPT7L is the SAGA-specific partner of TAF10 (Herbst et al., 2021; Papai et al., 2020). As expected, the realization of immunoprecipitation followed by mass spectrometry analysis showed that the loss of SUPT7L prevents the assembly of SAGA in a similar way as the loss of TAF10. Furthermore, it has recently been published that within mESC *Supt7l*<sup>-/-</sup>, TAF10 cannot be assembled into SAGA but only into TFIID (Fischer et al., 2021). Therefore, SUPT7L loss is a good model to study the impact of SAGA loss induced by TAF10 loss in mESCs. The loss of SUPT7L does not impact either the cell survival nor the colony growth. Moreover, unlike the loss of TAF7 and/or TAF10, *Supt7l*<sup>-/-</sup> mESC can be maintained in culture. Finally, the additional depletion of TAF7 in this *RT7; Supt7l*<sup>-/-</sup> cell line shows the same phenotype as *RT7* mutants. Therefore, the differences of phenotype observed between *RT7*, *RT10* et *RT710* cannot be explained by the loss of SAGA induced by TAF10 loss.



Another hypothesis to explain these differences in phenotypes is that the loss of TAF7 or TAF10 does not impair TFIID assembly at the same step. In particular, TAF10 plays an important role in the transition from the core-TFIID to the later assembly steps in which TAF7 is recruited (Bieniossek et al., 2013). It has already been shown that in the absence of TAF7, a nearly complete TFIID complex can be formed in which only TAF7 is absent (Gegonne et al., 2012). The latter could initiate transcription. In the absence of TAF10, it is possible that only the core-TFIID can be formed and cannot initiate transcription. To answer these questions, different experiments of TAF12 immunoprecipitation followed by mass spectrometry analysis at different days of treatment were performed. I confirmed that in the absence of TAF7, a TAF7-less TFIID is formed. I also showed that in the absence of TAF10 the core-TFIID is formed on which at least TAF1 assembles, hereafter referred to as 7-TAFs, indicating that subunits of TFIID can assemble to the core-TFIID not rearranged by TAF2-8-10. Note that the almost complete loss of TFIID for these partial complexes occurs after three days of treatment, i.e. just before the appearance of phenotype differences between the different mutant cell lines. Then, I showed by TBP immunoprecipitation followed by mass spectrometry analysis that TAF7-less TFIID can recruit TBP, in agreement with the study of Gegonne et al., 2012. These results were confirmed by analysis of fractions obtained by gel filtration. TBP plays an important role in initiation by modifying DNA topology and by interacting with other GTFs (Chen et al., 2021). We would therefore like to know if TBP can be loaded onto the promoter in the absence of TAF7 as it should be under control conditions. For this purpose, chromatin immunoprecipitation experiments via TBP followed by sequencing (TBP ChIP-seq) will be performed in the near future.

Since the loss of TAF10 has a more dramatic effect on TFIID assembly than the loss of TAF7, we hypothesize that the loss of TAF10 impacts more the transcription than the loss of TAF7. This could also explain the differences in phenotypes observed between the mutants. With that in mind, we first wanted to look at the global level of transcription in almost complete absence of canonical TFIID. Analysis of the phosphorylation levels of RNA pol II, RPB1<sup>pSer2</sup>, RPB1<sup>pSer5</sup>, which indicate the state of RNA pol II transcription, as well as analysis of the integration of 5-ethynyl uridine (EU) after a short exposure, indicating the level of RNA synthesis, showed that RNA polymerase II still transcribes at a normal level in almost complete absence of canonical TFIID, whether in the presence of TAF7-less TFIID or 7-TAFs. To have a more precise view of the nascent transcriptome in these particular contexts, we did a transient transcriptome sequencing (TT-seq (Schwalb et al., 2016)) which is currently being analyzed.

In conclusion, my thesis aims to better understand whether TFIID partial complexes that differ from the TFIID complex can initiate transcription, and if so, if this is important in setting up a particular gene expression program. My results show that the canonical TFIID complex is required both during embryonic development and in mESC maintenance. However, when TFIID assembly is affected, the severity of the phenotypes depends on both the cellular context and the resulting partial TFIID



complexes. In the case of mESC, the TAF7-less TFIID complex appears to be more able to compensate for the loss of TFIID than the 7-TAFs.

The compensation achieved seems to be more fine-grained than just maintaining the overall level of transcription. Indeed, although a cellular phenotype is visible after almost total loss of TFIID, this does not affect the level of RNA polymerase II phosphorylation or the transcription rate. Analysis of the results of TBP ChIP-seq, TT-seq and SLIC-CAGE will provide a better understanding of the importance of TFIID and how TFIID partial complex may or may not maintain this activity.

It is important to keep in mind that mESCs also express TBPL1, a paralogue of TBP involved in TFIID-independent RNA pol II transcription. TBPL1 initiates transcription primarily from promoters lacking a TATA box and with the collaboration of TFIIA (Martianov et al., 2016; Ohbayashi et al., 2003). However, my preliminary results from gel filtration analysis showed that TBPL1 is not associated with either TFIIA or TFIID in mESCs. The importance of TBPL1 in RNA pol II transcription remains to be further investigated.

Finally, although a cellular phenotype is visible after almost complete loss of TFIID, this does not affect the level of phosphorylation of RNA polymerase II or the level of active transcription. Thus, in the early stages of TFIID loss, transcription is not completely abrogated but may only be disrupted at the levels of particular genes and/or at particular stages of transcription initiation. Analysis of the TBP ChIP-seq, TT-seq data will provide a better understanding of the importance of TFIID in transcription initiation and how partial TFIID complexes may or may not maintain this activity. Cap Analysis of Gene Expression (CAGE) (Shiraki et al., 2003) will also be performed to determine the transcription start site at the level of the whole transcriptome. This will reveal whether, in these different mutant cell lines, transcription is initiated at the appropriate sites, and therefore whether the fundamental role of TFIID is not to initiate transcription but to initiate it specifically at the right site.



# Sommaire

Introduction .....	5
I. L'ADN contient l'information génétique : description générale de sa composition, de sa structure et de son organisation.....	7
1. L'ADN est un duplex de polymères nucléotidiques formant une double hélice.....	7
2. Dans le noyau, l'ADN est présent au sein de la chromatine.....	12
3. La chromatine interagit avec un ensemble de protéines dont le recrutement est régulé.....	17
II. La transcription au sein du noyau est réalisée par trois ARN polymérases différentes.....	23
1. Transcription médiée par l'ARN pol I.....	24
2. Transcription médiée par l'ARN pol II.....	25
3. Transcription médiée par l'ARN pol III .....	28
III. Mécanismes de la transcription médiée par l'ARN pol II .....	29
1. Promoteur minimum ( <i>core promoter</i> ) et séquence activatrice .....	29
2. Accessibilité des promoteurs .....	31
3. Mise en évidence de l'importance des GTFs dans la transcription médiée par l'ARN pol I, II et III. ....	33
4. Promoteurs et GTFs contrôlent la spécificité de l'initiation de la transcription dépendant de l'ARN pol II .....	41
5. Impact du <i>RNA buffering</i> sur l'étude des complexes impliqués dans la transcription.....	44
IV. Description du ' <i>Transcription Factor IID</i> ' (TFIID) : son évolution, sa composition et son rôle dans la transcription .....	47
1. Récapitulatif brève de la découverte de TFIID.....	47
2. Histoire évolutive de TFIID.....	48
3. Structure de TFIID .....	50
4. Assemblage de TFIID .....	53
5. TFIID et SAGA proche parenté .....	54
6. Rôle de TFIID dans l'initiation de la transcription :.....	59
7. Importance fonctionnelle des TAFs.....	64

8.	Existence de complexes TFIID partiels fonctionnels dans la transcription. ....	67
9.	Le chargement de TBP au promoteur n'est pas toujours requis pour initier la transcription : rôle des paralogues à TBP dans l'initiation de la transcription.....	68
V.	But de la thèse .....	75
<b>Résultats .....</b>		<b>77</b>
I.	Article.....	79
1.	Abstract .....	85
2.	Introduction.....	86
3.	Results.....	90
4.	Discussion .....	111
5.	Supplementary figures .....	119
6.	References article.....	125
7.	Material & Method: .....	133
II.	Validation de l'anticorps anti-TAF12 (22TA-2A1) pour l'étude de l'assemblage de TFIID par expérience d'immunoprécipitation. ....	147
III.	Etude de l'importance de TAF7 dans l'assemblage de TFIID.....	152
IV.	Etude de l'importance de TAF7 et TAF10 dans le recrutement des sous-unités de TFIID à la chromatine.....	154
V.	Etude de l'importance de TAF7 et TAF10 dans la transcription médiée par l'ARN pol II.	155
VI.	Potentiel rôle d'un complexe TAF10 dans le recrutement de TBPL1 aux promoteurs.....	158
VII.	Matériels et méthodes. ....	161
1.	Réalisation d'extrait de chromatine .....	161
2.	Protocol TT-seq.....	161
<b>Discussion .....</b>		<b>173</b>
I.	La réorganisation du core-TFIID par le module TAF2-8-10 n'est pas une étape requise pour le recrutement des autres TAFs. ....	175
II.	TAF10 joue un rôle important dans l'assemblage de TFIID en stabilisant ses partenaires.	177
III.	TAF7 est impliqué dans le recrutement de TAF1 à TFIID.....	179

IV. TAF1 est requis pour le recrutement optimal de TBP sur TFIID : possible mise en évidence d'un complexe TFTC contenant l'ensemble des sous-unités de TFIID à l'exception de TAF1 et TBP.	179
V. Potentielle mise à jour du modèle d'assemblage de TFIID.	180
VI. La transcription dépendante de l'ARN pol II est maintenue en absence de TFIID	182
VII. La double déplétion TAF7 et TAF10 affecte la transcription médiée par l'ARN pol II : de possibles effets indirects qui demandent de choisir le bon jour de traitement pour les analyses transcriptomiques.	182
VIII. Rôle fonctionnel de TAF7 dans la régulation du couplage des premières étapes de la transcription : mise en place du PIC, initiation et élongation de la transcription.	183
IX. Rôle fonctionnel de TFIID dans l'activation et la répression de la transcription	185
X. L'induction de la dégradation des sous-unités de TFIID régule la mise en place des programmes d'expressions génétiques.	188
XI. Potentielle mise à jour du modèle expliquant l'importance de TFIID dans la transcription.	189
XII. Expériences pour l'étude de l'importance fonctionnelle de TFIID, et de ses complexes partiels, dans la mise en place et l'expression d'un programme d'expression génétique particulier	192
XIII. Les sous-unités TAFs sont partagées avec d'autres complexes que TFIID ou peuvent avoir des fonctions indépendantes à TFIID, complexifiant l'étude de la fonction TFIID	193
1. La perte de TAF10 perturbe l'assemblage de SAGA : quelle est l'impact de la perte de SAGA, par rapport à la perte de TFIID, sur la transcription ?	193
2. TAF10 est potentiellement dans un autre complexe contenant TBPL1	194
3. Possible présence et rôle du complexe TFTC dans la transcription	195
4. TAF7 est impliqué dans des fonctions indépendantes à TFIID	195
XIV. Conclusion	195
Références	197
Références	198
Annexes	217



---

# Introduction

---

---

# Introduction

---



## I. L'ADN contient l'information génétique : description générale de sa composition, de sa structure et de son organisation

L'acide désoxyribonucléique (ADN) contient l'information génétique, c'est-à-dire l'ensemble des instructions nécessaires à la croissance d'un organisme, à sa réaction face aux signaux provenant de l'environnement et à sa reproduction. Dans ce chapitre, nous détaillerons la composition biochimique, la structure et l'architecture de l'ADN.

### 1. L'ADN est un duplex de polymères nucléotidiques formant une double hélice.

#### *i. Un acide nucléique est un polymère de nucléotide.*

Deux types d'acide nucléique coexistent dans la cellule, les ADN (acide désoxyribonucléique) et les ARN (acide ribonucléique). Ils sont tous les deux constitués de nucléotides contenant un groupement phosphate, un pentose et une base azotée (Figure 1A). Néanmoins, l'ADN est composé de désoxyribonucléotides dont le pentose est un désoxyribose alors que l'ARN est composé de ribonucléotides dont le pentose est un ribose (Figure 1B). De plus, ces deux acides nucléiques se différencient par leur composition en bases azotées. Bien que l'adénine (A), la guanine (G) et la cytosine (C) sont retrouvées au niveau de l'ADN et de l'ARN, la thymine (T) est spécifique de l'ADN et l'uracile (U) de l'ARN (Figure 1C).

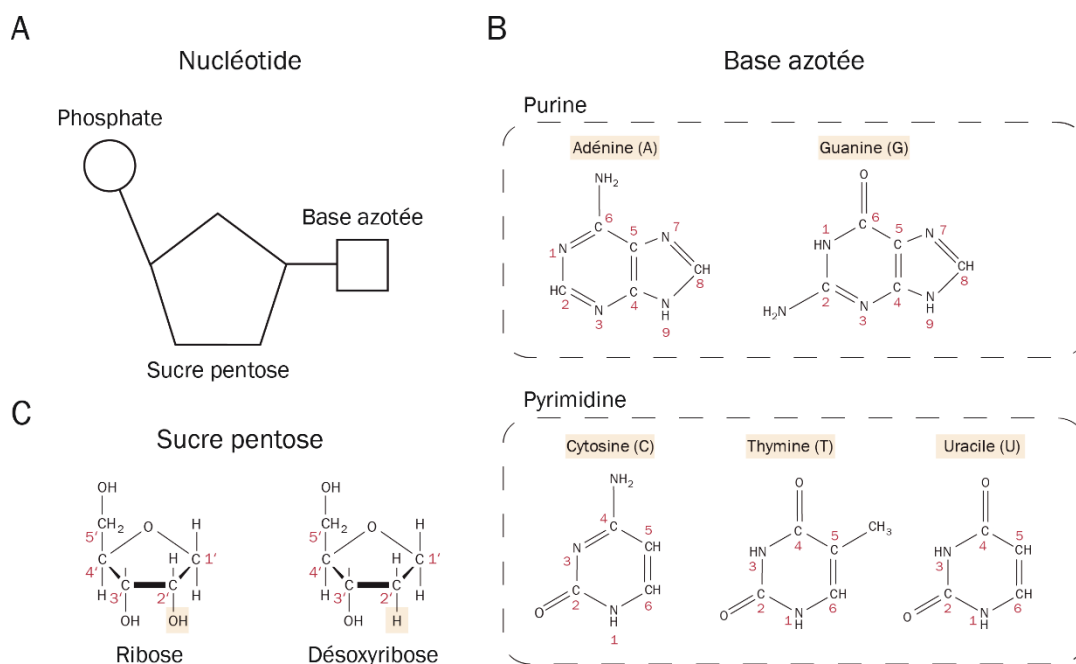


Figure 1 : Schéma du nucléotide : brique élémentaire des acides nucléiques. A) Le nucléotide est composé d'un groupe phosphate, d'un sucre pentose et d'une base azotée. B) Schéma des différentes bases azotées. C) Schéma des formes riboses et désoxyriboses du sucre pentose. Adapté à partir de Human Molecular Genetics – 4th edition.

Les nucléotides présent au sein de l'ADN ou de l'ARN peuvent être modifiés par l'ajout de nouveaux groupements chimiques au niveau des bases azotées ou des pentoses (Roundtree et al., 2017). Les modifications des nucléotides au sein des ARNs sont bien plus diversifiées que ceux présents au niveau de l'ADN. Parmi les modifications connues au sein de l'ADN, il existe la méthylation des C en 5-méthylcytosines (5mC) et des A en N6-méthyladénines (6mA). Les 5mC peuvent par la suite être oxydées en 5-hydroxyméthylcytosines, 5-formylcytosines et 5-carboxycytosines (Wu and Zhang, 2017). Récemment, il a été aussi montré que les 6mA peuvent être oxydées en N6-hydroxyméthyladénines (Xiong et al., 2019) (Figure 2).

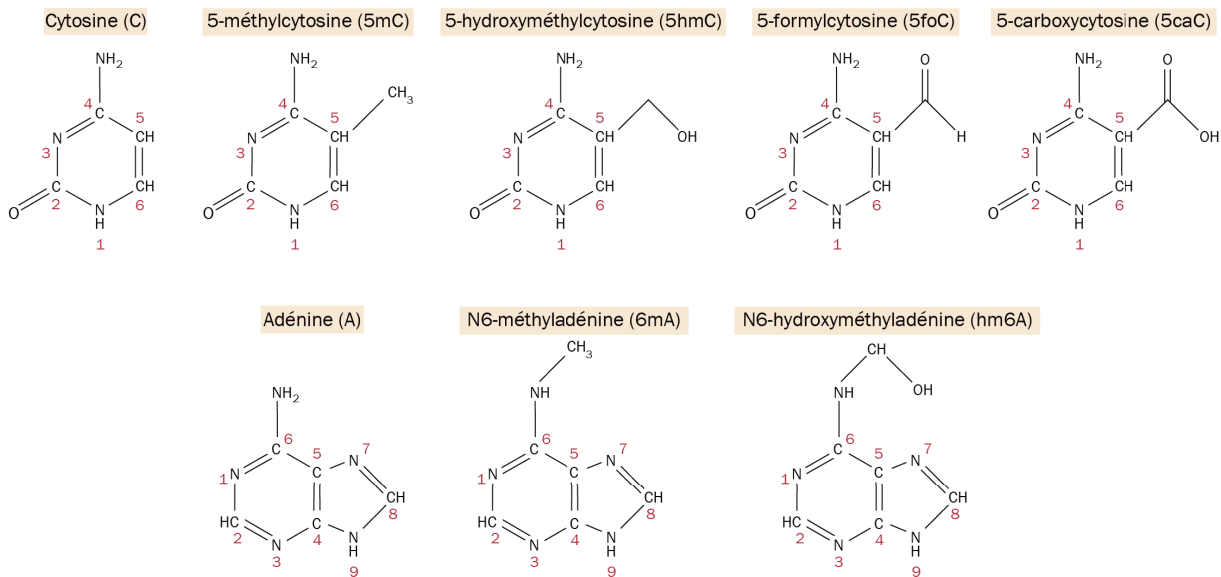


Figure 2 : Schéma de quelques nucléotides modifiés. Adapté à partir de *Human Molecular Genetics – 4th edition* et de Xiong et al., 2019.

Au sein des ARNs et des ADNs, les nucléotides sont liés les uns aux autres par des liaisons 3'-5'-phosphodiester (Figure 3A). Le squelette carboné correspond à l'ensemble des pentoses reliés par leur liaison 3'-5'-phosphodiester. L'extrémité 5' d'un acide nucléique correspond à l'extrémité du polymère où le phosphate présent en C'5 du sucre pentose est libre. L'extrémité 3' correspond à l'extrémité du polymère où le groupement alcool présent en C'3 est libre (Figure 3A).

## *ii. L'ADN est organisé sous la forme d'une double hélice*

L'ADN se présente sous la forme d'un duplex de chaînes polydésoxyribonucléotidiques, appelées « brin ». Ces brins sont appariés entre eux par des liaisons hydrogènes selon le modèle de l'appariement des bases mis en évidence par Watson et Crick (Watson and Crick, 1953). Les paires A-T forment 2 liaisons hydrogènes alors que les paires G-C forment 3 liaisons hydrogènes (Figure 3B).

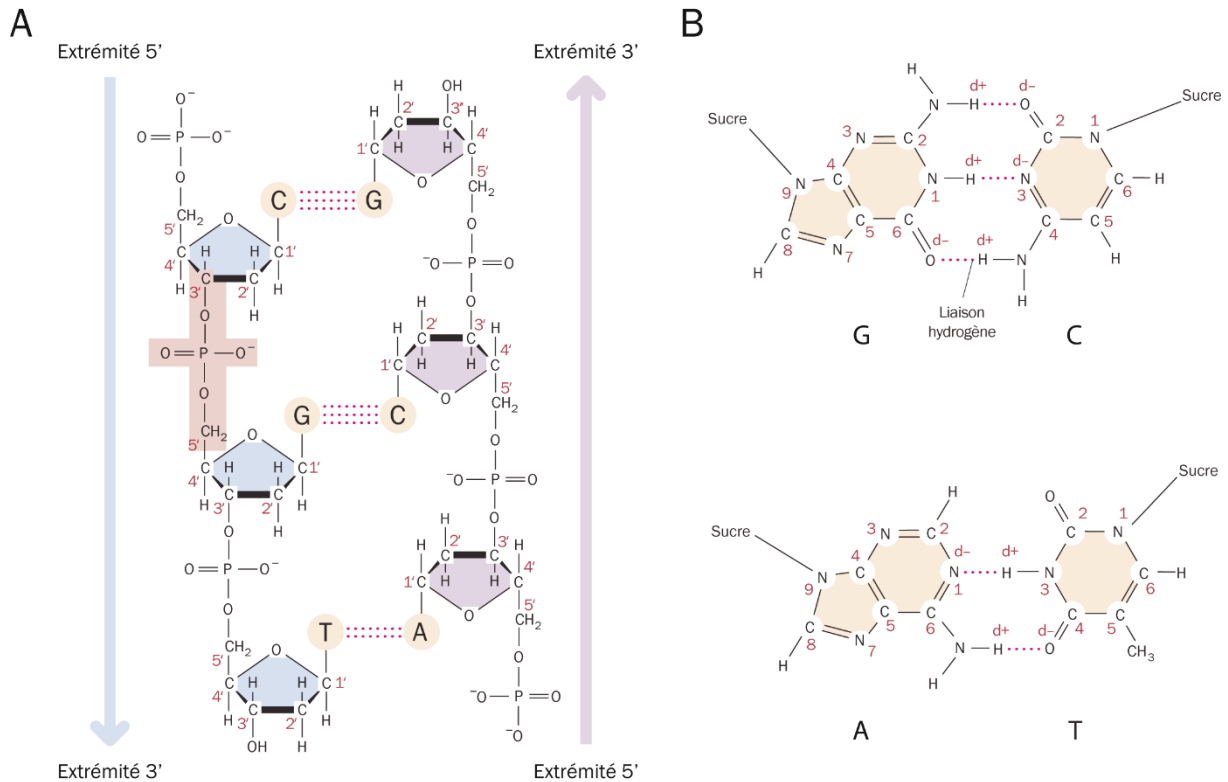


Figure 3 : L'ADN est un polymère de nucléotides pouvant s'apparier via les appariements de bases azotées. A) Schéma de deux brins d'ADN appariés l'une à l'autre. d+ donneur de liaisons hydrogènes, d- accepteur de liaisons hydrogènes. La croix rose correspond à une liaison 3'-5' phosphodiester. B) Schéma d'appariement des nucléotides dit de « Watson-Crick ». Adapté à partir de *Human Molecular Genetics – 4th edition*.

Plusieurs conséquences découlent de cet appariement.

- Premièrement, les deux brins formant un ADN double brin ne sont pas identiques mais complémentaires : chaque A se trouvant en face d'un T, chaque C devant un G et vice versa. Une paire de base (pb) correspond à deux nucléotides appariés
- Deuxièmement, les deux brins sont antiparallèles : un brin va dans le sens 5' vers 3' tandis que l'autre va dans le sens 3' vers 5' (Figure 3A).
- Troisièmement, cet appariement spécifique donne une structure en double hélice à l'ADN : chacun des brins tournant autour de l'autre, formant deux hélices. L'appariement des bases se produit à l'intérieur de l'hélice tandis que le squelette carboné est présent à la périphérie de l'hélice. Entre chaque hélice se trouvent deux types de sillons : l'un plus étroit et moins profond qui est le sillon mineur et l'autre plus large et plus profond qui est le sillon majeur (Figure 4B).

### iii. Polymorphisme de la double hélice de l'ADN et structure secondaire

La structure la plus connue, et pensée comme la plus physiologique, de la double hélice est la forme B (ADN-B). Cette structure a été élucidée par James Watson et Francis Crick sur la base de la découverte

de Rosalind Franklin, d'Erwin Chargaff et Maurice Wilkins (Cox et al., 2015; Franklin, 1953; Waston and Crick, 1953). D'autres formes d'hélices coexistent avec la forme B, chacune d'entre elle se caractérisant entre autres par le sens de l'hélice, son diamètre ou encore le nombre de nucléotides par tour d'hélice (Figure 4 A, B) (Harteis and Schneider, 2014).

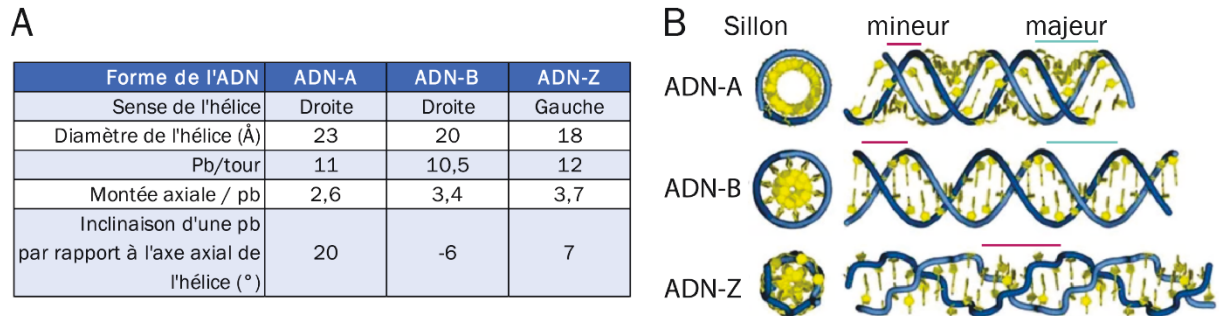


Figure 4 : Les différentes structures de la double hélice d'ADN. A) Quelques caractéristiques de chacune des formes A (ADN-A), B (ADN-B) et Z (ADN-Z) de la double hélice d'ADN avec leur représentation structurale en (B). Les bases azotées sont en jaune et le squelette carboné en bleu. Le trait horizontal violet montre le sillon mineur, celui turquoise le sillon majeur. Adapté de Molecular Biology – Principles and Practice – Second édition et de microbenotes.com.

Par exemple, la forme A (ADN-A) se caractérise par un sillon mineur plus profond et plus étroit et un sillon mineur plus large et moins profond. La forme Z (ADN-Z) se caractérise par un enroulement de la double hélice par la gauche.

Plusieurs facteurs sont à l'origine de la forme de la double hélice :

- 1) La séquence nucléotidique : il a été montré *in vitro* que les séquences ADN riches en G et en C adoptent préférentiellement la forme A (Koudelka et al., 2006; Rohs et al., 2010).
- 2) Les processus moléculaires se déroulant sur l'ADN : la transcription (voir introduction II & III) entraîne le super enroulement négatif de l'ADN-B, ce qui ferait basculer la topologie de la double hélice de la forme B à la forme Z (Shin et al., 2016).
- 3) La modification chimique des nucléotides : des simulations informatiques ont suggéré que la méthylation des C influencerait la largeur des sillons de l'ADN-B (Rao et al., 2018). D'autre part il a été montré qu'*in vitro* les polymères de nucléotides méthylés au C adoptent plus facilement une forme Z qu'une forme B (Behe and Felsenfeld, 1981).

En plus de la forme en double hélice, l'ADN peut se replier sur elle-même formant des structures secondaires. Dans ces cas particuliers, les bases ne s'apparient plus selon les principes définis par Watson et Crick. L'appariement se fait soit selon les principes définis par Hoogsteen (Nikolova et al., 2013), où A-T et C-G s'apparient via d'autres liaisons hydrogènes. Soit il se fait via des liaisons hydrogènes permettant l'appariement des G et des C entre eux (Figure 5A). Bien que plusieurs structures secondaires de l'ADN ont été suggérées, leurs existences *in vivo* est difficile à mettre en évidence, pour des raisons techniques (Poggi and Richard, 2021). La présence *in vivo* des motif G-quadruplex (G4) et des motifs I est relativement bien établie grâce à de récentes études (Poggi and Richard, 2021; Tang et al., 2020; Zeraati et al., 2018)

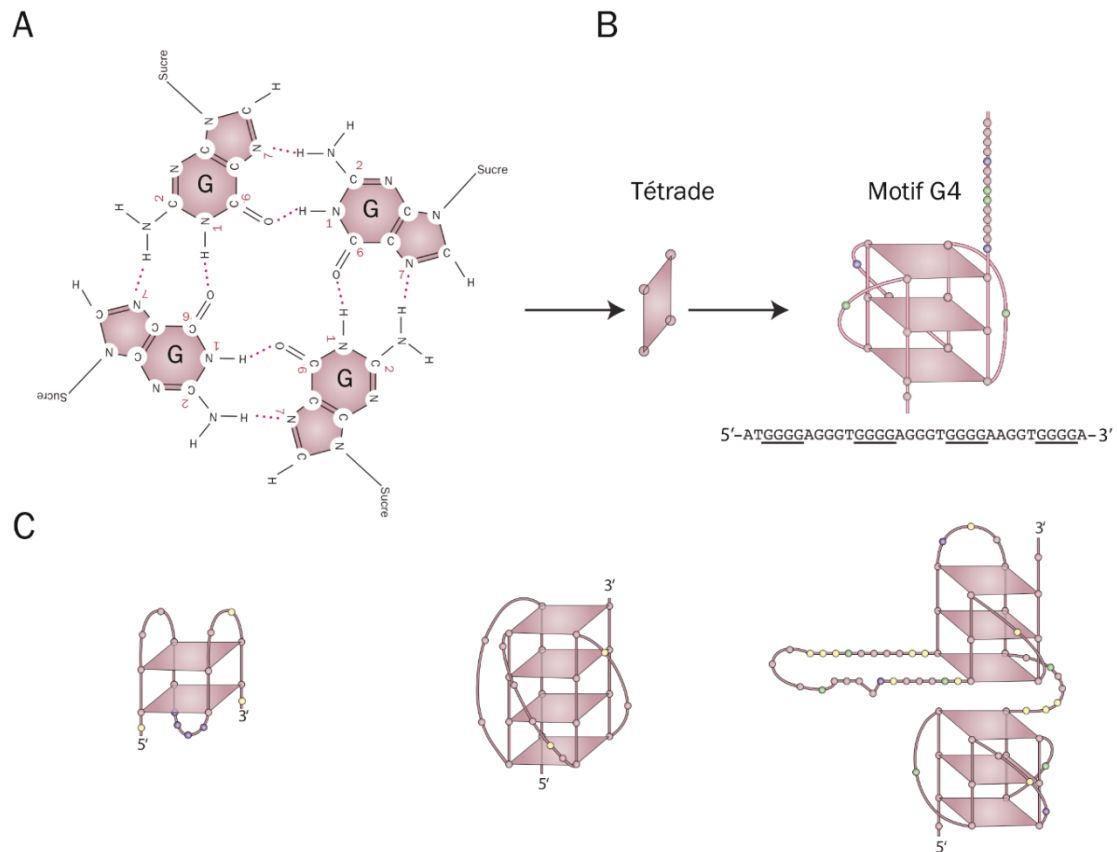


Figure 5 : Les motifs G4 comme exemple de structure secondaire de l'ADN. A) Possible appariement des guanines entre elles. B) Lorsque des séquences GpGpGpG sont suffisamment espacées, les guanines de chaque séquence peuvent s'apparier formant une tétrade. La superposition des tétrades forme un motif G4. C) Structure de quelques G4 supposés. La structure d'un G4 à l'autre varie en fonction de la séquence dont ils sont issus. Adapté de Brooks and Hurley, 2009.

Les motifs G4 se forment au niveau de régions riches en CG sous contrainte physique médiée par des processus moléculaires tel que la transcription (Brooks and Hurley, 2009). Cette structure est composée d'un empilement de deux ou plusieurs G-tétrades, reliées entre elles par la présence de boucles (Figure 5B). Chaque G-tétrade est constituée de quatre guanines dans lesquelles chaque guanine interagit avec les autres guanines adjacentes. Ces motifs G4 présentent des structures très différentes en fonction du nombre de G-tétrades empilées et de la longueur et de la composition des séquences liant les G-tétrades (Figure 5C). Les motifs I se forment sur le brin complémentaire au brin où se forme le motif G4 via un empilement de paires C-C appariées l'une à l'autre (Brooks and Hurley, 2009). La formation d'un motif G4 stabilise et/ou conduit à la formation du motif I sur le brin complémentaire, et réciproquement (Dhakal et al., 2012 ; Sutherland et al., 2016).

## 2. Dans le noyau, l'ADN est présent au sein de la chromatine

### *i. Description de l'ADN génomique*

Chez les eucaryotes, l'ADN est présent dans différents organites : dans le noyau, dans les plastes, dans les mitochondries et dans d'autres organites moins connus comme le nucléomorphe (Anisimova, 2019). Dans le noyau, l'ADN génomique (ADNg) correspond à plusieurs molécules d'ADN non circulaires, chacune d'entre elles étant appelée chromosome. Chaque espèce se caractérise par un nombre spécifique de chromosomes : la levure *Saccharomyces cerevisiae* possède 16 paires de chromosomes (dans son état diploïde) tandis que la souris possède 20 paires de chromosomes et l'homme possède 23 paires de chromosomes. Le nombre de chromosomes peut varier d'un type cellulaire (les gamètes sont haploïdes, les cellules musculaires contiennent plusieurs noyaux) à l'autre mais également entre les différentes étapes du cycle cellulaire.

Dans le noyau, chaque chromosome occupe un territoire défini appelé territoire chromosomique (TC) (Cremer et al., 2006) (Figure 6b). Dans chaque TC, l'ADN est organisé sous forme de chromatine (voir ci-dessous) présentant différent niveau de compaction : faiblement condensée on parle d'euchromatine (compartiment A), fortement condensée on parle d'hétérochromatine (compartiment B) (Figure 6a, c). Pour finir, des séquences d'ADN non consécutives peuvent se retrouver rapprochées dans l'espace formant des domaines topologiquement associés (Ernst et al., 2011) (Figure 6d).

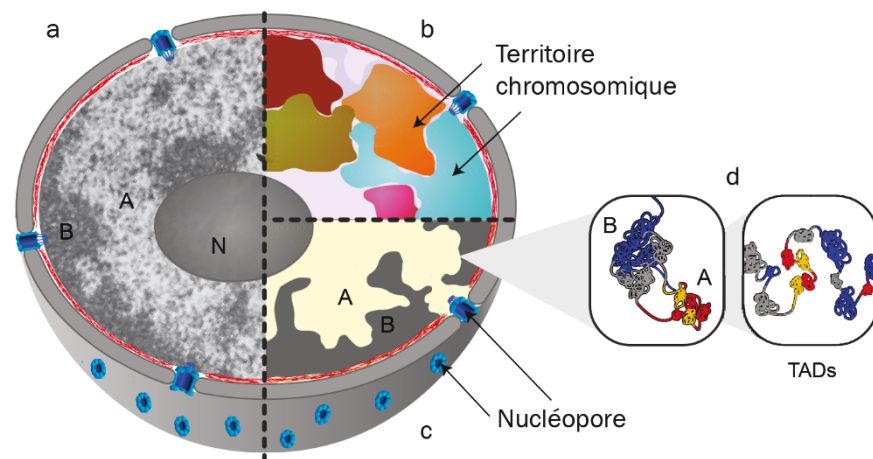


Figure 6 : Organisation du noyau : il est délimité par une double membrane de phospholipide (gris) dans laquelle les nucléopores sont insérés (bleu), régulant les transports des molécules entre le noyau et le cytoplasme. Les territoires chromosomiques (b) sont occupés par des chromosomes distincts. Dans le noyau la chromatine se trouve sous forme d'hétérochromatine (B) ou d'euchromatine (A) (a-c). (a) est une photographie au microscope électronique d'un noyau de cellule gliale de rat. A noter que dans certain type cellulaire, l'hétérochromatine est principalement condensée au centre du noyau. d) Schématisation d'un TAD. Adapté de Poeschel R, et al. 2016

### *ii. L'ADN génomique est présent sous forme de chromatine*

La chromatine est une structure retrouvée au sein du noyau des cellules eucaryotes. Celle-ci est constituée d'une suite de nucléosomes liés les uns aux autres. Chaque nucléosome correspond à une

séquence d'ADN entourant un octamère d'histone. L'ADN au sein du nucléosome est nommé ADN nucléosomal et l'ADN entre deux nucléosomes est nommé ADN linker. La taille de cette ADN linker varie de quelques pb à une centaine de pb selon la localisation du génome et l'espèce (Chen et al., 2010).

#### 1) Structure du nucléosome

La première structure radiographique à haute résolution du nucléosome a été obtenue par Luger et al. en 1997 (Luger et al., 1997). Autour d'un octamère d'histone sont enroulés environ 145-147 pb d'ADN formant une super hélice plate s'enroulant sur la gauche. L'octamère d'histone est formé de deux histones H2A, deux H2B, deux H3 et deux H4. Ces histones sont caractérisées par leur domaine de repliement d'histone (*histone fold domain, HFD*) qui correspond à une hélice  $\alpha$  courte N-terminale (Nterm) suivie d'une hélice  $\alpha$  longue et enfin d'une hélice  $\alpha$  courte C-terminale (Cterm) séparées les unes des autres par des boucles (Figure 7A). Ces HFD permettent la formation des hétérodimères H3-H4 et H2A-H2B. H3 et H2B présentent une quatrième hélice  $\alpha$  courte à leur extrémité N-terminale et C-terminale respectivement. La quatrième hélice  $\alpha$  de H3 permet aux deux hétérodimères H3-H4 d'interagir ensemble et former un tétramère. La quatrième hélice  $\alpha$  de H2B de chaque dimère H2A-H2B interagit avec l'un des H4 présents dans le tétramère. L'octamère assemblé est stabilisé par d'autres interactions histone-histone. L'ADN s'enroule autour de l'octamère d'histone par le biais de nombreuses interactions telles que des contacts électrostatiques et non polaires, ainsi que des liaisons hydrogènes (Figure 7B). Cela se produit principalement entre d'un côté les HFD et de l'autre côté le squelette carboné ainsi que le sillon mineur de l'ADN. Selon la séquence nucléotidique, certaines interactions peuvent également se produire à l'intérieur du sillon majeur. Cela indique que la séquence nucléotidique peut avoir un impact sur la formation et la stabilité des nucléosomes. De part et d'autre de leur HFD, les histones présentent une queue N-terminale et une queue C-terminale, qui est plus ou moins longue selon l'histone. Les queues d'histone sortent des nucléosomes soit par un canal formé par le sillon mineur, dans le cas de la queue H3, soit sur le côté de l'octamère dans le cas de la queue H4. Vu de face, le nucléosome présente une symétrie axiale, au niveau de l'axe dit dyadique (*dyad axis*) où se trouvent ~70 pb de part et d'autre (Luger et al., 1997).



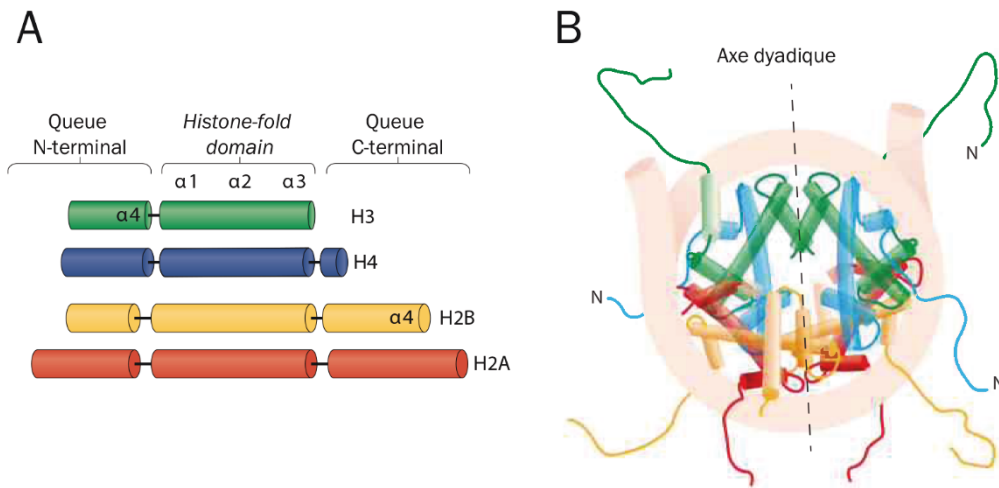


Figure 7 : Les histones au cœur des nucléosomes. A) Schéma des 4 histones canoniques. Les hélices  $\alpha 1$ ,  $\alpha 2$  et  $\alpha 3$  se trouvent au niveau de l'HFD de chaque histone. Une quatrième hélice ( $\alpha 4$ ) se trouve respectivement sur la queue N terminale et C terminale de H3 et H2B spécifiquement. B) Représentation schématique d'un nucléosome : au centre se trouve l'octamère d'histone entouré de l'ADN (orange pale). Adapté de Martire and Banaszynski, 2020 et *Molecular Biology of the cell* – sixth edition.

## 2) Diversité des nucléosomes

Au sein de la chromatine, il existe une diversité de nucléosome qui se caractérise par une composition en histone particulière et par la présence de modifications post-traductionnelles (PTMs).

- Variation de la composition des nucléosomes

Des paralogues, nommés variants, aux histones H2A, H2B, H3 et H4 ont été découverts, dont certains ont une expression spécifique à un type-cellulaire (Figure 8A). Certains de ces variants possèdent eux-mêmes des isoformes. Ainsi macroH2A correspond à un ensemble de protéines comprenant macroH2A1.1 et macroH2A1.2 (deux isoformes du gène *MACROH2A1*) et macroH2A2 (seule isoforme du gène *MACROH2A2*) (Ghiraldini et al., 2021). Ils peuvent être intégrés au sein des nucléosomes modifiant la structure du nucléosome (Ghiraldini et al., 2021) (Figure 8B). Les nucléosomes sont alors nommés « homotypiques » s'ils sont constitués des deux mêmes paralogues (contenant par exemple deux H3.3) ou « hétérotypiques » s'ils sont constitués de deux paralogues différents (contenant par exemple un H3.3 et un H3.Y.1) (Martire and Banaszynski, 2020). Il a aussi été rapporté l'existence d'« héminucléosome » n'ayant pas l'un ou les deux hétérodimères H2A-H2B (Castillo et al., 2017).

L'intégration des histones au sein des nucléosomes est majoritairement dépendant de la répllication de l'ADN. C'est le cas de H2A, H2B, H3, H4 ainsi que certaines variantes d'histones, comme H3.1 et H3.2. Pour d'autres variants d'histone, comme H3.3 et H2A.Z, l'intégration au sein des nucléosomes n'est pas réalisée durant la répllication de l'ADN. Ils sont alors incorporés dans les nucléosomes déjà formés, modifiant ainsi leur composition, via des machineries moléculaires appelées remodeleur de la chromatine (voir introduction III.2.ii). Dans tous les cas, l'assemblage des nucléosomes, ou la modification de leur composition, fait appel à des protéines chaperonnes (Talbert and Henikoff, 2017).



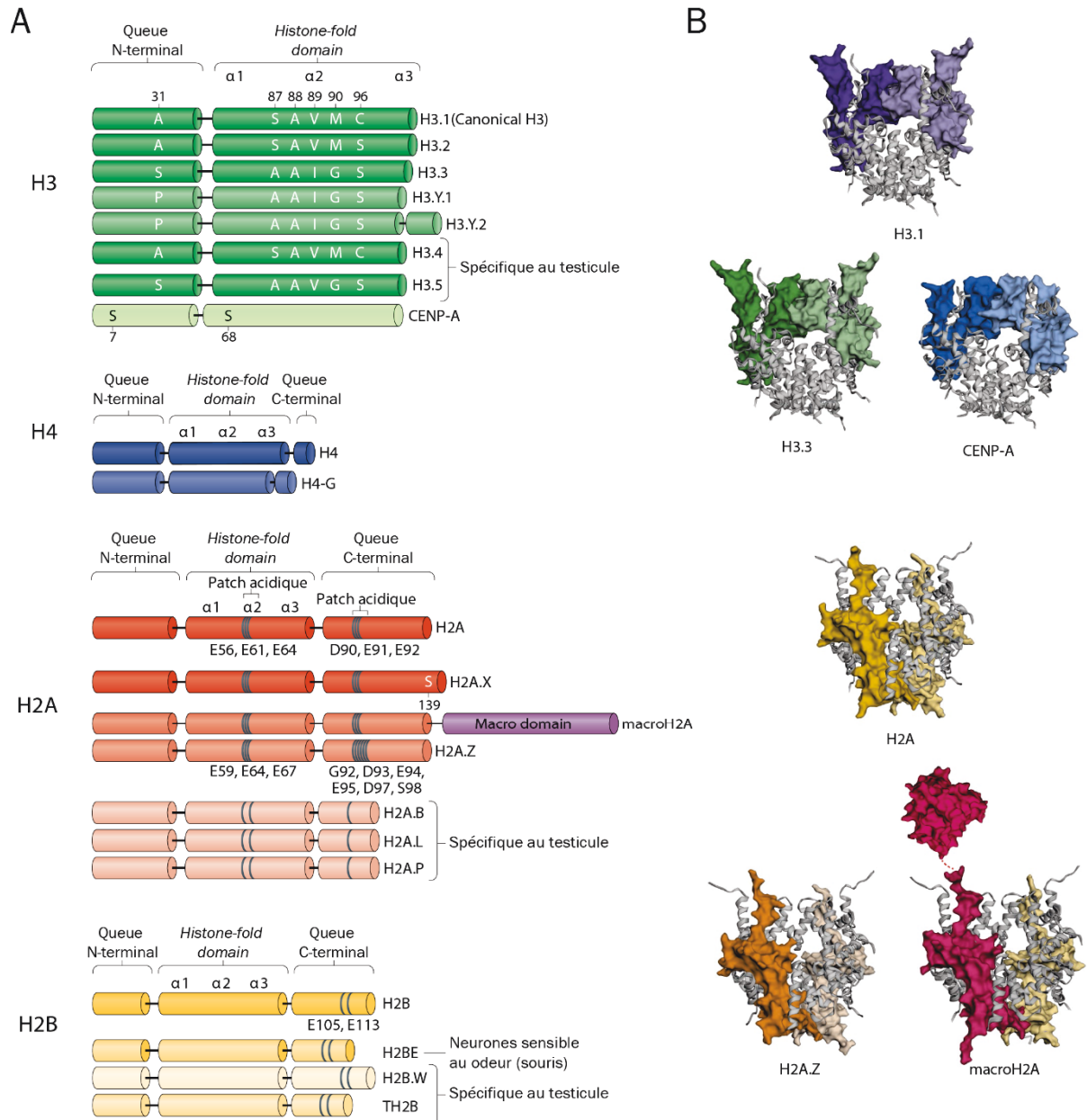


Figure 8 : Présentation des variants histones modifiant la structure des nucléosomes et leur possibilité d'être modifiés covalemment. A) Représentation schématique des différents variants d'histones retrouvés chez les mammifères. Pour certains d'entre eux est mis en exergue leur différence au niveau de la séquence en acides aminés. Ces différences sont à l'origine de la spécificité d'une histone pour une protéines chaperonnes particulière (CAF1 pour H3.1 et H3.2 et ATRX/DAXX pour H3.3 via la reconnaissance spécifique de la séquence au niveau de l'HFD). De plus, par ces différences certaines PTMs ne sont pas possible au niveau de certains variants. Contrairement à H3.1 et H3.2 n'ayant pas de sérine en position 31, H3.3 peut être phosphorylé à cette sérine. Enfin, les différences entre variants d'histone modifient la stabilité des nucléosomes. Les histones H2A.B, H2A.L et H2A.P présentent un patch acide moins important que H2A.Z au niveau de la queue C-terminale. Les patches acides favorisent les interactions entre nucléosomes. B) Variant d'histone au sein des nucléosomes. En fonction du variant intégré, la structure du nucléosome est faiblement modifiée (H3.1 VS CENP-A ou H2A VS H2A.Z) ou fortement modifiée (H2A VS macroH2A). Adapté de Ghiraldini et al., 2021 & Martire and Banaszynski, 2020.

- **Modification post-traductionnelle des histones**

En plus de ces variants d'histone, les nucléosomes varient également en fonction des modifications post-traductionnelles (PTMs) des histones. A l'heure actuelle, une vingtaine de PTMs ont été mise en évidence (Bannister and Kouzarides, 2011; Castillo et al., 2017; Cavalieri, 2021) (Figure 9). Les lysines (K) sont la cible d'une quinzaine de PTMs différentes, dont la méthylation, l'ubiquitination

ou encore l'acétylation (Figure 9). Les complexes qui mettent en place ou suppriment les PTMs sont respectivement appelés « *writers* » et « *erasers* » (Castillo et al., 2017). Le complexe coactivateur SAGA (*Spt-Ada-Gcn5-acetyltransferase*) est un exemple de complexe ayant à la fois une activité *writer* et *eraser*. Via son domaine acétyltransférase, SAGA acétyle les histones H3 et avec son domaine déubiquitinase, il déubiquitine les histones H2B.

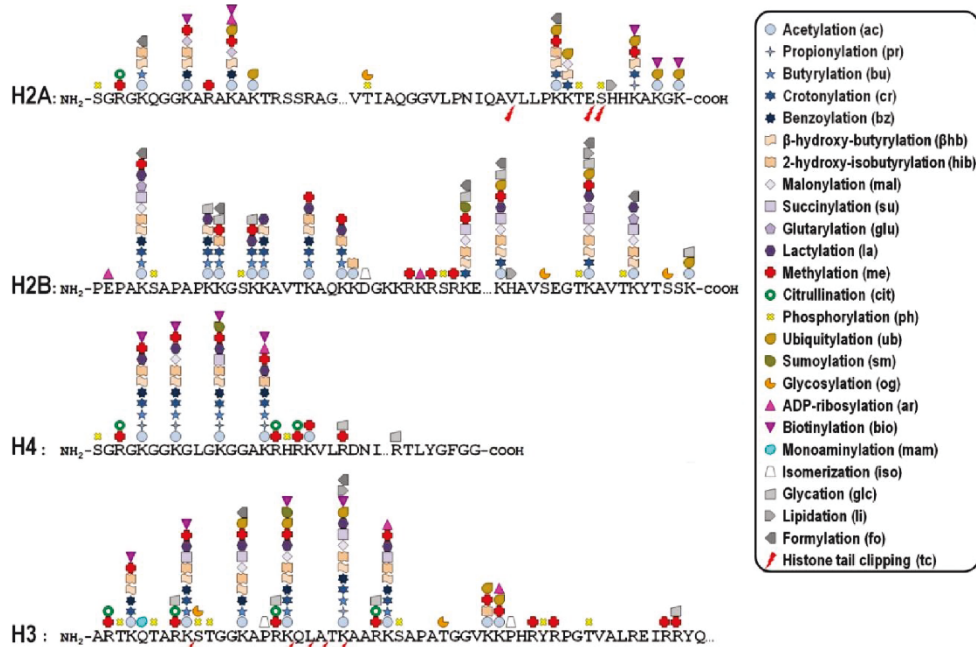


Figure 9 : Diagramme montrant les différentes PTMs répertoriées pouvant être mises en place sur les 4 histones canoniques. La première méthionine a été omise des séquences. Illustration issue de Cavalieri et al. 2021.

- *Histone modification cross-talk*

La mise en place des PTMs ou des variants d'histones dépend des PTMs et des variants d'histones déjà présent au niveau de la chromatine. Ce processus est appelé « *Histone modification cross-talk* » et est régulé à trois grands niveaux (Bannister and Kouzarides, 2011) (Figure 10):

- Un résidu particulier d'une histone peut-être la cible de différents complexes, il existe donc une compétition entre les PTMs
- Le recrutement d'un complexe *writer*, *eraser* ou remodeleur de la chromatine au niveau de nucléosomes particuliers est régulé par la présence d'un ou de plusieurs PTMs ou variants d'histone présent au niveau de ces nucléosomes, agissant parfois en synergie (voir introduction I.3.iii).
- La mise en place d'une PTM ou d'un variant d'histone est influencée par la structure/composition du nucléosome.

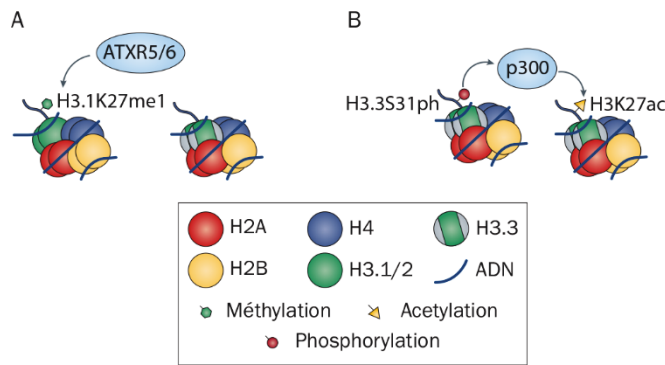


Figure 10 : Histone modification cross-talk. *Arabidopsis thaliana* trithorax-related protein 5/6 (ATXR5/6) méthyle seulement les histones H3.1 mais pas les histones H3.3 en raison de l'encombrement stérique de H3.3 Ser31 (A). Contrairement à H3.1 qui n'a pas de Ser31, H3.3 peut être phosphorylé au niveau de cette sérine, entraînant le recrutement de l'histone acétylase p300 menant à la phosphorylation de la lysine 27 de H3 (B). Adapté de Martire and Banaszynski, 2020.

### 3) Les histones linker pour former les chromatosomes

Les histones *linker* H1 sont des histones qui se lient aux nucléosomes en interagissant avec l'ADN nucléosomal et l'ADN *linker*, au niveau de l'axe dyadique (Bednar et al., 2017; Travers, 1999). L'ensemble est alors nommé chromatosome. La présence des histones *linker* protège environ 20 pbs supplémentaires par rapport au nucléosome, pour un total d'environ 165-168 pb (Bednar et al., 2017; Travers, 1999). Ces histones sont impliqués dans la condensation de la chromatine et la formation des TADs (Bednar et al., 2017; Fan et al., 2003; Fyodorov et al., 2018).

Il existe une variété de chromatosomes dû l'existence de multiples variants d'histones H1 pouvant porter différentes PTMs. Tous comme les histones H2A, H2B, H3 et H4, les histones H1 correspondent à une famille de protéines paralogues. Le génome de l'humain ainsi que de la souris présentent 11 paralogues. Certains ne sont exprimés que dans certains types cellulaires. Les variants H1 sont impliqués dans l'*Histone modification cross-talk*. Ils sont sujets à des PTMs tel que la phosphorylation, méthylation, acétylation, citrullination ou ubiquitination. Ces modifications vont réguler le taux de compaction de l'ADN et leur affinité pour les nucléosomes. De plus ces PTMs permettent de recruter différentes protéines tel que des *writers* ou *erasers* (Fyodorov et al., 2018; Kostova et al., 2013).

### 3. La chromatine interagit avec un ensemble de protéines dont le recrutement est régulé.

Les interactions protéine-chromatine sont importantes pour la mise en place et la régulation des processus moléculaires impliquant l'ADN. La régulation de l'organisation, de la structure et de la composition chimique de la chromatine permet le recrutement des protéines appropriées sur leurs sites de liaison. Le recrutement des protéines dépend d'une part de l'ADN et d'autre part des histones.

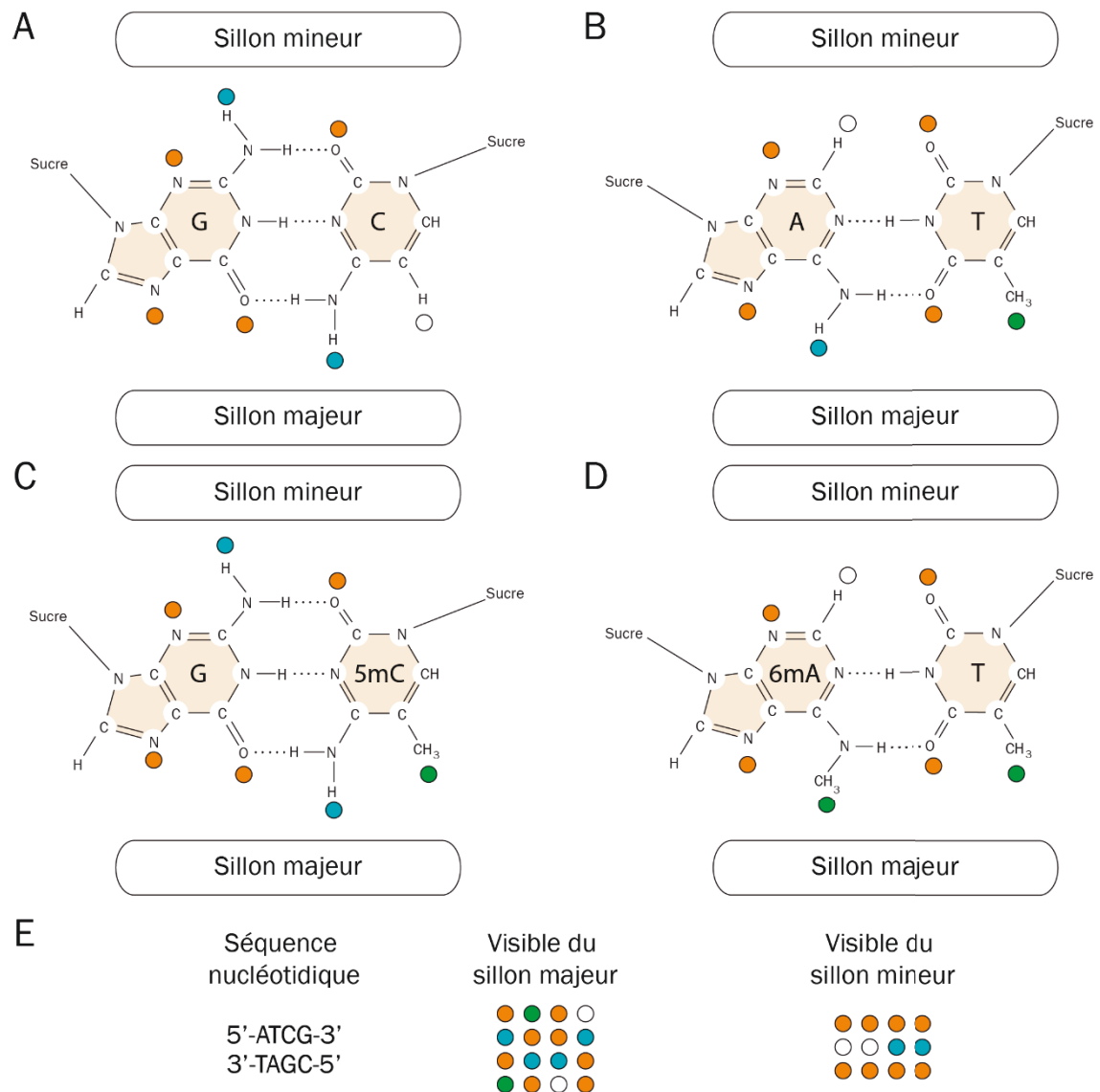
#### *i. Interactions protéine-chromatine dépendant de l'ADN*

Dans le cas du recrutement des protéines via l'ADN, deux mécanismes non exclusifs sont impliqués : 1) la lecture des bases (*base readout*), dite lecture directe, est fondée sur l'identité chimique des séquences nucléotidiques, 2) La lecture de forme (*shape readout*), dite indirecte, fait référence à une

reconnaissance partielle de la séquence nucléotidique en l'absence d'interaction directe entre les acides aminés et les nucléotides (Koudelka et al., 2006; Rohs et al., 2010).

### 1) Base readout

La reconnaissance d'une séquence par une protéine est fondée principalement sur des liaisons hydrogènes (liaisons H) et des contacts hydrophobes entre le domaine de liaison à l'ADN de la protéine (DBD, *DNA binding domain*) et les bases azotées. Pour une même séquence, les caractéristiques chimiques "lisibles" sont différentes entre le sillon majeur et le sillon mineur, ce qui peut affecter la spécificité de liaison de la protéine (Chiu et al., 2017).



**Figure 11 : Base readout : le sillon majeur apporte plus d'information que le sillon mineur.** (A & B) Représentation des groupements chimiques « lisibles » (rond de couleur) dans le sillon mineur (haut) et majeur (bas) pour le pb G-C (A) et A-T (B). (C & D) Même représentation que (A & B) pour des pb G-5mC (C) et 6mA-T (D). Les méthylations ne sont lisibles que dans le sillon majeur. C) Diagramme montrant les groupements chimiques lisibles dans le sillon majeur et le sillon mineur pour la séquence 5'-ATCG-3'. Plus de groupements chimiques sont lisibles dans le sillon majeur que dans le sillon mineur. De plus, au sein du sillon majeur, les pb A-T, T-A, C-G et G-C sont bien reconnaissables, ce qui n'est pas le cas dans le sillon mineur. Adapté de Chiu, T. et al. 2017.

- Lecture par le sillon majeur

Dans le sillon majeur, le profil des accepteurs et des donneurs de liaisons hydrogènes est spécifique pour chaque nucléotide (Figure 11 A, B, E). De plus, alors que les interactions protéine-purine sont médiées uniquement par des liaisons hydrogènes, les liaisons protéine-pyrimidine sont également appelées à des interactions hydrophobes (Figure 11 A, B). Pour ces raisons, la reconnaissance d'une séquence par le sillon majeur est spécifique (Rohs et al., 2010). La modification des A et C (méthylation et oxydation) font apparaître de nouveaux groupements chimiques dans le sillon majeur (Kribelbauer et al., 2020) (Figure 11 C, D). Ces groupements pourraient réguler le recrutement des protéines par le sillon majeur. En particulier, les MBPs (*Methyl-CpG binding protein*) reconnaissent spécifiquement les séquences contenant des 5mC (Hodges et al., 2020) (Figure 11 C). A l'heure actuelle, l'impact de la 6mA ou des différentes formes oxydées de 5mC sur le recrutement des protéines via le sillon majeur n'a pas été montré.

- Lecture par le sillon mineur

Le côté du sillon mineur est chimiquement monotone. Le schéma des donneurs et accepteurs de liaisons hydrogène ne permet pas de distinguer les paires de bases A-T de T-A ou G-C de C-G (Chiu et al., 2017; Rohs et al., 2010) (Figure 11 A, B, E). Seul le groupe amine exocyclique dépassant de la guanine à la position G2 permet de distinguer les paires de bases A-T/T-A et G-C/C-G (Patikoglou et al., 1999) (Figure 10 E). Les 5mC et 6mA, et leurs possibles formes oxydées, ne sont pas chimiquement lisibles de ce côté (Figure 11 C, D). Par conséquent, les protéines interagissant avec le sillon mineur reconnaissent les séquences nucléotidiques avec une spécificité moindre. Cela leur permet de se lier à une large gamme de séquences d'ADN.

Ainsi, une même séquence n'est pas reconnue de la même façon en fonction du sillon utilisé pour l'interaction. Une petite séquence comme 5'-ATCG-3' présente dans le sillon majeur une grande variabilité de groupement chimique à utiliser pour recruter et stabiliser une protéine sur l'ADN. Au contraire, dans le sillon mineur, le nombre et la richesse des groupement chimique sont moindres et par conséquent le recrutement d'une protéine est bien moins spécifique (Figure 11 E). Les protéines interagissent dans les sillons par leur domaine de liaison à l'ADN (*DNA-binding domain, DBD*) (Rohs et al., 2010). Ces domaines contiennent des hélices  $\alpha$  et/ou des feuillets  $\beta$ . Ils envahissent le sillon majeur ou le sillon mineur pour interagir avec les nucléotides. La force et la stabilité d'une interaction dépendent de la proximité et de l'angle entre les acides aminés et les nucléotides, qui dépendent eux-mêmes de la capacité d'un DBD à entrer dans le sillon.

1) *Shape readout*

Dans ce mécanisme, la protéine est sensible à la topologie et à la déformabilité de l'hélice d'ADN qui dépend elle-même de la séquence nucléotidique. De ce fait, il a été montré que certaines protéines présentent une affinité pour des structures secondaires particulières de l'ADN comme les motifs G4 et les motifs I (Brooks and Hurley, 2009; Lannes et al., 2017; Niu et al., 2018; Sutherland et al., 2016) (voir I, 1, iii) . Pour d'autres protéines, leur recrutement au niveau d'une séquence nucléotidique nécessite que l'ADN soit courbé (Patikoglou et al., 1999). La déformabilité de l'ADN est liée à sa séquence nucléotidique.

En conclusion, le *base readout* et le *shape readout* peuvent agir en synergie et augmenter la spécificité de l'interaction. D'une part, une structure d'ADN spécifique favorisera ou non la mise en exergue des groupements chimiques d'une séquence nucléotidique. D'autre part, une séquence spécifique favorisera l'acquisition d'une forme ou d'une flexibilité d'ADN particulière. De plus, la liaison d'une protéine ou la réalisation de processus moléculaire comme la transcription peut également avoir un impact sur la forme/flexibilité de l'ADN et peut réguler le recrutement d'autres protéines.

*ii. Interaction protéine-chromatine dépendant des histones*

Le recrutement d'une protéine au niveau des histones dépend à la fois de la composition des histones et à la fois des PTMs (Figure 10). En particulier, les PTMs sont reconnues par des domaines protéiques particuliers tels que les chromodomaines, WD40, MBT et Tudor reconnaissent les histones méthylées, alors que les domaines 14-3-3 reconnaissent les histones phosphorylées. Enfin les bromodomaines et les domaines PHD (*Plant homeodomain*) reconnaissent les histones acétylées (Yun et al., 2011). Cependant le recrutement d'une protéine au niveau d'une PTM peut également être perturbé, ou sa liaison déstabilisée, par la présence d'une PTM voisine. Ainsi, la protéine HP1 est recrutée au niveau des histones H3 méthylés, tel que les histones H3 méthylés au niveau de la lysine 9 (H3K9me3). La phosphorylation de H3 (H3S10) décroche HP1 des nucléosomes (Fischle et al., 2005). Le phénomène par lequel la méthylation et la phosphorylation ont un effet antagoniste dans le recrutement/ maintien d'une protéine à la chromatine s'appelle le « *phospho-methyl switch* » (Poleshko et al., 2019). Celui-ci a été démontré comme régulant le recrutement de différentes protéines ou complexes protéiques (Poleshko et al., 2019).

En plus du code mis en place par les PTMs, les variants d'histones agissent en synergie avec les PTMs et régulent le recrutement des protéines. Ainsi le recrutement de la protéine BRD2 (*bromodomain-containing protein 2*) dépend à la fois de l'acétylation de H4 et de la présence de H2A.Z (Martire and Banaszynski, 2020).



### *iii. Rôle de la compaction de la chromatine dans le recrutement des protéines*

La compaction de la chromatine influence également le recrutement des protéines au niveau de celle-ci. Les PTMs des histones et leur variant jouent un rôle important dans le processus de compaction.

#### 1) Impact des PTMs des histones sur la condensation de la chromatine :

La modification de l'environnement électrostatique par des PTMs module la compaction de l'ADN. Des expériences *in vitro* ont montré que la chromatine se replie sur elle-même en augmentant la concentration en sels de la solution. Ce phénomène a été expliqué par la neutralisation des charges négatives de l'ADN qui, à faibles concentrations en sels, repoussent le contact des nucléosomes avec l'ADN et maintiennent la chromatine étendue (Tolsma and Hansen, 2019). De même, l'acétylation ou la phosphorylation des histones réduisent la charge positive globale des histones, ce qui altère les interactions électrostatiques entre histones ou entre histones et ADN. De ce fait, l'acétylation et la phosphorylation maintiennent la chromatine décondensée (Bannister and Kouzarides, 2011). Par exemple, dans des conditions *in vitro*, l'acétylation de résidus de lysine spécifiques, comme H4K16ac et H4K12ac, est suffisante pour maintenir la chromatine tendue sous une forte concentration de sels (Tolsma and Hansen, 2019).

D'autres PTMs modifient le repliement de la chromatine en impactant la structure du nucléosome. Par exemple, l'ubiquitination et la sumoylation correspondent à la formation d'une liaison covalente entre respectivement un résidu lysine et un peptide ubiquitine ou SUMO (*Small Ubiquitin-like Modifier molecule*). L'ajout d'un peptide supplémentaire entraîne un encombrement stérique défavorisant le contact nucléosome-nucléosome (Bannister and Kouzarides, 2011; Tolsma and Hansen, 2019).

#### 2) Impact de la composition des nucléosomes sur la condensation de la chromatine :

Comme vu précédemment, il existe une diversité de protéines histones pouvant intégrer les nucléosomes ou chromatosomes. Certains nucléosomes/chromatosomes favorisent ainsi la formation d'hétérochromatine (H3.3, H2A.Z, macroH2A, CENP-A, H5). H5 présente d'une part une meilleure affinité pour les nucléosomes et d'autre part une composition riche en arginine capable de neutraliser les charges négatives de l'ADN, ces deux facteurs favorisant la compaction de l'ADN (Kowalski and Pałyga, 2011). D'autres, comme l'histone H3.Y.1, favorisent la déstabilisation des nucléosomes et donc l'état décondensé de la chromatine (Martire and Banaszynski, 2020).

#### 3) Impact des protéines se liant à la chromatine sur la condensation de la chromatine :

Un certain nombre de protéines présentent une capacité à réguler la condensation de la chromatine. Par exemple, HP1 est une famille de protéines liant l'un à l'autre deux nucléosomes (Machida et al., 2018) et favorisant alors la formation d'hétérochromatine (Zenk et al., 2021). Dans d'autres cas, les

protéines permettent d'ancrer la chromatine dans une zone favorisant le maintien de l'hétérochromatine. Ainsi, certaines protéines, comme CEC-4, se lient d'une part à la périphérie du noyau et d'autre part à H3K9me2 permettant de maintenir la chromatine dans une zone favorisant la condensation de la chromatine (Finlan et al., 2008; Holla et al., 2020; Poleshko et al., 2017, 2019).



## II. La transcription au sein du noyau est réalisée par trois ARN polymérases différentes.

L'expression génétique est l'ensemble des mécanismes de transcription et de traduction mettant en place le transcriptome et le protéome d'une cellule. La transcription est le processus moléculaire conduisant à la production d'une molécule d'ARN complémentaire à la molécule d'ADN utilisée comme patron. Cette production est assurée par des enzymes appelées ARN polymérases ADN dépendante (ARN pol). Les ARN pol :

- 1) assurent la sélection des nucléotides complémentaires à la séquence d'ADN lue et donc assurent la fidélité de l'ARN produit (Cheung and Cramer, 2012),
- 2) catalysent la synthèse des ARNs (Cheung and Cramer, 2012).

Les eucaryotes se caractérisent par 3 ARN polymérases nucléaires (ARN pol I, ARN pol II et l'ARN pol III), à l'exception des plantes qui ont deux ARN pol supplémentaires (IV et V). L'ARN pol I, II et III sont constituées de 14, 12 et 17 sous-unités respectivement (environ 0,5 à 1MDa). Cinq sous-unités sont partagées entre les trois ARN pol et 2 supplémentaires sont partagées uniquement entre l'ARN pol I et l'ARN pol III. (Martinez-Fernández and Navarro, 2018; Vannini, 2012, ensemble) (Tableau 1). Ces sous-unités partagées sont présentes au sein du corps catalytique. Ce corps forme une griffe composée d'une « pince » et d'une « mâchoire » au sein de laquelle l'ADN est engagé (Gamarrà and Narlikar, 2021). Au cours de la transcription, le corps catalytique permettent le positionnement correct du nucléotide triphosphate, catalyse la réaction de phosphodiesterification et le déplacement du duplex ARN-ADN (Cheung and Cramer, 2012; Griesenbeck et al., 2017). Parmi les autres sous-unités, certaines ont des homologues présents dans les autres ARN pol. C'est le cas des sous-unités du module *stalk* qui interagissent avec des régulateurs transcriptionnels et l'ARN en cours de synthèse (Gamarrà and Narlikar, 2021) (Tableau 1).

Chacune des ARN pol est responsable de la transcription de gènes particuliers. Un gène peut être défini par " une séquence d'ADN qui spécifie un ou plusieurs ARNs/protéines dont l'expression est d'une part régulée par les réseaux de régulation génétique (GRNs, *Genetic Regulatory Networks*) et d'autre part participent en tant qu'éléments des GRNs, souvent avec des effets indirects, ou en tant que résultante des GRNs, ces dernières produisant des effets phénotypiques plus directs." (Portin and Wilkins, 2017). Un GRNs est défini par l'ensemble des régulateurs moléculaires (séquence d'ADN, ARN, protéines) d'une cellule qui interagissent entre eux directement ou indirectement et avec d'autres substances de la cellule, le tout permettant de réguler les processus moléculaires au sein d'une cellule (Liu et al., 2018). La transcription médiée par l'ARN pol I, II et III sont régulées par un ensemble de facteurs de transcriptions et des cofacteurs et par conséquent par l'épigénome et la structure de la chromatine (Sharifi and Bierhoff, 2018). Dans le chapitre suivant sera décrit les mécanismes régulant la transcription médiée par l'ARN pol II. Mais ces mécanismes sont partagés entre les trois machineries

de transcriptions. Des perturbations de la transcription médiée par l'ARN pol I, II ou III sont impliquées dans de nombreuses maladies (Park et al., 2017; Sharifi and Bierhoff, 2018).

	ARN polymérase I		ARN polymérase II		ARN polymérase III	
	<i>S.cerevisiae</i>	<i>H.sapiens</i>	<i>S.cerevisiae</i>	<i>H.sapiens</i>	<i>S.cerevisiae</i>	<i>H.sapiens</i>
Corps catalytique des ARN pol	A190	RPA190/POLR1A	<b>RPB1/B220</b>	<b>RPB1/POLR2A</b>	C160	RPC1/POLR3A
	A135	RPA127/POLR1B	<b>RPB2/B150</b>	<b>RPB2/POLR2B</b>	C128	RPC2/POLR3B
	AC40	RPA5/POLR1C	RPB3/B44	RPB3/POLR2C	AC40	RPA5/POLR1C
	AC19	RPA9/POLR1D	RPB11/B12.5	RPB11/POLR2J2	AC19	RPA9/POLR1D
	RPB6/ABC23	RPB6/POLR2F	RPB6/ABC23	RPB6/POLR2F	RPB6/ABC23	RPB6/POLR2F
	RPB5/ABC27	RPB5/POLR2E	RPB5/ABC27	RPB5/POLR2E	RPB5/ABC27	RPB5/POLR2E
	RPB8/ABC14.5	RPB8/POLR2H	RPB8/ABC14.5	RPB8/POLR2H	RPB8/ABC14.5	RPB8/POLR2H
	RPB10/ABC10 $\alpha$	RPB10/POLR2L	RPB10/ABC10 $\alpha$	RPB10/POLR2L	<b>RPB10/ABC10<math>\alpha</math></b>	<b>RPB10/POLR2L</b>
	RPB12/ABC10 $\beta$	RPB12/POLR2K	RPB12/ABC10 $\beta$	RPB12/POLR2K	RPB12/ABC10 $\beta$	RPB12/POLR2K
	A14	?	RPB4/B32	RPB4/POLR2D	<b>C17</b>	<b>RPC9/POLR3I</b>
stalk module	<b>A43</b>	<b>RPA43/POLR1F</b>	RPB7/B16	RPB7/POLR2G	C25	RPC8/POLR3H
	A12.2	RPA12/POLR1H	<b>RPB9/B12.6</b>	<b>RPB9/POLR2I</b>	C11	RPC10/POLR3K
	<b>A49</b>	<b>RPA49/POLR1E</b>			<b>C53</b>	<b>RPC4/POLR3D</b>
	<b>A34.5</b>	<b>RPA34/POLR1G</b>			C37	RPC5/POLR3E
					<b>C82</b>	<b>RPC3/POLR3C</b>
					<b>C34</b>	<b>RPC6/POLR3F</b>
					C31	RPC7/POLR3G

Tableau 1 : Récapitulatif des sous-unités présentes dans les différentes ARN pol. Uniquement les ARN pol eucaryotiques I, II et III sont présentées ici. Pour les eucaryotes, le nom des sous-unités est donné pour la levure *S.cerevisiae* et *H.sapiens*. Sur une même ligne sont écrites les sous-unités identiques (bleu) ou homologues (noir). Les sous-unités en gras correspondent à celles utilisées pour recruter spécifiquement l'ARN pol à ses promoteurs. Tableau réalisé à partir des informations issu de Schramm et al 2002; Griesenbeck et al 2017; Russell et al 2006; site internet : ensembl et SGD

## 1. Transcription médiée par l'ARN pol I

L'ARN pol I est responsable de la transcription des précurseurs des ARNr ribosomiques (ARNr) 45S. Les séquences à l'origine des précurseurs sont répétées en tandem, séparées les unes des autres par des séquences intergéniques (Figure 12, haut) (Locati et al., 2018; Sharifi and Bierhoff, 2018). Au sein du noyau, cette transcription se fait spécifiquement au niveau du nucléole, lieu où sont également préformés les ribosomes (Pederson, 2011; Sharifi and Bierhoff, 2018). L'ARNr 45S après maturation donnera le 28S, 18S et 5.8S qui sont trois des quatre ARNr composant les ribosomes (Figure 12). Les voies de maturation de l'ARNr 45S varient d'un organisme à un autre (Tomecki et al., 2017). De plus, chez la plupart des insectes, l'ARNr 28S est clivé en deux parties de taille plus ou moins identiques proche du 18S (Winnebeck et al., 2010).

Via la maturation des ARNr sont produits des petits ARN ribosomiques (*RNA<sub>s</sub>r* : *small rRNA-derived RNAs*) dont la fonction n'est pas encore connue (Locati et al., 2018). De plus, l'ARN pol I transcrit également des longs ARNs non codant (lARNnc) à partir des « espaces intergéniques » comme l'ARNnc promoteur (*pRNA* : *promoteur RNA*) impliqué dans la régulation de la transcription des ARNr (Sharifi and Bierhoff, 2018) (Figure 12).

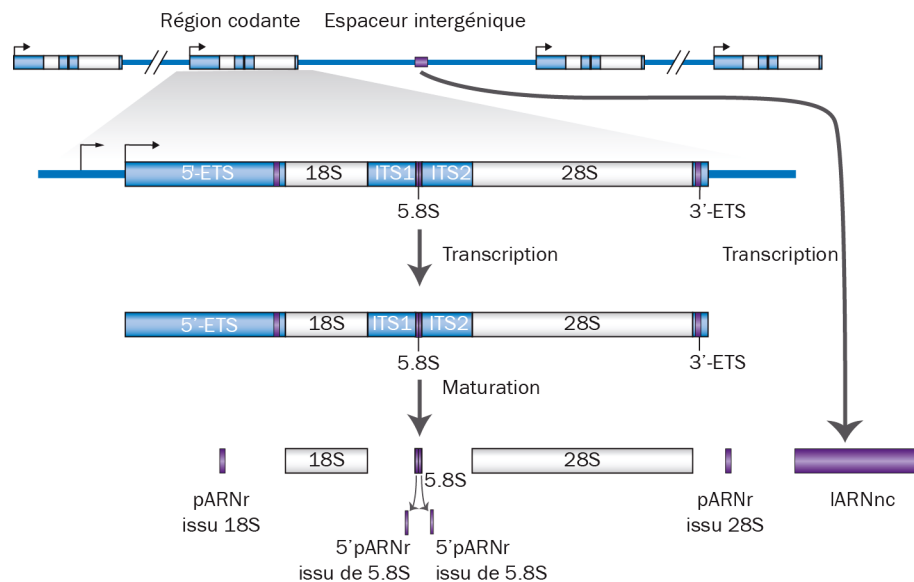


Figure 12: Synthèse des ARNr et l'ARNnc transcrits par l'ARN pol I. Les gènes codant les ARNr sont répétés en tandem. Après la transcription de pré-ARNr 45S, il est clivé en plusieurs ARNr donnant les ARN 5.8S, 18S et 28S ainsi que des pARNr. L'ARNr 5.8 peut par la suite être clivé en son centre pour donner deux autres pARNr. La fonction de ces pARNr n'est pas encore connue. La transcription au niveau des espaces intergéniques synthétise des l'ARNnc impliqués dans la régulation de la transcription médié par l'ARN pol I. Adapté de Sharifi and Bierhoff, 2018.

## 2. Transcription médiée par l'ARN pol II

L'ARN pol II transcrit l'ensemble des gènes codant des protéines produisant les ARNs messagers (ARNm). Ils en existent environ 22 000 au sein du génome de la souris [Ensembl, référence du génome de la souris GRCm39]. Les ARNm sont traduits dans le cytoplasme par les ribosomes pour produire des protéines (Figure 13c). Ces gènes se caractérisent par un ensemble d'exon et d'intron. Les introns sont transcrits puis retirés par épissage, alors que les exons sont conservés au sein des l'ARNm. L'épissage permet également de produire plusieurs ARNm matures à partir d'un seul gène codant. De ce fait, un gène peut être à l'origine de plusieurs protéines (Figure 13, isoforme 1 VS isoforme 2). Il est intéressant de noter que les régions introniques et exoniques correspondent à environ 41% du génome : 3,4% de séquences exoniques et 37,7% de séquences introniques (Yue et al., 2014).

L'ARN pol II transcrit également un grand nombre d'ARN non-codant (ARNnc). Cette famille correspond à un ensemble hétérogène d'ARNs répartis dans une vingtaine de sous-famille (Tableau 2, liste non exhaustive). Certains ARNnc, tel que les *snRNA* (*small nuclear RNA*), sont transcrits à partir de séquence d'ADN qui leur sont propres [Ensembl, référence du génome de la souris GRCm39]. Dans d'autres cas, les ARNnc, tel que certains *miRNA* (*microRNA*), sont le produit de maturation d'intron épissé. Grâce à l'amélioration des techniques de séquençage à haut débit, il est maintenant admis que l'ARN pol II peut transcrire en réalité une grande partie du génome : que ce soit dans des régions intergéniques ou à l'intérieur des gènes, que ce soit sur un ou l'autre des brins d'ADN, donnant naissance

à une myriade de longs ou courts ARNnc aux fonctions variées (Griesenbeck et al., 2017; Tudek et al., 2015).

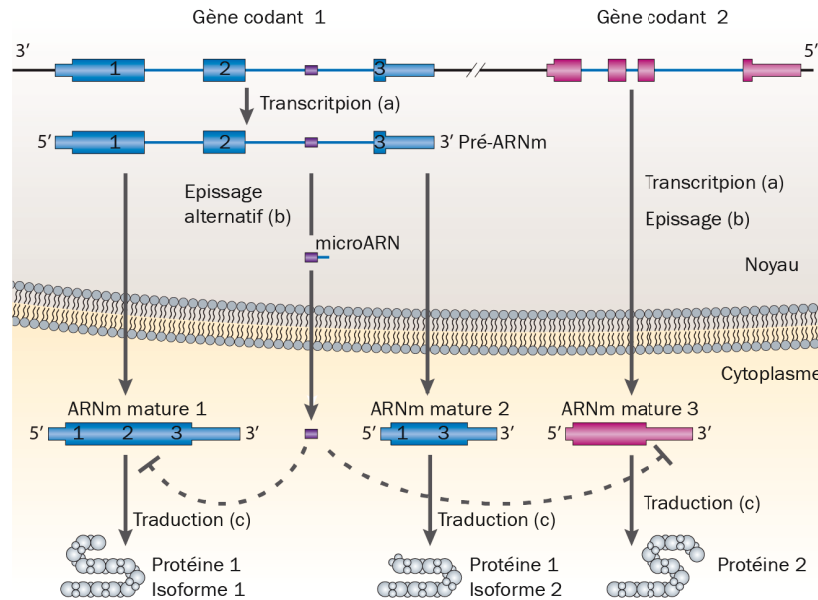


Figure 13 : Transcription, épissage et traduction des ARNm. En bleu et en rose, deux gènes codant pour des protéines. La transcription des gènes (a) est à l'origine des pré-ARNm contenant les exons (rectangle de couleur) et les introns (trait de couleur). Durant l'épissage (b) les introns sont enlevés par coupure et ligature du pré-ARNm. L'épissage est dit alternatif lorsqu'un ou plusieurs exons également sont enlevés. Ainsi un pré-ARNm peut être à l'origine de plusieurs ARNm matures, et par conséquent de plusieurs protéines, nommées isoformes. Après l'épissage, certains introns peuvent devenir des ARNnc ayant des fonctions à part entière. C'est le cas ici d'un miRNA qui, en se liant à des séquences particulières des ARNm inhibe leur traduction (trait pointillé). Adapté de Kapranov, P. et al. 2007

Acronyme	Nom entier	Rôle biologique	Références
snoRNAs	small nucleolar RNAs	Impliqué dans la maturation de divers ARN	(Taft et al., 2009)
sdRNAs	sno-derived small RNAs	Régule la traduction des ARNm	(Taft et al., 2009)
miRNAs	microRNAs	Régule la transcription et traduction des ARNm	(O'Brien et al., 2018)
siRNAs	small interfering RNA	Régule la traduction des ARNm	(Dana et al., 2017)
piRNAs	PIWI-associated RNA	Régule la transcription	(Ozata et al., 2019)
snRNAs	small nuclear RNAs	Sous-unité de la machinerie d'épissage	(Dergai and Hernandez, 2019)
circRNAs	circular RNAs	Impliqué dans la régulation entre autres l'épissage alternatif, de la transcription et de la traduction	(Tang et al., 2021)
lincRNAs	long intergenic non-coding RNAs	Impliqué dans la régulation de la transcription et de la condensation de la chromatine	(Ransohoff et al., 2018)
lncRNAs	long non-coding RNAs	Impliqué dans la régulation de la transcription et de la condensation de la chromatine	(Yao et al., 2019)
eRNAs	enhanceur RNA	Impliqué dans la régulation de la transcription	(Bose and Berger, 2017)
moRNAs	miRNA-offset-RNAs	Impliqué dans la régulation de la transcription	(Agnelli et al., 2019)

Tableau 2 : Liste non exhaustive des ARNnc transcrits par l'ARN pol II

The diagram illustrates the C-terminal domain (CTD) of RNA Polymerase II (ARN Pol II). At the top left, a schematic shows the ARN Pol II enzyme (green) transcribing DNA (ADN, blue lines) into RNA (ARN, red line). The CTD is represented as a series of repeating units (green and orange rectangles) extending from the enzyme. A consensus sequence box labeled 'Répétition consensus' shows the sequence: Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub>. Below this, two columns of amino acid sequences are shown for *S. cerevisiae* and *H. sapiens*, with positions 1 and 26 marked. The *S. cerevisiae* sequence is shown in green boxes, and the *H. sapiens* sequence is shown in orange boxes. The CTD is labeled 'CTD' at the end, with positions 1, 26, and 52 indicated. A blue arrow points from the *H. sapiens* sequence to the *S. cerevisiae* sequence, indicating a comparison or alignment.

**ARN Pol II**  
ADN  
ARN

Répétition consensus  
Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub>

CTD

1 26 52

*S. cerevisiae*

*H. sapiens*

1 26 52

QKHNENENSR

ISPDDSEEN

27

### 3. Transcription médiée par l'ARN pol III

L'ARN pol III transcrit un grand nombre d'ARNnc (Dieci et al., 2007; Yeganeh and Hernandez, 2020; Yu et al., 2020b) (Tableau 3). Les transcrits synthétisés par l'ARN pol III sont impliqués dans la régulation de la transcription médiée par l'ARN pol II ainsi que la synthèse des protéines (Tableau 3). Par exemple, l'ARNnc BC200/NC1 est exprimé quasi-exclusivement dans les neurones chez les mammifères. Cet ARNnc est impliqué dans le transport et la répression de la traduction des ARNm qui sont transcrits qu'au niveau des synapses (Sosińska et al., 2015).

Nom	Rôle biologique
<i>tRNA (transfert RNA)</i>	Apporte les acides aminés au niveau des ribosomes
<i>tRF (tRNA-derived fragments)</i>	Impliqué dans la régulation de la transcription et de la traduction
<i>rRNA 5s (ribosomal RNA 5S)</i>	Sous-unité des ribosomes
<i>snRNA U6 (U6 small nuclear RNA)</i>	Sous-unité de la machinerie d'épissage
<i>RNase P RNA (RNA component of RNase P)</i>	Impliqué dans la maturation des pré-ARNt
<i>Rnase MRP RNA (The RNase mitochondrial RNA processing)</i>	Impliqué dans la maturation des pré-ARNr
<i>7SL RNA</i>	Impliqué dans relocalisation des ribosomes au réticulum endoplasmique
<i>Vault RNA</i>	Impliqué dans l'assemblage et le transport des complexes protéiques
<i>Y RNA</i>	Non connu
<i>7SK RNA</i>	Régule l'action de p-TEFβ et par conséquent la transcription médiée par l'ARN pol II.
<i>BC1/BC200 RNA (Brain cytoplasmic1/200 RNA)</i>	Régule la traduction des ARNm
<i>SINEs (short interspersed nuclear elements)</i>	Rétrotransposon non autonome

Tableau 3 : Liste non exhaustive des ARNnc transcrits par l'ARN pol III

### III. Mécanismes de la transcription médiée par l'ARN pol II

#### 1. Promoteur minimum (*core promoter*) et séquence activatrice

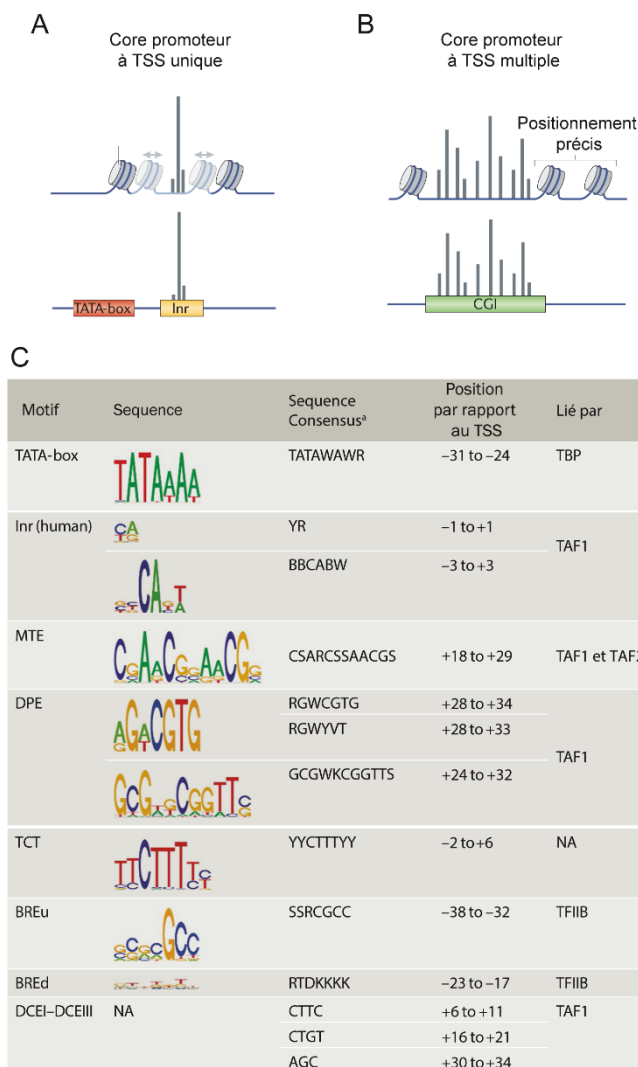
Le promoteur minimum est défini comme une séquence présente en amont du site de début de transcription (TSS), suffisante pour assembler les machineries de transcription et initier la transcription. (Haberle and Stark, 2018).

##### *i. Promoteur minimum des gènes codant des promoteurs.*

Au niveau des gènes codant des protéines, H3K4me2/me3, H3K27ac et H3K9ac ainsi que les variants H2A.Z et H3.3 sont retrouvés au niveau des promoteurs dont la transcription est activée (Ernst et al., 2011; Haberle and Stark, 2018). Ils existent deux classes de promoteur : les promoteurs minimums à TSS unique et les promoteurs minimums à TSS multiples.

- Les promoteurs minimums à TSS unique :

Ces promoteurs se définissent par la présence de séquences *cis* particulières (Figure 15) telles que la boîte TATA (*TATA-box*) et le motif initiateur (Inr) (Haberle and Stark, 2018). Néanmoins, rares sont



les promoteurs présentant la séquence canonique d'un de ces motifs. Ainsi, chez l'humain seulement ~2% des promoteurs des gènes ont la séquence consensus de la boîte TATA (TATAWAWR, W = A ou T, R = A ou G) et ~20% ont une séquence qui ressemble à une boîte TATA (*TATA-like*) (Yang et al., 2007). De plus, il est également rare qu'un gène cumule plusieurs motifs. Par exemple, alors que ~46% des promoteurs contiennent le motif Inr consensus (YYANWYY, Y = C ou T, N = A, C, T ou G), seulement ~15% contiennent à la fois le motif Inr et une séquence TATA (boîte TATA ou *TATA-like*) (Yang et al., 2007). Enfin que ce soit chez la levure ou chez l'homme, un peu moins de 50% des gènes ne contiennent ni boîte TATA ni séquence Inr (Yang et al., 2007).

Chez les mammifères et le modèle *Drosophila*, ce type de promoteur sont



Figure 15 : Promoteur minimum des gènes codant des protéines chez les mammifères. A-B) Schéma représentant l'organisation des nucléosomes (haut) et des séquences (bas) au niveau de promoteur minimum à TSS unique (A) et à TSS multiples (B). C) Tableau récapitulatif des séquences présentes au promoteur à motif défini. TATA-boîte : Boîte TATA. Inr : motif initiateur. MTE : Motif Ten element. DPE : Downstream promoteur element. BRE : TFIIB recognition element, upstream (u), downstream (d). DCE : Downstream core element. Adapté de Haberle and Stark, 2018.

spécifique de gènes exprimés dans les cellules différenciées des tissus adultes. Chez la Drosophile, ce type de promoteur est également retrouvé au niveau de gènes codant des facteurs de transcription impliqués dans l'embryogenèse. Le promoteur minimum de ces gènes se caractérise par un élément promoteur en aval (DPE, *Downstream promoteur element*) et un motif Inr (Haberle and Stark, 2018).

- Les promoteurs minimums à TSS multiple :

Ces promoteurs se trouvent dans une région déplétée en nucléosome (NDR : *nucleosome depleted region*). Les nucléosomes présents en aval du TSS sont précisément positionnés. Chez les mammifères, ces promoteurs minimums se caractérisent par une haute densité de dinucléotide CpG que l'on nomme îlot CpG. Ils sont retrouvés au niveau des gènes dits de ménage et de gènes codant des facteurs de transcription durant l'embryogenèse. Chez les drosophiles, ces promoteurs minimums ont tendance à avoir des séquences Ohler1, Olher6 et l'élément lié à la réplication (DRE ; *DNA replication-related element*). Ces promoteurs minimums ne sont retrouvés qu'au niveau des gènes de ménage (Haberle and Stark, 2018).

## ii. Promoteur minimum des gènes codant des ARNnc

Certains ARNnc sont transcrits à partir de promoteur qui leur son propre. Par exemple, le promoteur minimum des gènes codant les *snRNA* se caractérise par la présence d'un unique élément de séquence proximal (*proximal sequence element, PSE*) (Dergai and Hernandez, 2019) (Figure 26).

Dans d'autre cas, le promoteur minimum des gènes codant peut être utilisé pour synthétiser des ARNnc. C'est le cas pour les ARNnc issu de la maturation des introns (Kapranov et al., 2007). C'est également le cas d'un certain nombre de *lncRNAs* et *lincRNAs*. Leurs transcriptions sont initiées au niveau des promoteurs minimums de gènes codant des ARNm mais se déroulent en anti-sens, transcrivant la séquence se trouvant en amont du gène (Haberle and Stark, 2018; Ransohoff et al., 2018; Yao et al., 2019). Ce phénomène est appelé « transcription bidirectionnelle » (Haberle and Stark, 2018).

Enfin, dans certain cas, la transcription ne semble ne pas être réalisée au niveau de séquences particulières. La transcription des *eRNA* à partir des séquences *enhancers* est corrélée avec le niveau d'activité de l'*enhancer* ainsi que la présence de NDR, de H3K4me1/me2/me3, de H3K4me3 et de H3K9ac (Ernst et al., 2011; Haberle and Stark, 2018).

## iii. Séquence cis-régulatrice

Il existe deux types séquences activatrices. Les promoteurs en amont et les *enhancers*. Les promoteurs en amont se trouvent, comme le nom l'indique, en amont du promoteur minimum. Ensemble, le promoteur minimum et le promoteur en amont forment le promoteur. Les *enhancers*, quant à eux, sont



éloignés des promoteurs minimums, jusqu'à plusieurs milliers de bases, se trouvant soit en amont soit en aval du promoteur. Ces séquences activatrices recrutent des facteurs de transcription et des cofacteurs qui d'une part active le promoteur minimum pour la transcription en, par exemple, remodelant la chromatine et d'autre part régule le recrutement et le fonctionnement de la machinerie de transcription (Haberle and Stark, 2018).

Les *enhancers* activent la transcription des gènes qui se trouvent dans le mêmes TADs qu'eux. Ainsi, la formation des boucles au sein du TADs favorise le contact entre le *enhancer* et le promoteur minimum. La mise en place des TADs est régulée par des séquences dites isolatrices. Au niveau de ces séquences, un ensemble de protéines, dont CTCF, est recruté permettant d'isoler les TADs les uns par rapport aux autres (Kim et al., 2015).

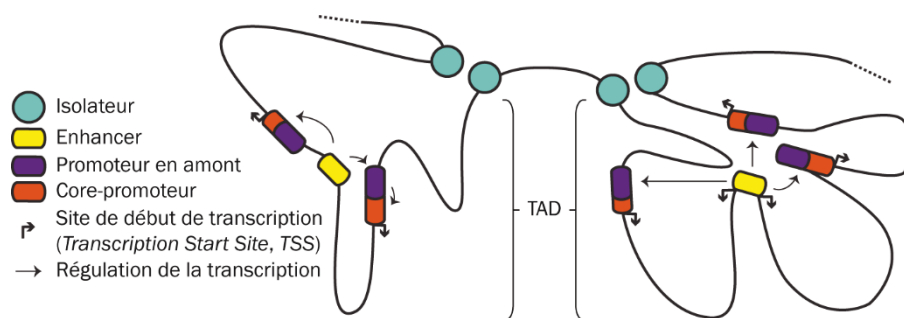


Figure 16 : Rôle fonctionnel des séquences cis-régulatrices dans la transcription. Au sein d'un TAD, les *enhancers* stimulent la transcription au niveau des promoteurs minimums. Chaque TAD est délimité par des séquences isolatrices.

## 2. Accessibilité des promoteurs

### *i. L'hétérochromatine et la zone périnucléaire : deux zones non permissives pour initier la transcription*

La localisation des gènes au niveau de la zone périnucléaire réprime la transcription (Finlan et al., 2008). La méthylation de H3K9 et H3K27 favorise la localisation de la chromatine au niveau de la périphérie du noyau interagissant indirectement avec la membrane du noyau (Gonzalez-Sandoval et al., 2015; Poleshko et al., 2017, 2019; Towbin et al., 2012). En particulier, H3K9me2 est présent uniquement au niveau de la périphérie nucléaire (Poleshko et al., 2017 ; Poleshko et al., 2019 ; See et al., 2019). Dans cette région, des marques épigénétiques sont mises en place réprimant la transcription et favorisant la condensation de la chromatine (Finlan et al., 2008; Holla et al., 2020; Peric-Hupkes et al., 2010; Poleshko et al., 2017; Poeschel et al., 2016).

L'activation de la transcription d'un gène est corrélée directement à la libération de ce gène de la périphérie nucléaire et à l'acquisition de marques épigénétiques activatrices de la transcription (Peric-Hupkes et al., 2010; Poleshko et al., 2017; See et al., 2019). Les nucléopores pourraient jouer un rôle important dans ce processus puisqu'ils sont capables d'activer la transcription (Buchwalter et al., 2019).

Cela peut également se faire pendant la mitose ou juste après celle-ci puisque toute l'organisation nucléaire est remodelée (Buchwalter et al., 2019 ; Poleshko et al., 2019). Ainsi la différenciation cellulaire a souvent lieu en même temps que la prolifération cellulaire (Halley-Stott et al., 2014). Il a aussi été démontré que les cellules en mitoses sont plus faciles à reprogrammer que les cellules en interphase (Halley-Stott et al., 2014).

## ii. Réorganisation des nucléosomes pour la mise en place du PIC

L'initiation de la transcription au niveau d'un promoteur minimum nécessite que ce dernier soit accessible à la machinerie de transcription. La présence de nucléosome au niveau du promoteur minimum inhibe la transcription (Lorch et al., 1987). Ainsi, les promoteurs minimums activés pour la transcription sont soit présents au sein d'une NDR (Haberle and Stark, 2018; Yuan et al., 2005), soit constitués de nucléosomes labiles tel que H2A.Z (Haberle and Stark, 2018).

La régulation de la localisation des nucléosomes au niveau des promoteurs est régulée par les « remodeleur de la chromatine ». Ces complexes permettent de déplacer, éjecter, restructurer ou modifier la composition des nucléosomes (Figure 17). Pour cela, ils utilisent l'énergie délivrée par l'hydrolyse de l'ATP (Clapier and Cairns, 2009; Kobayashi and Kurumizaka, 2019). Les remodeleurs sont répartis dans 4 familles, chacune se définissant par leur ATPase. Tous ces ATPase partagent le même domaine ATPase qui est divisé en deux parties : DExx et HELICc (Figure 17). Mais ils diffèrent les uns des autres par les domaines qui se trouvent de part et d'autre, ou à l'intérieur, du domaine ATPase (Clapier and Cairns, 2009; Clapier et al., 2017) (Figure 17)




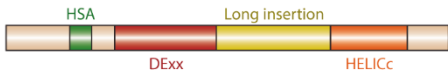
Famille	Domaine de l'ATPase	Complexe	Fonction
SWI/SNF		BAF PBAF	Glissement Ejection
ISWI		CHRAF ACF NURF	Glissement Ejection Activité <i>writers</i>
CHD		CHD1 NuRD ⋮	Assemblage Edition Glissement
INO80		INO80 SRCAP TIP60	Glissement Edition Activité <i>writers</i>

Figure 17 : Récapitulatif des différentes familles de remodeleurs de la chromatine présente chez les mammifères. Sont notifiées les caractéristiques de leur ATPase, les complexes que l'on retrouve dans chaque famille et leur rôle. « Edition » correspond au fait de modifier la composition du nucléosome.

Ces familles sont :

- Les complexes de la famille SWI/SNF (*switching defective/sucrose non-fermenting*) qui permettent le glissement et l'éjection des nucléosomes permettant l'activation de la transcription. Ils ne peuvent pas assembler les nucléosomes. Cette famille contient deux

complexes différents nommés chez les métazoaires BAF et PBAF (Clapier and Cairns, 2009; Clapier et al., 2017; Lorch and Kornberg, 2017). Cependant la composition des complexes peuvent varier en fonction du type cellulaire (Ho and Crabtree, 2010).

- La famille ISWI (*imitation switch*) contient trois complexes nommés chez les métazoaires CHRAF, ACF et NURF. Alors que les deux premiers complexes ont été décrits comme optimisant l'espacement des nucléosomes (Barisic et al., 2019), réprimant la transcription, le complexe NURF randomise l'espacement des nucléosomes favorisant la transcription (Clapier and Cairns, 2009; Clapier et al., 2017).
- Les complexes de la famille CHD (*chromodomain, helicase, DNA binding*) sont aussi décrits dans le glissement, l'éjection des nucléosomes et l'incorporation de H3.3. Certains d'entre eux présentent une activité enzymatique permettant la mise en place de marques épigénétiques. Chez les mammifères, il existe un grand nombre de complexes CHD tels que CHD1 et NuRD (Clapier and Cairns, 2009; Clapier et al., 2017; Kobayashi and Kurumizaka, 2019).
- Les complexes de la famille INO80 (*inositol requiring 80*) présentent différentes fonctions permettant l'activation de la transcription, la réparation de l'ADN et la réplication de l'ADN. Cette famille se distingue des autres par sa capacité à changer le dimer H2A-H2B dans les nucléosomes par le dimer H2A.Z.-H2B et vice-versa. INO80, SRCAP et TIP60 sont les noms des trois complexes que l'on retrouve chez l'humain (Clapier and Cairns, 2009; Clapier et al., 2017; Kobayashi and Kurumizaka, 2019). A noter que TIP60 a également une activité acétyl transférase.

### 3. Mise en évidence de l'importance des GTFs dans la transcription médiée par l'ARN pol I, II et III.

L'initiation de la transcription débute par le recrutement de facteur généraux de la transcription (GTFs ; *General Transcription Factor*) et de l'ARN pol II au niveau du promoteur, formant le complexe de pré-initiation (PIC ; *Pre Initiation Complex*).

Les GTFs sont, pour la plupart, des complexes multiprotéiques. Ils sont requis pour :

- 1) la reconnaissance spécifique des promoteurs,
- 2) le recrutement de l'ARN pol approprié au niveau de ce promoteur,
- 3) déterminer le sens et le TSS,
- 4) initier une transcription efficace.

L'importance des GTFs a été premièrement observée par le fait que les ARN pol ont besoin de facteurs additionnels pour initier la transcription au niveau d'un site particulier. (Weil et al., 1979). Par la suite les GTFs spécifiques à chaque ARN pol ont commencé à être caractérisés, dans un premier temps, en fractionnant des extraits cellulaires et en testant quelles fractions étaient requises pour initier la

transcription pour chaque ARN pol (Learned et al., 1985; Matsui et al., 1980; Mishina et al., 1982; Segall et al., 1980). Après 40 ans de recherches, la composition et le rôle de chaque GTFs dans la transcription médiée par l'une des trois ARN pol sont relativement bien connues. Dans le tableau 4 sont récapitulés les différents GTFs spécifiques à chaque ARN pol (Chen et al., 2021a; Hantsche and Cramer, 2017; Sainsbury et al., 2015; Schramm and Hernandez, 2002; Vannini, 2013).

	<i>S.cerevisiae</i>		<i>H.sapiens</i>	
	GTFs	Sous-unités	GTFs	Sous-unités
RNA polymérase I	UAF	Histone H3, H4, UAF30, RRN5, RRN9 et RRN10	UBF	Dimère de protéines UBF
	CF	TBP, <i>TFIIB-like factor</i> RRN7, RRN6 et RRN11	SL1	TBP, <i>TFIIB-like factor</i> TAF1B, TAF1A, TAF1C et TAF1D
	RRN3	RRN3	TIF-IA	TIF-IA
RNA polymérase II	TFIIA	TOA1 et TOA2	TFIIA	TFIIA $\alpha$ , TFIIA $\beta$ , TFIIA $\gamma$
	TFIIB	TFB	TFIIB	TFIIB
	TFIID	TBP + 14 différents TAFs (de TAF1 à TAF14)	TFIID	TBP + 13 différents TAFs (de TAF1 à TAF13)
	TFIIE	TFA1 et TFA2	TFIIE	TFIIE $\alpha$ et TFIIE $\beta$
	TFIIF	TFGI, TFG2 et TAF14	TFIIF	TFIIF $\alpha$ /RAP74 et TFIIF $\beta$ /RAP30
	TFIIH	Corps : p62, p52, p44, p34, p8, XPB et XPD Module kinase : CDK7, Cyclin H et MAT1	TFIIH	Corps : SSL2, TFB1, TFB2, TFB4, TFB5 et SSL1 Module kinase : KIN28, CCL1 et TFB3
RNA polymérase III	TFIIIA	TFIIIA/PZF1	TFIIIA	TFIIIA
	TFIIB	TBP <i>TFIIB-like factor</i> BRF1 (promoteur type 1) ou BRF2 (promoteur type 2)	TFIIB	TBP <i>TFIIB-like factor</i> BRF1 (promoteur type 1) ou BRF2 (promoteur type 2)
	BDP1	BDP1	BDP1	BDP1
	TFIIIC	TFC1, 2, 3, 6, 7 et 8.	TFIIIC	TFIIIC63, 220, 102, 110, 35 et 90

Tableau 4 : Récapitulatif des GTFs spécifiques à chaque ARN pol avec les noms de leur sous-unités pour la levure *S.cerevisiae* à gauche et l'humain à droite. Quelques différences apparaissent entre ces deux organismes, comme pour TFIID constitué de 21 sous-unités chez la levure et de 20 sous-unités chez l'homme. Réalisé à partir de Dergai and Hernandez, 2019; Sainsbury et al., 2015; Schier and Taatjes, 2020; Sharifi and Bierhoff, 2018; Vannini, 2013; Vannini and Cramer, 2012.

Lors de l'initiation de la transcription dépendante de l'ARN pol II, 6 GTFs (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH) vont s'assembler pour former le PIC au niveau du promoteur (Chen et al., 2021a; Sainsbury et al., 2015). Ce modèle d'assemblage a été mis en place d'abord par la reconstitution de systèmes de transcription *in vitro*, essayant l'addition aléatoire de différentes fractions purifiées (Matsui et al., 1980). Puis par l'analyse de complexe reconstitué à partir de complexes purifiés et analysés par cryo-microscopie électronique (Cryo-EM) (He et al., 2013).

### i. TFIID : premier GTFs à l'origine de la formation du PIC

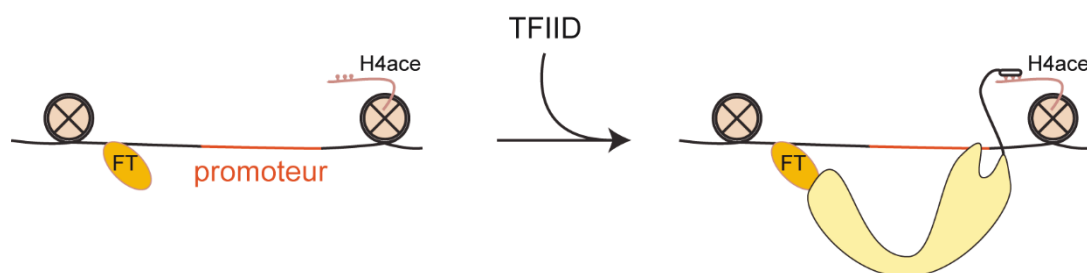


Figure 18 : Recrutement de TFIID par le promoteur minimum, des facteurs de transcription et des marques épigénétiques.

Chez les mammifères, TFIID est un complexe de ~1.2MDa composé de la *TATA-binding protein* (TBP) ainsi que 13 *TBP-associated factors* (TAFs). Chaque TAF est numéroté de 1 à 13 en fonction de son poids moléculaire (Tora, 2002). Au sein du PIC il est le plus gros complexe, suivi de l'ARN pol II et de TFIIH. Le recrutement de TFIID est régulé par les séquences nucléotidiques, les marques épigénétiques et les facteurs de transcription présents au niveau des promoteurs (Chen et al., 2021a; van Ingen et al., 2008; Kalogeropoulou et al., 2010; Louder et al., 2016; Papadopoulos et al., 2015; Patel et al., 2018). Après son recrutement, TFIID va charger TBP au niveau de l'ADN entraînant sa courbure (Chen et al., 2021a; Guzikovich-guerstein and Shakked, 1996; Patel et al., 2018). D'autre part, il va réguler la formation du PIC (Chen et al., 2021a, 2021b). Comme TFIID est le point central de mon sujet de thèse, dans le chapitre suivant (voir introduction, IV) sera développé avec de plus amples détails sa structure, son assemblage, son recrutement et sa stabilisation sur le promoteur ainsi que son importance dans l'initiation de la transcription.

Le recrutement de TFIID comme premier GTFs n'est pas une règle générale. Dans le cas de la transcription des *snRNA*, le complexe SNAPc (*snRNA activating protein complex*) reconnaît le PSE et recrute les GTFs pour initier la transcription (Dergai et al., 2018). Des essais de transcription *in vitro* ont montré que la transcription efficace de l'*snRNA* U1 nécessite la présence de TBP, TFIIA, TFIIB, TFIIF et TFIIIE (Kuhlman et al., 1999).

*ii. TFIIA et TFIIB : deuxième et troisième GTFs recrutés auprès de TFIID.*

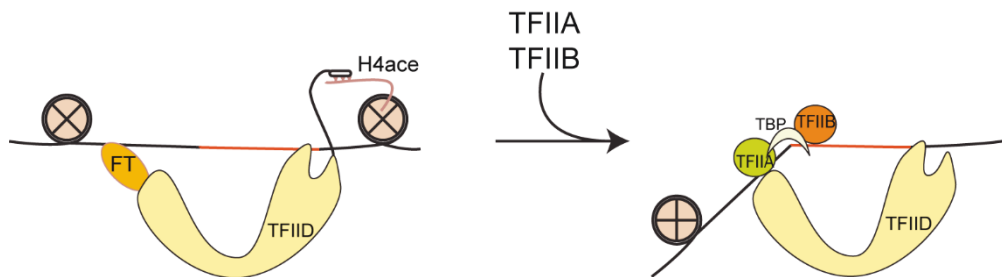


Figure 19 : Recrutement de TFIIA et B au niveau de TBP.

- **TFIIA**

TFIIA est constitué des sous-unités nommées TFIIA $\alpha$ , TFIIA $\beta$  et TFIIA $\gamma$  (Høiby et al., 2007). TFIIA $\alpha$  et TFIIA $\beta$  sont issus d'une unique protéine qui est clivée via la Taspase 1 au niveau de la séquence conservée QVD|G. Au sein du PIC, TFIIA interagit avec l'hétérodimère TAF4-TAF12 de TFIID ainsi qu'avec TBP (Chen et al., 2021a; Kraemer et al., 2001; Patel et al., 2018; Schier and Taatjes, 2020). TFIIA est important pour décrocher TBP de TFIID permettant son chargement au sein du sillon mineur induisant la courbure de l'ADN (Chen et al., 2021a; Patel et al., 2018). Néanmoins, dans des conditions de transcription *in vitro*, l'importance de TFIIA varie d'un système à un autre (Fant et al., 2020; Orphanides et al., 1996). Dans un contexte *in vivo*, TFIIA est important pour contrecarrer l'effet

du complexe NC2 (nom chez la levure, orthologue de DR1/DRAP1). Ce complexe, en interagissant avec TBP, régule la formation du PIC au promoteur (Kleijman et al., 2004; Koster et al., 2014; Thomas and Chiang, 2006). En particulier, au niveau de promoteurs contenant une séquence Inr, le complexe TFIIA $\alpha$ - $\beta$ - $\gamma$ , où  $\alpha$  et  $\beta$  sont séparés, inhibe l'activité de NC2 et favorise le recrutement de TFIID et le chargement de TBP au promoteur (Malecová et al., 2015). De ce fait, l'activité de la Taspase 1 à cliver TFIIA $\alpha\beta$  permet de réguler la transcription (Zhou et al., 2006). La mutation de TFIIA $\alpha\beta$  empêchant son clivage entraîne la mort de la majorité des souris mutantes après la naissance (Oyama et al., 2013). Parmi les survivantes, l'ensemble des mâles sont infertiles dû à la perturbation de la différenciation des spermatozoïdes (Oyama et al., 2013). Ainsi, la coupure de TFIIA $\alpha\beta$  en TFIIA $\alpha$  et TFIIA $\beta$  est essentielle pour la mise en place du programme génétique requis pour la spermatogenèse.

Pour finir TFIIA $\alpha\beta$  a un paralogue, nommé ALF, qui s'assemble avec TFIIA $\gamma$  (Upadhyaya et al., 2002). Il est exprimé uniquement dans les cellules germinales (Martianov et al., 2016; Xiao et al., 2006). *In vitro*, ALF est capable de stabiliser TBP sur une séquence TATA (Upadhyaya et al., 2002). Durant la spermatogenèse chez la souris, ALF interagit avec la paralogue à TBP, TBPL1 (*TBP-like 1*). La présence et le rôle de ALF au sein du PIC ne sont pas connues.

- TFIIB

Dans des conditions de transcription *in vitro*, TFIIB est requis pour initier la transcription au niveau d'un TSS unique (Matsui et al., 1980). TFIIB peut également se lier à TBP sans la présence de TFIIA (Kraemer et al., 2001). Contrairement à TFIIA, TFIIB est requis pour la transcription. La déplétion de TFIIB au sein de *S.cerevisiae* empêche l'expression de la quasi-totalité des gènes codant pour des ARNm ou ARN non codant. L'expression de 5007 gènes, sur un total de 5080, est significativement réduite (Tourigny et al., 2021).

TFIIB est le seul des 6 GTFs à n'être constitué que d'une unique protéine de 316 acides aminés chez l'humain. TFIIB vient contacter TBP d'une part et d'autre part l'ADN au niveau du *TFIIB recognition element* (BRE), stabilisant le complexe de pré-initiation en formation (Sainsbury et al., 2015). TFIIB contacte également la grosse sous-unité RPB1 de l'ARN pol II (Bushnell et al., 2004) participant à son recrutement et à son positionnement au niveau du promoteur (Liu et al., 2010)

### iii. TFIIF : quatrième GTFs recruté apportant l'ARN pol II

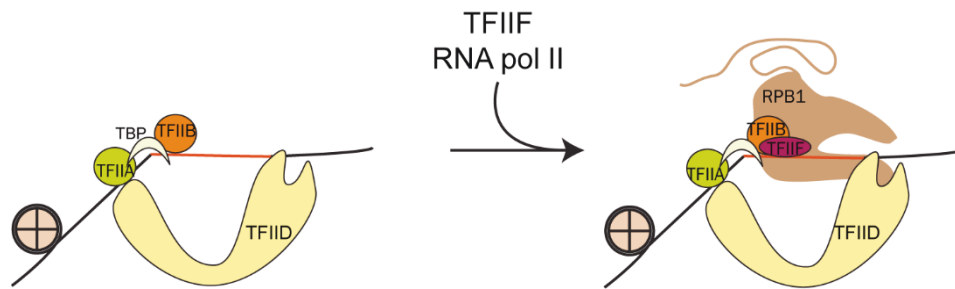


Figure 20 : Recrutement du complexe TFIIF-RNA pol II

TFIIF est constitué de 2 sous-unités TFIIF $\alpha$  et TFIIF $\beta$ . TFIIF est lié à l'ARN pol II au niveau de la fente de l'ARN pol II avec des points de contact proches du lobe RPB2, de RPB9 et de l'extrémité C terminale de RPB1 (Schier and Taatjes, 2020). L'interaction de TFIIF à l'ARN pol II prévient la liaison non spécifique de l'ARN pol II à l'ADN et par conséquent inhibe l'initiation de transcription aberrante (Sainsbury et al., 2015).

TFIIF est recruté au niveau du PIC en formation en contact avec TFIIA, TFIIB et TFIID (Chen et al., 2021a). De ce fait, il permet, avec l'aide de TFIIB, de recruter l'ARN pol II au niveau du promoteur en plus de stabiliser le PIC en formation (Bushnell et al., 2004; Chen et al., 2021a; Schier and Taatjes, 2020)

### iv. TFIIIE et TFIIH : Cinquième et sixième GTF recrutés

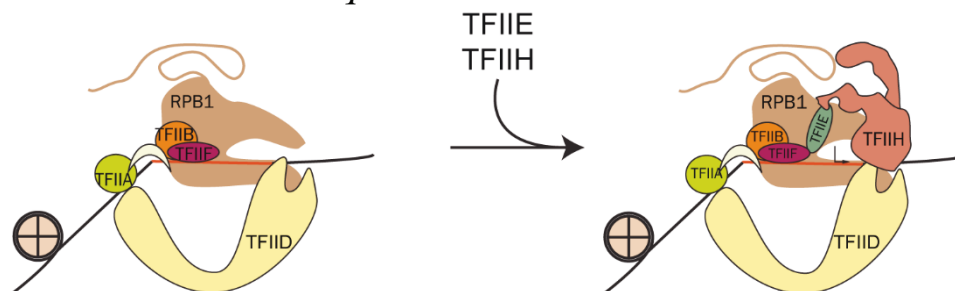


Figure 21 : Recrutement du complexe TFIIIE et TFIIH

- TFIIIE

TFIIIE est composé deux sous-unités TFIIIE $\alpha$  et TFIIIE $\beta$ . Une fois recruté, il contacte principalement différentes sous-unités de l'ARN pol II ainsi que TFIIF, TFIIB, TFIID (Chen et al., 2021a; Thomas and Chiang, 2006). Son principal rôle est de recruter et de positionner TFIIH, avec l'aide de TFIID, au sein du PIC (Chen et al., 2021a) et de stimuler l'activité ATPase de TFIIH (Sainsbury et al., 2015).

- TFIIH



TFIIH est le troisième plus gros complexe à être recruté, avec ces 10 sous-unités pour un poids total d'environ 0.5 MDa (Greber et al., 2019). En plus de son rôle important dans la mise en place de la transcription, TFIIH est impliqué dans les mécanismes de réparation des dommages à l'ADN (Kokic et al., 2019). Le complexe central est constitué de 7 sous-unités. Parmi eux, XPB est une ATPase ADN-dépendant et XPD est une 5'-3' hélicase. Les trois sous-unités restantes forment un module kinase, dont CDK7 est l'enzyme (Greber et al., 2019).

TFIIH est recruté par TFIIE. Ces derniers interagissent ensemble à de nombreux points de contact. L'intégration de TFIIH au complexe va réarranger sa structure conduisant à son interaction avec l'ARN pol. L'arrivée de TFIIH dans le complexe va d'une part finaliser la formation du complexe de pré-initiation (PIC) (Chen et al., 2021a; Greber et al., 2019)

#### v. *Le médiateur : nouveau composant du PIC*

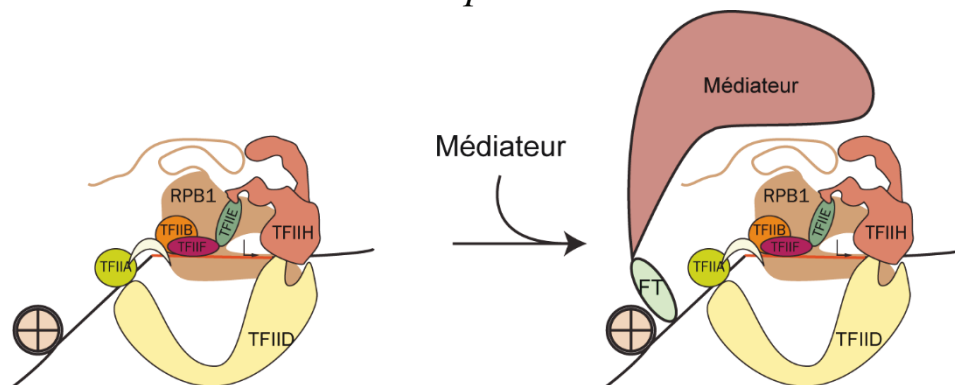


Figure 22 : Recrutement du médiateur

Le médiateur est considéré comme partie intégrante du PIC (Schier and Taatjes, 2020). Ce complexe est constitué de 25 et 30 sous-unités chez respectivement *S. cerevisiae* et l'humain. Sa structure générale a été découpée en 4 modules : une tête, un corps, une queue et un module kinase qui n'est associé que transitoirement aux 3 autres. Les sous-unités MED14 et MED17 sont les deux sous-unités principales permettant de lier la tête, le corps et la queue.

Le médiateur est recruté au niveau des promoteurs via la liaison de sa queue au niveau de facteurs de transcription (Allen and Taatjes, 2015; Haberle and Stark, 2018; Poss et al., 2013; Soutourina, 2018). Il interagit avec certains coactivateurs pour réguler la transcription et la mise en place du PIC. Dans les cellules mammifères, CHD1 est recruté par le médiateur aux niveaux des gènes activés par la transcription (Lin et al., 2011). Dans le modèle *S.cerevisiae*, SWI/SNF et SAGA sont recrutés par le médiateur au niveau des gènes (Lemieux and Gaudreau, 2004). D'autre part, il entre en contact avec le PIC pour le stabiliser. Les mutations nulles de *Med14* et *Med17* entraînent la délocalisation de TFIIB, TBP, TFIIH et l'ARN pol II et réduit drastiquement l'expression des ARNm, *snRNA* et *snoRNA* chez la



levure. Chez le mammifère, le médiateur n'est pas requis pour la formation du PIC, mais est néanmoins important pour l'expression du programme génétique (Jaeger et al., 2020; Sun et al., 2021).

### vi. Ouverture de l'ADN

La deuxième étape de cette initiation de la transcription est le passage du PIC fermé au PIC ouvert. L'activité ATPase de la sous-unité XPB de TFIIH va relaxer l'ADN et l'insérer dans le centre active de l'ARN pol II (Figure 23 A). Ainsi le triptolide (TLP), en se liant à XPB, inhibe l'activité ATPase de TFIIH et empêche la transcription (Bensaude, 2011; Forero-Quintero et al., 2021). TFIID et le médiateur favorise l'activité de TFIIH dans l'ouverture de la double hélice (Chen et al., 2021b). Cette structure composée du PIC avec l'ADN ouverte, maintenue à l'aide de TFIIB et TFIIF, est la bulle de transcription (Sainsbury et al., 2015). A partir de cette étape, la synthèse de l'ARN va débuter.

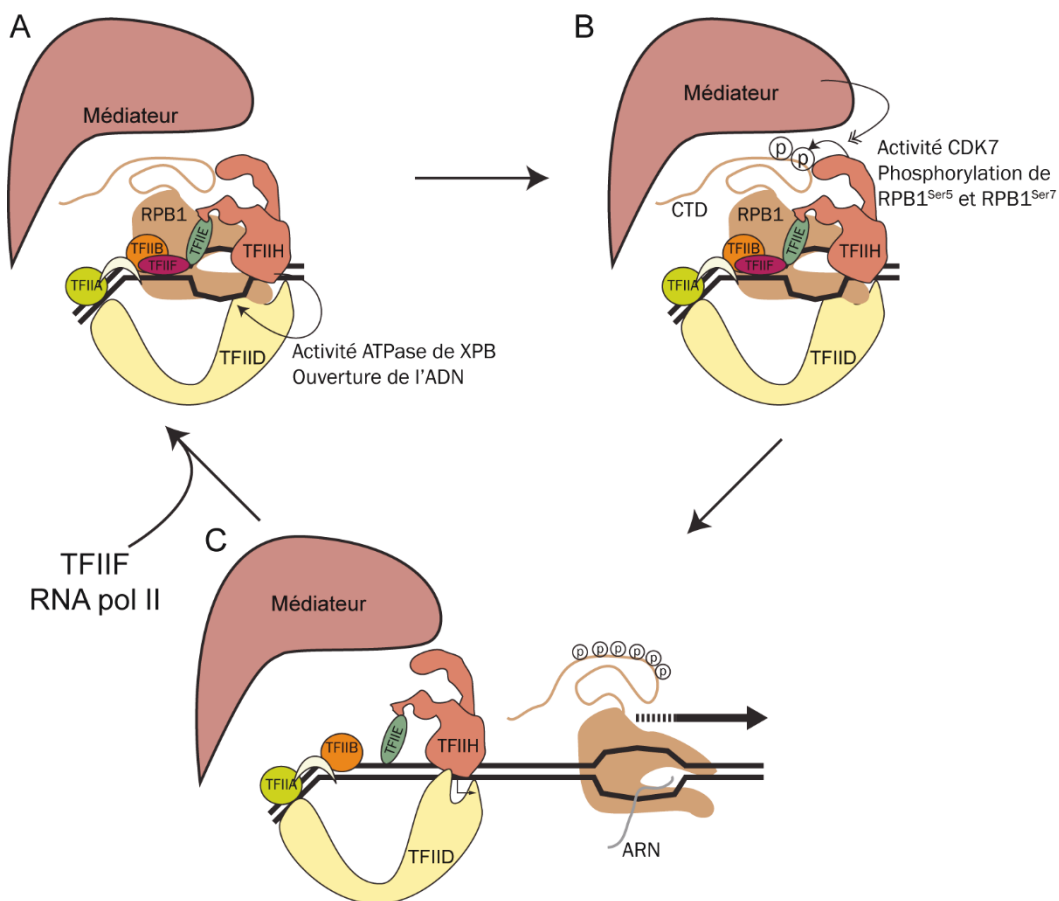


Figure 23 : Début d'élongation et réinitiation de la transcription. A) Après la formation du PIC, l'activité XPB permet l'ouverture de la bulle de transcription. B) L'activité de CDK7 permet la phosphorylation de la sérine 5 et 7 de RPB1. Le médiateur favorise la phosphorylation du CTD. C) Après le départ de l'ARN pol II, une partie du PIC reste au promoteur pour faciliter la réinitiation.

### vii. *Première élongation précoce :*

Après la mise en place du PIC, la kinase CDK7 de TFIIH, régulé par TFIIIE, TFIID et le médiateur, phosphoryle les serine 5 et 7 de l'extrémité C-terminale (CTD) de la sous-unité RPB1 de l'ARN pol II (RBP<sup>pSer5</sup>, RBP<sup>pSer7</sup>) (Chen et al., 2021b; Harlen and Churchman, 2017) (Figure 23 B). En particulier, le médiateur contacte le CTD non phosphorylé et le dirige vers CDK7 facilitant sa phosphorylation. Une fois phosphorylé, le médiateur présente une bien moindre affinité pour le CTD. Le traitement de cellule par THZ1, un inhibiteur spécifique de CDK7 abroge également la transcription (Forero-Quintero et al., 2021). Ces phosphorylations libèrent l'ARN pol II du PIC (Haberle and Stark, 2018).

Lors de cette première phase de transcription, TFIIB stimule la synthèse de l'ARN naissant, permet à l'ARN de sortir de l'ARN polymérase II et favorise la séparation du dimer ADN-ARN (Bushnell et al., 2004; Sainsbury et al., 2015). Lorsque l'ARN atteint 12 à 13 nucléotides TFIIB est relâché du PIC.

En ce qui concerne le complexe de pré-initiation, il a été suggéré que TFIID, TFIIA, TFIIH, TFIIIE ainsi que le médiateur resterait au niveau du promoteur. Ceci favoriserait la réinitiation de la transcription en facilitant le recrutement d'une nouvelle ARN pol II (Yudkovsky et al., 2000) (Figure 23 C).

### viii. *La pause et reprise de la transcription :*

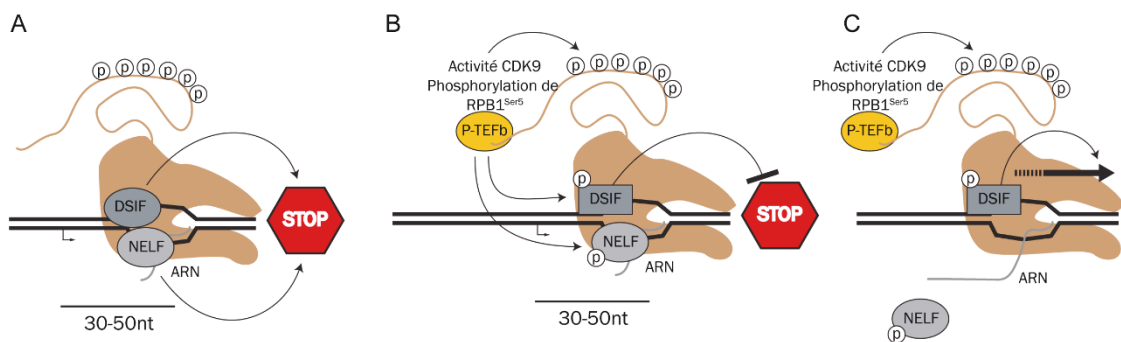


Figure 24 : Pause et reprise de la transcription. A) Le recrutement de DSIF et NELF au niveau de l'ARN pol II bloque sa progression de long du gène. B) Le recrutement de P-TEFβ au CTD de l'ARN pol II entraîne la phosphorylation de DSIF, NELF, et de la sérine 2 du CTD. C) NELF phosphorylé est décroché de l'ARN pol II et DSIF phosphorylé favorise l'élongation.

Après la synthèse de 30 à 50 de nucléotides, les complexes NELF (*Negative ELongation Factor*) et DSIF (*DRB-Sensitivity Inducing Factor*) se lie à l'ARN pol II bloquant la transcription (Schier and Taatjes, 2020) (Figure 24 A). En particulier, NELF stabilise la liaison de l'ARN à l'ADN, défavorise l'arrivé de nouveaux nucléotides dans le site catalytique de l'ARN pol II, et empêche le recrutement de TFIIS (Schier and Taatjes, 2020). TFIIS est impliqué dans la reprise de la transcription, lorsque celle-ci est bloquée, en se liant à l'ARN pol II et en clivant l'ARN naissant (Izban and Luse, 1992). La pause de la transcription pourrait également être favorisée par l'apparition rapide de super enroulement négatif

de l'ADN empêchant physiquement la progression de l'ARN pol II et la présence des nucléosomes (Haberle and Stark, 2018).

La pause de l'ARN pol II permet :

- 1) de maintenir la chromatine ouverte pour faciliter la réinitiation d'une nouvelle transcription,
- 2) de faciliter la transition du complexe de pré initiation au complexe d'élongation,
- 3) être un nouveau point de régulation de la transcription permettant l'expression rapide et synchrone d'un certain nombre de gènes (Adelman and Lis, 2012; Pimmitt et al., 2021; Schier and Taatjes, 2020).

Le recrutement de P-TEF $\beta$  (*Positive transcription elongation factor b*) par le médiateur cause la reprise de la transcription (Allen and Taatjes, 2015; Jaeger et al., 2020) (Figure 24 B). La kinase CDK9 de P-TEF $\beta$  phosphoryle (Figure 24 C) :

- 1) NELF, le dissociant de l'ARN pol II,
- 2) DSIF, le rendant activateur de l'élongation,
- 3) la sérine 2 du CTD de RBP1 (RBP1<sup>pSer2</sup>).

Cette nouvelle phosphorylation du CTD favorise le recrutement d'un ensemble de complexes permettant la maturation de l'ARN naissant ou la modification de la composition de la chromatine (Adelman and Lis, 2012). Le flavopiridole, inhibiteur de CDK9, perturbe la phosphorylation RBP1 à la sérine 2 et maintient l'ARN pol II en pause.

Durant l'élongation, l'ARN pol II transcrit à une vitesse variant entre 0.4 et 5 kb/min (Ben-Ari et al., 2010; Darzacq et al., 2007; Jonkers et al., 2014; Singh and Padgett, 2009; Wada et al., 2009). La vitesse de transcription a tendance à augmenter le long du corps du gène et la méthylation d'histone (H3K79me2). La présence d'une importante densité d'exon ou de régions riches en CG réduit la vitesse de transcription (Jonkers et al., 2014). Il est intéressant de noter que seulement 1% des transcriptions initiées mène à la production d'un ARNm totale. Plus précisément, seulement 13% des ARN pol II interagissant avec le promoteur sont libérés du PIC, et parmi eux seulement 8.6% réaliseront l'élongation (Darzacq et al., 2007).

#### 4. Promoteurs et GTFs contrôlent la spécificité de l'initiation de la transcription dépendant de l'ARN pol II

Environ 85% du génome d'humain est transcrit par l'ARN pol II et est à l'origine d'un grand nombre d'ARNnc (voir introduction, II). Bien qu'il existe de nombreuses régulations pour réguler sa spécificité, l'ARN pol II est présent au niveau de gène réprimé pour la transcription ainsi que des zones

intergéniques (Federico et al., 2020). La mutation de facteurs régulant la spécificité de transcription de l'ARN pol II entraîne l'apparition de transcription aberrante. Ainsi, chez *S.cerevisiae*, en absence de MOT1 et NC2, deux complexes régulateurs du chargement de TBP sur l'ADN, entraînent la formation du PIC au niveau de promoteur cryptique et la transcription de séquence intergénique (Koster et al., 2014). De plus, des mutations dans les machineries transcriptionnelles de l'ARN pol I ou III, ou des promoteurs de leurs gènes, entraînent le recrutement de l'ARN pol II à leurs promoteurs.

### i. Transcription des ARNr par l'ARN pol II

La transcription médiée par l'ARN pol I est initiée à partir d'un promoteur minimum constitué de deux éléments : un élément en amont (*upstream element* : UE) et un élément corps (*core element* : CE) (Figure 25 A). UE et CE sont reconnus tout d'abord par le GTFs UBF, vient alors le recrutement de SL1 (aussi nommé TIF-IB) et de TIF-IA au niveau de CE (Figure 25 A). Il a été montré que la mutation nulle des sous-unités *Uaf30* ou de *Rrn9* de UBF dans *S.cerevisiae*, empêche la liaison efficace de UBF au promoteur. Dans ce contexte mutant particulier, les ARNr sont alors transcrits par l'ARN pol II (Figure 25 B).

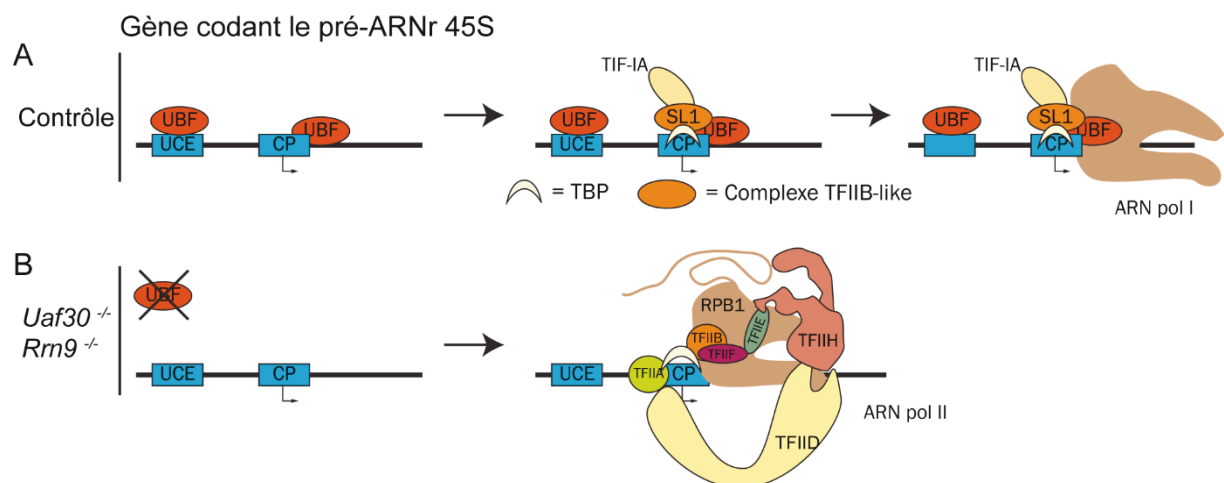


Figure 25 : Initiation de la transcription au promoteur des gènes 45S en condition normale (A) et en condition mutante nulle pour *Uaf30* et *Rrn9* entraînant la perte de UBF (B). En absence UBF, le pré-ARNr 45S est transcrit par la machinerie de transcription pol II. A noter que le PIC dessiné est hypothétique.

### ii. Transcription du snRNA U6 par l'ARN pol II

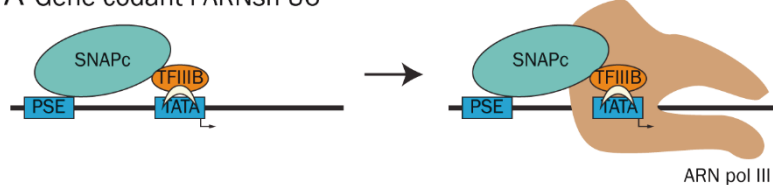
Le promoteur minimum des *snRNA U6* est composé d'une séquence PSE et une boîte TATA. La transcription est initiée par le recrutement d'une part de SNAPc au niveau de PSE et d'autre part de TFIIB, contenant TBP, au niveau de la boîte TATA. S'ensuit alors le recrutement de l'ARN pol III.

Contrairement au promoteur minimum du U6, le promoteur minimum situé en amont des gènes *snRNA* transcrit par l'ARN pol II sont dépourvus de boîte TATA. La présence ou non de la boîte TATA confère la spécificité de recrutement de l'ARN pol II ou III. L'ajout d'une boîte TATA dans le promoteur

minimum de U2 entraîne le recrutement de l'ARN pol III. La mutation de la boîte TATA dans le promoteur minimum de U6 entraîne le recrutement de l'ARN pol II (Dergai and Hernandez, 2019; Lobo and Hernandez, 1989).

Les mécanismes expliquant l'importance de la boîte TATA dans le recrutement spécifique de l'ARN pol III ne sont pas complètement élucidés. Il a cependant été montré que TFIIA et BRF2 interagissent avec TBP au même point de contact, et donc que leurs interactions sont mutuellement exclusives. Quand TBP est recruté par BRF2, il favorise le recrutement de BRF2 au promoteur à boîte TATA et empêche le recrutement de BRF2 au niveau des promoteurs sans boîte TATA. La présence simultanée de BRF2 au niveau de la boîte TATA et de SNAPc au niveau de PSE favorise la formation du PIC spécifique à l'ARN pol III. *A contrario*, quand TBP est recruté par TFIIA, TFIIA favorise le recrutement de TBP au niveau de SNAPc lié au PSE et empêche l'interaction de TBP avec BRF2. TFIIIB est également recruté au niveau de SNAPc au même point de contact utilisé par BRF2, ce qui permet 1) d'inhiber le recrutement de BRF2, 2) de stabiliser le complexe PSE-SNAPc-TFIIA-TBP, 3) la mise en place du PIC spécifique à l'ARN pol II (Dergai and Hernandez, 2019; Dergai et al., 2018).

**A Gène codant l'ARNsn U6**



**B Gène codant un ARNsn (ex : U1)**

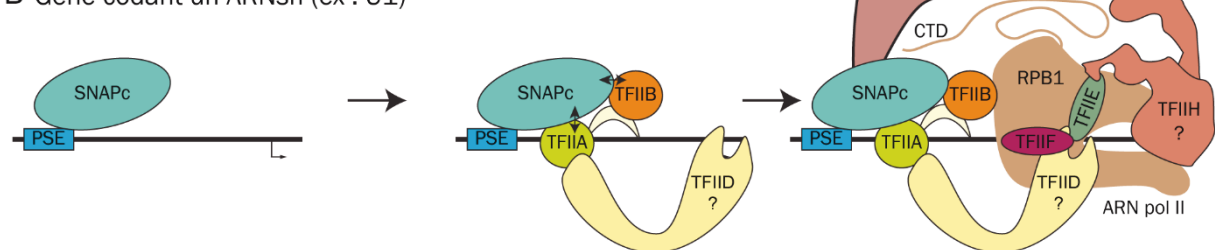


Figure 26 : La présence de la boîte TATA spécifie le recrutement de la machinerie de transcription ARN pol III (A). Dans le cas des snRNA U1 à U5, l'absence de boîte TATA entraîne le recrutement de la machinerie de transcription ARN pol II. A noter que le dessin du PIC spécifique à l'ARN pol II est hypothétique. En particulier, l'importance de TFIID dans la transcription des snRNA n'a pas été montrée.

## 5. Impact du *RNA buffering* sur l'étude des complexes impliqués dans la transcription

Au sein de la cellule l'abondance d'ARNs particuliers dépend de leur synthèse et de leur dégradation. Ce principe s'appelle le tamponnage des ARNs, ou *RNA buffering* (Hartenian and Glaunsinger, 2019). Dans ce cas, la synthèse et la dégradation des ARNs se régulent l'un l'autre : l'altération de la transcription est compensée par une diminution de la dégradation des ARNs et inversement. Ainsi, le niveau global d'ARNs total reste constant malgré les possibles perturbations de la transcription (Sun et al., 2012) (Figure 27, A-B). Ainsi, l'analyse des ARN totaux pour étudier le rôle d'un complexe ne permet pas d'avoir qu'une vision partielle de son importance dans la transcription (Figure 27 C-D). De nouvelles approches permettent dorénavant d'étudier le transcriptome naissant (Wissink et al., 2019).

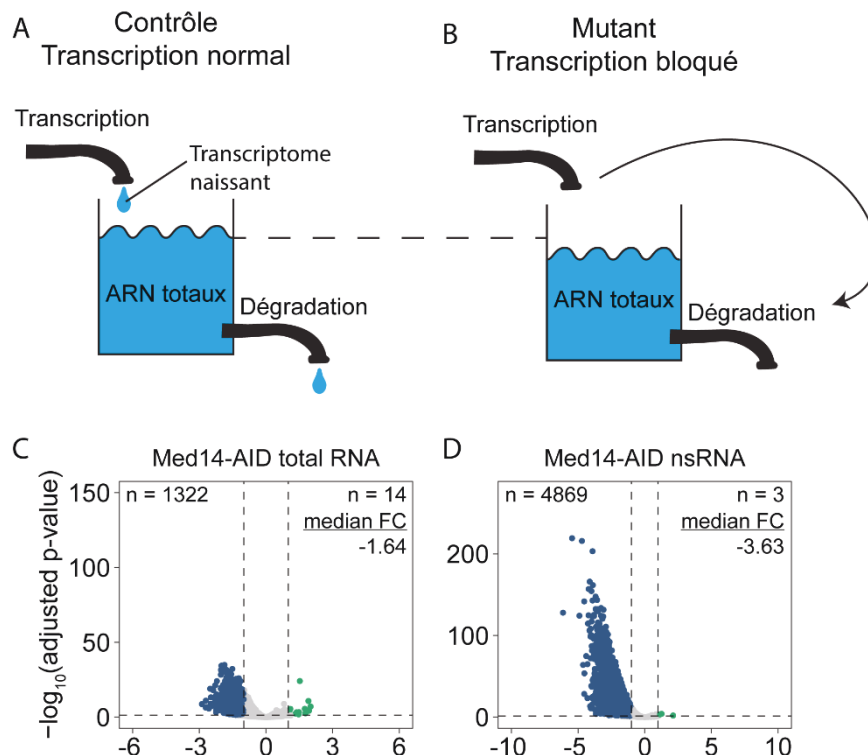


Figure 27 : Mécanisme du RNA buffering. Le niveau d'ARN total est un équilibre entre le niveau de transcription et le niveau de dégradation (A). Lorsqu'une mutation entraîne l'arrêt de la transcription, des mécanismes de régulation entraînent la diminution de la dégradation. Ainsi, le niveau total d'ARN reste constant. Issus de la thèse de Tiago Baptista. C-D) Analyse par séquençage, des ARN totaux (C) ou nouvellement synthétisés (D) après la déplétion conditionnelle de la sous-unité du médiateur MED14 chez *S.cerevisiae*. A gauche se trouvent les transcrits sous représentés par rapport au contrôle et à droite ceux sur représentés. Issu de Tourigny et al. 2021.

L'analyse des ARNs naissant est réalisée à partir de l'extraction des ARNs en cours de synthèse se trouvant à l'intérieur de l'ARN pol II. Pour cela, la technique est d'isoler la chromatine puis de fortement la laver afin de ne récupérer que les ARNs associés à la chromatine (*chromatine-associated RNA*). Cependant beaucoup d'ARNs n'étant pas associés aux ARN pol sont également récupérés. Une

autre technique repose sur l'immunoprécipitation des ARN pol II puis l'extraction des ARNs (*Pol II-associated RNA*) (Wissink et al., 2019).

L'analyse des ARN nouvellement synthétisés repose sur le marquage métabolique des ARNs en cours de synthèse durant un court intervalle de temps. Pour cela, plusieurs types de nucléotides modifiés peuvent être utilisés tels que le 4-thiouridine (4sU), le 5-bromouridine (BrU) ou le 5-ethenyl-uridine (EU) (Duffy et al., 2019). Par la suite, les ARNs marqués sont purifiés et séquencés. Dans le cas du marquage au 4sU, il est également possible de traiter chimiquement les ARNs marqués entraînant la conversion des 4sU incorporés en une base azotée ressemblant à la cytosine (C). Après le séquençage des ARNs totaux, la mutation de U en C permettra de séparer les ARNs marqués de ceux non marqués (Wissink et al., 2019). Dans la review (Wissink et al., 2019) est décrit un grand nombre d'approches permettant d'étudier la transcriptome naissant (Table 1).

Les mécanismes du *RNA buffering* reposent sur la communication entre les machineries de transcription d'une part et de dégradation et de traduction d'autre part. Chacun des facteurs de ces machineries sont capables d'être transloqué du cytoplasme au noyau et inversement. De ce fait, ils pourraient agir en tant que régulateur de la transcription ou de la dégradation en fonction du niveau globale d'ARN dans la cellule. Néanmoins, à l'heure actuelle, les facteurs impliqués dans le *RNA buffering* restent à être élucidés (Timmers and Tora, 2018).



## Introduction

### Mécanismes de la transcription médiée par l'ARN pol II

	Principe	Avantages	Considération
<b>caRNA-seq</b>	Lavage intensive de la chromatine Extraction des ARNs Séquençage	Peut être utilisé pour isoler toutes les espèces de chromatine-ARN associées à la chromatine Peut être combiné avec des méthodes qui testent les processus de co-transcription, y compris la méthylation et l'édition de l'ARN.	Séquence également les ARN non naissants qui sont stables et liés à la machinerie de transcription (ex : <i>snRNA</i> )
<b>Start-seq</b>	Lavage intensive de la chromatine Extraction des ARNs Trie des ARNs en fonction de la taille et de leur coiffe 5' Séquençage	Identifie simultanément les TSS et de pause Permet d'identifier de nouveaux <i>enhancers</i>	Ne détecte pas de transcription au-delà des premiers ~100 nucléotides
<b>Yeast NET-seq</b>	Immunoprécipitation de l'ARN pol II Extraction des ARNs	Détecte les ARNs présent dans l'ARN pol II. Identifie les positions de la Pol II à une résolution nucléotidique à l'échelle du génome.	Nécessite d'avoir une lignée de cellules avec l'ARN pol II marqué par un épitope.
<b>mNET-seq</b>	Immunoprécipitation de l'ARN pol II Extraction des ARNs	Détecte les ARNs présent dans l'ARN pol II. Identifie les positions de la Pol II à une résolution nucléotidique à l'échelle du génome. Peut isoler l'ARN pol II ayant différentes PTMs	Séquence également les ARN non naissants qui sont stables et liés à la machinerie de transcription (ex : <i>snRNA</i> ) Ne détecte pas les ARNs ayant moins de 30 nucléotides. Ne détecte que les eRNA d'enhancers déjà connus
<b>Pro-seq</b>	Perméabilisation des cellules Marquage avec des nucléotides biotinylés Extraction des ARNs Fragmentation Purification Séquençage	Capture les ARNs des polymérases compétentes sur le plan transcriptionnel. Identifie les positions de transcription active à une résolution nucléotidique et à l'échelle du génome. Permet d'identifier de nouveaux <i>enhancers</i>	Ne mesure pas le retour en arrière des ARN pol ( <i>backtracking</i> ) Capture également les ARN en cours de transcription à partir des ARN pol I et ARN pol III.
<b>Pro-cap</b>	Protocole pro-seq avec sélection des ARNs coiffés en 5'	Identifie les TSS Permet d'identifier de nouveaux <i>enhancers</i>	Ne détecte pas de transcription au-delà des premiers ~100 nucléotides
<b>CoPRO</b>	Variant du PRO-cap	Identifie simultanément les TSS et de pause Mesure l'état de coiffage de l'ARN	Ne mesure pas la transcription au-delà du site de pause proximal
<b>SMIT-seq</b>	Protocole similaire au caRNA seq mais avec une étape pour amplifier des séquences introniques	Mesure l'état de l'épissage pendant la transcription	Limité aux espèces à introns courts
<b>TT-seq</b>	Marquage des ARNs avec du 4-thiouridine Extraction et sonication des ARNs Purification des ARNs marqués Séquençage	Capture des ARNs des polymérases en cours de transcription. Peut être utilisé pour déterminer la stabilité de l'ARN Identifie les sites de terminaison de la transcription	Ne détecte pas les pauses de l'ARN pol II Ne détecte que les eRNA d'enhancers déjà connus
<b>SLAM-seq TimeLapse-seq</b>	Marquage des ARNs avec du 4-thiouridine Extraction des ARNs Traitement chimique permettant l'oxydation des 4sU incorporés en C. Séquençage : l'analyse bio-informatique permet de différencier les ARNs marqués porteur de la mutation des autres ARNs non marqués.	Capture les ARN des polymérases en transcription active. Peut être utilisé pour déterminer la stabilité de l'ARN	Nécessite un séquençage profond pour mesurer le taux de conversion chimique Les longs temps de marquage ne permettent pas de capturer les ARNs qui viennent d'être synthétisés

Table 1 : Récapitulatif de quelques techniques associées au séquençage permettant d'analyser le transcriptome naissant. *caRNA-seq*, chromatin-associated RNA sequencing; *CoPRO*, coordinated precision run-on and sequencing; *eRNA enhancer RNA*; *mNET-seq*, mammalian native elongating transcript sequencing; *NET-seq*, native elongating transcript sequencing; *PRO-cap*, precision run-on with cap selection; *PRO-seq*, precision run-on sequencing; *SLAM-seq*, thiol (SH)-linked alkylation for the metabolic sequencing of RNA; *SMIT-seq*, single-molecule intron tracking sequencing; *TT-seq*, transient transcriptome sequencing. Issu de (Wissink et al., 2019).



## IV. Description du 'Transcription Factor IID' (TFIID) : son évolution, sa composition et son rôle dans la transcription

### 1. Récapitulatif brève de la découverte de TFIID

Les premières évidences de l'apparition de TFIID, ainsi que des autres GTFs, sont issues de l'étude de Weil *et al.* en 1979, soit 10 ans après la mise en évidence des trois ARN pol (Roeder and Rutter, 1969; Weil et al., 1979). Dans cette étude, les auteurs mettent en évidence l'extrait cellulaire S-100 contenant des facteurs permettant d'initier précisément la transcription à partir d'un fragment d'ADN de l'Adénovirus Ad2. Un an plus tard, cet extrait a été séparé en 4 fractions A, B, C et D par chromatographie sur colonne de phosphocellulose (Figure 28). Des essais de transcription *in vitro* permettent de mettre en évidence l'importance des fractions A, C et D pour que l'ARN pol II initie la transcription à un TSS unique (Matsui et al., 1980). Par la suite, il a été montré que la fraction D contenait d'autres facteurs tel que TFIID (Flores et al., 1992).

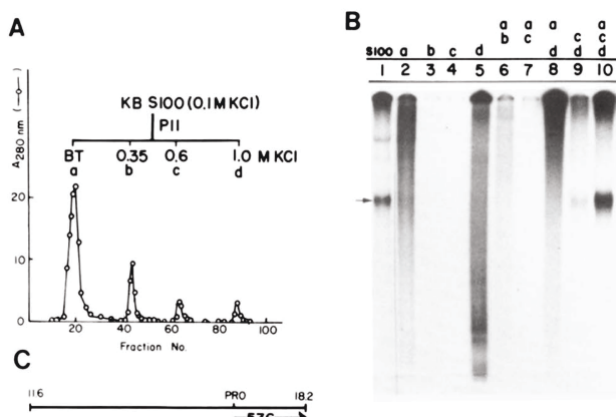


Figure 28 : Première évidence de la présence d'un complexe TFIID. A) Profils d'absorbance A280nm pour les fractions A, B, C et D obtenus par chromatographie sur colonne de cellulose de l'extrait de cellules humaines KD S100 B) Essai de transcription *in vitro* utilisant l'extrait S100 (1) ou différentes combinaisons de fractions A, B, C et D. C) Schéma du fragment d'ADN Ad2 contenant le promoteur à boîte TATA Major Late Promoter (PRO). Issu de Matsui et al. 1980

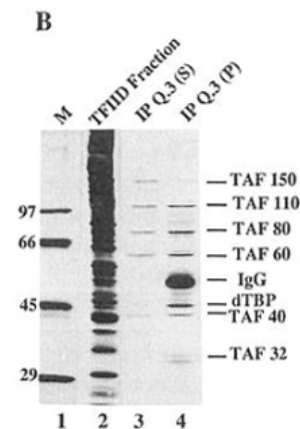


Figure 29 : Première évidence du caractère multiprotéique de TFIID. La fraction TFIID avant immunoprécipitation (2) et l'élution de l'immunoprécipitation anti-TBP (3-4) ont été analysées par silver staining. Issu de Dynlacht et al., 1991

De 1980 à 1991, la composition de TFIID n'est pas sûre. Ainsi, il a été même considéré comme une protéine de liaison à la boîte TATA (Kao et al., 1990; Peterson et al., 1990). En 1991, les travaux de Dynlacht *et al.* ont découvert que TFIID est un complexe de protéines (Dynlacht et al., 1991) (Figure 29). TFIID a alors été décrit comme constitué de TBP (*TATA-binding protein*) et des protéines TAFs (*TBP-associated factors*). Les TAFs ont été décrits comme des co-facteurs nécessaires pour activer la transcription au niveau de certains promoteurs via l'induction des facteurs de transcription (Dynlacht et al., 1991). Dans les années qui ont suivi, les gènes des *Tafs* dans différents modèles biologiques ont été découverts. Dans chaque modèle des noms différents ont été données aux TAFs complexifiant la

compréhension générale de TFIID dans les différents modèles. En 2002, une nomenclature a été définie par Dr Laszlo Tora dans le but d'unifier le nom des protéines TAFs (Tora, 2002) (Tableau 6).

**Table 1.** New Pol II TAF nomenclature including the corresponding known orthologs and paralogs

New name	<i>H. sapiens</i> (hs)	<i>D. melanogaster</i> (dm)	<i>C. elegans</i> (ce)		<i>S. cerevisiae</i> (sc)	<i>S. pombe</i> (sp)
			previous name	new name		
TAF1	TAF <sub>II</sub> 250	TAF <sub>II</sub> 230	<i>taf-1</i> (W04A8.7)	<i>taf-1</i>	Taf145/130	TAF <sub>II</sub> 111
TAF2	TAF <sub>II</sub> 150	TAF <sub>II</sub> 150	<i>taf-2</i> (Y37F11B.4)	<i>taf-2</i>	Taf150 or TSM1	(T38673)
TAF3	TAF <sub>II</sub> 140	TAF <sub>II</sub> 155 or BIP2	(C11G6.1)	<i>taf-3</i>	Taf47	
TAF4	TAF <sub>II</sub> 130/135	TAF <sub>II</sub> 110	<i>taf-5</i> (R119.6)	<i>taf-4</i>	Taf48 or MPT1	(T50183)
TAF4b	TAF <sub>II</sub> 105					
TAF5	TAF <sub>II</sub> 100	TAF <sub>II</sub> 80	<i>taf-4</i> (F30F8.8)	<i>taf-5</i>	Taf90	TAF <sub>II</sub> 72
TAF5b						TAF <sub>II</sub> 73
TAF5L	PAF65β	Cannonball				
TAF6	TAF <sub>II</sub> 80	TAF <sub>II</sub> 60	<i>taf-3.1</i> (W09B6.2)	<i>taf-6.1</i>	Taf60	(CAA20756)
TAF6L	PAF65α	(AAF52013)	<i>taf-3.2</i> (Y37E11AL.8)	<i>taf-6.2</i>		
TAF7	TAF <sub>II</sub> 55	(AAF54162)	<i>taf-8.1</i> (F54F7.1)	<i>taf-7.1</i>	Taf67	TAF <sub>II</sub> 62/PTR6
TAF7L	TAF2Q		<i>taf-8.2</i> (Y111B2A.16)	<i>taf-7.2</i>		
TAF8	(BAB71460)	Prodos	(ZK1320.12)	<i>taf-8</i>	Taf65	(T40895)
TAF9	TAF <sub>II</sub> 32/31	TAF <sub>II</sub> 40	<i>taf-10</i> (T12D8.7)	<i>taf-9</i>	Taf17	(S62536)
TAF9L	TAF <sub>II</sub> 31L (AAG09711)					
TAF10	TAF <sub>II</sub> 30	TAF <sub>II</sub> 24	<i>taf-11</i> (K03B4.3)	<i>taf-10</i>	Taf25	(T39928)
TAF10b		TAF <sub>II</sub> 16				
TAF11	TAF <sub>II</sub> 28	TAF <sub>II</sub> 30β	<i>taf-7.1</i> (F48D6.1)	<i>taf-11.1</i>	Taf40	(CAA93543)
TAF11L			<i>taf-7.2</i> (K10D3.3)	<i>taf-11.2</i>		
TAF12	TAF <sub>II</sub> 20/15	TAF <sub>II</sub> 30α	<i>taf-9</i> (Y56A4.3)	<i>taf-12</i>	Taf61/68	(T37702)
TAF13	TAF <sub>II</sub> 18	(AAF53875)	<i>taf-6</i> (C14A4.10)	<i>taf-13</i>	Taf19 or FUN81	(CAA19300)
TAF14					Taf30	
TAF15	TAF <sub>II</sub> 68					

Tableau 5 : Tableau issu de Tora, L. 2002 proposant une nomenclature unifiée pour les TAFs pour l'ensemble des modèles.

## 2. Histoire évolutive de TFIID

TFIID est constitué d'un ensemble de TAFs dont l'origine provient de pre-LECA (*Last Eukaryotic Common Ancestor*) (Antonova et al., 2019). On suppose que le complexe TFIID ancestrale (aTFIID), si il existait, était composé des TAFs ancestraux (aTAF) 1, 2, 4, 5, 6, 7, 8, 9, 10, 12 ainsi que aSPT3 et aTBP. Des duplications génétiques successives ont permis de faire apparaître au fur et à mesure l'ensemble des TAFs que l'on connaît à l'heure actuelle. Du fait de l'apparition progressive des homologues TAFs, certaines ne sont présent que dans un embranchement particulier. *Taf1l* est spécifique des *Cercopithecidae* aussi connus sous le nom de « singes de l'ancien monde », *Taf9b* des placentaires, *Taf7l* des mammifères, *Taf4b* des vertébrés. *Taf4x* est issu de la duplication de *Taf4b* mais a été perdu chez les tétrapodes. Ainsi ils ne sont présents que chez les *Actinopterygii* et *Sarcopterygii*, les poissons à nageoires rayonnées et charnues respectivement. *Taf3* provient de la duplication de *aTaf8* de tel sorte que les plantes (*Archeplastidia*) n'ont pas de protéine TAF3. Les plantes à fleurs ont quant à elle une protéine EER4 homologue à TAF12 (Antonova et al., 2019). La mouche *D.melanogaster* semble être l'une des seule à présenter un paralogue à *Taf10* : *Taf10b* (Pahi et al., 2017) (ensembl). Les clades des souris et des rats ont la particularité d'avoir un paralogue *Taf7l2*, indiquant une possible duplication de *Taf7* chez l'ancêtre commun *Myomorpha* (Kojima et al., 2019) (ensembl) (Figure 30).



TFIID partage son histoire évolutive avec le complexe coactivateur SAGA. Chez les mammifères, SAGA partage 3 sous-unités avec TFIID, contre 8 estimés chez pré-LECA (Figure 30). Cette différenciation entre SAGA et TFIID est due à l'apparition d'un ensemble de paralogues se spécialisant soit comme sous-unités de SAGA ou de TFIID. Entre le pré-LECA et LECA, *aTaf4* s'est dupliqué faisant apparaître la sous-unité spécifique de SAGA *Ada1*, TAF4 se spécialisant pour TFIID. Durant la même période, *Spt3*, *Taf11* et *Taf13* sont apparus du même gène ancestral, le premier étant spécifique à SAGA et les deux autres à TFIID. La création de *Supt7l* à partir de *Taf8* s'est réalisée soit avant l'apparition des eucaryotes soit chez l'ancêtre commun *Unikonta*. Enfin, à l'apparition des métazoaires, la duplication de *aTaf5* et *aTaf6* donnèrent naissance aux sous-unités *Taf5* et *Taf6* spécifiques à TFIID et *Taf5l* et *Taf6l* spécifiques à SAGA (Antonova et al., 2019). Ainsi, chez les métazoaires, seul TAF9, TAF10 et TAF12 sont partagés entre TFIID et SAGA (Bardot et al., 2017; Herbst et al., 2021). Chez la levure *S.cerevisiae* en plus de ces trois TAFs, TAF5 et TAF6 sont également partagés entre ces deux complexes (Papai et al., 2020; Taatjes, 2017) (Figure 30)

### 3. Structure de TFIID

Les TAFs sont des protéines principalement structurales dépourvus d'activité enzymatique (Wassarman and Sauer, 2001). Grâce à un ensemble de domaines protéiques (Figure 31) (Müller et al., 2010), les TAFs interagissent ensemble et sont à l'origine de la structure flexible de TFIID (Chen et al., 2021a; Louder et al., 2016; Patel et al., 2018) (Figure 33 A). En particulier, 8 TAFs ont des HFDs permettant de former spécifiquement les paires TAF3-TAF10 (TAF3-10), TAF8-10, TAF6-9, TAF4-13 et TAF11-13 (Bieniossek et al., 2013a; Chen et al., 2021c; Louder et al., 2016; Patel et al., 2018; Soutoglou et al., 2005) (Figure 31).

TFIID est formé de trois lobes nommés A, B et C (Figure 33 B, C, D). Le lobe A et B sont tous deux formés d'un tétramère composé de TAF4-12, TAF6-9. Ce tétramère est en contact avec les domaines NTD2 et WD40 de TAF5 (Bieniossek et al., 2013a; Chen et al., 2021a; Patel et al., 2018) (Figure 32). De ce fait, les lobe A et B sont structurellement similaires (Figure 32 A et B). Néanmoins dans le lobe A contient l'hétérodimère TAF3-10 alors que le lobe B contient TAF8-10 (Chen et al., 2021c; Patel et al., 2018) (Figure 32 A et B). Au sein du lobe A, TAF11-13 est également présent contactant d'une part TBP et de l'autre TAF4-12 (Figure 32 B). L'interaction TAF6-9, TAF4-12, TAF3-10 et TAF11-13 forment un octamère à la structure proche de l'octamère d'histone (Figure 32 B et C). Quelques différences visibles tels que le désaxement de TAF11-13 est légèrement désaxé par rapport à H2A et H2B (Patel et al., 2018) (Figure 32 B et C).

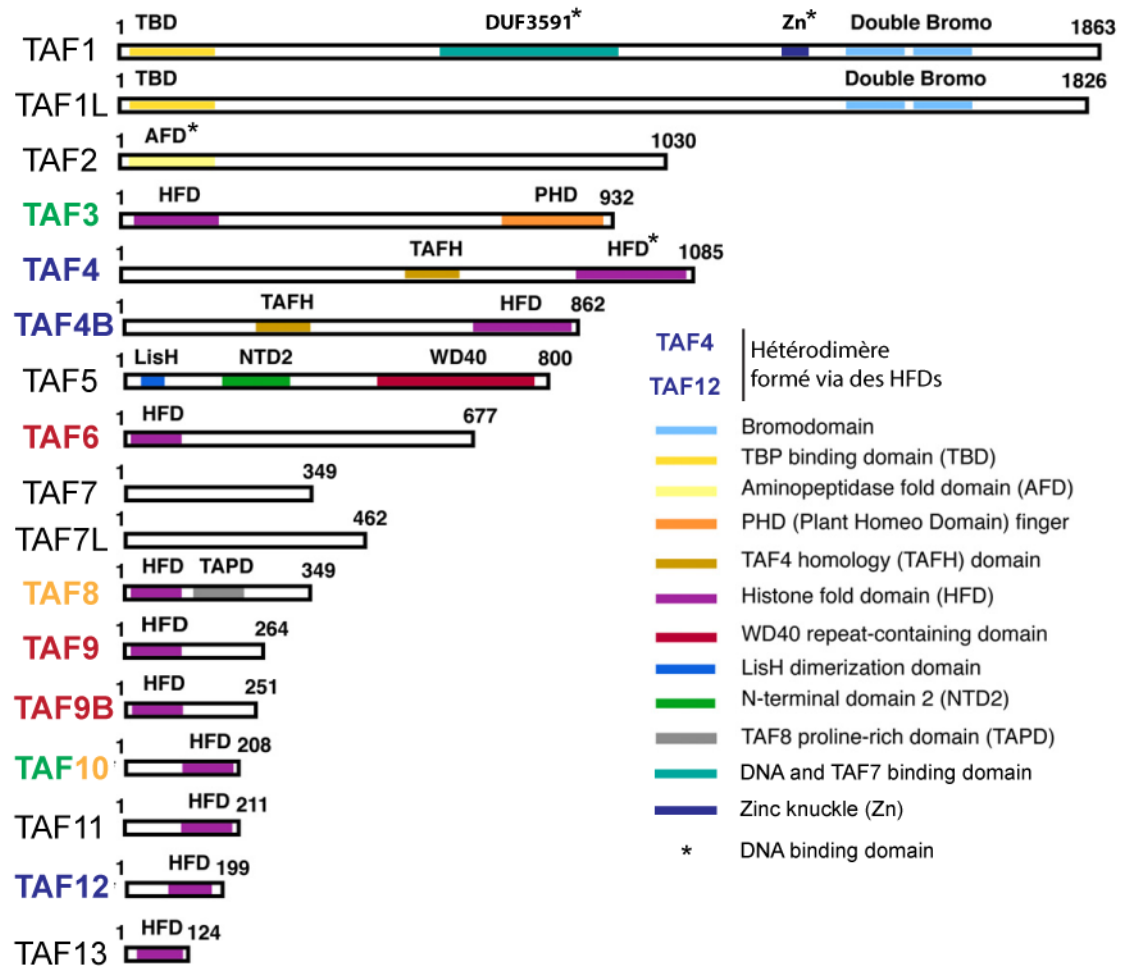


Figure 31 : Récapitulatif des domaines présents au niveau de chaque TAFs présent dans TFIID chez les métazoaires. Les TAFs de même couleur s'associe via les HFD pour former des hétérodimères. Adapté de Müller et al. 2010.

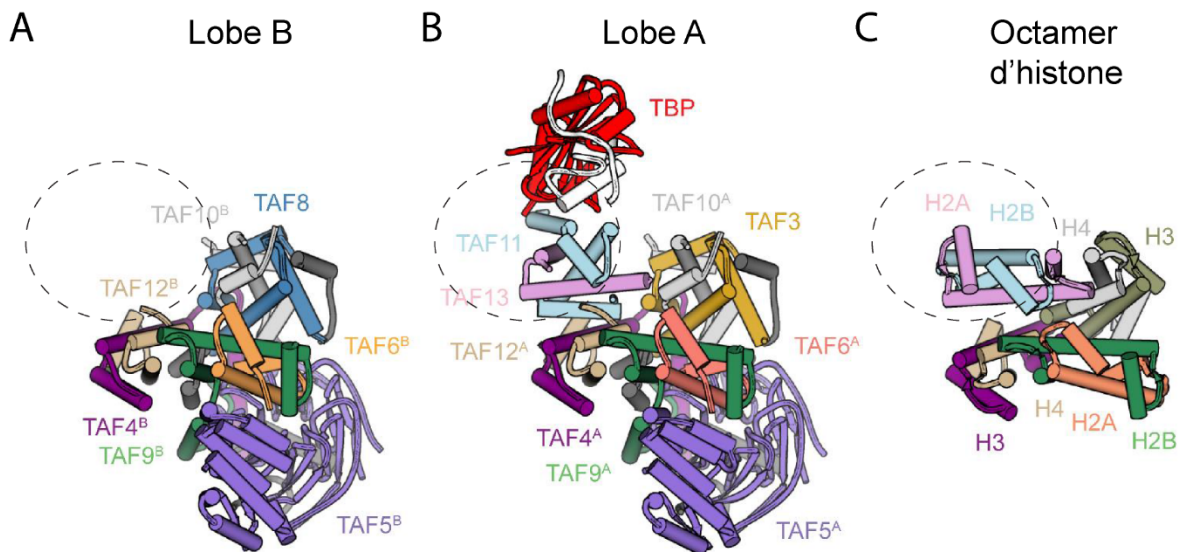


Figure 32 : Le lobe B (A) et B (A) et l'octamer d'histone présentent des similitudes structurales. Le cercle en pointillé montre une des différences entre ces trois structures. Issu de Patel et al. 2018.



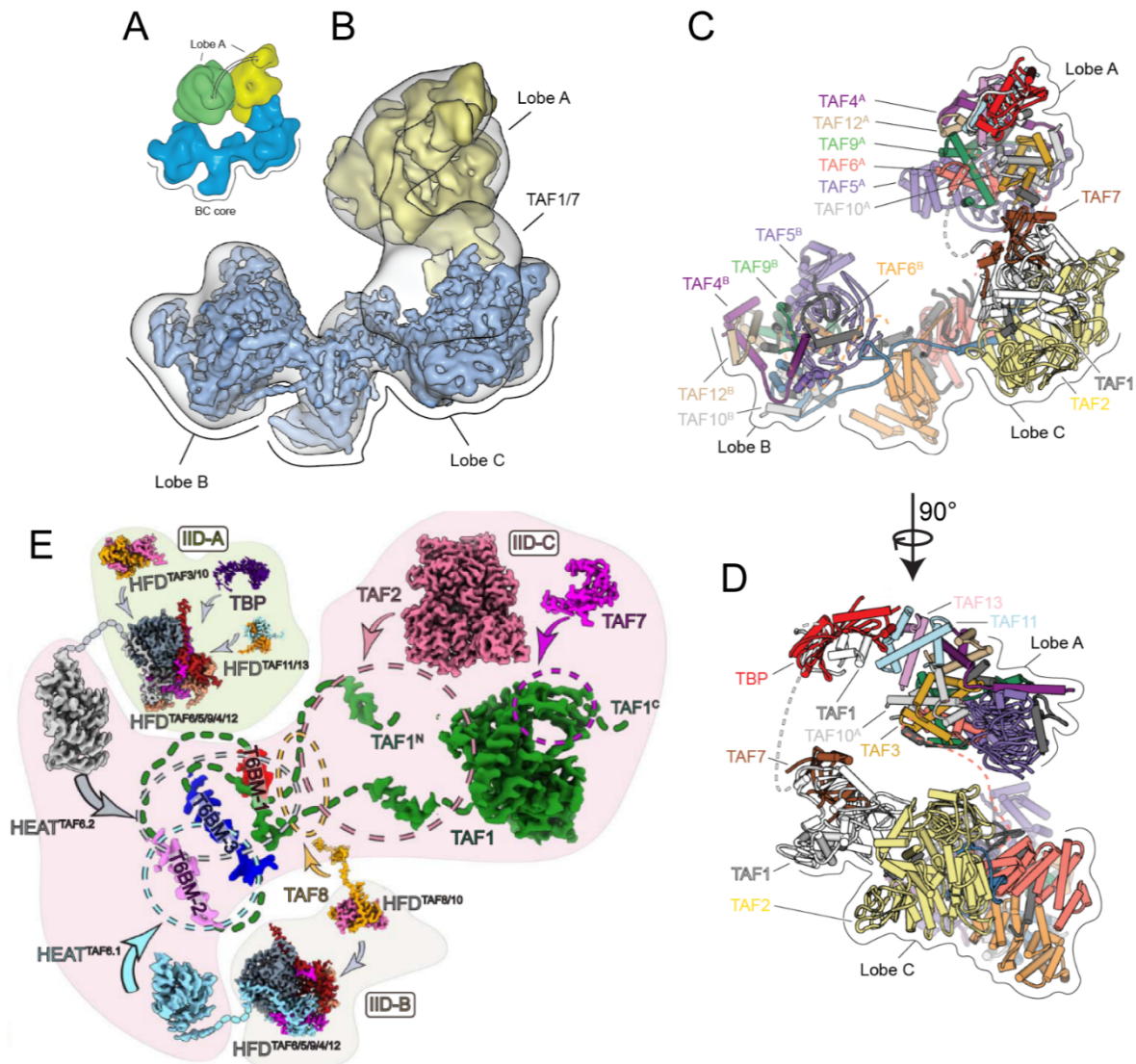


Figure 33 : Vue d'ensemble de la structure de TFIID. A-D) Structure de TFIID immunoprécipité à partir de cellules HeLa et résolue par Cryo-EM. En (A) est illustré la mobilité du lobe A par rapport au lobe B et C. (D) est une vue à 90° dans le sens indiqué par la flèche de (C). Issu de Patel et al. 2018. E) Représentation de l'ensemble des interactions au sein de TFIID. Issu de Chen et al. 2021.

Le lobe A et B sont liés via les deux domaines HEAT de chaque TAF6 qui les compose (HEAT<sup>TAF6.A</sup> issu de TAF6 du lobe A et HEAT<sup>TAF6.B</sup> issu du TAF6 du lobe B) (Chen et al., 2021a; Kolesnikova et al., 2018; Louder et al., 2016; Patel et al., 2018) (Figure 33 C et E). Ces HEAT dimérisés sont stabilisés par TAF8 et TAF1 (Chen et al., 2021a; Patel et al., 2018) (Figure 33 E). TAF8 a une expansion qui va le long des domaines HEAT dimérisés (Chen et al., 2021a; Louder et al., 2016) (Figure 33 C et E). Son extrémité C-terminal est associée à TAF2 présent dans le lobe C (Chen et al., 2021a; Louder et al., 2016; Patel et al., 2018; Trowitzsch et al., 2015) (Figure 33 D et E). TAF1 contacte les deux domaines HEAT via trois motifs de liaison à TAF6 (*TAF6-binding motifs*, *T6BMs*) (Figure 33 E) (Chen et al., 2021a). TAF1 contacte par ailleurs TBP via ses domaines TAND1 et 2 (Anandapadamanaban et al., 2013; Liu et al., 1998) et avec TAF7 via son domaine DU3591 (Bhattacharya et al., 2014; Wang et al., 2014)

(Figure 33 C, D et E). Par ces deux points de contact (TAF1-HEAT<sup>TAF6</sup> et TAF8-TAF2), le double lobe B-C (bilobe BC) forme une structure rigide alors que lobe A conserve une certaine flexibilité par rapport au reste du complexe (Chen et al., 2021a; Patel et al., 2018) (Figure 33 A).

## 4. Assemblage de TFIID

D'après le modèle actuel, l'assemblage de TFIID est séquentiel. En premier lieu, un ensemble de modules TAF se forme au sein du cytoplasme. Les dimérisations de certains sous-unités de TFIID sont réalisés soit co-traductionnelles telles que TAF8-TAF10 (TAF8-10), TAF6-9 ou TAF1-TBP (Figure 34, Etape 1), ou de façon indépendante à la traduction, telles que TAF3-10 (Kamenova et al., 2019). D'autre part, certains TAFs vont être repliés à l'aide de chaperonne tels que le domaine WD40 de TAF5 via le complexe TRiC/CCT (Antonova et al., 2018).

L'assemblage de TFIID suit avec la formation de sous module TFIID (Figure 34, Etape 2). La formation du core-TFIID débute par le recrutement de TAF6-9 au niveau du WD40 de TAF5. Deux modules TAF5-6-9 ainsi que deux hétérodimères TAF4-12 s'assemblent alors pour former le core-TFIID, présentant une symétrie axiale (Bieniossek et al., 2013b). D'autre part, TAF2 est recruté au niveau de l'extrémité C-terminal de TAF8 de l'hétérodimère TAF8-10 formant le module TAF2-8-10 (Soutoglou et al., 2005; Trowitzsch et al., 2015). Il a été montré que TAF7 est associé à TAF11-13 dans le cytoplasme (Gupta et al., 2017). TAF7 interagit avec TAF1 (Bhattacharya et al., 2014; Wang et al., 2014). Il a du coup été suggéré qu'un complexe TAF1-7-11-13 existe peut-être comme sous-module mais cela reste à être montré (Patel et al., 2020).

L'assemblage du complexe TFIID continue par la formation du module 8-TAFs. Par TAF8, le module TAF2-8-10 est associé au core-TFIID (Demény et al., 2007a; Scheer et al., 2021) (Figure 34, étape 3). Le recrutement spécifique de TAF8-10 au niveau du lobe B est induit par les contraintes physiques imposés par l'asymétrie de l'interaction des domaines HEATs<sup>TAF6</sup> (Chen et al., 2021a). Cette incorporation entraîne une réorganisation du core-TFIID : le pré-lobe B se balance vers le pré-lobe A d'environ 40 Å° et le pré-lobe A de 14 Å° vers le pré-lobe B (Bieniossek et al., 2013b). Cette réorganisation a été suggérée comme important pour le recrutement des autres modules. Ainsi, TAF3-10, TAF11-13, TAF1-TBP et TAF7 (avec TAF1 ou TAF11-13) s'assemble sur le module 8TAFs formant TFIID (Figure 34, étape 3).

Certaines études suggèrent que l'assemblage final de TFIID est réalisé dans le noyau (Bieniossek et al., 2013a; Patel et al., 2020; Trowitzsch et al., 2015). Cela concorde avec le fait que TAF10 requière TAF8 ou TAF3 pour être importé dans noyau (Soutoglou et al., 2005). Cependant, seul TAF3, TAF8 et TAF1 semblent présenter une signal de localisation nucléaire (*nuclear localization signal, NLS*) (Soutoglou et al., 2005, uniprot). De ce fait, l'importation de TAF7, TAF11-13 et du coreTFIID au sein du noyau reste à être élucidé.

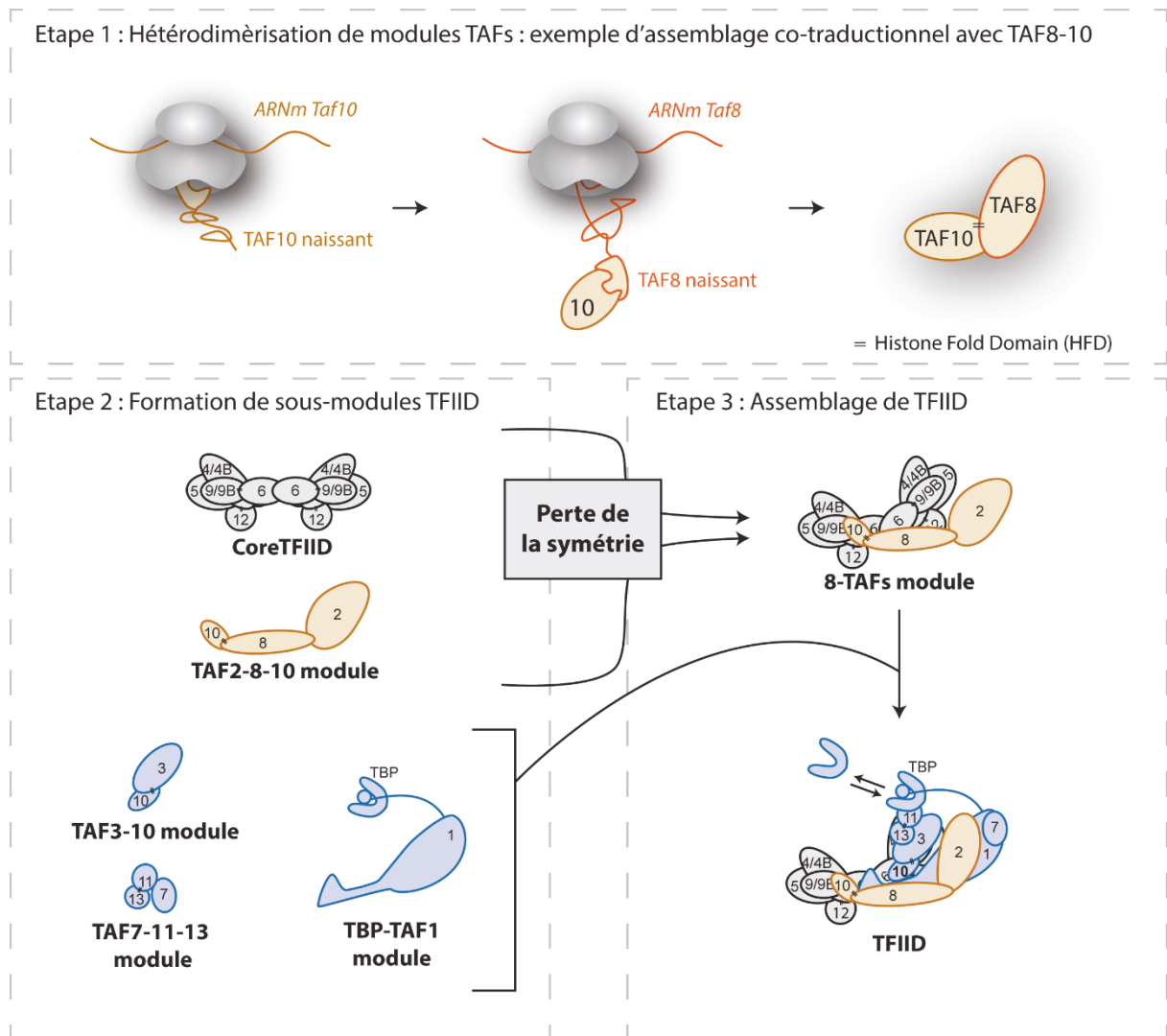


Figure 34 : Assemblage séquentiel de SAGA

## 5. TFIID et SAGA proche parenté

### i. Structure de SAGA

Le complexe SAGA (*Spt-Ada-Gcn5 acetyltransferase*) est composé de 19 à 20 sous-unités, en fonction des espèces, réparties en 5 modules : le module histone acetyltransferase (HAT), un module histone déubiquitineuse (DUB), le module TRRAP, et le module *splicing* (SPL) ainsi que le module corps autour duquel les 4 premiers modules sont positionnés (Helmlinger and Tora, 2017; Herbst et al., 2021) (Figure 35).

Ce module corps contient 3 TAFs partagés avec TFIID : TAF9/9B, TAF10 et TAF12 (Herbst et al., 2021). Ils forment les 3 hétérodimères TAF6L-9/9B, TAF10-SUPT7L, TAF12-TADA1, dont les interactions sont médiées par leurs HFDs. Dans ce corps est également présent SUPT3H, qui contient deux HFDs qui se dimérisent entre eux. Ces 4 appariements basés sur les HFDs forment un pseudo-octamère déformé à la structure différentes de l'octamère d'histone (Herbst et al., 2021). Tout comme dans TFIID, ce pseudo-octamère est en contact proche avec le domaine WD40 de TAF5L. Enfin, le



corps également formé de la protéine SUPT20H qui parcourt tout le corps SAGA servant de squelette. Sa présence est important pour le maintien de TRRAP, SPL et HAT dans SAGA (Herbst et al., 2021; Nagy et al., 2009) (Figure 35).

La composition et la structure de SAGA de *S.cerevisiae* sont différentes de celles du SAGA de mammifère (Herbst et al., 2021; Papai et al., 2020). En particulier, une protéine SPT8 est également présente au niveau du corps de SAGA, permettant le recrutement de TBP. Cette interaction est également favorisée par le contact de TBP avec les HFDs de SPT7 (orthologue de SUPT7L). L'absence de SPT8 ainsi que le fait que les HFDs de SUPT7L soient repliés l'un sur l'autre fait que TBP n'est pas recruté par SAGA chez les mammifères (Fischer et al., 2021a; Papai et al., 2020) (Figure 35).

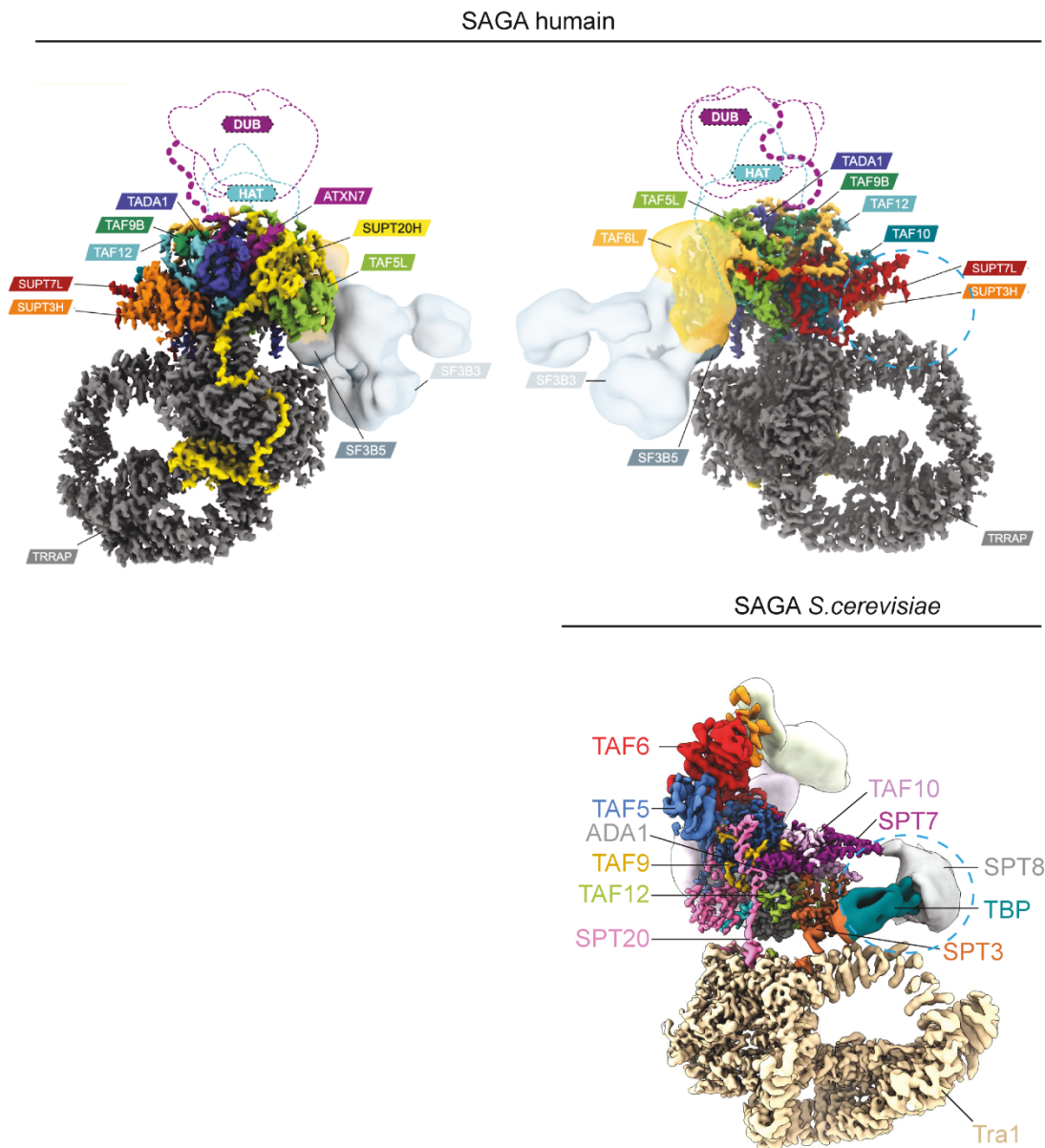


Figure 35 : Structure du complexe SAGA humain (haut) et de *S.cerevisiae* (bas). Le cercle pointillé bleu indique le lieu de liaison de TBP dans le complexe SAGA de *S.cerevisiae*. Adapté de Herbst et al., 2021 & Papai et al., 2020.

## ii. Fonction de SAGA

SAGA régule la transcription par sa capacité à acétyler les histones, via son domaine HAT, ou à les déubiquitiner, via son domaine DUB (Fischer et al., 2019; Helmlinger and Tora, 2017). Chez *S.cerevisiae* SAGA a tout d'abord été décrit comme régulant la transcription de gènes en lien avec la réponse au stress (Huisinga, 2005). Néanmoins, de plus récentes études ont montré que SAGA est requis pour l'expression de l'ensemble des gènes chez *S.cerevisiae*. D'une part, SAGA est retrouvé aux promoteurs de tous les gènes transcrits. D'autre part, les mutations nulles de *Stp20<sup>-/-</sup>* et *Spt7<sup>-/-</sup>*, deux sous-unités du module corps, entraînent une réduction globale de la transcription des ARNm (Baptista et al., 2017).

Chez les mammifères, SAGA ne semble pas être requis pour l'expression de l'ensemble des gènes. La mutation de différentes sous-unités de SAGA entraîne la morte tardive des embryons (Tableau 7). Chez les mESC, la perte conditionnelle de SUPT20H (orthologue de SPT20) ou de SUPT7L (orthologue de SPT7) n'entraîne pas la réduction totale de la transcription. SAGA a été décrit comme important pour stabiliser le programme d'expression génétique nécessaire au maintien de la pluripotence des mESC (Fischer et al., 2021b). De plus les cellules mutantes présentent des phénotypes différents en fonction du milieu de culture utilisé (Fischer et al., 2021b). En particulier, en condition 2i+lif+serum les cellules mutantes présentent un phénotype moins important que lorsqu'elles sont cultivées en milieu lif+serum. Le 2i (2 *inhibitor*) correspond à deux molécules : CHIR99021 et PD0325901. Elles inhibent les voies de signalisations de MEK/ERK et GSK3 impliquées la répression de l'expression des facteurs de transcriptions de pluripotences (Martello and Smith, 2014). D'autre part, le 2i entraîne une hypométhylation de la chromatine et augmente la quantité de 5hmC (Huisinga, 2005). Ainsi, la fonction de SAGA semblerait être dépend de l'épigénome. Ensemble ces données suggèrent que, contrairement à SAGA chez *S.cerevisiae*, la fonction de SAGA chez les mammifères est spécifique à l'expression de gènes particuliers dans un contexte particulier.

	Mutant	Se développe jusqu'à	Référence
<b><i>Supt20h</i></b>	Mutant hypomorphe	9,5 jpc	(Warrier et al., 2017; Zohn et al., 2006)
<b><i>Supt3h</i></b>	Mutant nul	14,5 jpc	(Perez-Garcia et al., 2018)
<b><i>Usp22</i></b>	Mutant nul	13,5 jpc	(Wang et al., 2021)
<b><i>Atxn7l3</i></b>	Mutant nul	7,5 jpc.	(Wang et al., 2021)
<b><i>Trrap</i></b>	Mutant nul	7,5 jpc	(Herceg et al., 2001)

Tableau 6 : Récapitulatif de quelques mutants murins

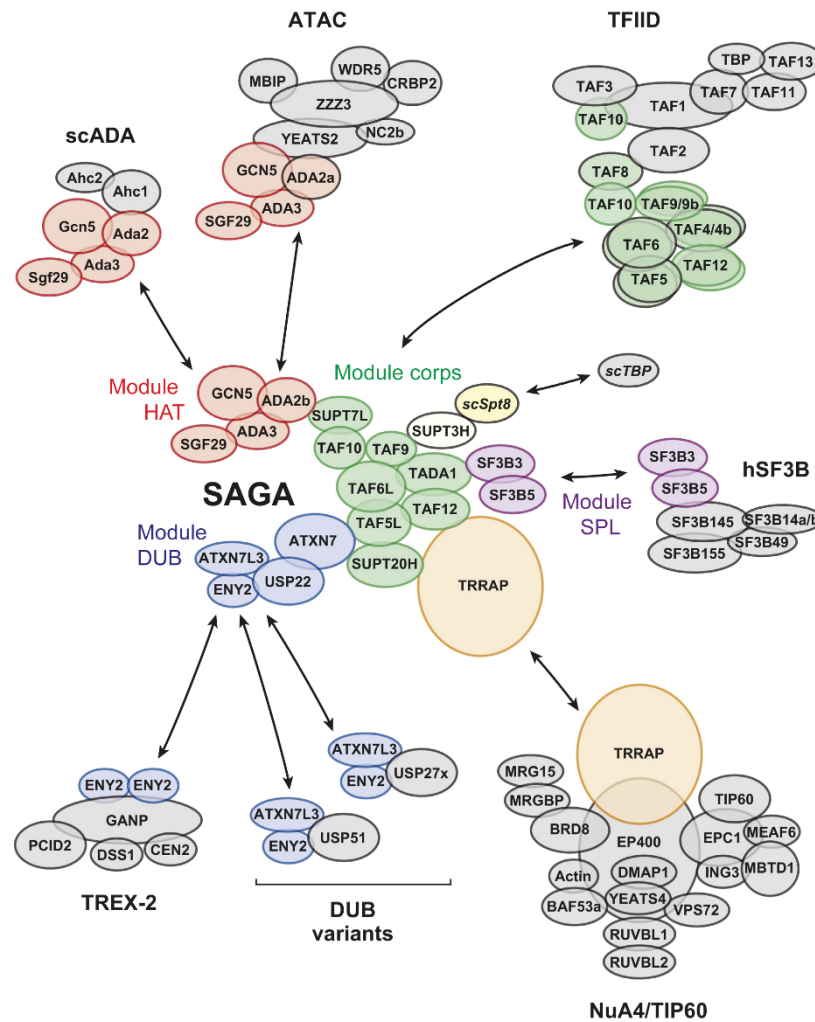


Figure 36 : SAGA partage ses sous-unités avec de nombreux complexes participant eux aussi à la mise en place et/ou la régulation de la transcription. Issu de Helmlinger and Tora, 2017.

L'étude du rôle de SAGA dans la transcription est difficile car d'une part ce dernier partage de nombreuses sous-unités avec d'autres complexes, et d'autre part d'autres complexes sont impliqués dans l'acétylation et la déubiquitination des histones.

- Le module DUB

Le module DUB de SAGA entraîne la déubiquitination de H2B mono-ubiquitiné (Helmlinger and Tora, 2017). La sous-unités catalytique du module DUB à l'origine de la déubiquitination est yUbp8/hUSP22 (Zhang et al., 2008). Chez les levures, *Ubp10* est une autre déubiquitinase indépendante de SAGA. La mutation nulle de *Ubp10* ou *Ubp8* entraîne une augmentation importante H2Bub (Bonnet et al., 2014; Emre et al., 2005). Néanmoins, alors que l'activité de UBP10 est impliquée dans la répression de la transcription, l'activité de UBP8 permet l'activation de la transcription.

Chez les cellules mammifères, l'activité de déubiquitination est également supportée par deux autres complexes DUB indépendants de SAGA. Ces deux complexes contiennent ATXN7L3 et ENY2 ainsi

que la déubiquitineuse USP27X ou USP51. (Helmlinger and Tora, 2017) (Figure 36). La mutation nulle de *hUsp22*, *hUsp27x* ou *hUsp51* n'entraîne pas une modification importante du niveau de H2Bub. Seul la mutation nulle d'*Atxn7l3* ou des trois déubiquitineuses entraîne une augmentation significative de H2Bub (Atanassov et al., 2016). Il semblerait que les trois complexes agiraient en concert, régulant l'activité de déubiquitination aux niveaux de loci particuliers. Par conséquent ils auraient des activités différentes que ce soit au niveau de la régulation de la transcription, de la maturation des ARNs ou bien de la réplication de l'ADN (Atanassov et al., 2016).

- Le module HAT

Le module HAT de SAGA acétyle principalement les histones H3 aux résidus K9 et K14 (Helmlinger and Tora, 2017). Chez *S.cerevisiae*, GCN5 est l'acétylase du module HAT de SAGA, mais il est également retrouvé dans le complexe ADA (Helmlinger and Tora, 2017) (Figure 36). La mutation nulle de *Gcn5*<sup>-/-</sup> entraîne une diminution de H3K9ac (Baptista et al., 2017; Bonnet et al., 2014). En absence de GCN5 le niveau globale de la transcription est diminué ce qui suggère le rôle important de l'acétylation des histones par GCN5 pour l'initiation de la transcription (Baptista et al., 2017).

Chez les cellules mammifères, GCN5 a un paralogue nommé PCAF qui peut également se trouver dans SAGA (Helmlinger and Tora, 2017). La double mutation nulle *Gcn5*<sup>-/-</sup> ; *Pcaf*<sup>f/-</sup> chez des MEFs entraîne une disparition complète de H3K9ac mais pas H3K14ac (Jin et al., 2011). GCN5 et PCAF ne sont pas équivalents dans leur activité acétylase : alors que les mutants nuls *Pcaf*<sup>f/-</sup> se développent sans problème apparent, indiquant qu'il n'est pas requis pour l'embryogenèse, la mutation nulle de *Gcn5* est létale au stade embryonnaire, des problèmes développementaux apparaissant à partir de E7.5 (Xu et al., 2000). GCN5 et PCAF sont par ailleurs partagés avec le complexe ATAC (Helmlinger and Tora, 2017). ATAC et SAGA ne présentent pas eux non plus de redondance complète au niveau de leur activité d'acétylation et de régulation de l'expression des gènes (Fischer et al., 2021b; Helmlinger and Tora, 2017) (Figure 36). Bien que les deux complexes semblent importants pour maintenir l'état d'auto-renouvellement des mESC, l'induction de la perte conditionnelle de sous-unités spécifiques à l'un ou l'autre des complexes diminue l'accessibilité de la chromatine, le niveau globale de H3K9ac et l'expression de différents gènes (Fischer et al., 2021b). Ainsi chez les mESC SAGA et ATAC seraient impliqués dans la régulation de l'expression de gènes particuliers.

- TRRAP

scTra1/hTRRAP est la sous-unité la plus grosse du complexe (~420kDa). Il est partagé avec le complexe de scNuA4/hTIP60 à l'origine de l'acétylation de H4, H2A et H2A.Z (Figure 36). Que ce soit chez *S.cerevisiae* ou l'humain, scTra1/hTRRAP a été montré comme étant capable de se lier à différents facteurs de transcription (Brown et al., 2001; McMahon et al., 1998). Chez *S.cerevisiae*, la mutation *scTra1* l'empêchant de se lier au facteur de transcription GCN4p entraîne à la fois une diminution de l'acétylation de H3 et de H4 au niveau des promoteurs des gènes dont l'expression est stimulée par

GCN4P ainsi qu'une diminution de l'expression de ces gènes (Brown et al., 2001). De la même façon, chez les mESC TRRAP semble être recruter au niveau de OCT4 pour maintenir l'expression de gènes de pluripotence (Sawan et al., 2013). La mutation nulle conditionnelle de *Trrap* diminue l'acétylation de H3K9 et H4, favorisant la mise en place de H3K27me3, une marque répressive pour la transcription, au niveaux des promoteurs des gènes de pluripotences, entraînant une diminution de l'expression de ces gènes et favorisant par conséquent la différenciation des mESCs (Sawan et al., 2013).

- Le module SPL

Le module SPL n'est présent que dans les complexes SAGA présents chez la Drosophile ainsi que ceux présents dans les mammifères (Fischer et al., 2021a; Helmlinger and Tora, 2017). Ce module pourrait donc être apparu chez l'ancêtre des métazoaires. La fonction de ce complexe n'a pour l'instant pas été étudiée.

## 6. Rôle de TFIID dans l'initiation de la transcription :

TFIID est considéré comme le premier GTFs à être recruté au niveau des gènes transcrit par l'ARN pol II (Voir chapitre III).

### 1) Cibler le site d'initiation de la transcription.

Le recrutement de TFIID au niveau des promoteurs dépend d'une part de l'épigénome. H3K4me3 et l'acétylation de H3 et H4 sont des marques d'activation de la transcription retrouvées au niveau du nucléosome +1 (Ernst et al., 2011; Heintzman et al., 2007). Chez les mammifères, H3K4me3 permet de recruter TFIID via TAF3 et facilite la mise en place du PIC à tout type de promoteur (van Ingen et al., 2008; Lauberth et al., 2013; Vermeulen et al., 2007). La phosphorylation de H3T3 entraîne la dissociation de TAF3 de H3K4me3 (Varier et al., 2010). Cette découverte permet de comprendre l'importance de la régulation de l'épigénome dans le recrutement de TFIID. Les deux bromodomaines de TAF1 permettraient également le recrutement de TFIID au promoteur (Jacobson et al., 2000). En particulier, la présence de H3K9ac et de H3K14ac au niveau de H3K4me3 favorise le recrutement de TFIID (Vermeulen et al., 2007).

D'autre part les facteurs de transcription favorisent le recrutement de TFIID. Par exemple, le facteur de transcription p53 permet de recruter et de stabiliser TFIID au niveau des promoteurs des gènes dont p53 active la transcription (Coleman et al., 2017). Pour autre exemple, l'interaction de TAF10 avec GATA1 est requise l'expression du programme d'expression génétique nécessaire à l'érythropoïèse chez la souris (Papadopoulos et al., 2015). L'interaction entre TAF4B et le facteur de transcription ZFP628 est requis pour la spermatogenèse chez la souris (Gustafson et al., 2020). Dans les cellules cancéreuses de cancer colorectale, TAF4B interagit avec différents facteurs de transcription de la famille *Jun* régulant l'expression de l'*integrin α6* (Varier et al., 2010).

Enfin, plusieurs TAFs sont impliqués dans la liaison à l'ADN. En particulier, TAF1 reconnaît les promoteurs minims des régions aux niveaux des séquences MTE, DPE et DCE (Lee et al., 2005; Louder et al., 2016; Wang et al., 2014). Le domaine *winged-helix* (WH) de TAF1 contacte DPE/MTE (Louder et al., 2016; Wang et al., 2014). TAF1 a également un domaine *CCHC zinc knuckle* (ZnK) qui contacte l'Inr (Curran et al., 2018; Louder et al., 2016). Chez *S.cerevisiae*, l'extrémité C-terminal de TAF1, non conservée chez l'humain, est également une zone de liaison à l'ADN (Mencía and Struhl, 2001). L'interaction de TFIID avec l'Inr est également réalisée par TAF2 (Chalkley and Verrijzer, 1999; Louder et al., 2016). D'autres TAFs ont été suggérés comme interagissant avec l'ADN tels que l'hétérodimère TAF4-12 (Gazit et al., 2009)

## 2) Chargement de TBP dans la double hélice d'ADN

L'analyse par Cryo-EM de complexes PIC entiers ou partiellement assemblés a permis de mettre en évidence l'importance fonctionnelle de TFIID dans le chargement de TBP dans l'ADN. Deux études différentes ont permis de mettre en évidence deux modèles pour le chargement de TBP relativement similaire (Chen et al., 2021a; Patel et al., 2018).

- Le modèle Nogales (Patel et al., 2018)

Le modèle Nogales explique comment TFIID charge TBP au niveau d'un promoteur contenant une boîte TATA, une séquence Inr, DPE et MTE. Dans ce cas, 6 états de TFIID ont été décrits correspondant à six conformations du lobe A par rapport au reste du bilobe BC : 'canonique', 'étendu', 'scanning', 'réarrangé' et 'engagé'.

- Canonique et étendu

Dans l'état canonique et étendu, le lobe A se déplace librement entre le lobe B et C. TBP est en contact avec d'une part TAF11-13 sur un côté et d'autre part avec le domaine TAND2 de TAF1 de l'autre côté. De plus, le domaine TAND1 de TAF1 se trouve à l'intérieur du domaine de liaison à l'ADN (Gupta et al., 2017; Liu et al., 1998; Patel et al., 2018). Sous cette forme, TBP ne peut donc ni interagir à l'ADN ni avec TFIIA (Patel et al., 2018) (Figure 37 A a et b).

En présence de séquences ressemblant à une séquence Inr, MTE et DPE, le lobe C, formé de TAF1-2-7 rentre en contact avec l'ADN. Le *high mobility group* (HMG) de TAF1 contacte l'ADN au niveau de l'Inr dans le sillon mineur. TAF1 contacte également, avec TAF2, le MTE et le DPE au niveau des deux sillons de la double hélice (Figure 37 B c et d, étoile). TAF1 est rapporté comme vraiment flexible par rapport à son point d'accrochage qui est TAF6, optimisant sa liaison à l'ADN (Chen et al., 2021a; Louder et al., 2016). Ce contact du lobe C avec l'ADN va être consolidé avec la liaison du lobe B à l'ADN. TAF4 contacte l'ADN juste en amont de la boîte TATA via une boucle située entre la troisième et la quatrième l'hélice  $\alpha$  (respectivement  $\alpha_3$  et  $\alpha_4$ ) (Figure 37 D, I, TAF4).



- *Scanning* et réarrangé

L'interaction de TFIID avec l'ADN déplace le TAND1 du domaine de liaison à l'ADN de TBP. Ainsi TBP est en contact proche avec l'ADN (Figure 37 A, c). TFIID est dans l'état *scanning* (Figure 37 A c). TFIIA est alors recruté au niveau de TBP, déplaçant le TAND2 de TAF1. TFIID est alors sous l'état réarrangé (Figure 37 A d). Cette nouvelle interaction stabilise le contact entre le lobe A et le lobe B. TBP s'insère dans le sillon mineur de l'ADN au niveau de la séquence TATA.

- Engagé

TFIID est alors sous la forme engagée : TBP se trouve au niveau du lobe B, interagit avec TFIIA et est inséré dans l'ADN plié. L'insertion de TBP dans l'ADN rompt le contact de TBP avec TAF11-13 (Figure 37 B e). Ceci permet le recrutement de TFIIB et le début de la formation du PIC (Figure 37 A f).

Le modèle Nogales a été mis en lumière en utilisant un promoteur synthétique (SCP : *Super Core Promoter*) contenant la boîte TATA, l'Inr, le DPE et le MTE (Louder et al., 2016). Or ce genre de promoteur n'est pas retrouvé, ou très rarement, au sein du génome (voir chapitre III.1.)

- Le modèle Xu (Chen et al., 2021a)

Le modèle de Xu a déterminé comment TFIID charger TBP dans l'ADN à partir de promoteur endogène ne contenant qu'un ou deux séquences permettant le recrutement de TFIID. Le modèle est assez proche du modèle de Nogales, à quelques exceptions près.

Tout d'abord, TFIIA est indiqué comment capable de se lier à TBP sans la présence d'ADN (Figure 37 B a, b). TFIID passe alors d'une conformation canonique dans laquelle le lobe A est proche du lobe C (Figure 37 B a), à une conformation réarrangée dans laquelle le lobe A est proche du lobe B et TBP interagit avec TFIIA (Figure 37 B b). Sous conformation canonique ou réarrangée, TFIID est recruté au niveau du corps promoteur. TFIID est alors juste posé sur l'ADN. Contrairement au modèle de Nogales, le lobe A est décrit comme pouvant se déplacer librement du lobe B à C.

Le chargement de TBP ainsi que l'assemblage du PIC dépend par la suite du type de promoteur présent. Dans tous les cas, la formation du PIC est initiée avant que TBP soit chargée dans l'ADN. TFIIB est alors recruté au niveau de TBP au niveau du lobe B. Puis, TFIIF et l'ARN pol II sont également recrutés, formant un PIC intermédiaire (Figure 37 C).

Au niveau d'un SCP, le lobe C est verrouillé au niveau des séquences Inr, MTE et DPE et TBP est chargé au niveau de la boîte TATA. L'ADN est alors plié d'un anglais de 45°. TFIID reste en contact avec l'ARN pol II. TBP et le lobe C sont séparés de ~32pb (Figure 37 C I et D I).

Dans le cas d'une absence de boîte TATA, le complexe TFIID est repositionné de tel façon que TBP va scanner la séquence en amont et se lier à une séquence similaire à une boîte TATA (Figure 37 C IIa et

D IIa). De la même façon, en absence des séquences Inr, DPE, le lobe C va scanner la séquence se trouvant en aval, jusqu'à trouver une séquence permettant au lobe C de s'ancrer (Figure 37 C IIb et D IIb). Dans ces deux cas, l'engagement de TBP dans l'ADN va courber l'ADN d'un angle de 100°. TFIID n'est plus en contact direct avec l'ARN pol II. TBP et le lobe C sont séparés de ~42pb.

Dans les trois cas, TBP et TFIID n'ont plus qu'un contact indirect via l'interaction entre TAF4-TFIIA-TBP.

### 3) Recruter le GTFs

TFIID joue un rôle important dans le recrutement de GTFs. Ainsi la déplétion conditionnelle de TAF12 au sein des mESC perturbe la localisation de TFIIA, TFIIB et le médiateur au niveau des promoteurs (Sun et al., 2021). Il est intéressant de noter que la modification post-traductionnelle des TAFs permet de réguler la mise en place du PIC. Par exemple, lors de la différenciation des cellules F9 en endoderme pariétal, la méthylase SET9 est présent au niveau des promoteurs des gènes dont l'expression est activée durant la différenciation. Lors du recrutement de TFIID à ces promoteurs, TAF10 va être méthylé par SET9 ce qui favorise la transcription de ces gènes (Kouskouti et al., 2004). Ceci peut suggérer que la méthylation de TAF10 favorise la mise en place du PIC.

### 4) Favorise la phosphorylation de la Ser2 et Ser7 de CTD<sup>RPB1</sup>

La présence de TFIID au sein du PIC favorise le positionnement du module CAK de TFIIH, contenant CDK7, près de l'ARN pol II. La résolution de la structure de PIC contenant TBP seul, sans TFIID, montrait précédemment un autre positionnement ne favorisant pas ce contact. La présence de TFIID favorise donc le contact de CTD<sup>RPB1</sup> au niveau de CDK7 et par conséquent favorise sa phosphorylation au niveau de la RPB1<sup>Ser2</sup> et de la RPB1<sup>Ser7</sup>, ce que TBP seul ne fait pas (Chen et al., 2021a).

En conclusion, TFIID intervient tout au long de la mise en place du PIC, de la reconnaissance du promoteur à la phosphorylation de RPB1 nécessaire à l'initiation de l'élongation. De ce fait, il permet de coordonner les différentes étapes de l'initiation de la transcription.



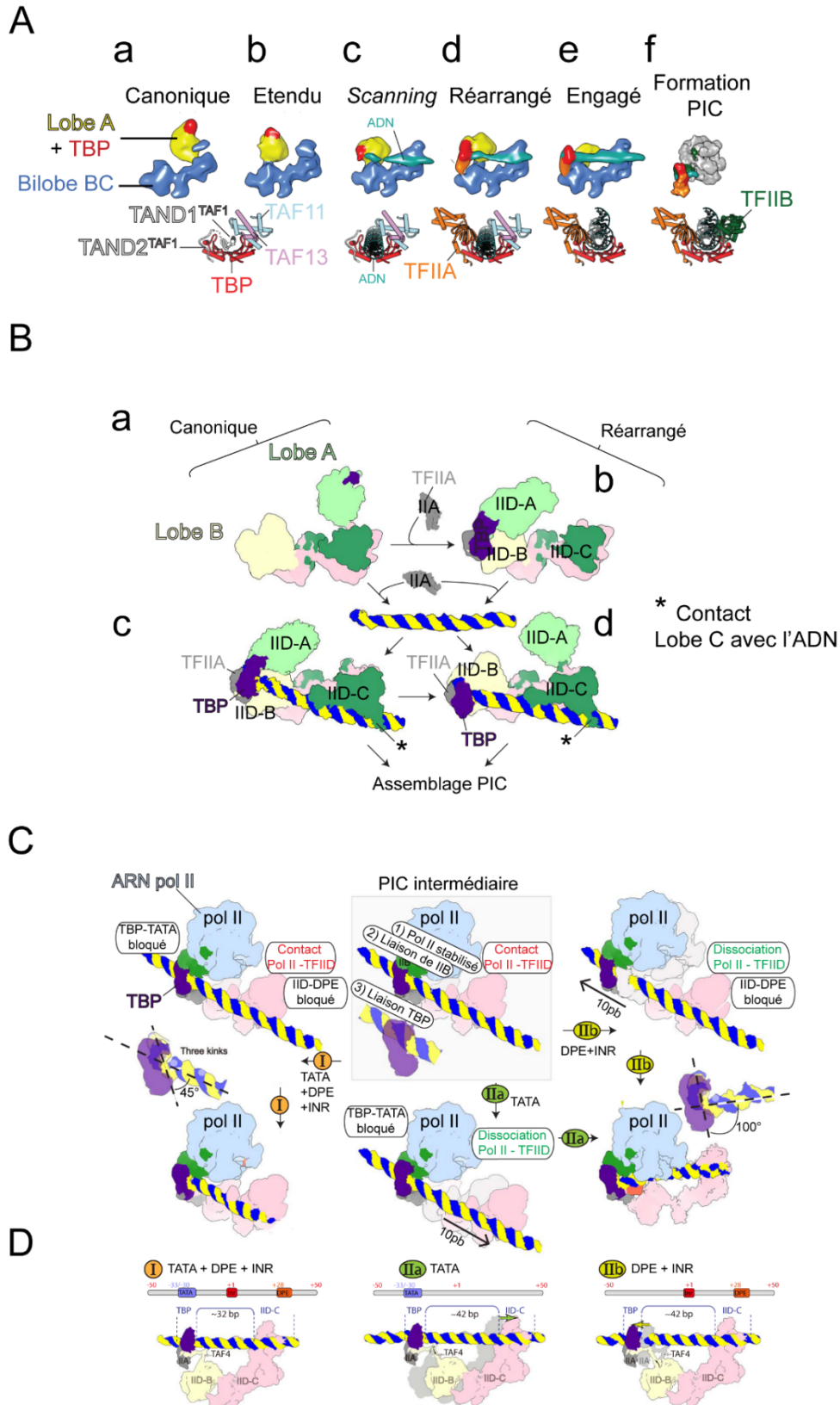


Figure 37 Schéma expliquant le recrutement de TFIID au niveau des promoteurs, le chargement de TBP dans l'ADN et le recrutement des premiers GTFs en fonction du modèle de Nogales (A) ou de Xu (B, C et D). Adaptés de Patel et al. 2018 et Chen et al. 2021.

## 7. Importance fonctionnelle des TAFs.

### *iii. Les TAFs sont requis pour le développement embryonnaire chez la souris*

L'importance des TAFs *in vivo* a été montrée par l'étude d'embryons de souris mutés pour *Taf7*, *Taf8* et *Taf10* (Gegonne et al., 2012; Mohan II et al., 2003; Voss et al., 2000). La mutation nulle de chacune de ces protéines bloque le développement embryonnaire 4,5 jours après la fécondation (E). Il est intéressant de noter que les embryons murins délétés en *Tbp* arrêtent également de se développer vers E4,5 (Martianov et al., 2002a) indiquant l'importance d'un TFIID complet pour les premières étapes de du développement embryonnaire murins.

Les mutants *Taf8* et *Taf10* présentent des phénotypes relativement similaires (Mohan II et al., 2003; Voss et al., 2000). Ils éclosent de la zonae pellucidae et s'implantent. Les boutons embryonnaires de ces mutants présentent un taux important d'apoptose et par conséquent la production de mESC à partir des blastocystes mutants n'est pas possible. La similitude des phénotypes des mutants *Taf8*<sup>-/-</sup> et des mutants *Taf10*<sup>-/-</sup> peut être mis en lien avec le fait que TAF10 est requis pour maintenir l'intégrité de TAF8 (Kamenova et al., 2019). Les mutants nuls zygotiques *Taf7*<sup>-/-</sup> ont été bien moins décrits. Contrairement aux deux autres mutants, les mutants *Taf7*<sup>-/-</sup> ne semblent pas être capable de s'implanter. Les auteurs n'ont pas regardé si le niveau d'apoptose était augmenté chez les mutants au niveau du bouton embryonnaire ou si il était possible d'en générer des mESC (Gegonne et al., 2012).

### *iv. Rôle des paralogues au TAFs dans l'établissement d'un programme d'expression génétique particulier.*

#### 1) TAF4 vs TAF4B

La génération de mutants zygotiques *Taf4*<sup>-/-</sup> ou *Taf4b*<sup>-/-</sup> a permis de mettre en évidence que ces deux protéines n'ont pas de fonction complètement redondante. Les embryons *Taf4*<sup>-/-</sup> sont capables de se développer jusqu'à E8.5 (Langer et al., 2016). Cependant dès E6,5, les embryons mutants commencent à se distinguer par une taille plus petite, indiquant que le développement commence à être perturbé. L'analyse de la composition de TFIID à partir des mESC dérivés des mutants a permis de montrer que TAF4B est intégré dans TFIID en absence de TAF4. Ainsi, il a été suggéré que TAF4B peut compenser la perte de TAF4 pour maintenir la viabilité des mESC et initier les premières étapes précoces du développement embryonnaire. Par la suite, dans les stades tardifs du développement, TAF4 est requis pour la mise en place des programmes d'expression génétique nécessaire à la différenciation cellulaire (Langer et al., 2016).

La mutation nulle de *Taf4b* est viable. Cependant, les souris deviennent rapidement infertiles. Les mâles *Taf4b*<sup>-/-</sup> deviennent progressivement infertile vers 11 semaines. TAF4B est exprimé dans les spermatogonies. Il est nécessaire pour leur prolifération et l'expression des gènes impliquées dans la spermatogénèse (Falender et al., 2005a). Chez les femelles, TAF4B est exprimé dans les ovocytes. Les femelles *Taf4b*<sup>-/-</sup> sont infertiles dès la naissance. TAF4B a été montré comme requis pour la mise en place des follicules et la reprise de la méiose après la fécondation (Falender et al., 2005b).

## 2) TAF9 vs TAF9B

Chez les mammifères, TAF9 et TAF9B ont en partie des fonctions spécifiques. La perte de TAF9B chez les mESC n'affecte pas leur capacité d'auto-renouvellement et les souris *Taf9b*<sup>-/-</sup> sont viable et fertile (Herrera et al., 2014). TAF9B semble spécifiquement important pour mettre en place le programme d'expression génétique lors de la différenciation neuronale. Il est cependant important de noter que dans ce processus de différenciation, TAF9B fonctionne principalement au sein de SAGA (Herrera et al., 2014). A l'heure actuelle, il n'existe pas de souris *Taf9*<sup>-/-</sup>.

Au sein des cellules HeLa, TAF9 et TAF9B sont tous deux requis pour maintenir la viabilité cellulaire. Néanmoins, la fonction de TAF9 et TAF9B ne se chevauchent pas complètement : la réduction de leur quantité au sein des cellules perturbe l'expression de gènes différents (Frontini et al., 2005).

## 3) TAF7 vs TAF7L

TAF7 et TAF7L ont des fonctions spécifiques. Alors que la mutation nulle *Taf7*<sup>-/-</sup> est létale au stade blastocyste, la mutation nulle de *Taf7l*<sup>-/-</sup> est viable. TAF7L est requis d'une part pour la spermatogénèse. Durant la spermatogénèse, TAF7L est transloqué dans le noyau à un moment précis de la différenciation cellulaire et remplace TAF7 dans TFIID (Pointud, 2003). La perte de TAF7L perturbe l'expression de gènes spécifiquement exprimé dans les cellules germinales et réduit la production et la mobilité des spermatozoïde (Zhou et al., 2013). TAF7L est également requis pour l'adipogenèse. Il est impliqué dans la mise en place du programme d'expression génétique nécessaire à la différenciation. Dans ce contexte, TAF7L est lié à TBP, et par conséquent potentiellement à TFIID (Zhou et al., 2013).

## v. *L'importance de TAF10 et de TAF7 dépend du contexte cellulaire et du programme d'expression génétique.*

### 1) TAF7

TAF7 semble être requis pour la transcription dans des contextes cellulaires particuliers. La différenciation des lymphocytes T à partir de leurs progéniteurs se fait en deux étapes. TAF7 a été montré comme requis pour la première étape de différenciation, mais pas pour la deuxième étape. Il est cependant important à noter que la déplétion de TAF7 n'était pas totale lors de ces expériences. Les progéniteurs ayant passé la première étape de différenciation ne présentaient qu'un faible nombre de gènes dérégulés. Ainsi dans ce contexte TAF7 n'est pas requis pour l'expression de l'ensemble des gènes (Gegonne et al., 2012). Au sein des mESC, TAF7 n'est pas requis pour maintenir leur état de pluripotence. Il est intéressant de noter que cette particularité lui est spécifique, les autres TAFs 1, 3, 4, 4B, 5, 6, 9, 11, 12 et 13 ont été montrés eux comme requis pour maintenir cet état pluripotent (Pijnappel et al., 2013).

### 2) TAF10

Chez les mammifères, TAF10 n'a pas de paralogue. Les études de mutants dans lesquelles *Taf10* est délété ont démontré que l'importance fonctionnelle de TAF10 dans la survie et la différenciation cellulaire dépend du contexte cellulaire. Ainsi, alors que la mutation nulle zygotique de *Taf10* entraîne la mort du bouton embryonnaire, les cellules trophoblastiques de ces mêmes embryons survivent un certain temps en culture (Mohan II et al., 2003). De plus, TAF10 a été montré comme facultatif pour certaine différenciation cellulaire. En absence de TAF10, les cellules F9 de carcinome embryonnaire meurent. Néanmoins, le traitement de ces cellules à l'acide rétinoïque les maintient en vie durant la différenciation en type endodermique primitive (Metzger et al., 1999). *A contrario*, ces mêmes cellules mutantes ne peuvent pas se différencier en type endoderme pariétale lorsqu'elles sont traitées avec de l'acide rétinoïque et du AMP cyclique dibutyryle (AMPC). Dans un contexte *in vivo* il a également été montré que la délétion conditionnelle de *Taf10* dans les kératinocytes chez la souris adulte n'a pas d'effet sur sa capacité de régénération de la peau. Mais là encore, la même délétion réalisée chez le fœtus empêchait la différenciation des kératinocytes (Indra et al., 2005). Le dernier exemple en date est la délétion conditionnelle de *Taf10* au sein des mésodermes embryonnaires. La délétion est initiée peu avant 7 jours après la fécondation et TAF10 est absent des mésodermes à partir de 8,5 jours après la fécondation (Bardot et al., 2017; Perantoni et al., 2005). Ces mutants sont capables de se développer jusqu'à 9.5 jours après la fécondation. Chez ces mutants, le mésoderme des plaques latérales n'est pas capable de former de bourgeon de membres alors que le mésoderme présomitique continue à produire des somites, tous les deux heures, processus appelé somitogenèse (Bardot et al., 2017).

Il est intéressant de noter que la perte de TAF10 n'abroge pas complètement la transcription. La somitogenèse est un processus hautement régulé qui requiert l'expression cyclique de gènes. Du fait que la somitogenèse continue de se réaliser pendant une journée en absence de TAF10 indique que TAF10 n'est pas requis pour la transcription durant cet intervalle de temps. Ainsi, *Lfng*, un gène

présentant une expression cyclique au cours de la somitogenèse, conserve cette expression cyclique en absence de TAF10 (Bardot et al., 2017). De plus, des analyses sur micropuce à partir de mésoderme présomitique ne présentaient pas une réduction globale du niveau d'abondance des ARNs. D'autres études ont montré que TAF10 n'est pas requis pour l'expression de l'ensemble des gènes. Ainsi, TAF10 est requis au sein des cellules F9 pour l'expression de la cycline E mais pas des cyclines A, B1, C, D1, D3 et F (Metzger et al., 1999).

TAF10 est requis pour maintenir l'intégrité de TFIID (Bardot et al., 2017; Tatarakis et al., 2008). Par conséquent, le fait que la transcription est maintenue en absence de TAF10 indique que la transcription est initiée sans la présence de TFIID.

## 8. Existence de complexes TFIID partiels fonctionnels dans la transcription.

Dans différents modèles des complexes TAFs contenant ou non TBP, différent de TFIID, ont été mises en évidence (Tableau 8).

A partir d'extrait de cellules HeLa, plusieurs populations de complexes TAFs ont été découverts présentant des capacités différentes à activer la transcription (Brou et al., 1993). Par la suite des approches de co-immunoprécipitation a permis de définir la composition d'un complexe TAFs contenant TAF10 et d'un autre complexe TAFs ne contenant pas TAF10 (Jacq et al., 1994). Il est important de noter que les complexes purifiés ont été analysés par coloration à l'argent (*silver staining*). L'identité de chaque peptide indiqué reste par conséquent à être confirmée. Néanmoins, il semble que les deux complexes soient formés du core-TFIID ainsi que de TAF1, TAF7 et TBP. Le complexe contenant TAF10 semble avoir en plus les TAFs 3, 8, 11 et 13, et correspondrait donc au complexe TFIID (Tableau 8). Plus tard, d'autres analyses à partir de cellules HeLa ont montré l'existence d'un complexe appelé TFTC (*TBP-Free Taf-containing Complex*). Il est au moins constitué de TAF4, TAF5, TAF6, TAF7, TAF8, TAF10 et TAF12 (Demény et al., 2007b; Wieczorek et al., 1998). Il ne contient ni TAF1, ni TBP, ni TAF11. La présence de TAF3, TAF2, TAF9 et TAF13 au sein du complexe n'a pas été étudié dans ces études (Tableau 8). Néanmoins, il s'est révélé par la suite que la fraction purifiée contenait également le complexe SAGA, qui était alors non connu (Brand et al., 1999; Demény et al., 2007b; Wieczorek et al., 1998) (Tableau 8). Cette fraction est capable d'induire une transcription à partir de promoteur contenant ou non une boîte TATA en condition *in vitro* (Wieczorek et al., 1998).

Dans les cellules souches embryonnaires humaines (hESC : *human embryonic stem cell*) H9, seuls les TAFs 2, 3, 5, 6, 7 et 11 sont exprimés (Maston et al., 2012). Ces cellules ont un complexe TAFs d'environ 443 kDa et composé seulement des TAFs 1, 7, 11, 13 ainsi que TBP (Maston et al., 2012) (Tableau 8). Ce complexe a été nommé récemment S-TAFs (*Small TAFs*) (Fant et al., 2020). Alors que TAF3 et TAF5 sont retrouvés au niveau de quasi l'ensemble des promoteurs des gènes exprimés, le complexe S-TAFs n'est retrouvé qu'au niveau des promoteurs d'un sous-ensemble de gènes. Chacun

des TAFs exprimés a été montré comme important pour maintenir l'état indifférencié des cellules. De la même manière, l'expression ectopique des TAFs non exprimés dans ces hESC perturbe leur expression génétique et induit leur différenciation. Le traitement des hESC à l'acide rétinoïque, les engageant dans une voie de différenciation neuro-ectodermique, induit l'expression des TAFs 1, 4, 8, 9, 10, 12, 13 (Maston et al., 2012). La fonctionnalité du complexe S-TAF dans la transcription a été étudiée par des essais de transcription *in vitro* (Fant et al., 2020). A partir d'un SCP, le complexe S-TAF a été montré comme capable d'initier la transcription et de supporter la pause de l'ARN pol II. Cependant, il est important de noter qu'une autre étude (Langer et al., 2016) a montré au contraire que dans ces mêmes hESC H9, la plus part des sous-unités de TFIID sont exprimées (Tableau 8).

Embryon de souris		Cellules HeLa		hESC	
Bardot et al. 2017		Jacq et al. 1994	Wieczorek et al. 1998 Demény et al. 2007	Maston et al. 2012	Langer et al. 2016
Holo-TFIID		Complexe TAFs contenant TAF10 = holo-TFIID	Complexe TAFs ne contenant pas TAF10	TFTC	S-TAF
Core-TFIID	TAF4	TAF4	TAF4		TAF4B
	TAF5	TAF5	TAF5		
	TAF6	TAF6	TAF6		TAF6
	TAF9	TAF9	TAF9	TAF9?	TAF9 ?
	TAF12	TAF12	TAF12	TAF12	TAF12
TFIID	TAF2	TAF2 ?		TAF2 ?	TAF2
	TAF8	TAF8			TAF8 ?
	TAF10	TAF10		TAF10	TAF10
	TAF3	TAF3		TAF3 ?	
S-TAF	TAF1	TAF1	TAF1	TAF1	
	TAF7	TAF7	TAF7	TAF7	TAF7
	TAF11	TAF11		TAF11	TAF11 ?
	TAF13 ?	TAF13		TAF13	TAF13
SAGA	TBP	TBP	TBP	TBP	TBP
				SUPT7L	
				SUPT3H	
				GCN5	
				SAP130	

Tableau 7 : Récapitulatif de la composition des complexes TAFs purifiés dans les différents articles cités ci-dessus. Le ? indique les TAFs qui n'ont pas été analysés mais que l'on peut penser comme présent dans le complexe compte tenu des connaissances actuelles sur l'assemblage de TFIID. Pour information, SUPT7L et SUPT3H sont deux sous-unités du corps SAGA, GCN5 est la kinase du module HAT et SAP130 est une sous-unité du complexe SF3b du module SPL

## 9. Le chargement de TBP au promoteur n'est pas toujours requis pour initier la transcription : rôle des paralogues à TBP dans l'initiation de la transcription.

Le corps TBP a une structure à deux lobes en fer à cheval (Figure 38 A). Le premier TBP est retrouvé chez les archées et on le retrouve chez l'ensemble des eucaryotes (Ravarani et al., 2020). Bien que les orthologues chez les protistes présentent une grande diversité en séquence d'acide aminé (Parra-Marín et al., 2020), le corps des orthologues TBP de l'humain à la souris en passant par les plantes présentent un haut taux d'identité (Figure 38B). L'ensemble des acides aminés impliqués dans la liaison



de TBP au boîte TATA en condition *in vitro* sont également fortement conservé (Ravarani et al., 2020). Les orthologues TBP se différencient les uns des autres par l'extrémité N-terminal. Chez les vertébrés,

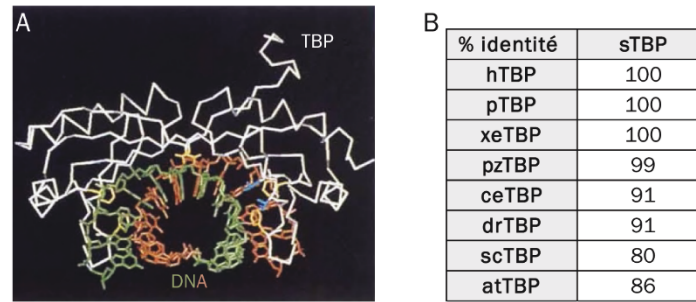


Figure 38 : TBP est conservé au sein des eucaryotes. A) Structure de TBP d'*Arabidopsis thaliana* lié à une boîte TATA. Issu de Kim et al. 1993. B) Alignement de différents orthologues TBP. L'alignement a été fait sur Kalign. Les numéro d'accension des protéines TBP sont : souris (sTBP), P29037 ; humain (hTBP), P20226 ; poule (pTBP), Q6JLB1 ; xénope (xeTBP), P27633 ; poisson-zèbre (pzTBP), Q7SXL3 ; *C.elegans* (ceTBP), P32085 ; *D. melanogaster* (drTBP), P20227 ; *S.cerevisiae* (scTBP), P13393 ; *A.thaliana* (atTBP), P28148

cette extrémité présentent une taille variable avec nombre en répétitions poly-glutamine différent (Ravarani et al., 2020).

TBP est impliqué dans la transcription médiée par l'ARN pol I, II et III (Cormack and Struhl, 1992; Veenstra et al., 2000) (voir introduction, II). De ce fait, la mutation nulle de TBP est toujours létale quelques soit l'organisme vertébré (Jallow et al., 2004; Martianov et al., 2002a; Müller et al., 2001; Veenstra et al., 2000). Dans le modèle levure, il a été montré que la transcription de certains gènes médiée par l'ARN pol II requière la liaison de TBP aux promoteurs de ces gènes. Néanmoins, TBP n'est pas requis pour l'expression de l'ensemble des gènes. Ainsi, chez le modèle Xénope, la transcription d'un certain nombre de gènes ne varient pas en absence de TBP (Veenstra et al., 2000). Dans les embryons de souris *Tbp*<sup>-/-</sup>, une transcription médiée par l'ARN pol II est toujours détectable peu après à 3,5 jpc au sein du bouton embryonnaire du blastocyste (Martianov et al., 2002a). Une récente étude a également montré que la déplétion de TBP au sein des mESC, qui sont issu du bouton embryonnaire des blastocystes, ne perturbe pas la transcription des gènes transcrits par l'ARN pol II (Kwan et al., 2021). Néanmoins, la perte de TBP affecte l'expression des ARNt.

### *i. Initiation de la transcription par TBPL1*

Dans cette dernière étude, il a été montré que TBPL1 (*TBP-like 1*, nommé aussi TRF2, TLF ou TRP pour respectivement *TBP-related factor 2*, *TBP-like factor* et *TBP-related protein*) est présent au niveau de l'ensemble des gènes transcrits par l'ARN pol II dans les mESC (Kwan et al., 2021).

TBPL1 est un paralogue de TBP apparu chez l'ancêtre commun des bilatériens (Ravarani et al., 2020). A l'heure actuelle la structure de TBPL1 n'a pas été résolue, mais des prédictions bio-informatiques ont indiqué que TBPL1 devrait avoir la même structure que le cTBP (Di Pietro et al., 2007). Le cTBP et TBPL1 présentent très peu de similarité : 39% d'identité chez l'humain et la souris, et 42% chez le

poisson zèbre (Bartfai et al., 2004; Ohbayashi et al., 2001). Notamment, parmi les 35 acides aminés de TBP impliqué dans sa liaison au boîte TATA (Kim et al., 1993), 21 sont différents chez TBPL1. Il est notable que parmi ces résidus différents, 13 n'ont pas les même caractéristiques hydrophile/hydrophobe (Di Pietro et al., 2007) (Figure 39 A). Enfin, TBPL1 se différencie de TBP également par le fait qu'il ne présente pas d'extrémité N terminal désordonnée. Par ailleurs, TBPL1 est très conservés au sein des vertébrés : TBPL1 de la souris à 100% d'identités avec celui humain, 97% d'identité avec celui de la poule, 94% d'identité avec celui du xénope et 82% d'identité avec celui du poisson-zèbre (Figure 39 B).

A

Position sur le lobe TBP			S1.3	S1.4	S1.6	S1.8	S2.2	L3.5	L3.27	L3.28	L3.31	S3.4	L4.54	S4.2	S4.4	L5.1	L5.3	S5.1	S5.3	S5.5	S5.6	L6.1
Zone d'interaction			r	b	b	r	w	p	i	r	r	p	w	p	b	r	p	w	b	b	w	w
N-terminal	TBP	Souris	Q	N	V	T	E	R	F	A	I	R	R	T	L	F	S	K	V	T	G	K
N-terminal	TBPL1	Souris	T	N	V	V	I	V	G	K	M	L	R	T	T	W	S	K	I	T	G	T
N-terminal	TBPL1	Humain	T	N	V	V	I	V	G	K	M	L	R	T	T	W	S	K	I	T	G	T
N-terminal	TBPL1	Poule	T	N	V	V	I	V	G	K	M	L	R	T	T	W	S	K	I	T	G	T
N-terminal	TBPL1	Xénope	T	N	V	V	I	V	G	K	M	L	R	T	T	W	S	K	I	T	G	T
N-terminal	TBPL1	Poisson-zèbre	T	N	V	V	I	V	G	K	M	L	R	T	S	W	S	K	I	T	G	T
N-terminal	TBPL1	Drosophile	N	N	V	S	E	N	G	M	M	L	Y	T	S	W	S	R	T	T	G	T
N-terminal	TBPL1	C.elegans	R	N	V	N	T	K	G	V	K	K	G	Y	K	Y	S	K	Y	V	G	F

Position sur le lobe TBP			S1.3	S1.4	S1.6	L3.27	L3.28	L3.31	S3.4	S4.2	S4.4	L5.1	L5.3	S5.1	S5.3	S5.5	L6.1
Zone d'interaction			r	b	b	i	b	r	p	r	b	r	p	p	rb	b	r
C-terminal	TBP	Souris	Q	N	V	F	P	I	R	V	L	F	S	K	V	T	K
C-terminal	TBPL1	Souris	V	N	L	H	P	C	R	T	Q	F	T	S	T	T	-
C-terminal	TBPL1	Humain	V	N	L	H	P	C	R	T	Q	F	T	S	T	T	-
C-terminal	TBPL1	Poule	V	N	L	H	P	C	R	T	Q	F	T	S	T	T	-
C-terminal	TBPL1	Xénope	V	N	L	H	P	C	R	T	Q	F	T	S	T	T	-
C-terminal	TBPL1	Poisson-zèbre	V	N	L	H	P	S	R	T	Q	F	T	N	T	T	-
C-terminal	TBPL1	Drosophile	V	N	L	H	P	T	K	T	K	F	T	S	T	T	-
C-terminal	TBPL1	C.elegans	N	N	L	S	V	V	R	T	R	H	T	S	T	T	Q

B

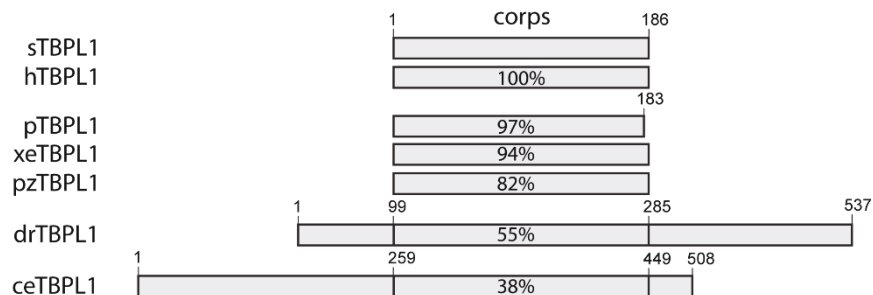


Figure 39 : Les orthologues TBPL1, très conservés au sein des vertébrés, ont divergé par rapport à TBP et ne présentent plus la même séquence en acide aminé au niveau de leur domaine de liaison à l'ADN. A) Les acides aminés de TBP entrant en contact avec l'ADN ne sont pas conservés chez les orthologues TBPL1. Les positions des acides aminés de TBP entrant en contact avec l'ADN ont été défini par Kim, J. et al. 1993. Les zones d'interaction des acides aminés sur l'ADN sont également notées : r : désoxyribose, b : base azotée, i, résidu intercalant, w : liaison hydrogène médiée par l'eau. Les acides aminés de mTBP et de différents orthologues TBPL1 présent aux positions des acides aminés de TBP entrant en contact avec l'ADN ont été alignés via la base de données mise en place par Ravarani, C. et al. 2020. En gras et italique les acides aminés des TBPL1 qui diffèrent de ceux présent chez TBP. En rouge sont les acides aminés hydrophobes et en bleu ceux hydrophiles. B) Comparaison des séquences en acide aminé de TBPL1 de souris (sTBPL1) par rapport à TBPL1 humain (hTBPL1), de poulet (pTBPL1), de xénope (xeTBPL1), et poisson-zèbre (zeTBPL1), de drosophile (drTBPL1) et de C.elegans (ceTBPL1). Les séquences d'acide aminé des 'corps' ont été alignées via le logiciel Kalign. Les numéros d'accès des protéines TBPL1 sont : souris, P62340 ; humain, P62380 ; drosophile, Q7KA83 ; C.elegans, B2D6P4.

TBPL1 est impliqué uniquement dans la transcription médiée par l'ARN pol II. En effet, il a été montré que TBPL1 ne fait pas parti de la machinerie de transcription liée à l'ARN pol I ou III (Teichmann et al., 1999; Veenstra et al., 2000). Dans le cas de la transcription médiée par l'ARN pol II,



TBPL1 est impliqué dans la mise en place de programme d'expression génétique particulier. La mutation nulle de *Tbpl1* dans le poisson-zèbre ou le xénope est létale à l'état embryonnaire. Dans ces deux modèles, TBPL1 est requis pour l'expression d'un sous-ensemble de gène nécessaire au développement embryonnaire (Müller et al., 2001; Veenstra et al., 2000). Dans le modèle souris, TBPL1 n'est pas exprimé de façon ubiquitaire. On le retrouve entre autres dans le cœurs et les testicules (Chong et al., 2005; Martianov et al., 2001; Ohbayashi et al., 1999). En particulier, TBPL1 est requis pour la spermatogenèse. La perte de TBPL1 altère le processus de différenciation et mène les cellules à l'apoptose (Martianov et al., 2001, 2002b). *A contrario*, les femelles *Tbpl1*<sup>-/-</sup> sont fertiles (Martianov et al., 2001). Lors de la spermatogenèse, TBPL1 n'est exprimé qu'à des moments particuliers : lorsque les progéniteurs sont bloqués en prophase I de méiose et lors de la dernière phase de différenciation lorsque les cellules acquièrent entre autre leur morphologie de spermatozoïde (Martianov et al., 2002b; Zhang et al., 2001).

Contrairement à TBP, TBPL1 n'est pas recruté au niveau des boîtes TATA (Chong et al., 2005; Ohbayashi et al., 1999). Son impossibilité à se lier au boîte TATA peut être corrélé au fait que la majorité des acides aminés présent dans TBP impliqué dans sa liaison au boîte TATA ne sont pas présent dans TBPL1. Néanmoins, il est recruté et initie la transcription à partir de promoteur ne contenant pas de boîte TATA (Chong et al., 2005; Ohbayashi et al., 1999; Sorscher et al., 1994). Il semble que la transcription faisant appelle à TBPL1 est réalisée via un PIC différant de celui impliqué dans la transcription dépendant de TBP. En particulier, TFIID ne semble pas être impliqué dans le chargement de TBPL1 aux promoteurs (Martianov et al., 2016), mais l'interaction entre TFIIA et TBPL1 est nécessaire pour initier la transcription (Chong et al., 2005; Martianov et al., 2016; Ohbayashi et al., 1999).

Chez la *Drosophila melanogaster* et *Caenorhabditis elegans* ont la particularité d'avoir 1) des expansions en N terminal et C terminal non présent chez les orthologues TBPL1 des mammifères (Figure 39 B) 2) de nombreuses variations dans les acides aminés interagissant avec l'ADN (Figure 39 A). De ce fait, les orthologues TBPL1 ont divergé chez les non vertébrés et l'on peut supposer qu'ils ont acquis des fonctions différentes.

Chez la drosophile, dmTBPL1 est requis pour le développement embryonnaire de différents organes ainsi que la différenciations des gamètes (Isogai et al., 2007; Kopytova et al., 2006). Il est impliqué dans l'expression de l'histone H1 et des protéines ribosomales (PR) (Baumann and Gilmour, 2017; Isogai et al., 2007). Son recrutement au niveau des promoteurs des gènes de PR se fait via le facteur de transcription M1BP ainsi que peut-être TFIID (Baumann and Gilmour, 2017). Il a été suggéré que dmTBPL1 ne puisse pas se lier par lui-même au promoteur due à sa grande différence en acide aminé au niveau du domaine de liaison à l'ADN (Baumann and Gilmour, 2017). Dans le modèle *C.elegans*, ceTBPL1 est exprimé dans toutes les cellules durant le développement et est requis pour la transcription dépendant de l'ARN pol II (Dantonel et al., 2000; Kaltenbach et al., 2000).

## ii. Initiation de la transcription par TBPL2

TBPL2 (nommé aussi TBP2 ou TRF3) est un autre paralogue de TBP capable également d'initier la transcription. TBPL2 est apparu chez l'ancêtre commun des vertébrés (Bartfai et al., 2004; Ravarani et al., 2020). Tout comme TBP, TBPL2 est constitué d'un corps en fer à cheval ainsi qu'une extrémité N-terminal désordonnée et non conservée. Alors que l'extrémité N terminal de TBP se caractérise par des expansions glutamines, celle de TBPL2 se caractérise par sa richesse en arginine (Figure 40 A). Comparé à TBPL1, cTBPL2 est relativement proche de cTBP : ils présentent 93% d'identité chez le poisson zèbre (Bartfai et al., 2004) et 81% chez la souris (Figure 40 A). Les orthologues TBPL2 présents chez l'humain, le poulet, le xénope et le poisson-zèbre ont plus de 90% d'identité, seul TBPL2 de souris présente 'seulement' 88% d'identité (Figure 40 B). Il est intéressant de noter que le TBPL2 humain est plus proche du TBP souris que l'est le TBPL2 de souris (Figure 40 C). De ce fait, bien que le corps TBPL2 ait subi une forte pression de sélection, mTBPL2 a divergé de la protéine ancestrale comparée aux autres orthologues TBPL2. Néanmoins, l'ensemble des acides aminés présent dans TBP lui permettant de se lier à la boîte TATA sont présents au sein de TBPL2 (Di Pietro et al., 2007). Ainsi, tout comme TBP, TBPL2 présente une forte affinité pour les boîtes TATAs et faible affinité pour les séquences qui divergent de la séquence consensus TATA *in vitro* (Akhtar and Veenstra, 2009; Jallow et al., 2004).

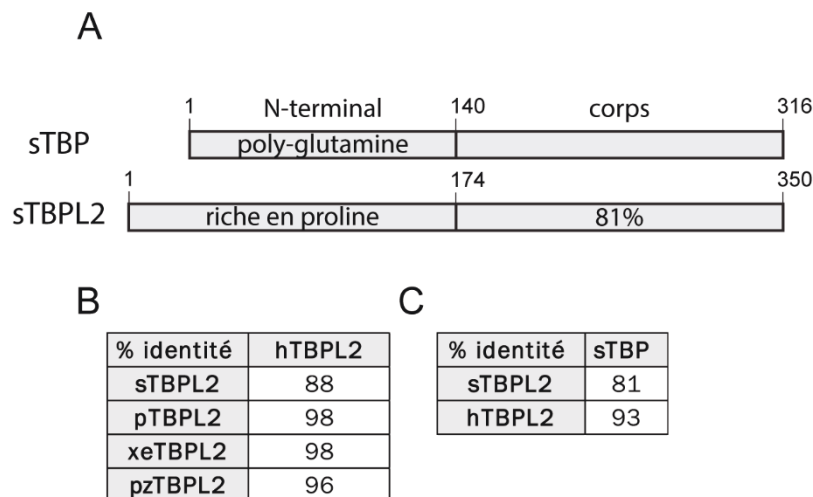


Figure 40 : Les orthologues TBPL2 présentent un haut taux d'identité entre eux et avec les orthologues TBP. A) Schéma de l'alignement des séquences d'acide aminé de TBP de souris (sTBP) et de TBPL2 de souris (sTBPL2). B) Alignement des séquences en acides aminés des orthologues TBPL2 de souris (sTBPL2), de poulet (pTBPL2) de xénope (xeTBPL2) et de poisson zèbre (pzTBPL2) ainsi que C) sTBP par rapport à sTBPL2 et hTBPL2. Les alignements ont été réalisés sur Kalign. Les numéros d'accèsion des protéines TBPL2 sont : souris, Q6SJ95 ; humain, Q6SJ96 ; poule, A6H910 ; xénope, Q5UE94 ; poisson-zèbre, Q1JPY4. Numéro d'accèsion de mTBP, P29037.

Dans le modèle Xénope et le modèle poisson-zèbre, TBPL2 est requis pour le développement embryonnaire (Akhtar and Veenstra, 2009; Jallow et al., 2004). Parallèlement à TBP, TBPL2 est requis durant le développement pour l'expression de certains gènes transcrits par l'ARN pol II (Akhtar and

Veenstra, 2009; Jallow et al., 2004). Chez l'adulte, TBPL2 n'est exprimé qu'au sein du testicule et des ovaires.

Dans le modèle souris, TBPL2 n'est exprimé que dans les ovocytes (Bartfai et al., 2004; Gazdag et al., 2007). Par conséquent, la mutation nulle de *Tbpl2* n'est pas létale mais est à l'origine d'une infertilité uniquement chez les femmes (Gazdag et al., 2009). Durant la croissance ovocytaire, TBP n'est pas présent et est remplacé par TBPL2 (Gazdag et al., 2007). Après la fertilisation TBPL2 n'est plus présent et est remplacée par TBP (Gazdag et al., 2007, 2009). L'expression ectopique de TBPL2 au sein du zygote perturbe les premières étapes du développement embryonnaire (Gazdag et al., 2009). Par conséquent, TBPL2 est spécifiquement important pour la croissance ovocytaire.

Durant ce processus, TBPL2 est impliqué dans la transcription médiée par l'ARN pol II. Les ovocytes *Tbpl2*<sup>-/-</sup> présentent une réduction globale de l'ARN pol II en élongation (Gazdag et al., 2009). La transcription chez les ovocytes est réalisée via un complexe TBPL2-TFIIA indépendant de TFIID. Ce complexe est recruté principalement à partir de boîte TATA et initie la transcription à partir de TSS autres que ceux mis en place par la machinerie de transcription TFIID/TBP (Yu et al., 2020a). Ensemble ces résultats indiquent que la mise en place et l'expression du programme génétique nécessaire à la croissance ovocytaire requière une machinerie transcriptionnelle particulière faisant appelle au moins à TBPL2-TFIIA. Il a été suggéré que cette machinerie est importante pour mettre en place au sein de l'ovocyte un transcriptome particulier qui sera utilisé durant les premières étapes du développement embryonnaire.



## V. But de la thèse

Les approches *in vitro* telles que la résolution par cryo-EM de complexes PIC et TFIID entiers ou partiels ainsi que des essais de transcription *in vitro* ont permis de mettre au point un modèle expliquant comment la transcription médiée par l'ARN pol II est initiée. TFIID est recruté spécifiquement sur les promoteurs qui ont été rendus accessibles. Par la suite sont recrutés les autres GTFs et l'ARN pol II pour mettre en place le PIC. En conséquence, TFIID est au cœur de la mise en place de la transcription au niveau de gènes particuliers. De ce fait, la présence de tels ou tels paralogues TAFs au sein de TFIID est impliquée dans l'expression de certains programmes génétiques. TFIID est donc considéré comme requis pour initier la transcription au niveau de gènes particuliers dans un contexte cellulaires particuliers.

Néanmoins, de nombreuses données provenant d'approches *in cellulo* et *in vivo* indique que la déplétion de TAF7 ou TAF10 n'est pas délétère pour la survie et la différenciation cellulaire dans certains contextes bien particuliers. D'autre part, certains complexes TAFs, ne contenant qu'une partie des sous-unités de TFIID, ont été mis en évidence. Dans des situations particulières, certains de ces complexes TAFs permettent l'expression de sous-ensemble de gènes et ne sont sensibles qu'à certains facteurs de transcription. Cependant, à l'heure actuelle, le lien entre ces complexes TAFs et les données *in cellulo* et *in vivo* n'a pas été fait. De plus, TFIID s'assemble de façon séquentielle où chaque TAFs semble jouer un rôle important pour le maintien de l'intégrité du complexe TFIID entier. Ainsi, la formation de ces complexes ne contenant qu'une partie des sous-unités de TFIID restent mystérieux.

Le but de thèse est de :

- 1) Etudier la formation et la composition de complexes TAFs différent de TFIID
- 2) Etudier le potentiel rôle fonctionnel de ces complexes TAFs dans l'initiation de la transcription
- 3) Etudier les différences fonctionnelles du complexe TFIID par rapports aux complexes TAFs dans le but de mieux comprendre le rôle de TFIID dans la transcription médiée par l'ARN pol II.

Dans ce but, j'ai utilisé deux modèles biologiques : comme modèle *in vivo* le développement embryonnaire murin et comme un modèle *in cellulo* les cellules souches embryonnaires murines (mESC)



---

# Résultats

---

---

# Résultats

---



## I. Article

### **RNA polymerase II transcription with partially assembled TFIID complexes**

**Vincent Hisler<sup>1,2,3,4</sup>, Paul Bardot<sup>1,2,3,4</sup>, Matthieu Stierle<sup>1,2,3,4</sup>, Lynda Hadj Arab<sup>1,2,3,4</sup>,  
Cynthia Ehrhard<sup>1,2,3,4</sup>, Bastien Morlet<sup>1,2,3,4</sup>, Luc Négroni<sup>1,2,3,4</sup>, László Tora<sup>1,2,3,4</sup> and  
Stéphane D. Vincent<sup>1,2,3,4,\*</sup>**

<sup>1</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France;

<sup>2</sup>Centre National de la Recherche Scientifique (CNRS), UMR7104, 67404 Illkirch, France;

<sup>3</sup>Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, 67404 Illkirch,  
France;

<sup>4</sup>Université de Strasbourg, 67404 Illkirch, France;

\*Correspondence should be addressed to S.D.V. [vincent@igbmc.fr](mailto:vincent@igbmc.fr)

Running title: TFIID independent transcription initiation

Key words: TFIID, TBP-associated factors (TAFs), TAF7, TAF10, RNA polymerase II, TFIIA,  
transcription, mass spectrometry, mESCs



## RNA polymerase II transcription with partially assembled TFIID complexes

Vincent Hisler [...] Laszlo Tora & Stéphane D. Vincent

Cette publication est directement en lien avec le but de ma thèse et les questions biologiques qui en découlent. Il a été montré que l'importance fonctionnelle de TAF7 et TAF10 varie en fonction du contexte cellulaire. Cependant à l'heure actuelle l'importance de ces deux protéines n'a pas été étudiée dans le même contexte cellulaire suivant une même stratégie de délétion. Dans ce but, dans un premier temps j'ai étudié l'impact de la perte de TAF7 ou TAF10 sur le développement embryonnaire. Pour cela j'ai comparé le phénotype d'embryons mutants dans lesquels *Taf7* ou *Taf10* sont conditionnellement délétés uniquement dans le mésoderme précoce résultant à la déplétion de TAF7 ou TAF10 dans le mésoderme présomitique et somitique postérieur ainsi que dans le mésoderme des lames latérales postérieures. La validation de la perte de TAF10 dans le mésoderme ainsi que l'impact de sa déplétion sur le développement embryonnaire avaient préalablement été publiés dans Bardot et al. 2017. J'ai validé la déplétion des TAF7 dans les tissus attendus. Les embryons conditionnels mutants *Taf7* meurent plus tardivement que les mutants conditionnels *Taf10*. Contrairement aux mutants conditionnels *Taf10*, les mutants conditionnels *Taf7* forment des bourgeons de membres antérieurs et produisent des globules rouges. Ils présentent néanmoins un défaut de développement cardiovasculaire. Ensemble ces résultats indiquent que dans un contexte embryonnaire identique, l'importance fonctionnelle de TAF7 et TAF10 n'est pas similaire.

Pour mieux étudier l'impact de la perte de TAF7 et TAF10 sur d'une part l'assemblage du complexe holo-TFIID et d'autre part l'initiation de la transcription, j'ai dérivé des cellules souches embryonnaires murines (mESC) dans lesquelles le gène *Taf7* ou *Taf10* ou les deux gènes peuvent être délétés par l'ajout de 4-hydroxitamoxifène dans le milieu de culture. Dr. Stéphane Vincent avait déjà dérivé des mESC dans lesquelles *Taf10* peut être délété et Dr. Paul Bardot avait commencé à utiliser ces cellules pour étudier l'impact de la perte de TAF10 sur l'assemblage du holo-TFIID et la transcription. J'ai confirmé que la déplétion de protéines TAF7 et TAF10 est efficace dès le deuxième jour de traitement (jour 2). Avec l'aide de Bastien Morlet et Luc Négroni qui ont réalisé les analyses par spectrométrie de masse, j'ai montré qu'à partir du jour 3 la majeure partie des complexes holo-TFIID ont été déplétés. En absence

de TAF7, deux populations de complexes partiels TFIID sont formées : une ne contenant pas TAF7 et l'autre ne contenant ni TAF7 ni TAF1. En absence de TAF10, un complexe partiel TFIID constitué de TAF4-5-6-9-12-1-7-TBP est formé. Enfin, en absence des deux protéines TAF7 et TAF10, seulement le complexe core-TFIID est formé, constitué de TAF4-5-6-9-12. Un ensemble d'analyses phénotypiques (croissance de la population cellulaire, analyse de l'apoptose et de la viabilité, capacité à former de nouvelles colonies après avoir passage) ont montré que les trois mutants présentent des phénotypes différents en termes de sévérité dès le jour 4. Alors que la croissance et la viabilité cellulaire ne sont que peu affectées par la perte de TAF7, la double déplétion TAF7 et TAF10 induit fortement l'apoptose des cellules et affecte fortement leur capacité à croître. Le mutant *Taf10* présentent un phénotype intermédiaire entre ces deux mutants.

TAF10 est également présent dans le complexe coactivateur de la transcription SAGA. Il est lié à SUPT7L via des interactions du type *Histone fold domain*. Il avait déjà été montré que la déplétion de TAF10 affecte l'assemblage de SAGA. De ce fait, la différence de phénotype entre les mESC mutants *Taf7* et *Taf10* pourrait s'expliquer par la perte de SAGA uniquement dans les mESC mutantes *Taf10*. Pour cela, Matthieu Stierle a généré une lignée *Supt7l<sup>-/-</sup>* à partir des mESC où *Taf7* peut être conditionnellement délété. J'ai montré que la déplétion de TAF10 entraîne la perte de SUPT7L mais que l'inverse n'est pas vrai. Avec l'aide de Lyndia Hadj Arab, j'ai montré que la perte de SUPT7L, additionné ou non à la perte de TAF7, n'affecte ni la croissance et ni la viabilité cellulaire. Ainsi, la différence de phénotype entre les mutants *Taf7* et *Taf10* n'est pas due à la perte de SAGA dans les mutants *Taf10*. Ensemble ces résultats indiquent que dans le contexte des mESC l'importance fonctionnelle de TAF7 et TAF10 n'est pas similaire, mais que cette différence n'est pas due à la perte de SAGA en absence de TAF10. Au niveau de l'assemblage du complexe holo-TFIID, cette différence se traduit par l'importance plus précoce et plus importante de TAF10 par rapport à TAF7 et par conséquence à la formation de complexes partiels moins aboutis en absence de TAF10 qu'en absence de TAF7.

Afin de mieux comprendre l'impact de la perte de TAF7 et/ou TAF10 sur la transcription médiée par l'ARN pol II, j'ai analysé l'activité de l'ARN pol II à trois et quatre jours de traitement. La perte de

TAF7 ou TAF10 n'affecte pas l'activité globale de l'ARN pol II. De la même manière, à jour 3 la double déplétion TAF7 et TAF10 n'affecte pas non plus l'activité globale de l'ARN pol II. La transcription n'est affectée uniquement dans le cas de la double déplétion après quatre jours de traitement. Ensemble, ces résultats suggèrent que dans les premières instants suivant la perte de la quasi-totalité de TAF7 et/ou TAF10, et par conséquence des complexes holo-TFIID, la transcription est maintenue. Ainsi le holo-TFIID ne semble pas être requis pour maintenir la transcription, du moins à court terme. Je suis actuellement en train d'étudier plus précisément le transcriptome naissant par TT-seq.

Dans le futur, avec l'aide de Matthieu Stierle, Dr. Dylane Detilleux, et l'équipe du Pr. Ferenc Müller, j'étudierai plus précisément si et comment les complexes partiels TFIID résultants de la perte de TAF7 ou TAF10 initient la transcription. Pour cela, la localisation des complexes holo-TFIID ou des complexes partiels résultants sera analysée par des expériences de CUT&TAG contre TBP, TAF12, TAF7 et TAF10. De plus, les sites de début de la transcription (TSS : *Transcriptional Start Site*) sera analysés au niveau du transcriptome globale par CAGE pour savoir l'importance du complexe holo-TFIID dans le choix du TSS par rapport aux complexes partiels.

J'ai participé à la rédaction du papier en écrivant le matériel et méthodes, relisant l'article écrit par Dr. Stéphane Vincent et en créant et rectifiant les figures du papier.

A noter que d'autres expériences ont été testées mais ne seront pas publiées car non concluantes. Cela comprend la réalisation de lignée degron permettant de dépléter conditionnellement TAF7 ou TAF10 ou encore la réalisation d'électrophorèse à partir d'extrait cellulaire en 2D, en condition non dénaturante puis dénaturante, suivi de western blot permettant d'étudier rapidement et avec un minimum de matériels les complexes protéiques formés après la déplétion de TAF7 et/ou TAF10.



## 1. Abstract

RNA polymerase II (Pol II) transcription initiation starts with the assembly of the preinitiation complex (PIC) on core promoters. The PIC is composed of six general transcription factors (GTFs). The recognition of the core promoter sequences by the TFIID GTF complex is the first step of the PIC assembly. In metazoans, TFIID is composed of the TATA binding protein (TBP) and of 13 TBP associated factors (TAFs). Genetic depletion of different murine TAFs have shown that TAFs such as TAF7 or TAF10, can be either absolutely required, or dispensable for Pol II transcription, depending on the cellular contexts. In this report, we depleted TAF7 and/or TAF10 in the same cellular system. According to the step-wise assembly model of TFIID, the absence of TAF10 should lead to the arrest of TFIID assembly at an early stage, while absence of TAF7 should have more limited effects. Our immunoprecipitation coupled to mass spectrometry-based analyses indeed indicated that TAF10 depletion results in TFIID disassembly in to building blocks, while the lack of TAF7 results in a TAF7-less TFIID. As TAF10 is also a subunit of the transcriptional co-activator Spt-Ada-Gcn5 acetyl transferase (SAGA), the more severe phenotype in *Taf10* mutant cells could have been the consequence of additional TFIID and SAGA defects. However, this is not the case as depletion of a crucial SAGA-specific structural subunit, SUPT7L, does not further increase the severity of TAF7 depletion. Thus, the difference in phenotype severity correlates with the degree of TFIID disassembly. Surprisingly however, when Pol II transcription activity changes have been evaluated, no major global changes could be observed in the absence of either TAF7 and/or TAF10. Our data suggest that the inducible loss of fully assembled canonical TFIID does not correlate with the lack of global Pol II transcription activity changes suggesting that partially assembled TFIID complexes can participate in Pol II transcription initiation, but might impact the efficacy of Pol II transcription leading the observed phenotypes.



## 2. Introduction

RNA polymerase II (Pol II) is responsible for the transcription of all protein coding genes, as well as long non coding RNA and some short non coding RNA (for a review, see (Roeder, 2019)). Pol II transcription is finely regulated, allowing gene by gene variable levels of expression depending on the cellular context. As a result, dysfunction in Pol II transcription is associated with a wide range of pathologies such as cancer, metabolic or neural diseases. Pol II transcription is first regulated by the binding of specific transcription factors to enhancers [for a review, see (Zabidi and Stark, 2016)]. These transcription factors are able to recruit different classes of transcriptional co-activator complexes that modify the structure of the chromatin: i) ATP-dependent remodeling chromatin factors that mobilize nucleosomes allowing the accessibility of the promoter region and ii) histone modifying complexes that catalyze different reactions such as acetylation or methylation on the histones leading to chromatin opening (for a review, see (Zabidi and Stark, 2016)). The action of these co-activators is crucial to create a favorable context for transcription initiation. Transcription initiation is the obligatory step of Pol II recruitment via the formation of the pre-initiation complex on the promoter.

The pre-initiation complex (PIC) is composed of 6 factors, called general transcription factors (GTFs); TFIIA, TFIIB, TFIID, TFIIE, TFIIIF and TFIIH, and Pol II (for a review, see (Roeder, 2019)). TFIID is the first GTF to bind to the promoter, initiating the nucleation of the PIC. In metazoans, TFIID is composed of the TATA-binding protein (TBP) and of 13 TBP-associated factors (TAFs), among which 6 are present in 2 copies (Bieniossek et al., 2013; Patel et al., 2018). The TAF proteins are important for PIC assembly (Chen et al., 2021) and can also serve as interfaces to interact with transcription factors for activator-dependent transcription [for a review, see (Bhuiyan and Timmers, 2019)].

In TFIID, nine of the TAFs share a conserved histone fold (HF) domain that allows the formation of specific heterodimers. TFIID is composed of three lobes, called A, B and C. Lobes A contains a histone octamer-like structure composed of TAF4/TAF12, TAF6/TAF9, TAF11/13 and TAF3/TAF10 HF pairs and lobe B a histone hexamer containing histone TAF4/TAF12, TAF6/TAF9 and TAF8/TAF10 HF pairs associated with TAF5. Lobe A and B differ by the presence of the TAF10/TAF8 (in B) or TAF10/TAF3 (in A) HF heterodimers and the presence of the TAF11/TAF13 HF pair as well as TBP in the lobe A (Chen et al., 2021; Patel et al., 2018). The HEAT domains of the 2 copies of TAF6 connect these two lobes and, in association with TAF1, TAF7 and TAF2, constitutes the lobe C (Chen et al., 2021; Patel et al., 2018). It has been proposed that TFIID assembly is sequential, via the interaction with different submodules (Bieniossek et al., 2013; Gupta et al., 2017; Trowitzsch et al., 2015). The first step is the formation of the core-TFIID (TAF4/TAF12/TAF6/TAF9/TAF5) (Bieniossek et al., 2013). The core-TFIID has an axial symmetry that is then broken by the addition of the HF pair TAF8/TAF10 coupled to TAF2, forming the 8TAF complex, or lobe B (Trowitzsch et al., 2015). The next steps of TFIID assembly are not yet elucidated, but it is known that TAF1 can interact with TAF7 and TBP and that the HF pairs TAF11/TAF13 can interact with TBP (Anandapadamanaban et al., 2013; Bhattacharya et al., 2014; Gupta et al., 2017; Wang et al., 2014a). An interaction between TAF7 and TAF11/TAF13 has been identified in the cytoplasm and it was proposed to be an intermediate toward a TAF1/TAF7/TAF11/TAF13/TBP submodule also called S-TAF could integrate the 8TAF complex to form the holo-TFIID (Gupta et al., 2017). Among the HF-containing TAFs, TAF9, TAF10 and TAF12 are also part of the nucleosome-like core structure described in the transcription co-activator Supt-Ada-Gcn5-acetyl transferase (SAGA) (Fischer et al., 2019; Helmlinger and Tora, 2017; Herbst et al.). Remarkably, TAF10 nuclear importation depends on its interaction with its HF partners, either TAF8 or TAF3 in TFIID or SUPT7L in SAGA (Soutoglou et al., 2005).

TFIID composition is actually more complex in higher eukaryotes [for a review, see (Levine et al., 2014)]. First, incomplete TFIID complexes have been characterized as TAF2- and TAF10-lacking TFIID complexes have been described in yeast, *Drosophila* and human cells (Jacq et al., 1994; Kaufmann et al., 1996; Martin et al., 1999; Papai et al., 2009). Second, some TAFs have paralogs (Levine et al., 2014): drosophila testis specific TAF4 (Nht), TAF5 (Can), TAF6 (Mia), TAF8 (Sa) and TAF12 (Rye) (Hiller et al., 2004, 2001), vertebrates TAF4B (Freiman et al., 2001), TAF7L (Pointud et al., 2003) and TAF9B (Chen and Manley, 2003; Frontini et al., 2005), and primate specific TAF1L (Wang and Page, 2002). In the mouse, TAF4B, TAF9B and TAF7L are associated with cell-specific functions [for a review, see (Levine et al., 2014)] however, they are co-expressed with their relative paralogs (Gura et al., 2020; Herrera et al., 2014; Langer et al., 2016) and *Taf4* and *Taf4b* may have partial overlapping functions (Langer et al., 2016). TAF4B and TAF9B are part of TFIID (Bardot et al., 2017; Liu et al., 2008) and TAF4 and TAF9 are present in 2 copies in TFIID (Bieniossek et al., 2013). However, it remains to be determined whether the core-TFIID contains only one paralog or both. Three TBP paralogs have also been identified in metazoans by their high homology with the characteristic horse-saddle C-terminal DNA binding domain of TBP (Hernandez, 1993): the insect-specific TBP-related factor 1 (TRF1) (Hansen et al., 1997), the metazoan-specific TBP-like 1 (TBPL1/TRF2/TLF/TLP) (Dantonel et al., 2000; Kaltenbach et al., 2000; Ohbayashi et al., 1999a, 1999b; Rabenstein et al., 1999; Teichmann et al., 1999) and the vertebrate-specific TBPL2 (also called TBP2/TRF3) (Bártfai et al., 2004; Jallow et al., 2004; Persengiev et al., 2003). TRF1 is not a component of TFIID as it does not interact with core TAFs (Hansen et al., 1997) and is primarily associated with Pol I transcription (Takada et al., 2000). In the mouse, TBPL1 (also called TRF2) is the most distant of the TBP paralogs and it does not bind to the TATA box (Ohbayashi et al., 1999a) but is able to interact with the TCT polypyrimidine initiator motif associated in most ribosomal protein gene promoters (Wang et al., 2014b).

TBPL1 is required for spermatogenesis (Martianov et al., 2001, 2016; Zhang et al., 2001; Zhou et al., 2013) but it is not part of TFIID but is a germ-cell specific subunit of TFIIA, co-recruited with TFIID (Martianov et al., 2016). In the mouse, TBPL2 (also called TRF3/TBP2) is only expressed in growing oocytes (Gazdag et al., 2007; Xiao et al., 2006) and is important for oocyte growth (Gazdag et al., 2009). In growing oocytes, TBPL2 does not integrate a TFIID complex but rather strongly interacts with TFIIA without any TAFs (Yu et al., 2020). Overall, these data indicate that despite of the potential diversity in TBP paralogs, the TFIID complex is only assembled with TBP.

TFIID is important *in vivo* as mutations in *Taf7*, *Taf8* and *Taf10* lead to peri-implantation lethality in the mouse (Gegonne et al., 2012; Mohan et al., 2003; Voss et al., 2000). Interestingly, in both *Taf8*<sup>-/-</sup> and *Taf10*<sup>-/-</sup> blastocysts, inner cell masses fail to outgrowth *in vitro* while mutant trophoectoderm is not affected (Mohan et al., 2003; Voss et al., 2000), indicating that TAF8 and TAF10 requirement depends on the cellular context. This hypothesis was been further supported by conditional deletion of *Taf7* or *Taf10* in different cell types. While TAF7 is important for early thymocytes proliferation and differentiation, it is not required for the final step of thymocyte differentiation (Gegonne et al., 2012). Similarly, conditional deletion of *Taf10* in keratinocytes, liver cells and embryonic mesoderm has different effects depending on the developmental stage and on the cell types (Bardot et al., 2017; Indra et al., 2005; Tatarakis et al., 2008). Conditional deletion of *Taf10* in the early mesoderm leads to the absence of developing limb buds and to a growth arrest in the trunk region around E10 (Bardot et al., 2017). Remarkably, cyclic transcription involved in the elongation of the embryo [for a review, see (Hubaud and Pourquié, 2014)] is still active in the absence of TAF10 protein at E9.5, prior to the growth arrest, suggesting that active Pol II transcription does not rely on primarily on TAF10 in this context (Bardot et al., 2017). Depletion of TAF10 affects the assembly of TFIID

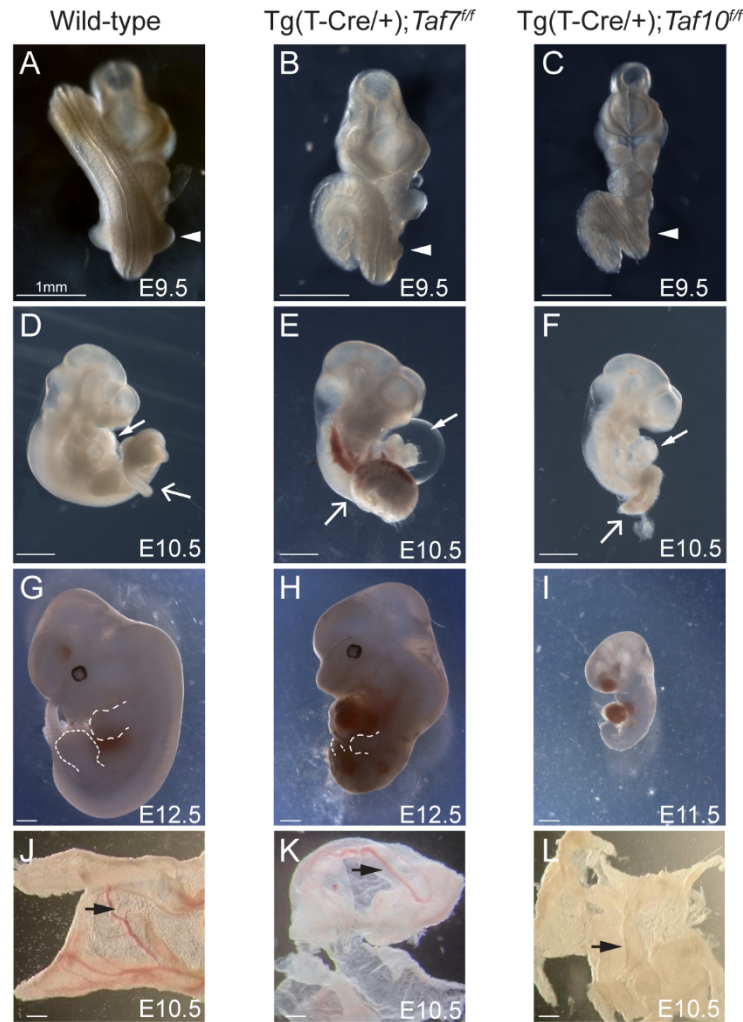
(Bardot et al., 2017; Mohan et al., 2003; Tatarakis et al., 2008), whereas TAF7 depletion does not (Gegonne et al., 2012).

Overall, it is established that TFIID composition and cellular requirement are variable depending on the context, however as the available data have been gathered from different systems it is not possible to draw comparative conclusions about the function of partial and/or holo-TFIID in cellular homeostasis. In this study, we analyzed the phenotype of the loss of *Taf7* and *Taf10* in comparable conditions, i.e., in the embryo and in mouse embryonic stem cells (mESCs) as well as the consequences in the assembly of TFIID and in Pol II transcription in order to test whether partial TFIID complexes are able to sustain Pol II transcription. Our data show that loss of *Taf7* has a less severe phenotype than the loss of *Taf10*. While TAF7 depletion leads to the assembly of a TAF7-less TFIID, TAF10 depletion strongly affects TFIID assembly, leading to only presence of the core-TFIID. However, in both situations, Pol II transcription is still active, strongly supporting the idea that partial TFIID modules are able to support transcription in the absence of holo-TFIID.

### 3. Results

#### ***Deletion of *Taf7* and *Taf10* in the same embryonic context induces similar yet different phenotypes***

To be able to compare the conditional deletion of *Taf10* in the early mesoderm (Bardot et al., 2017), we conditionally deleted *Taf7* using the same *T-Cre* transgenic line (Perantoni et al., 2005). Efficiency of TAF7 depletion was assessed by immunofluorescence (Sup. Fig. 1). As expected from the recombination pattern of the *T-Cre* line, the structures above the heart are not targeted and therefore normal (Perantoni et al., 2005). Tg(*T-Cre*/+);*Taf7*<sup>ff</sup> and Tg(*T-Cre*/+);*Taf10*<sup>ff</sup> mutant embryos (hereafter called *TCre;Taf7* and *TCre;Taf10*, respectively) are



**Fig. 1: Conditional deletion of *Taf7* or *Taf10* in the early mesoderm results in similar yet different phenotypes.** Tg(T-Cre/+);*Taf7<sup>ff</sup>* and Tg(T-Cre/+);*Taf10<sup>ff</sup>* embryos display different phenotypes. (A-I) Wholemount view of wild-type (A, D, G), Tg(T-Cre/+);*Taf7<sup>ff</sup>* mutant (B, E, H) and Tg(T-Cre/+);*Taf10<sup>ff</sup>* mutant (C, F, I) embryos at E9.5(A-C, n>4), E10.5 (E-F, n>4), and E12.5 (G, H, n>2). As no Tg(T-Cre/+);*Taf10<sup>ff</sup>* mutant embryos could be recovered at E12.5, a E11.5 Tg(T-Cre/+);*Taf10<sup>ff</sup>* mutant embryo is shown (I). White arrowheads (A-C) and dashed lines (G, H) indicate the limb buds. White arrows (D-F) indicate the tail and plain white arrow (D-F) the heart. (J-L) Wholemount views of wild-type (J), Tg(T-Cre/+);*Taf7<sup>ff</sup>* (K) and Tg(T-Cre/+);*Taf10<sup>ff</sup>* (L) yolk sacs embryos are showed (n>3). Black arrows indicate the presence of blood cells. Scale bars: 1 mm.





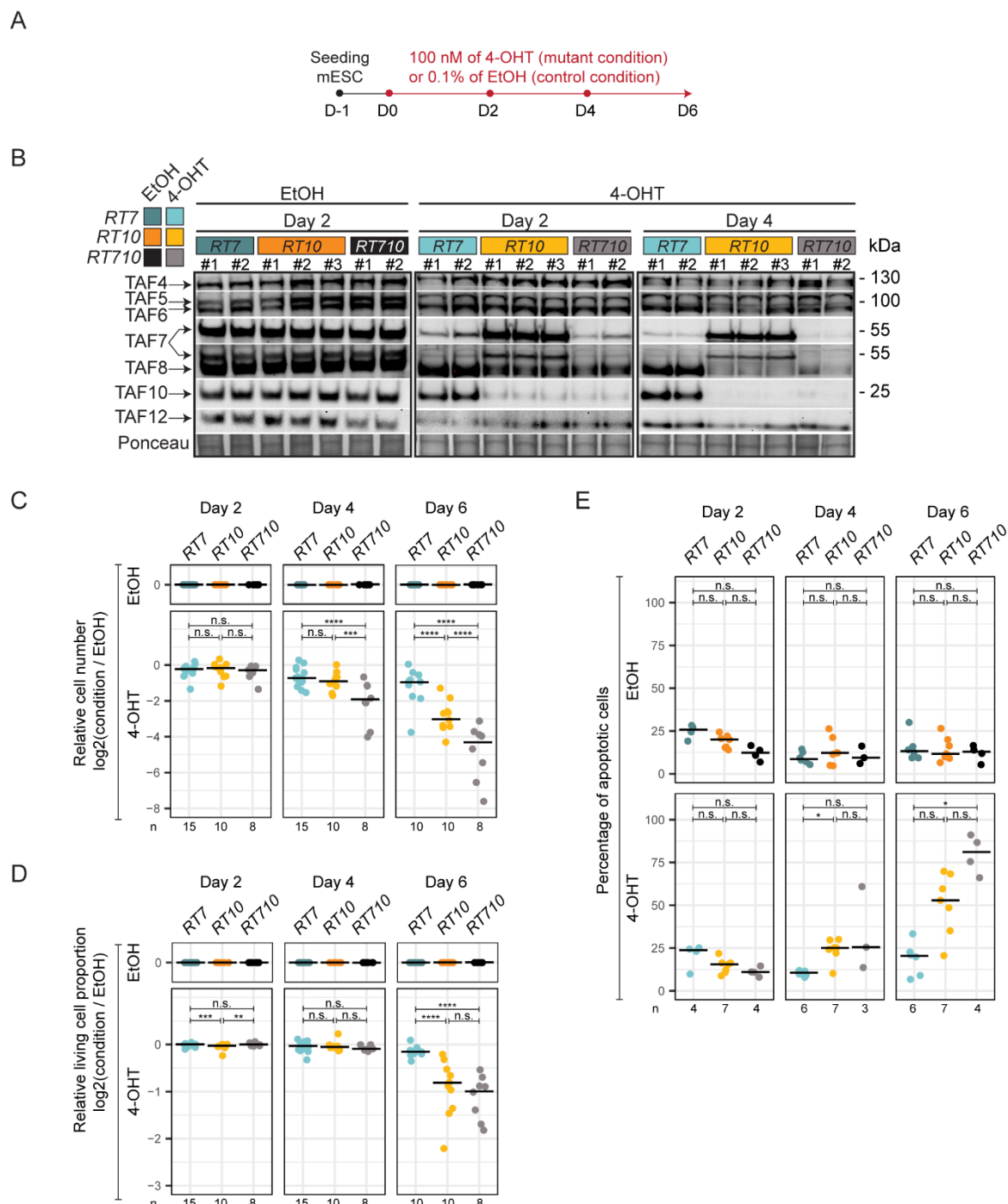
similar at embryonic day (E) 9.5. The only visible difference that we could detect was that no limb buds are present in *TCre;Taf10* mutant embryos, while *TCre;Taf7* embryos displayed forelimb buds although smaller when compared to the control (Fig. 1B compared to 1A, C, white arrowhead). At E10.5, both *TCre;Taf10* and *TCre;Taf7* mutant embryos display a growth arrest in the trunk and tail regions (Fig. 1E, F compared to D, white arrow). After E10.5, in addition to the presence of limb buds in the *TCre;Taf7* embryos, more differences could be observed between the two mutants. First, enlarged pericardia are observed in all the *TCre;Taf7* mutant embryos (Fig. 1E, white arrow) suggesting that *Taf7* deletion leads to cardiovascular defects not observed in *TCre;Taf10* mutant embryos (Fig. 1F, plain white arrow). Second, while *TCre;Taf10* mutants could not be retrieved after E11.5 due to placenta and allantois degeneration (Bardot et al., 2017), *TCre;Taf7* mutant embryos could be still detected at E12.5 characterized by hypotrophied limb buds, trunk and tail regions (Fig. 1H). Third, presence of blood cells was obvious in the vasculature of controls and *TCre;Taf7* mutant embryos (Fig. 1J, K, black arrowhead), but not of *TCre;Taf10* mutant embryos (Fig. 1L, black arrowhead). Altogether, these data indicate that conditional deletion of *Taf7* or *Taf10* in the same genetic context leads to a similar growth arrest phenotype, but different outcomes. *Taf10* conditional deletion leads to a more severe phenotype, while *Taf7* conditional deletion leads to a milder phenotype and is associated with specific cardiovascular defects.

***Deletion of Taf7 and Taf10 in mESCs recapitulate the difference in the severity of the phenotype observed in the mutant embryos***

In order to get more insight into the molecular consequences of the deletion of *Taf7* or *Taf10*, we derived mESC from blastocysts carrying the inducible ubiquitously expressed *R26<sup>CreERT2</sup>* allele (Ventura et al., 2007) associated with the *Taf7<sup>f</sup>*, *Taf10<sup>f</sup>*, or *Taf7<sup>f</sup>* as well as

*Taf10<sup>f</sup>* alleles (*R26<sup>CreERT2/+</sup>;Taf7<sup>ff</sup>*, *R26<sup>CreERT2/+</sup>;Taf10<sup>ff</sup>* and *R26<sup>CreERT2/+</sup>;Taf7<sup>ff</sup>;Taf10<sup>ff</sup>*). Hereafter the *R26<sup>CreERT2/+</sup>;Taf7<sup>ff</sup>* line will be called *RT7*; *R26<sup>CreERT2/+</sup>*, the *Taf10<sup>ff</sup>* line will be called *RT10* and the *R26<sup>CreERT2/+</sup>;Taf7<sup>ff</sup>;Taf10<sup>ff</sup>* line will be called *RT710*. The Cre recombinase induced deletion of the floxed alleles of *Taf7<sup>ff</sup>* and/or *Taf10<sup>ff</sup>* can be induced by addition of 4-hydroxy tamoxifen (4-OHT) in the culture medium (day 0) to activate CreERT2 (Fig. 2A, hereafter called mutant *RT7*, mutant *RT10* and mutant *RT710*). Efficiency of the deletion was monitored at days 2 and 4 by western blot analyses (Fig. 2B). No depletion was observed in the presence of the vehicle (ethanol), confirming the specificity of the induction of the deletion (Fig. 2B, left panel). However, in the presence of 4-OHT, depletion of TAF7 protein in *RT7* and *RT710* mESCs can be observed as early as day 2 and is almost complete at day 4 (Fig. 2B, middle and right panels). Similarly, depletion of TAF10 protein in *RT10* and *RT710* mESCs can be observed as early as day 2 and the corresponding TAF proteins are not detectable at day 4 (Fig. 2B, day 2 middle and day 4 right panels). The depletion of TAF7 and TAF10 proteins is homogenous within the cellular population as shown by immunolocalization (Sup. Fig. 2A, B). Depletion of the protein of interest does not affect the expression of the tested TAF4, TAF5, TAF6 and TAF12 TFIID subunits, and in agreement with our earlier studies, depletion of TAF10 lead to the destabilization of its HF partner TAF8 (Bardot et al., 2017; Kamenova et al., 2019) (Fig. 2B). Altogether, these data indicate that we have established an efficient cellular model to study the effects of the depletion of TAF7 and/or TAF10 proteins.

Next, we investigated the effect of the 4-OHT treatment on cell proliferation and viability at day 2, 4 and 6. In culture, mutant *RT7*, *RT10* and *RT710* cells form smaller colonies compared to the control cells (Sup Fig. 3A, B). Moreover, when the mutant cells were split at day 4 and analyzed 2 days after (D4+2), their capacity to form colonies was severely impaired, especially for the mutant *RT10* and *RT710* cells (Sup Fig. 3A, B), indicating that both TAF7 and TAF10 are required for mESC maintenance. These data are in agreement with the incapac-



**Fig. 2: Conditional depletion of either TAF7 or TAF10, or both, in mESC induces different phenotypes.** (A) Strategy to induce the depletion of the protein of interest in  $R26^{CreERT2/+};Taf7^{ff}$  (RT7),  $R26^{CreERT2/+};Taf10^{ff}$  (RT10),  $R26^{CreERT2/+};Taf7^{ff};Taf10^{ff}$  (RT710) mESCs. (B) Western blot analysis of TAF4, TAF5, TAF6, TAF7, TAF8, TAF10 and TAF12 expression in EtOH (control) or 4-OHT (mutant) treated RT7, RT10 and RT710 at day 2 and day 4. The Ponceau staining is displayed at the bottom of the panel. (C-D) Log<sub>2</sub> relative total cell number (C), log<sub>2</sub> relative living cell proportion (D) and percentage of apoptotic cells (E) in control and mutant RT7, RT10 and RT710 mESCs at day 2, 4 and 6. The black lines indicate the median. n indicates the number of replicates (n≥4 biological replicates). Kruskal-Wallis test with Benjamini Hochberg post-hoc test. ns; not significant, \* <0.05; \*\* <0.005; \*\*\* <0.0005; \*\*\*\* <0.00005. RT7;  $R26^{CreERT2/+};Taf7^{ff}$ , RT10;  $R26^{CreERT2/+};Taf10^{ff}$ , RT710;  $R26^{CreERT2/+};Taf7^{ff};Taf10^{ff}$ , 4-OHT; 4-hydroxy tamoxifen.



-ty of *Taf7*<sup>-/-</sup> null and *Taf10*<sup>-/-</sup> null inner cell masses to grow *in vitro* (Gegonne et al., 2012; Mohan et al., 2003).

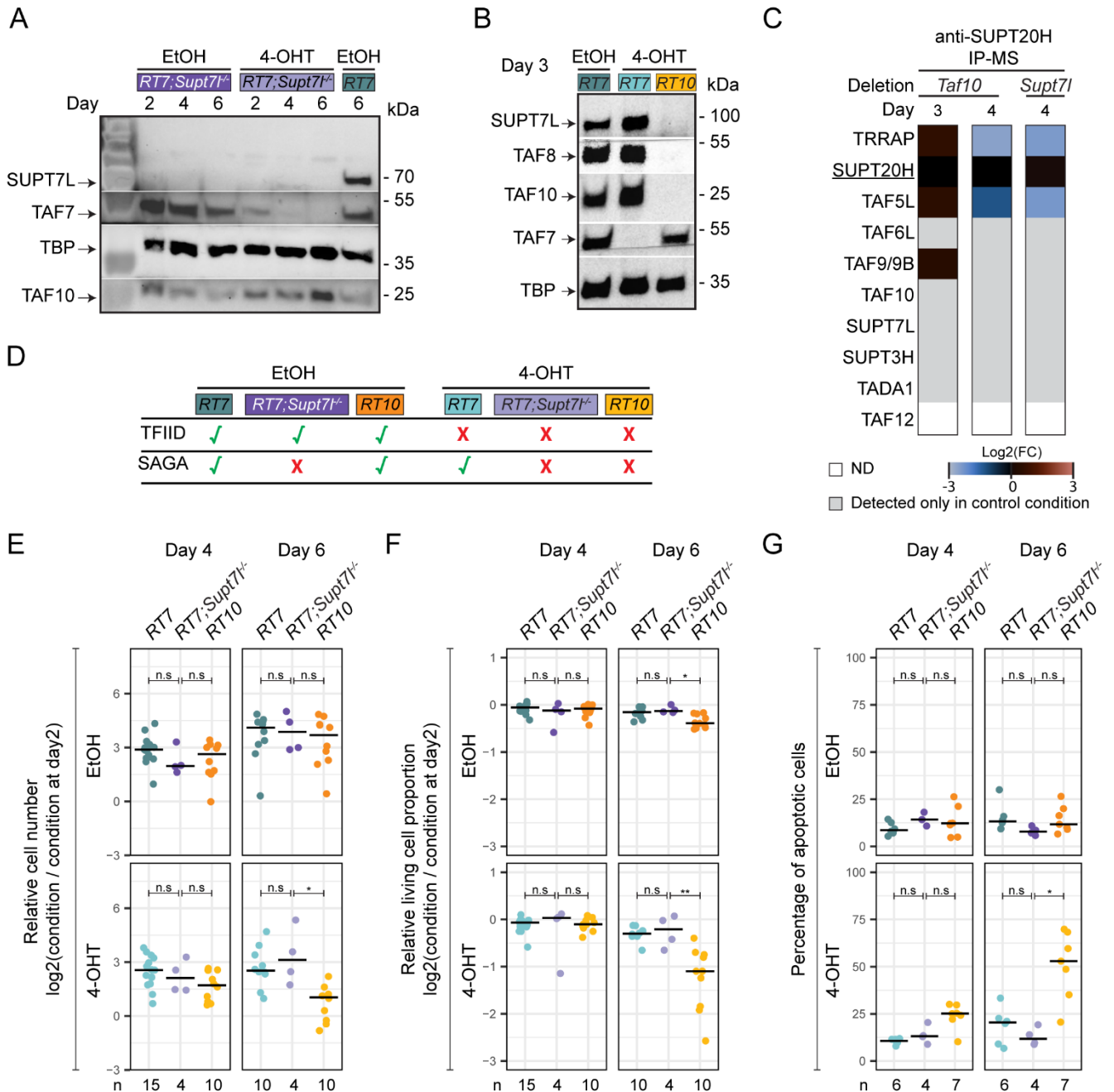
We then quantified the evolution of the number of cells in each mESC clones, under control and mutant conditions. After two days of 4-OHT addition there were no detectable differences in cell numbers between the control and mutant cells (Fig. 2C). While depletion of TAF7 has a limited effect at all tested days (days 2, 4 and 6), depletion of TAF10 had a minor effect at day 4, but a strong reduction in cell number was observed at day 6 (Fig. 2C). Interestingly, the combination of TAF7 and TAF10 depletion induces a more severe reduction of cell number already at day 4 (Fig. 2C).

We then investigated the evolution of the cell viability (Fig. 2D). No differences between the control and the mutant cells could be observed at day 2 and day 4. However, at day 6, while depletion of TAF7 did not have any significant effect, depletion of TAF10 or combined depletion of TAF7 and TAF10 reduce the viability about a 2-fold in mutant *RT10* and mutant *RT710* cells. Analysis of apoptosis by quantification of Annexin V signal showed that the percentage of apoptotic cells ranged from 10 to 25% from day 2 to day 6, in control and mutant *RT7* cells (Fig. 2E) indicating that TAF7 depletion does not induce apoptosis. To the contrary, although at days 2 and 4 no significant increase in apoptosis could be observed in mutant *RT10* and mutant *RT710* cells, at day 6, the apoptosis rate increased 3-fold in the absence of TAF10 and 5-fold in the combined absence of TAF7 and TAF10 when compared to control conditions (Fig. 2E). This indicates that TAF10 is important for the survival of the mESCs, as already suggested by the increased apoptosis observed in the inner cell mass of *Taf10*<sup>-/-</sup> mutant blastocysts (Metzger et al., 1999; Mohan et al., 2003). To the contrary, TAF7 is not essential for the survival of the mESCs: this is in agreement with the observation that *Taf7*<sup>-/-</sup> blastocyst can survive until hatching at E4.5 (Gegonne et al., 2012).

Altogether, these data indicate that the depletion of TAF10 in growing mESCs has a much more pronounced effect than the depletion of TAF7. TAF10 is absolutely required for the growth and survival of the mESCs. TAF7 depletion does not obviously affect mESCs growth and survival. This is not the consequence of the expression of TAF7 paralogs as they are not expressed in the controls nor in the mutant cells (Sup. Fig. 4A-B). However, combined depletion of TAF7 and TAF10 induces a much more severe phenotype than the single depletions, suggesting that TAF7 and TAF10 are involved in different aspects cellular homeostasis regulated by TFIID-mediated transcription.

***The severity of the TAF10 depletion phenotype is not caused by defects in SAGA assembly***

TAF10 is a shared subunit between TFIID and the SAGA coactivator complex, whereas TAF7 is a TFIID specific subunit (Helmlinger and Tora, 2017). SAGA assembly is defective in E9.5 *Taf10* conditional mutant embryos (Bardot et al., 2017). Therefore, we hypothesized that the difference in phenotype severity between mutant *RT7* and *RT10* mESCs could be associated with a combined impairment of TFIID and SAGA function in *RT10*. To test this hypothesis, we decided to affect SAGA assembly in the context of the *RT7* line and to compare the phenotypic consequence to the phenotype of mutant *RT7* and mutant *RT10* mESCs. SUPT7L is the partner of TAF10 within SAGA and is important for SAGA integrity (Fischer et al., 2021). In *Supt7l*<sup>-/-</sup> mESCs, SAGA assembly is disrupted but this is not associated with a strong phenotype when cultured in the presence of LIF and 2i inhibitors, the cell culture conditions we used in this study (Fischer et al., 2021). For these reasons, we engineered a *Supt7l* homozygous mutation in the *RT7* line (hereafter called *RT7;Supt7l*<sup>-/-</sup>) using the same CRISPR/Cas9 strategy as in (Fischer et al., 2021), in order to be able to deplete TAF7 in a context where SUPT7L is missing and SAGA assembly is seriously impaired. Absence of SUPT7L and TAF7, after 4-OHT treatment of *RT7;Supt7l*<sup>-/-</sup> cells, was validated by western blot



**Fig. 3: SUPT7L and TAF10 depletions in mESCs similarly impair SAGA assembly but do not have the same cellular phenotype.** (A) Western blot analysis of SUPT7L, TAF7, TAF8, TAF10 and TBP expression in control (EtOH) and mutant (4-OHT) *RT7* and *RT10* mESC at day 3. (B) Western blot analysis of SUPT7L, TAF7, TAF10 and TBP in control (EtOH) and mutant (4-OHT) *RT7;Supt7l<sup>-/-</sup>* mESCs at days 2, 4 and 6 compared to control (EtOH) *RT7* mESC at day 6. (C) Relative normalized XIC values for detected SAGA subunits after anti-SUP20H immunoprecipitation coupled to mass spectrometry analysis (IP-MS) from mutant versus control *RT10* mESCs at day 3 and 4, and *Supt7l<sup>-/-</sup>* versus day 4 control *RT7* mESCs nuclear enriched cell lysates. The color legend is indicated at the bottom of the panel. (n=1 biological x 3 technical replicates) (D) Summary of the expected state of TFIID and SAGA complexes assembly in the different mESCs. (E-G) Log2 relative total cell number (E), log2 relative living cell number (F) and percentage of apoptotic cells (G) in control and mutant *RT7*, *RT10* and *RT710* mESCs at day 4 and 6. The black lines indicate the median. n indicates the number of



replicates ( $n \geq 4$  biological replicates). Kruskal-Wallis test with Benjamini Hochberg post-hoc test. ns; not significant, \*  $<0.05$ ; \*\* $<0.005$ ; \*\*\*  $<0.0005$ ; \*\*\*\*  $<0.00005$ . *RT7*; *R26<sup>CreERT2/+</sup>;Taf7<sup>ff</sup>*, *RT10*; *R26<sup>CreERT2/+</sup>;Taf10<sup>ff</sup>*, *RT710*; *R26<sup>CreERT2/+</sup>;Taf7<sup>ff</sup>;Taf10<sup>ff</sup>*, 4-OHT; 4-hydroxy tamoxifen.

(Fig. 3A). Interestingly, SUPT7L protein expression in mutant *RT10* cells indicated that SUPT7L protein is lost in absence of TAF10 (Fig. 3B). This shows that in *RT10* cells TAF10 is also required for the stability of SUPT7L. We confirmed by immunoprecipitation coupled with mass spectrometry (IP-MS) using antibodies against SUPT20H, a main protein of SAGA, that SAGA is not assembled in absence of TAF10 and/or SUPT7L (Fig. 3C) as already reported (Bardot et al., 2017; Fischer et al., 2021).

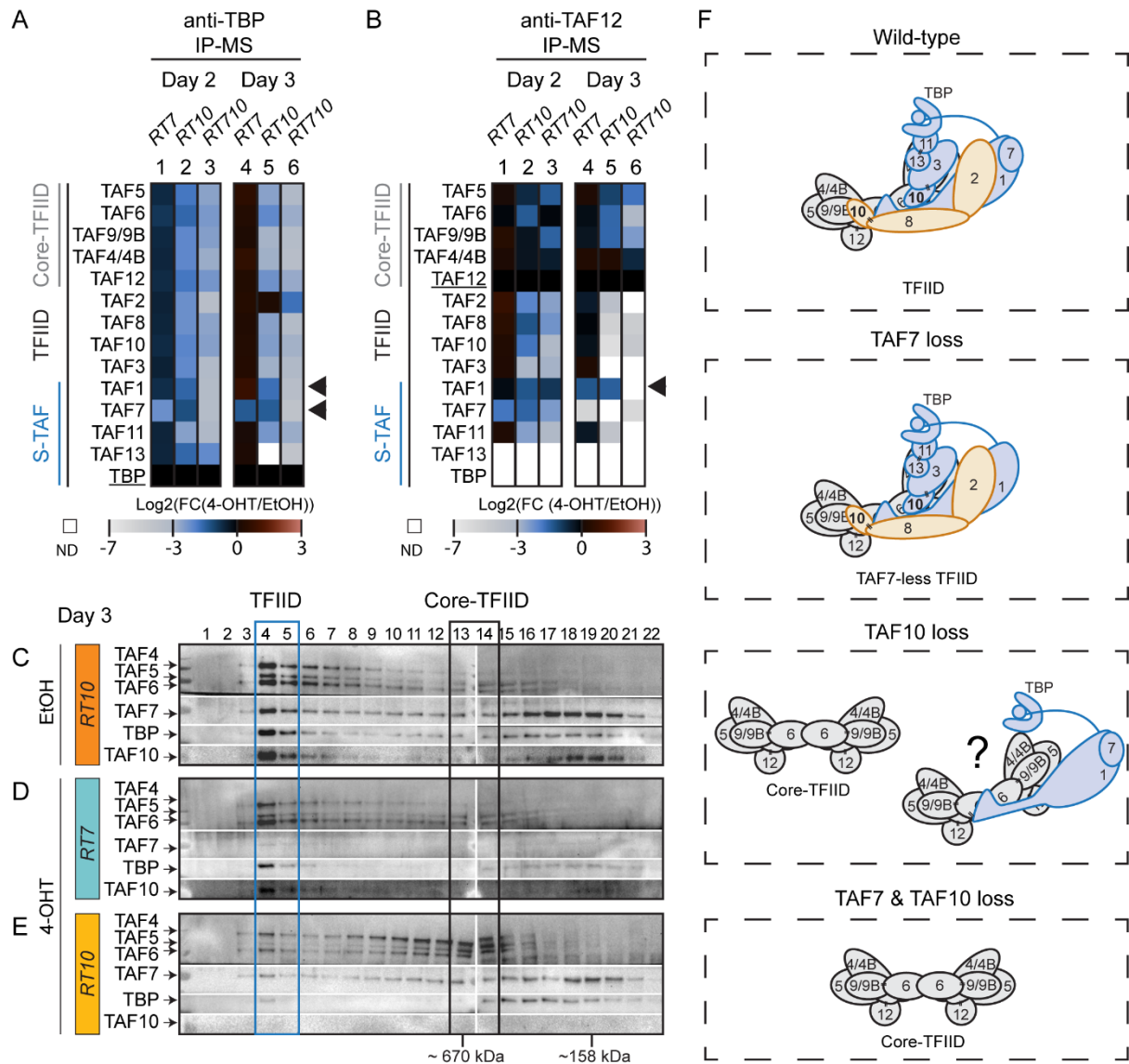
Our different cellular measurements indicated the growth and survival of *RT7;Supt7l<sup>-/-</sup>* mESCs are not affected (Fig. 3D-G). After 4-OHT induction of TAF7 depletion in the same cellular background, cell number (Fig. 3E), cell viability (Fig. 3F) and apoptosis (Fig. 3G) measurements indicate that the absence of SUPT7L does not increase the severity of the phenotype of TAF7 depletion alone. Therefore, SAGA disruption does not aggravate the phenotype of TAF7 loss, strongly suggesting that the difference in severity between mutant *RT7* and mutant *RT10* mESCs is not caused by the lack of SAGA in mutant *RT10*, but by the different molecular consequences associated to the loss of TAF7 or TAF10 in TFIID.

### ***Depletion of TAF7 or TAF10 leads to the formation of different partial TFIID complexes***

In order to get more insight into the status of TFIID complex in the mutant mESCs, we performed immunoprecipitation-coupled mass spectrometric (IP-MS) using anti-TAF12 and anti-TBP antibodies on nuclear enriched whole cell extracts (NWCE) at day 2 and 3. Depletion of TAF7 or/and TAF10 in the NWCE was checked by immunolocalization and western blot. While depletion was not complete at day2, at day 3, only traces of TAF7 or/and TAF10 could be detected (Sup Fig. 2A-B, Sup Fig. 6A-D, “In” lanes). As TAF12 is part of the core-TFIID and TBP only part of the canonical TFIID complex, these 2 types of IPs allow to discriminate between the different partial or complete TFIID complexes. TBP is also part of the SL1 and

TFIIIB complexes that are involved in Pol I and Pol III transcription initiation, respectively (Vannini and Cramer, 2012). TAF12 is also a subunit of SAGA, but the used anti-TAF12 does not IP SAGA (data not shown). We validated that the lack of TAF7 or TAF10 did not interfere with the interaction between TBP and SL1 subunits TAF1A, TAF1B, TAF1C and TAF1D as well as with TFIIIB subunit BRF1 (Sup Fig. 5), confirming the TAF7 and TAF10 depletion did not result in non-specific effects. Remarkably, as already observed in the *TCre;Taf10* mutant embryos (Bardot et al., 2017), TAF10 depletion resulted in increased interaction between TBP and its non-TFIID partners, indicating that TBP is relocated to these complexes under these conditions suggesting that TFIID assembly is indeed affected following TAF10 deletion (Sup Fig. 5).

TAF7 depletion did not induce significant changes in the TFIID partners of TBP (Fig. 4A, lanes 1 and 4, Sup Fig. 6A, C) and of TAF12 (Fig. 4B, lanes 1 and 4, Sup Fig. 6B, D), except for the loss of TAF7 itself. This indicates that the absence of TAF7 has a very limited effect on TFIID assembly, resulting in some TAF7-less TFIID complex (Fig. 4F). On the other hand, depletion of TAF10 caused a more severe impairment of TFIID assembly. First, anti-TBP IP-MS indicated that TBP incorporation in TFIID is severely impaired, as we observed a strong decrease in the interaction between TBP and all other TFIID subunits (Fig. 4A lanes 2 and 5, Sup. Fig. 6A, C). Second, anti-TAF12 IP-MS showed that in absence of TAF10, TAF12 maintains the interactions with TAF5, TAF6, TAF9/9B and TAF4/4B core-TFIID units, but interactions with the other TAFs or TBP were severely decreased (Fig. 4B, lanes 2 and 5, Sup. Fig. 6B, D). This indicates that the core-TFIID is the main partial TFIID complex present in the TAF10 depleted cells (Fig. 4F). The IP-MS results obtained on cell in which both TAF7 and TAF10 was depleted led to a TFIID dissociation similar to that obtained in TAF10 depleted cells, suggesting that in cells where TAF10 is depleted, the 5-subunit containing core-TFIID complex might be the most dominant partial TFIID complex present (Fig. 4A, B, F). However,



**Fig. 4: Conditional depletion of TAF7 or/and TAF10 differentially affects TFIID assembly.** (A-B) Relative normalized XIC values for TFIID subunits anti-TBP IP (A) or anti-TAF12 (B) immunoprecipitation coupled to mass spectrometry analyses (IP-MS) from mutant versus control *RT7*, *RT10* and *RT710* nuclear enriched cell lysates at day 2 and 3. The color legend is indicated at the bottom of the panels (n=1 biological x 3 technical replicates) (C-E) Gel filtration coupled to western blot analysis of TAF4, TAF5, TAF6, TAF7, TAF10 and TBP expression from control (EtOH) *RT10* (C), and mutant (4-OHT) *RT7* and *RT10* (E) at day 3. The number on the top correspond of the different fractions analyzed. The molecular weight of the 13 and 14 fractions is indicated at the bottom of the panel (n=1 biological x 2 technical replicates). (F) Summary of TFIID and partial TFIID complexes in wild-type cells and in absence of TAF7 or/and TAF10 proteins. *RT7*; *R26<sup>CreERT2/+</sup>;Taf7<sup>flf</sup>*, *RT10*; *R26<sup>CreERT2/+</sup>;Taf10<sup>flf</sup>*, *RT710*; *R26<sup>CreERT2/+</sup>;Taf7<sup>flf</sup>;Taf10<sup>flf</sup>*, 4-OHT; 4-hydroxy tamoxifen.



in the mutant *RT10* cells, TBP interactions with TAF1 and TAF7 were decreased in the anti-TBP IP, but much less compared to the other non-core-TFIID TAFs (Fig. 4A, lanes 2 and 5, arrowheads). Similarly, in the anti-TAF12 IP, TBP was still interacting with TAF1 (Fig. 4B, lanes 2 and 5, arrowhead). Interestingly, these interactions were no longer observed in the absence of both TAF7 and TAF10 (Fig. 4A-B, lanes 3 and 6, arrowheads) suggesting that TAF10 is stabilizing these interactions. As TAF1 is interacting with TAF7 but also with TAF6 in TFIID (Chen et al., 2021), this suggests that some core-TFIID complexes may be associated to TAF1, TAF7 and TBP (Fig. 4F). Finally, some TAF7 and TAF10 peptides were still detected in the mutant conditions at day 2 (Fig. 4A-B, lanes 1-3) and even less at day 3 (Fig. 4A-B, lanes 1-3), suggesting that a minor proportion of some very stable TFIID might still be present despite the efficient depletion at day 3 in the mutant cells.

In order to get more insight into the molecular consequences of the depletion of TAF7 and TAF10 on TFIID assembly, we performed gel filtration coupled to western blot analyses *RT7* or *RT10* mESCs NWCEs collected at day 3, in control and mutant conditions. In control cells (*RT10* without 4-OHT), the TFIID complex is collected in fractions 4-5 (Fig. 4C, F, blue box) as suggested by the molecular weight and the co-localization of TBP and all tested TAFs. In mutant *RT7* extracts, the profile is very similar to the control lysates, except for the absence of TAF7, strongly suggesting that a TFIID complex without TAF7 is present in these cells (Fig. 4D, F). In mutant *RT10* cells however, very little high molecular weight complex could be detected lacking TBP, and TAF4, TAF5 and TAF6 proteins were localized in lower molecular weight fractions peaking at 13-14 (Fig. 4E, F, grey box), indicating that the core-TFIID submodule is dominant in absence of TAF10 in these cells.

Altogether, these data demonstrate that TFIID assembly is differentially affected by the depletion in TAF7 and TAF10, and while an almost complete TFIID complex lacking TAF7 seem to persist in mutant *RT7* cells, in the mutant *RT10* and *RT710* cells, the canonical TFIID

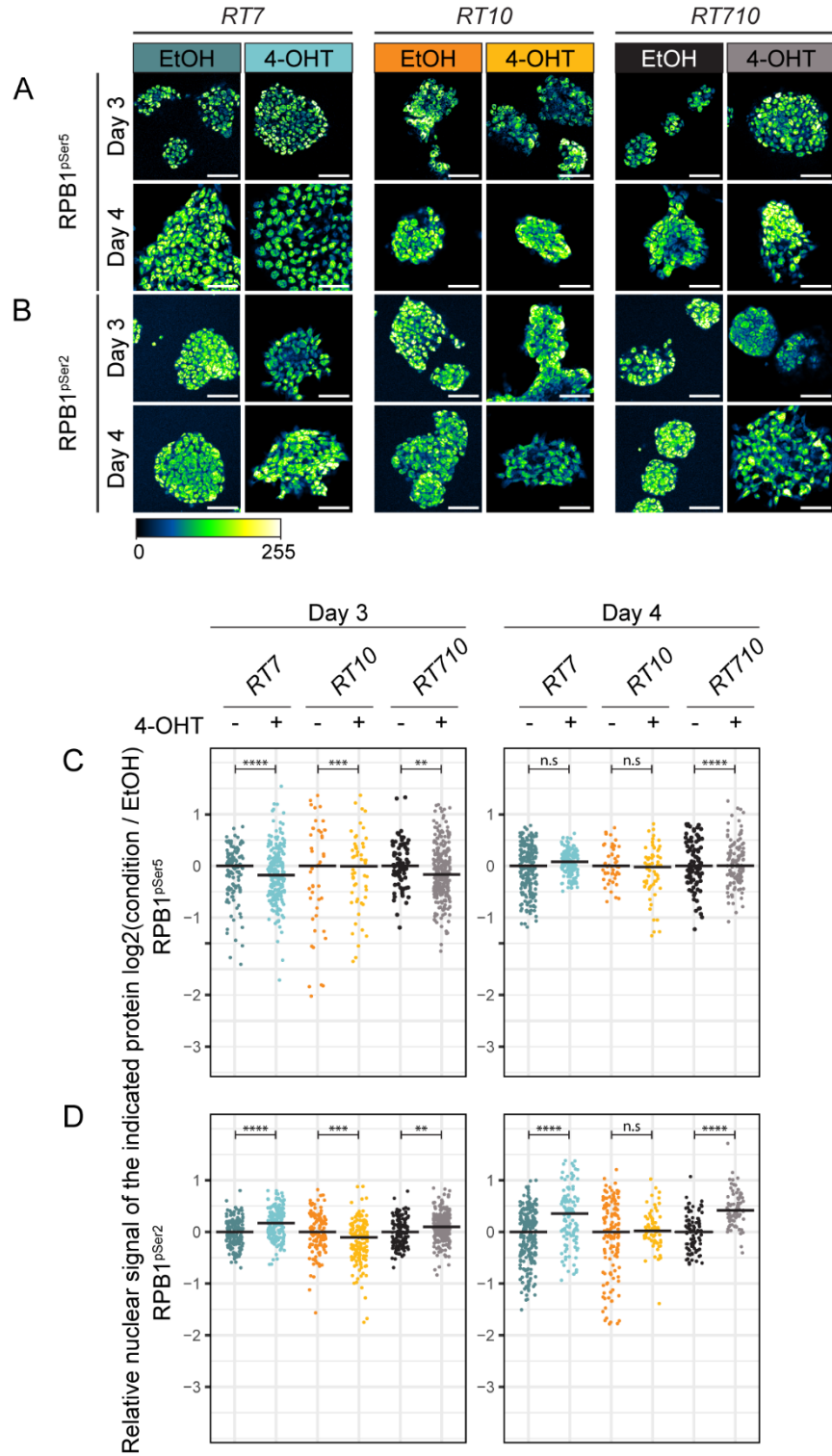
complex is not detected and only partial submodules such as the core-TFIID complex remain the dominant TAF complex (Fig. 4F).

### ***Active Pol II transcription in absence of holo-TFIID***

According to the test book view, TFIID is the first GTF to recognize the promoters and to start the formation of the PIC, we investigated the Pol II activity of by analyzing the phosphorylation status of C-terminal domain (CTD) of the Pol II subunit RPB1 in mutant *RT7*, *RT10* and *RT710* cells at day 3 and day 4 using immunofluorescence (IF) staining on fixed ESCs. Phosphorylation of the biggest Pol II subunit RBP1 on Serine 5 and Serine 2 is associated with the initiation and the elongation status of Pol II, respectively [for a review (Harlen and Churchman, 2017)]. Using this method, no difference in RPB1<sup>pSer2</sup> and in RPB1<sup>pSer5</sup> phosphorylation signal could be observed between control and mutant *RT7* cells (Fig. 5A-C, B-D) indicating that Pol II transcription initiation and elongation is active in the absence of TAF7. Surprisingly, similarly, no significant differences could be observed between control and mutant *RT10* and *RT710* cell lines (Fig. 5A-D) also suggesting that in the absence of TAF10, or TAF7 and TAF10 the global Pol II transcription initiation and/or elongation is not affected. Altogether, these data suggest that the absence of canonical TFIID does not impair global Pol II transcription activity.

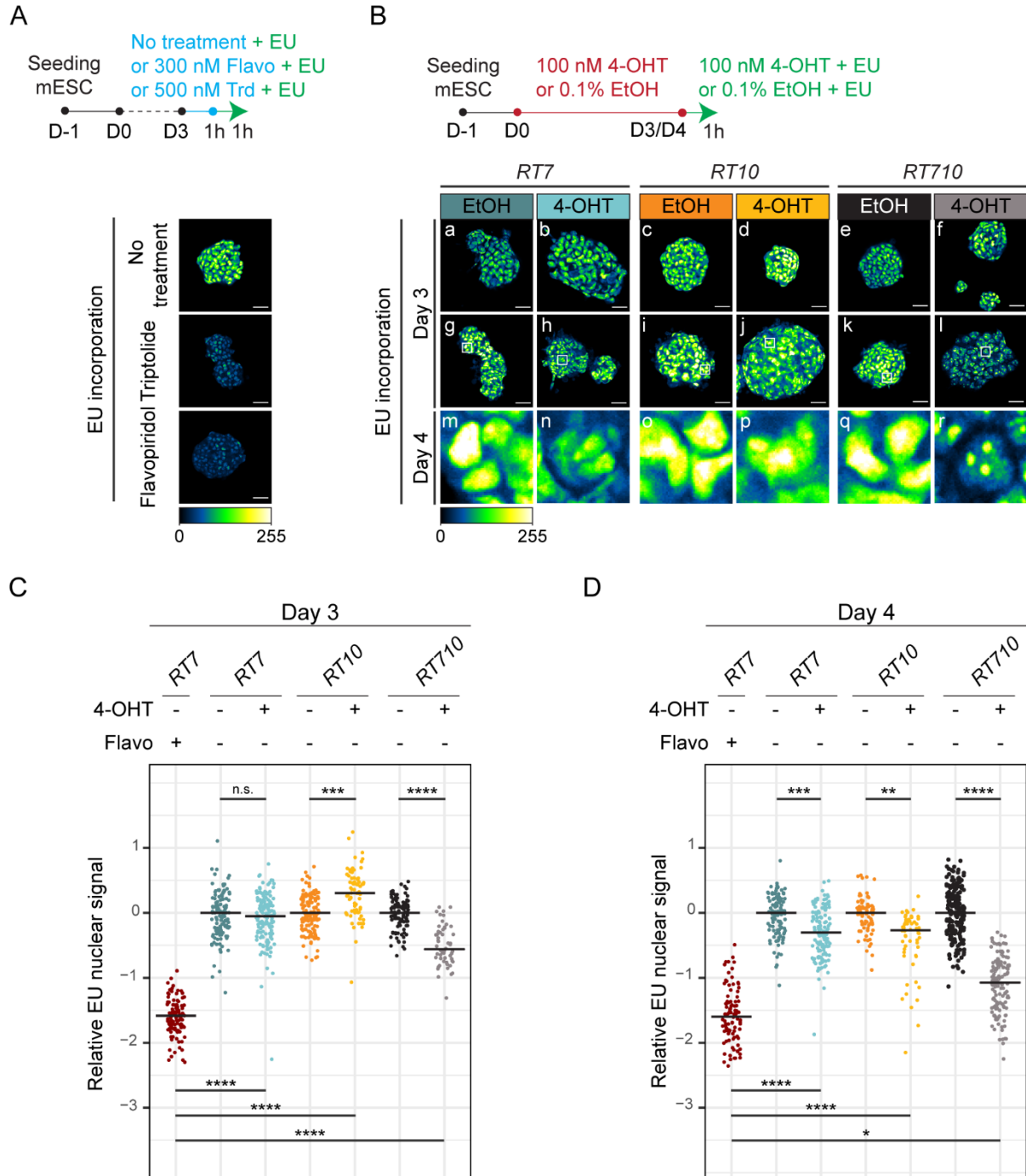
To further verify that Pol II is actively transcribing in our mutant cell lines, we analyzed the incorporation of 5-ethynyl uridine (EU) after one hour of incubation in the culture medium. To control the metabolic RNA labeling efficiency, we pre-incubated control cells with two different transcription inhibitors triptolide (transcription initiation inhibitor) or flavopiridol (transcription elongation inhibitor). As expected EU incorporation was readily detected in non-treated cells, and abolished in the triptolide or flavopiridol treated cells (Fig. 6A). Experiment





**Fig. 5: Active Pol II transcription in absence of canonical TFIID.** (A-D) Representative views of immunofluorescence using anti-RPB1<sup>pSer5</sup> (A, C) and anti-RPB1<sup>pSer2</sup> (B, D) antibodies on control (EtOH) and mutant (4-OHT) RT7, RT10 and RT710 mESCs at day 3 and 4. Color scale (Green Fire Blue LUT scale) is indicated at the bottom of the panel (B), scale bar: 50  $\mu$ m. (C-D) Quantification of RPB1<sup>pSer5</sup> (C) and RPB1<sup>pSer2</sup> (D) nuclear signal. The black lines indicate the median (n=1 biological replicate). Wilcoxon rank sum test: ns; not significant, \* <0.05; \*\*<0.005; \*\*\* <0.0005; \*\*\*\* <0.00005. RT7;

$R26^{CreERT2/+};Taf7^{ff}$ ,  $RT10$ ;  $R26^{CreERT2/+};Taf10^{ff}$ ,  $RT710$ ;  $R26^{CreERT2/+};Taf7^{ff};Taf10^{ff}$ , 4-OHT; 4-hydroxy tamoxifen.



**Fig. 6: Active RNA synthesis in absence of canonical TFIID.** (A, B) Control confocal imaging of 5-ethynyl uridine (EU) incorporation on control *RT7* and *RT10* mESCs in presence of 300 nM flavopiridol (Flavo) (*RT7*) or 500 nM triptolide (Trd) (*RT10*) (A) or in control (EtOH) and mutant (4-OHT) *RT7*, *RT10* and *RT710* mESCs at day 3 and 4 (B). In (B), panels m- r correspond to magnifications of representative nuclei highlighted in white squares in g-l, respectively. Color scale (Green Fire Blue LUT scale) is indicated at the bottom of the panels (C, D) Nuclear quantification of EU signal from (A, B). The black lines indicate the median. (n=1 biological replicate). Kruskal-Wallis test with Benjamini Hochberg post-hoc test. ns; not significant, \* <0.05; \*\*<0.005; \*\*\* <0.0005; \*\*\*\* <0.00005. *RT7*;

$R26^{CreERT2/+};Taf7^{ff}$ ,  $RT10$ ;  $R26^{CreERT2/+};Taf10^{ff}$ ,  $RT710$ ;  $R26^{CreERT2/+};Taf7^{ff};Taf10^{ff}$ , 4-OHT; 4-hydroxy tamoxifen.

was performed after 3 and 4 days of 4-OHT treatment (Fig. 6B). At day 3, EU incorporation is similar between all the control and mutant cell lines (Fig. 6B a-f, C). At day 4, no differences were observed between control and mutant *RT7* and *RT10* cells (Fig. 6B g-j, D). In mutant *RT710* cells, however, EU incorporation was significantly decreased and residual strong signal (Fig. 6B k, l, D) was restricted to some nuclear spots (Fig 6, r compared to m-q), indicating that in these mutant cells, Pol II transcription activity was reduced, despite the fact that no significant difference was observed in RPB1<sup>pSer2</sup> and RPB1<sup>pSer5</sup> levels (Fig. 5B, C).

Altogether, these data may suggest that despite the absence of canonical TFIID, Pol II transcription initiation and elongation is still active. Surprisingly, no obvious difference could be observed between *RT7* and *RT10* mutant cells despite the fact that TFIID assembly is much more perturbed in the absence of TAF10 than in the absence of TAF7.

#### 4. Discussion

In this study we have analyzed the depletion of two different TFIID subunits; TAF7 and TAF10 in the same cellular context. We showed that in the embryo and in mESCs, deletion of TAF7 is less severe than that of TAF10. As TAF10 is a shared subunit with the SAGA complex, we excluded the hypothesis that the increased severity in the *Taf10* loss of function phenotype is associated with a SAGA defect as loss of an essential SAGA protein; SUPT7L, does not aggravate the phenotype of the TAF7 depletion. Analysis of TFIID assembly by IP-MS indicates that TAF10 depletion has a more pronounced disassembly effect on TFIID assembly than TAF7 depletion. Indeed, while in the absence of TAF7, an almost fully assembled TAF7-less TFIID complex can be detected, in the absence of TAF10, an uncomplete TFIID complex mainly composed of the 5 TAF-containing (5TAF) core-TFIID, eventually associated partially

with TAF1/ TAF7 heterodimer, can be detected in the corresponding ESC extracts. Remarkably, combination of both TAF7 and TAF10 depletion in mESCs leads to a more severe phenotype and only the core-TFIID is detected. Surprisingly, activity of Pol II transcription assessed by RBP1<sup>pSer2</sup> and RBP1<sup>pSer5</sup> immunolocalization or by EU incorporation in nascent transcripts is not perturbed, except in ESCs where both TAF7 and TAF10 are depleted, correlating with a decreased Pol II activity. Altogether, our data suggest that Pol II transcription still occurs in the absence of fully assembled canonical TFIID. We propose that canonical TFIID can be replaced by TFIID submodules to promote transcription initiation. However, as suggested by the different phenotypes observed, Pol II transcription is not optimal in absence of canonical TFIID.

Recent structural studies have proposed that TFIID is playing a major role in PIC assembly and deposition of TBP on core promoters (Chen et al., 2021; Patel et al., 2018). While TBP has been initially characterized as the TATA-box binding protein (Hernandez, 1993), only a minority of mammalian promoters have a TATA box (Yang et al., 2007). TFIID is able to deposit TBP on TATA box-containing or TATA-less promoters, leading to a similar bending of the DNA (Chen et al., 2021). As a scaffold, TFIID is also offering several interaction interfaces with transcription factors via its different TAFs: for example, TAF10 has been shown to interact with the estrogen receptor as well as GATA1 (Jacq et al., 1994; Papadopoulos et al., 2015). Therefore, TFIID appears central to the process of Pol II transcription initiation. In all our mutant mESCs, TFIID assembly is affected but the structure of the TFIID scaffold is differentially perturbed. The loss of TAF7 has only limited effect as both anti-TBP and anti-TAF12 IPs indicate that a TAF7-less TFIID still containing TBP is formed. The loss of TAF10, however, has a much stronger effect as the core-TFIID submodule is the most advanced complex detected. Interestingly, we also observed a sub stoichiometric interaction with TAF1 and TAF7. TAF1 and TBP co-translationally assemble (Kamenova et al., 2019) via TAF1's

TAND domain (Kokubo et al., 1994; Nishikawa et al., 1997). TAF1 also interact with TAF7 via its DU3591 central domain (Bhattacharya et al., 2014; Wang et al., 2014a) and with the core-TFIID via its 3 TAF6-binding motifs (Chen et al., 2021). This could suggest the existence of an incomplete TFIID complex containing the core-TFIID associated with TAF1, TAF7 and TBP. Interestingly, it has been recently shown that an *in vitro* assembled lobe C-like single TAF (S-TAF) complex composed of TAF1, TAF7, TAF11, TAF13 and TBP is able to promote *in vitro* transcription and pausing from the *HSPA1B* human promoter in presence of purified PIC components (Fant et al., 2020). There is no evidence that support the existence of the S-TAF complex *in vivo* and in the *RT10* mutant cells, the anti-TBP IP did not show strong interaction with TAF11 and TAF13. Although it is not clear how the association between core-TFIID, TAF1 and TAF7 could be assembled, TAF7 appears to be important as in mutant *RT710* cells depleted in both TAF10 and TAF7, no such sub stoichiometric interactions are observed and only the 5TAF core-complex is detected.

Interestingly, our data indicate that Pol II transcription is still active despite the absence of TAF7 or TAF10 proteins. Only when the two proteins are depleted in mutant *RT710* cells, a significant decrease in global nascent RNA production was observed. Several non-mutually exclusive hypotheses can be proposed to explain the absence of effect on Pol II activity. A first hypothesis is that the residual low abundant canonical TFIID complexes, which are detected in the mutant mESCs despite the efficient depletion in TAF7 or TAF10 proteins, might be responsible for the maintenance of active transcription. Such stable TFIID complexes have already been observed in *Taf10* conditional E9.5 mouse mutants (Bardot 2017). Although we cannot exclude this hypothesis, it is unlikely to be the main reason as i) we did not detect a decrease in global Pol II activity in the mutant *RT7* and *RT10* cells, and ii) despite the maintenance of Pol II activity, there is a strong phenotype associated with the depletion of TAF10. A second hypothesis is that a full TFIID containing TAF10 is not required to maintain

active transcription as proposed in (Tatarakis et al., 2008). After initiation of transcription, TFIID and some GTFs remain bound to the promoter while other PIC components are released, forming a scaffold that facilitates the recruitment of new Pol II (Hawley and Roeder, 1987; Zawel et al., 1995). The minor proportion of canonical TFIID complexes that we detected in our mutant cells could be already stably interacting with the promoters of actively transcribed Pol II genes. Analysis of the distribution of TAF7 and TAF10 proteins on the chromatin of the mutant *RT7*, *RT10* and *RT710* cells would allow to test this hypothesis. A third interesting hypothesis is that the TAF7-less TFIID complex and the core-TFIID complex, eventually associated with TAF7 and TAF1, are able to support efficiently Pol II transcription activity in the *RT7* and *RT10* mutant cells in the absence of canonical TFIID although the molecular mechanisms of Pol II recruitment and transcription initiation are not clear. Our data are in the same line as several evidences in the literature that point out that Pol II transcription does not always rely on canonical TFIID nor TBP. First, alternative complexes to TFIID support Pol II transcription in specific conditions. It has been shown that the TFTC complex that contains some TAFs but not TBP is able to initiate transcription *in vitro* on TATA- containing and TATA-less promoters (Wieczorek et al., 1998). Ribosomal protein coding genes transcription initiation is controlled by a non-TFIID complex containing the TBP paralog TBPL1 (Baumann and Gilmour, 2017; Parry et al., 2010; Wang et al., 2014b) and active transcription during oocyte growth is mediated by an oocyte-specific complex lacking TBP and TAFs (Yu et al., 2020). Second, conditional deletions of *Taf7* or *Taf10* have shown that the requirement for these subunits is not absolute and is dependent on the cellular context (Bardot et al., 2017; Gegonne et al., 2012; Indra et al., 2005; Mohan et al., 2003; Tatarakis et al., 2008). In these analyses, however, the effect of the mutation on Pol II transcriptional activity has not been directly studied except in (Bardot et al., 2017) where it was shown that cyclic transcription waves that control somitogenesis (Hubaud and Pourquié, 2014), are still active in the absence of TAF10,



strongly supporting the fact that transcription initiation can occur without TAF10 and canonical TFIID in this particular context.

A striking observation is that the severity of the phenotypes associated with the depletion of TAF7 or TAF10 does not correlate with the global Pol II nascent transcription activity while it correlates with the degree of TFIID assembly defects. Indeed, mutant *RT7* cells which have TAF7-less TFIID complex have a much milder phenotype compared to mutant *RT10* cells, in which the core-TFIID, weakly associated with TAF1 and TAF7, is the most advanced assembled TFIID submodule. Our data show that the phenotype observed in mutant *RT10* cells is not associated with a SAGA complex defect as mutation in *Supt7l* coding for a protein important for SAGA assembly does not have an effect in this context (Fischer et al., 2021). Remarkably, mutant *RT710* cells that are devoid of TAF7 and TAF10 and mainly contain the core-TFIID have the most severe phenotype. These data suggest that TAF7 and TAF10 absence does not affect Pol II transcription similarly. The increased severity in phenotype suggests that there is an additive effect of the absence of TAF7 and TAF10 although we cannot exclude some indirect effects leading to an overall transcription defect as suggested by the decreased TBP – Pol I SL1 complex at day 3 and the strong reduction in EU incorporation. Nevertheless, it is tempting to speculate that TAF7 and TAF10 are associated with specific transcription programs. Analysis of nascent transcripts in the mutant *RT7*, *RT10* and *RT710* cells will be important to analyze precisely the effect on active transcription but analysis of the phenotype of the mutant embryos support this hypothesis. Using the embryonic mesoderm as a common same genetic context, conditional deletion of *Taf7* leads to a growth arrest in the trunk/tail region also reported for the deletion of *Taf10* in the same context (Bardot et al., 2017), however, while *TCre;Taf10* mutant embryos are resorbed at E11.5, *TCre;Taf7* mutant embryos are still developing at E12.5. Specific differences were also observed. On one hand, *TCre;Taf10* mutant embryos lack limb buds and are devoid of red blood cells. This latter phenotype is a

consequence of the specific interaction between TAF10 and GATA1 during erythropoiesis (Papadopoulos et al., 2015) and this is presumably the main reason for the more precocious lethality, in addition to placental and allantois defects (Bardot et al., 2017). On the other hand, *TCre;Taf7* mutant embryos have hypotrophied limb buds and enlarged pericardium suggesting that these mutant embryos also have cardiovascular defects. Altogether, these specific differences are in favor of specific transcription regulation mediated via TFIID.

Another non-exclusive hypothesis is that in the absence of a fully assembled TFIID, the precision of Pol II transcription initiation is affected. Two types of Pol II initiation have been described: i) promoters of highly active and tissue specific genes are associated with TSS concentrated on a very narrow region (sharp TSS usage) and the presence of TATA-box, ii) promoters of ubiquitously expressed genes but also developmental genes are associated with numerous TSS occupying a larger region (broad TSS usage) (Carninci et al., 2006). Moreover, during oocyte growth and during early development, a change in the TSS usage, associated with a shift in the transcriptional machinery has been described (Cvetesic et al., 2020; Haberle et al., 2014; Yu et al., 2020). In particular it was shown that the TBPL2/TFIIA oocyte-specific transcription initiation machinery has a specific sharp TSS signature coupled with the association with TATA-like containing promoter sequences, while the canonical TFIID PIC has a broad TSS signature associated with GC-rich sequences (Yu et al., 2020). These data indicate that TFIID is associated with specific core promoter recognition. It is possible that in the absence of fully assembled holo TFIID, especially in the situation where only a very primitive complex is present, Pol II is still recruited on the DNA to potentially cryptic initiation sites, not in known promoter-independent manner and less efficiently. This could explain the intriguing observation that no effect on Pol II CTD phosphorylation nor EU incorporation could be observed in the absence of TAF7 or TAF10 while strong phenotypic differences are present.

Analysis of the distribution of Pol II on the chromatin as well as Cap Analysis of gene Expression would allow to get more insight into this interesting hypothesis.

### **Material and methods**

See supplementary data.

### **Data availability**

Proteomic data data have been deposited in the proteomics identification database (ProteomeXchange PRIDE database, accession code: PXDxxxxx).

### **Acknowledgments**

We thank all members of the Tora lab for protocols, thoughtful discussions and suggestions, C. Ebel and M. Philipps for help with FACS sorting, S. Falcone and the IGBMC animal facility for animal caretaking and the IGBMC cell culture facility.

### **Funding**

This work was supported by funds from the Agence Nationale de la Recherche (PICen-19-CE11-0003-02, and EpiCAST-19-CE12-0029-01) and NIH grant 1RO1GM131626-01 to LT, and Fondation pour la Recherche Médicale (EQU202103012631) to SDV and LT as well as funds from CNRS, INSERM, Strasbourg University, and Investissements d'Avenir grants (ANR-10-IDEX-0002-02 and ANR-10-LABX-0030-INRT).

### **Author's contributions**

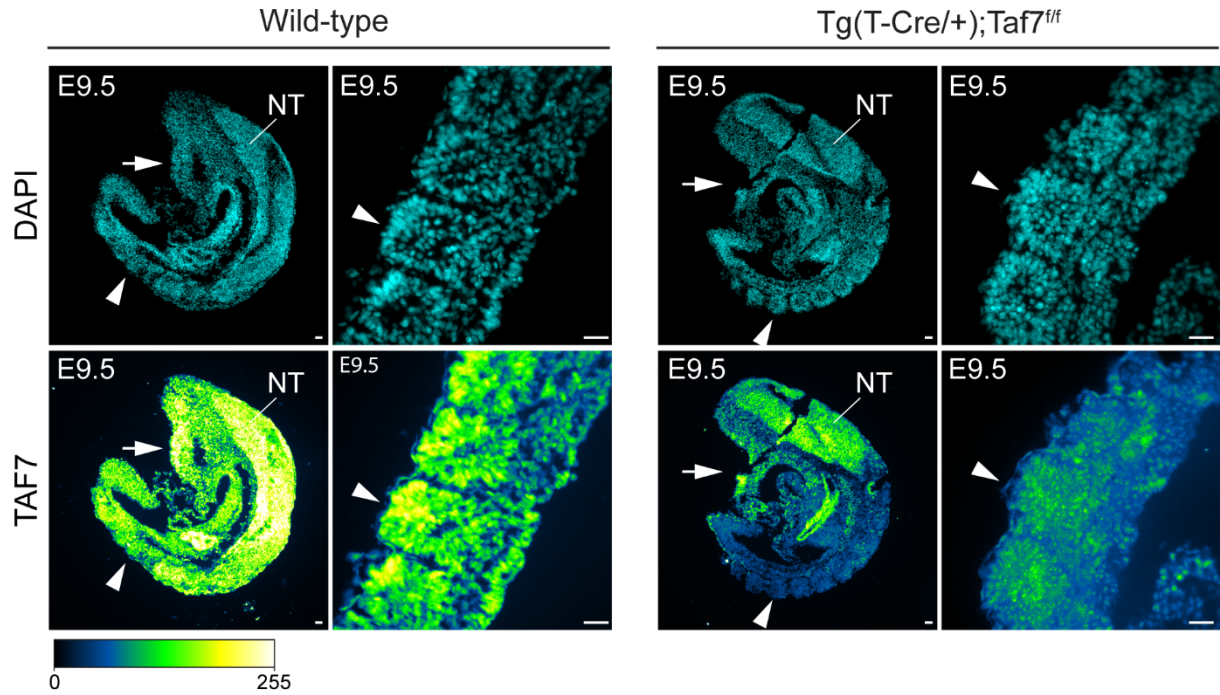
VH, PB, MS, CE, LHA performed experiments. BM and LN performed mass spectrometry analysis, VH, LT, and SDV designed the experiments. SDV conceived the project and wrote the paper with support from VH and LT.

### **Competing interests**

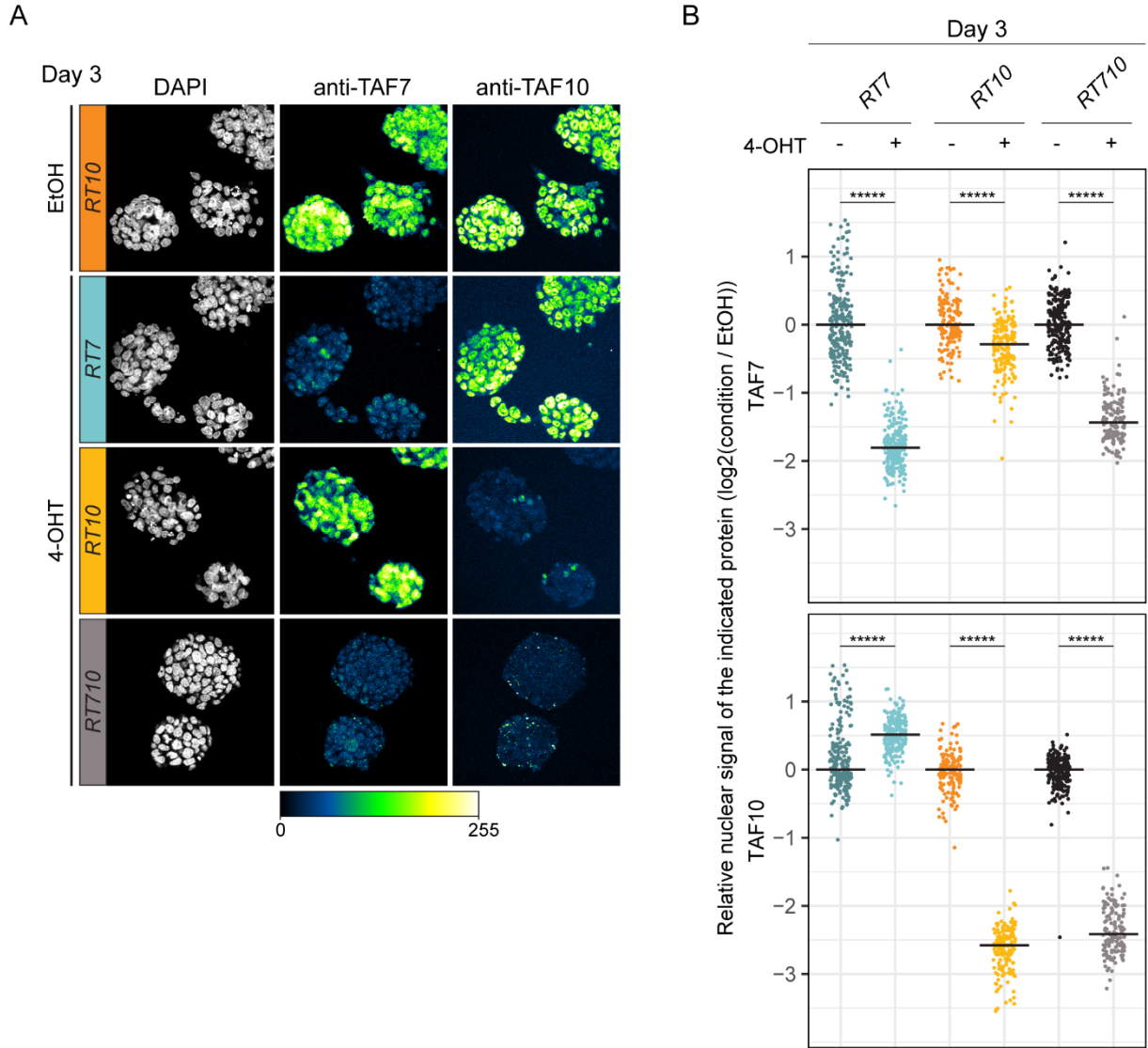
The authors declare that they have no competing interests.



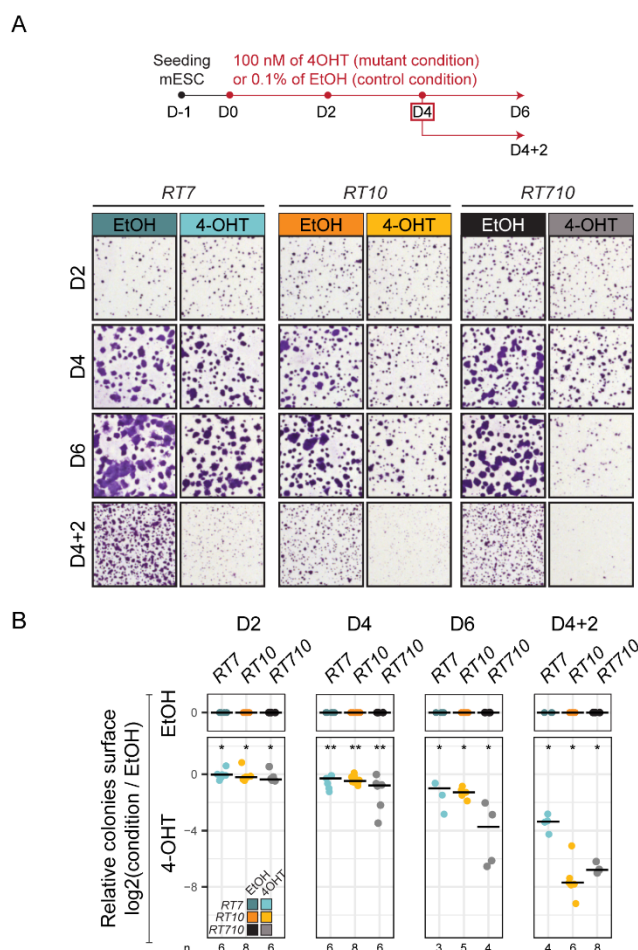
## 5. Supplementary figures



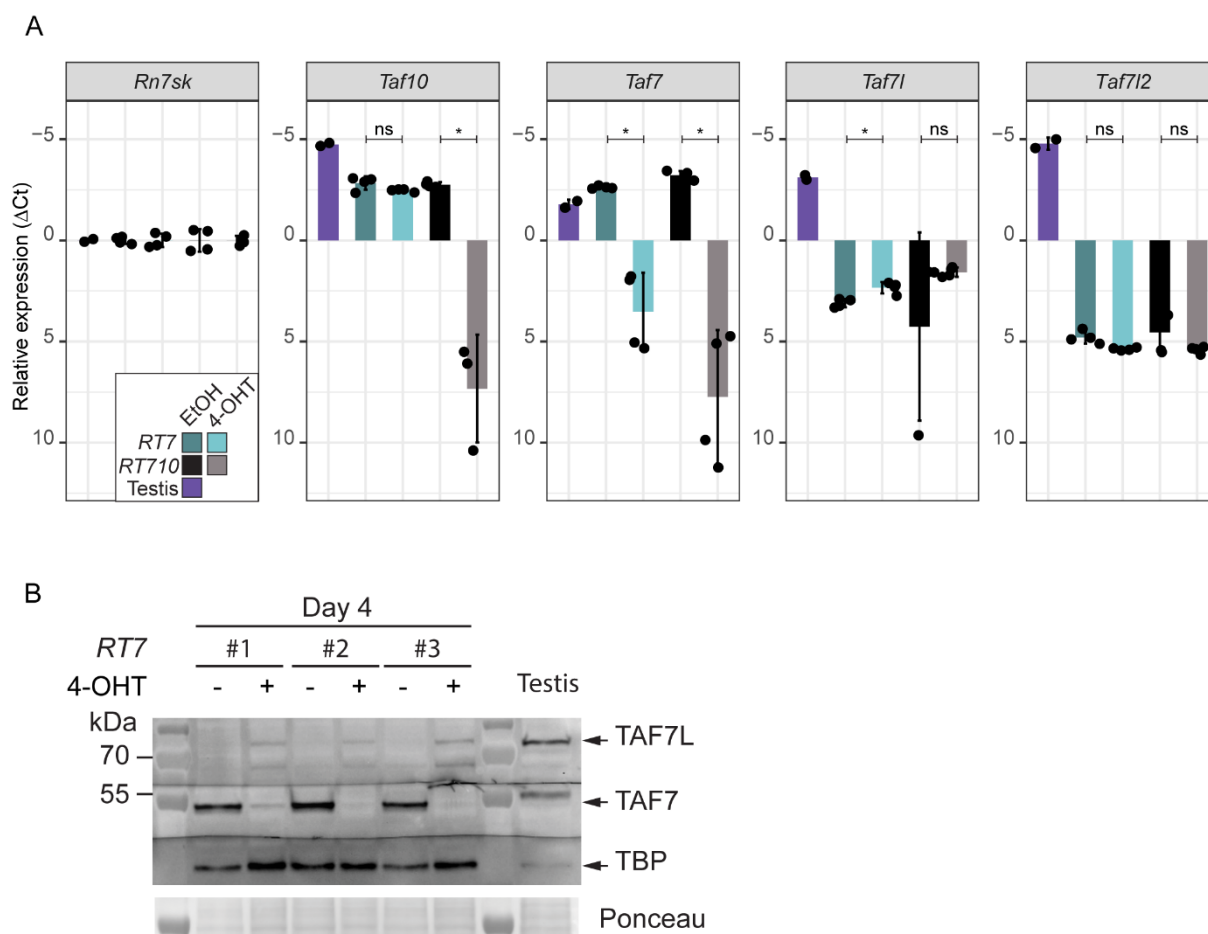
**Sup Fig. 1: Efficient TAF7 depletion in the paraxial mesoderm of embryonic day (E) 9.5 *Tg(T-Cre/+);Taf7<sup>fl/fl</sup>* mutant embryo.** DAPI counterstained anti-TAF7 immunofluorescence on sagittal section from wild-type (left panel) and *Tg(T-Cre/+);Taf7<sup>fl/fl</sup>* E9.5 mutant (right panel) embryos. In *Tg(T-Cre/+)* embryos, Cre is active in progenitors that contribute to mesoderm cells posterior to the heart (white arrow), including the paraxial mesoderm (unsegmented presomitic mesoderm and somites (arrowhead)) but not in the neural tube (NT). Color scale (Green Fire Blue LUT scale) is indicated at the bottom (n=2).



**Sup Fig. 2: Homogenous TAF7 or/and TAF10 depletion in mutant *RT7*, *RT10* and *RT710* mESCs.** (A) DAPI counterstained anti-TAF7 and anti-TAF10 immunofluorescence on control (EtOH) *RT10* and mutant (4-OHT) *RT7*, *RT10* and *RT710* mESCs at day 3. Color scale (Green Fire Blue LUT scale) is indicated at the bottom. (B) Nuclear quantification of anti-TAF7 and anti-TAF10 immunofluorescence assay in (A). The black lines indicate the median (n=1 replicate, except control and mutant *RT7* n=2 biological replicates). Wilcoxon rank sum test: \*\*\*\*\*, <0.000005. *RT7*; *R26<sup>CreERT2/+</sup>*; *Taf7<sup>fl/fl</sup>*, *RT10*; *R26<sup>CreERT2/+</sup>*; *Taf10<sup>fl/fl</sup>*, *RT710*; *R26<sup>CreERT2/+</sup>*; *Taf7<sup>fl/fl</sup>*; *Taf10<sup>fl/fl</sup>*, 4-OHT; 4-hydroxy tamoxifen.

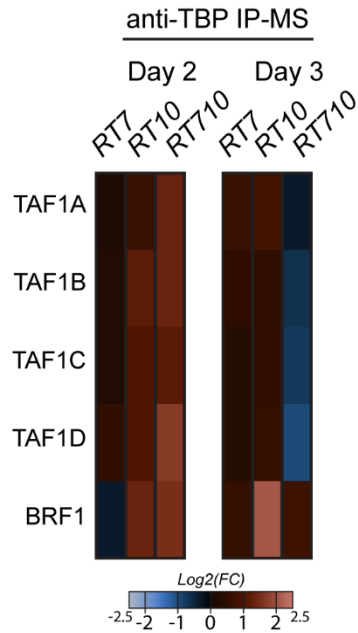


**Sup Fig. 3: Absence of TAF7 and/or TAF10 impact *RT7*, *RT10* and *RT710* mESC clonal growth after replating.** (A) Crystal violet-stained wells of control (EtOH) and mutant (4-OHT) *RT7*, *RT10* and *RT710* mESCs at day 3, day 4 and day 6, and 2 days after replating at day 4 (D4+2). The timing of the experiment is indicated at the bottom. (B) relative quantification of the total surface covered by the colonies from (A). The black lines indicate the median. n indicated the number of replicates (n≥2 biological replicates). Wilcoxon rank sum test done on mutant versus control for each clone. ns; not significant, \* <0.05; \*\*<0.005; \*\*\* <0.0005. *RT7*; *R26<sup>CreERT2/+</sup>;Taf7<sup>fl/fl</sup>*, *RT10*; *R26<sup>CreERT2/+</sup>;Taf10<sup>fl/fl</sup>*, *RT710*; *R26<sup>CreERT2/+</sup>;Taf7<sup>fl/fl</sup>;Taf10<sup>fl/fl</sup>*, 4-OHT; 4-hydroxy tamoxifen.

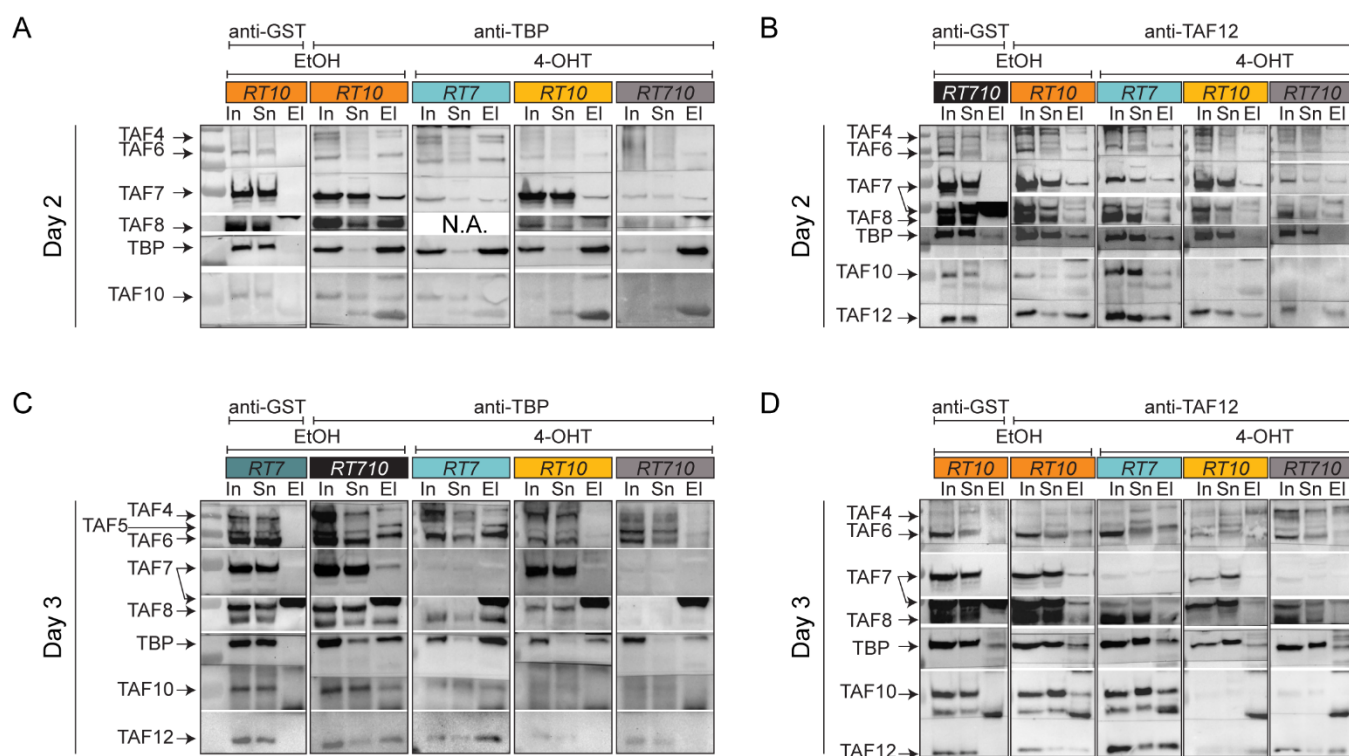


**Sup Fig. 4: *Taf7* paralogues do not rescue the TAF7 depletion in *RT7* and *RT710* mESCs.** (A) RT-qPCR analysis of *Taf10*, *Taf7*, *Taf7l* and *Taf7l2* expression from control (EtOH) and mutant (4-OHT) *RT7* and *RT710* mESCs RNA at day 4. The RNA polymerase III transcribed gene *Rn7sk* was used as reference gene. n=2 biological x 2 technical replicates, except for testis; n= 2 technical replicates. Wilcoxon rank sum test: ns; not significant, \* <0.05; \*\*<0.005. (B) Western blot analysis of TAF7L, TAF7 and TBP expression at day 4 from control (EtOH) and mutant (4-OHT) mESCs nuclear enriched cell lysates and testis whole cell lysate (n=3). *RT7*; *R26<sup>CreERT2/+</sup>;Taf7<sup>fl/fl</sup>*, *RT710*; *R26<sup>CreERT2/+</sup>;Taf7<sup>fl/fl</sup>;Taf10<sup>fl/fl</sup>*.





**Sup Fig. 5: Redistribution of TBP to SL1 and TFIIB complexes upon TAF10 depletion.** Relative normalized XIC values for SL1 complex subunits (TAF1A, TAF1B, TAF1C and TAF1D) and TFIIB subunit BRF1 from anti-TBP immunoprecipitation coupled to mass spectrometry analysis (IP-MS) from mutant versus control *RT7*, *RT10* and *RT710* mESCs nuclear enriched cell lysates at day 2 and 3. The color legend is indicated at the bottom of the panel (n= 3 technical replicates).



**Sup Fig. 6: Analysis of TFIID assembly in absence of TAF7 and/or TAF10 by immunoprecipitation coupled to western blot analyses.** (A-D) Western blot analysis of mock anti-GST immunoprecipitation (IP) (A-D), anti-TBP IP (A, C) and anti-TAF12 IP (B, D) on control (EtOH) and mutant (4-OHT) *RT7*, *RT10* and *RT710* mESCs at day 2 (A, B) and day 3 (C, D). The proteins analyzed are indicated on the left of each panel. In; input, Sn; supernatant, El; elution.

## 6. References article

- Anandapadamanaban, M., Andresen, C., Helander, S., Ohyama, Y., Siponen, M.I., Lundström, P., Kokubo, T., Ikura, M., Moche, M., and Sunnerhagen, M. (2013). High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nature Structural & Molecular Biology* 20, 1008–1014.
- Bardot, P., Vincent, S.D., Fournier, M., Hubaud, A., Joint, M., Tora, L., and Pourquié, O. (2017). The TAF10-containing TFIID and SAGA transcriptional complexes are dispensable for early somitogenesis in the mouse embryo. *Development (Cambridge, England)* 144, 3808–3818.
- Bártfai, R., Balduf, C., Hilton, T., Rathmann, Y., Hadzhiev, Y., Tora, L., Orbán, L., and Müller, F. (2004). TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Current Biology : CB* 14, 593–598.
- Baumann, D.G., and Gilmour, D.S. (2017). A sequence-specific core promoter-binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes. *Nucleic Acids Research* 45, 10481–10491.
- Bhattacharya, S., Lou, X., Hwang, P., Rajashankar, K.R., Wang, X., Gustafsson, J.-Å., Fletterick, R.J., Jacobson, R.H., and Webb, P. (2014). Structural and functional insight into TAF1–TAF7, a subcomplex of transcription factor II D. *Proc National Acad Sci* 111, 9103–9108.
- Bhuiyan, T., and Timmers, H.T.M. (2019). Promoter Recognition: Putting TFIID on the Spot. *Trends in Cell Biology* 29, 752–763.
- 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.
- 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., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics* 38, 626–635.
- Chen, Z., and Manley, J.L. (2003). In Vivo Functional Analysis of the Histone 3-like TAF9 and a TAF9-related Factor, TAF9L\*. *J Biol Chem* 278, 35172–35183.
- Chen, X., Qi, Y., Wu, Z., Wang, X., Li, J., Zhao, D., Hou, H., Li, Y., Yu, Z., Liu, W., et al. (2021). Structural insights into preinitiation complex assembly on core promoters. *Science* 372, eaba8490.
- Cvetesic, N., Borkowska, M., Hatanaka, Y., Leitch, H.G., Müller, F., Yu, C., Vincent, S.D., Tora, L., Hajkova, P., and Lenhard, B. (2020). Global regulatory transitions at core promoters demarcate the mammalian germline cycle. *BioRxiv* 11, 2020.10.30.361865.

Dantonel, J.C., Quintin, S., Lakatos, L., Labouesse, M., and Tora, L. (2000). TBP-like factor is required for embryonic RNA polymerase II transcription in *C. elegans*. *Molecular Cell* 6, 715–722.

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., et al. (2020). TFIID Enables RNA Polymerase II Promoter-Proximal Pausing. *Molecular Cell*.

Fischer, V., Schumacher, K., Tora, L., and Devys, D. (2019). Global role for coactivator complexes in RNA polymerase II transcription. *Transcription* 10, 29–36.

Fischer, V., Plassard, D., Ye, T., Reina-San-Martin, B., Stierle, M., Tora, L., and Devys, D. (2021). The related coactivator complexes SAGA and ATAC control embryonic stem cell self-renewal through acetyltransferase-independent mechanisms. *Cell Reports* 36, 109598.

Freiman, R.N., Albright, S.R., Zheng, S., Sha, W.C., Hammer, R.E., and Tjian, R. (2001). Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science (New York, NY)* 293, 2084–2087.

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.

Gazdag, E., Rajkovic, A., Torres-Padilla, M.-E., and Tora, L. (2007). Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction (Cambridge, England)* 134, 51–62.

Gazdag, E., Santenard, A., Ziegler-Birling, C., Altobelli, G., Poch, O., Tora, L., and Torres-Padilla, M.-E. (2009). TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes and Development* 23, 2210–2223.

Gegonne, A., Tai, X., Zhang, J., Wu, G., Zhu, J., Yoshimoto, A., Hanson, J., Cultraro, C., Chen, Q.-R., Guintier, T., et al. (2012). The general transcription factor TAF7 is essential for embryonic development but not essential for the survival or differentiation of mature T cells. *Molecular and Cellular Biology* 32, 1984–1997.

Gupta, K., Watson, A.A., Baptista, T., Scheer, E., Chambers, A.L., Koehler, C., Zou, J., Obong-Ebong, I., Kandiah, E., Temblador, A., et al. (2017). Architecture of TAF11/TAF13/TBP complex suggests novel regulation properties of general transcription factor TFIID. *ELife* 6, 1.

Gura, M.A., Mikedis, M.M., Seymour, K.A., Rooij, D.G. de, Page, D.C., and Freiman, R.N. (2020). Dynamic and regulated TAF gene expression during mouse embryonic germ cell development. *PLoS Genetics* 16, e1008515.

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., et al. (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* 507, 381–385.

- Hansen, S.K., Takada, S., Jacobson, R.H., Lis, J.T., and Tjian, R. (1997). Transcription properties of a cell type-specific TATA-binding protein, TRF. *Cell* *91*, 71–83.
- Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology* *18*, 263–273.
- Hawley, D.K., and Roeder, R.G. (1987). Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. *J Biol Chem* *262*, 3452–3461.
- Helmlinger, D., and Tora, L. (2017). Sharing the SAGA. *Trends in Biochemical Sciences* *42*, 850–861.
- Herbst, D.A., Esbin, M.N., Louder, R.K., Dugast-Darzacq, C., Dailey, G.M., Fang, Q., Darzacq, X., Tjian, R., and Nogales, E. Structure of the human SAGA coactivator complex: The divergent architecture of human SAGA allows modular coordination of transcription activation and co-transcriptional splicing.
- Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? *Gene Dev* *7*, 1291–1308.
- Herrera, F.J., Yamaguchi, T., Roelink, H., and Tjian, R. (2014). Core promoter factor TAF9B regulates neuronal gene expression. *ELife* *3*, e02559.
- Hiller, M., Chen, X., Pringle, M.J., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B., Lin, T.-Y., Marino, S., and Fuller, M.T. (2004). Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development (Cambridge, England)* *131*, 5297–5308.
- Hiller, M.A., Lin, T.Y., Wood, C., and Fuller, M.T. (2001). Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes and Development* *15*, 1021–1030.
- Hubaud, A., and Pourquié, O. (2014). Signalling dynamics in vertebrate segmentation. *Nature Reviews Molecular Cell Biology* *15*, 709–721.
- Indra, A.K., Mohan, W.S., Frontini, M., Scheer, E., Messaddeq, N., Metzger, D., and Tora, L. (2005). TAF10 is required for the establishment of skin barrier function in foetal, but not in adult mouse epidermis. *Developmental Biology* *285*, 28–37.
- Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., and Tora, L. (1994). Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* *79*, 107–117.
- Jallow, Z., Jacobi, U.G., Weeks, D.L., Dawid, I.B., and Veenstra, G.J.C. (2004). Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in *Xenopus*. *Proceedings of the National Academy of Sciences* *101*, 13525–13530.

- Kaltenbach, L., Horner, M.A., Rothman, J.H., and Mango, S.E. (2000). The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. *Molecular Cell* 6, 705–713.
- Kamenova, I., Mukherjee, P., Conic, S., Mueller, F., El-Saafin, F., Bardot, P., Garnier, J.-M., Dembélé, D., Capponi, S., Timmers, H.T.M., et al. (2019). Co-translational assembly of mammalian nuclear multisubunit complexes. *Nature Communications* 10, 93.
- Kaufmann, J., Verrijzer, C.P., Shao, J., and Smale, S.T. (1996). CIF, an essential cofactor for TFIID-dependent initiator function. *Gene Dev* 10, 873–886.
- Kokubo, T., Yamashita, S., Horikoshi, M., Roeder, R.G., and Nakatani, Y. (1994). Interaction between the N-terminal domain of the 230-kDa subunit and the TATA box-binding subunit of TFIID negatively regulates TATA-box binding. *Proc National Acad Sci* 91, 3520–3524.
- Langer, D., Martianov, I., Alpern, D., Rhinn, M., Keime, C., Dollé, P., Mengus, G., and Davidson, I. (2016). Essential role of the TFIID subunit TAF4 in murine embryogenesis and embryonic stem cell differentiation. *Nature Communications* 7, 11063–16.
- Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell* 157, 13–25.
- Liu, W.-L., Coleman, R.A., Grob, P., King, D.S., Florens, L., Washburn, M.P., Geles, K.G., Yang, J.L., Ramey, V., Nogales, E., et al. (2008). Structural Changes in TAF4b-TFIID Correlate with Promoter Selectivity. *Mol Cell* 29, 81–91.
- Martianov, I., Fimia, G.M., Dierich, A., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2001). Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Molecular Cell* 7, 509–515.
- Martianov, I., Velt, A., Davidson, G., Choukrallah, M.-A., and Davidson, I. (2016). TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. *Scientific Reports* 6, 32069.
- Martin, J., Halenbeck, R., and Kaufmann, J. (1999). Human Transcription Factor hTAFII150 (CIF150) Is Involved in Transcriptional Regulation of Cell Cycle Progression. *Mol Cell Biol* 19, 5548–5556.
- Metzger, D., Scheer, E., Soldatov, A., and Tora, L. (1999). Mammalian TAF(II)30 is required for cell cycle progression and specific cellular differentiation programmes. *The EMBO Journal* 18, 4823–4834.
- Mohan, W.S., Scheer, E., Wendling, O., Metzger, D., and Tora, L. (2003). TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice. *Molecular and Cellular Biology* 23, 4307–4318.
- Nishikawa, J., Kokubo, T., Horikoshi, M., Roeder, R.G., and Nakatani, Y. (1997). *Drosophila* TAFII230 and the transcriptional activator VP16 bind competitively to the TATA box-binding domain of the TATA box-binding protein. *Proc National Acad Sci* 94, 85–90.

- Ohbayashi, T., Makino, Y., and Tamura, T.A. (1999a). Identification of a mouse TBP-like protein (TLP) distantly related to the drosophila TBP-related factor. *Nucleic Acids Research* 27, 750–755.
- Ohbayashi, T., Kishimoto, T., Makino, Y., Shimada, M., Nakadai, T., Aoki, T., Kawata, T., Niwa, S., and Tamura, T. (1999b). Isolation of cDNA, chromosome mapping, and expression of the human TBP-like protein. *Biochemical and Biophysical Research Communications* 255, 137–142.
- Papadopoulos, P., Gutiérrez, L., Demmers, J., Scheer, E., Pourfarzad, F., Papageorgiou, D.N., Karkoulia, E., Strouboulis, J., Werken, H.J.G. van de, Linden, R. van der, et al. (2015). TAF10 Interacts with the GATA1 Transcription Factor and Controls Mouse Erythropoiesis. *Molecular and Cellular Biology* 35, 2103–2118.
- Papai, G., Tripathi, M.K., Ruhlmann, C., Werten, S., Crucifix, C., Weil, P.A., and Schultz, P. (2009). Mapping the initiator binding Taf2 subunit in the structure of hydrated yeast TFIID. *Structure (London, England : 1993)* 17, 363–373.
- 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 and 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* 362, eaau8872.
- Perantoni, A.O., Timofeeva, O., Naillat, F., Richman, C., Pajni-Underwood, S., Wilson, C., Vainio, S., Dove, L.F., and Lewandoski, M. (2005). Inactivation of FGF8 in early mesoderm reveals an essential role in kidney development. *Development* 132, 3859–3871.
- Persengiev, S.P., Zhu, X., Dixit, B.L., Maston, G.A., Kittler, E.L.W., and Green, M.R. (2003). TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proceedings of the National Academy of Sciences* 100, 14887–14891.
- Pointud, J.-C., Mengus, G., Brancorsini, S., Monaco, L., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2003). The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID subunit TAF7, is developmentally regulated during male germ-cell differentiation. *Journal of Cell Science* 116, 1847–1858.
- Rabenstein, M.D., Zhou, S., Lis, J.T., and Tjian, R. (1999). TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proceedings of the National Academy of Sciences* 96, 4791–4796.
- Roeder, R.G. (2019). 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nature Structural & Molecular Biology* 26, 783–791.
- Soutoglou, E., Demény, M.A., Scheer, E., Fienga, G., Sassone-Corsi, P., and Tora, L. (2005). The nuclear import of TAF10 is regulated by one of its three histone fold domain-containing interaction partners. *Molecular and Cellular Biology* 25, 4092–4104.



- Takada, S., Lis, J.T., Zhou, S., and Tjian, R. (2000). A TRF1:BRF complex directs *Drosophila* RNA polymerase III transcription. *Cell* 101, 459–469.
- Tatarakis, A., Margaritis, T., Martinez-Jimenez, C.P., Kouskouti, A., Mohan, W.S., Haroniti, A., Kafetzopoulos, D., Tora, L., and Talianidis, I. (2008). Dominant and redundant functions of TFIID involved in the regulation of hepatic genes. *Molecular Cell* 31, 531–543.
- Teichmann, M., Wang, Z., Martinez, E., Tjernberg, A., Zhang, D., Vollmer, F., Chait, B.T., and Roeder, R.G. (1999). Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. *Proceedings of the National Academy of Sciences* 96, 13720–13725.
- Trowitzsch, S., Viola, C., Scheer, E., Conic, S., Chavant, V., Fournier, M., Papai, G., Ebong, I.-O., Schaffitzel, C., Zou, J., et al. (2015). Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nature Communications* 6, 6011.
- Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Molecular Cell* 45, 439–446.
- Ventura, A., Kirsch, D.G., McLaughlin, M.E., Tuveson, D.A., Grimm, J., Lintault, L., Newman, J., Reczek, E.E., Weissleder, R., and Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. *Nature* 445, 661–665.
- Voss, A.K., Thomas, T., Petrou, P., Anastassiadis, K., Schöler, H., and Gruss, P. (2000). Taube nuss is a novel gene essential for the survival of pluripotent cells of early mouse embryos. *Development (Cambridge, England)* 127, 5449–5461.
- Wang, P.J., and Page, D.C. (2002). Functional substitution for TAF II 250 by a retroposed homolog that is expressed in human spermatogenesis. *Hum Mol Genet* 11, 2341–2346.
- Wang, H., Curran, E.C., Hinds, T.R., Wang, E.H., and Zheng, N. (2014a). Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. *Cell Research* 24, 1433–1444.
- Wang, Y.L., Duttke, S.H.C., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J., and Kadonaga, J.T. (2014b). TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes and Development* 28, 1550–1555.
- Wieczorek, E., Brand, M., Jacq, X., and Tora, L. (1998). Function of TAFII-containing complex without TBP in transcription by RNA polymerase II. *Nature* 393, 187–191.
- Xiao, L., Kim, M., and DeJong, J. (2006). Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice. *Gene Expression Patterns* 6, 409–419.
- 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.



- Yu, C., Cvetic, N., Hisler, V., Gupta, K., Ye, T., Gazdag, E., Negroni, L., Hajkova, P., Berger, I., Lenhard, B., et al. (2020). TBPL2/TFIIA complex establishes the maternal transcriptome through oocyte-specific promoter usage. *Nature Communications* 11, 6439.
- Zabidi, M.A., and Stark, A. (2016). Regulatory Enhancer–Core–Promoter Communication via Transcription Factors and Cofactors. *Trends Genet* 32, 801–814.
- Zawel, L., Kumar, K.P., and Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. *Gene Dev* 9, 1479–1490.
- Zhang, D., Penttilä, T.L., Morris, P.L., Teichmann, M., and Roeder, R.G. (2001). Spermiogenesis deficiency in mice lacking the Trf2 gene. *Science (New York, NY)* 292, 1153–1155.
- Zhou, H., Grubisic, I., Zheng, K., He, Y., Wang, P.J., Kaplan, T., and Tjian, R. (2013). Taf7l cooperates with Trf2 to regulate spermiogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 110, 16886–16891.



## 7. Material & Method:

### Mice

Animal experimentation was carried out according to animal welfare regulation and guidelines of the French Ministry of Agriculture and Ministry of Higher Education, Research and Innovation. The original mouse lines (*Taf7<sup>f</sup>*, *Taf10<sup>f</sup>*, *R26<sup>CreERT2</sup>*, Tg(T-Cre)) have all been described (Supplementary Table 1). The day of vaginal plug was scored as embryonic day (E) 0.5. Embryos were imaged with an macroscope Leica MZ16 coupled with a camera CoolSnap-Pro<sub>cf</sub> color (RS Photometrics).

### Embryo section and immunolocalization

Embryos were fixed in 4% paraformaldehyde (PFA) for 1 hour at 4°C under agitation, rinsed three times in PBS, equilibrated first in 15% sucrose/PBS (overnight, 4°C), then in 30% sucrose/PBS (3 h, 4°C) and finally in 15% sucrose/50% Cryomatrix/PBS (1hour, 4°C). Embryos were embedded in Cryomatrix in liquid nitrogen vapors and stored at -80°C. Fifteen micrometers-sections were obtained on a Leica cryostat.

Sections were rehydrated twice in PBS, 30 min each time, at room temperature (RT), permeabilized in 0.5% Triton X-100/PBS (Sigma-Aldrich) 40 min at RT, blocked in AB buffer (3% Bovine Serum Albumin (BSA), 1% goat serum, 0.1% Tween 20 in PBS) 40 min at RT and rewashed in 0.1% Triton X-100/PBS. Primary antibodies (Supplementary Table 2) were diluted 1/1000 in AB buffer and incubated overnight at 4°C in a humid atmosphere. Sections were then washed 3 times in 0.1% Triton X-100/PBS 40 min each. Secondary antibody (Goat anti-mouse or anti-rabbit, conjugated to Alexa 488 nm or 594 nm, see Supplementary Table 1) was diluted 1/1000 in 1 µg/mL DAPI (4',6-diamidino-2-phenylindole dihydrochloride)/AB buffer and incubated 1 h at RT. The sections were then washed in 0.1%Triton X-100/PBS 5 times 10 min each at RT. The slides were mounted in Vectashield® and sealed with nail polish. They were imaged with a LSM 510 laser-scanning microscope (20x Plan APO objective). The pictures are shown with the LUT “Green Fire Blue” scale.

### Generation and maintenance of mESCs

E3.5 Blastocysts were collected from *R26<sup>CreERT2/+</sup>;Taf7<sup>ff</sup>* x *Taf7<sup>ff</sup>* *R26<sup>CreERT2/+</sup>;Taf10<sup>ff</sup>* x *Taf10<sup>ff</sup>* or *R26<sup>CreERT2/+</sup>;Taf7<sup>ff</sup>;Taf10<sup>ff</sup>* x *Taf7<sup>ff</sup>;Taf10<sup>ff</sup>* breedings. The uterus were dissected in PBS (37°C) and flushed with M2 medium (37°C) to collect the embryos using a insulin syringe. Single blastocysts were placed in a 96 well plate coated with mouse embryonic fibroblast (MEF) feeders and grown in 2i+LIF medium (DMEM medium supplemented with 15% fetal calf serum ES-tested, 2 mM L-glutamine, 0.1% β-mercaptoethanol, 100 UI/ ml penicillin and 100 mg/ml streptomycin, 0.1 mM non-essential amino acids, 100 µL/50 mL of leukemia inhibitory factor (LIF), 3 µM CHIR99021 and 1 mM PD0325901) at 37°C under 5% CO<sub>2</sub>. After amplification, clones were frozen in 30% fetal calf serum, 20% DMSO in DMEM medium. Genotype of each clone was checked by PCR (see protocol thereafter, primers in

Supplementary Table 3). For maintenance and for experiments, cells were grown on plate coated with 0.1% gelatin/PBS solution. Efficiency of depletion of the protein of interest after 4-hydroxytamoxifene (4-OHT) treatment was checked by western blot. All established clones were mycoplasma free. The biochemical, immunofluorescence and nascent transcription experiments were carried out using the  $R26^{CreERT2/+};Taf7^{ff}$  clone #13 (RT7#13),  $R26^{CreERT2/+};Taf10^{ff}$  clone #41 (RT10#41) and  $R26^{CreERT2/+};Taf7^{ff};Taf10^{ff}$  clone #47 (RT710#47), except for the phenotypic analyses where at least 2 independent clones were studied (see below).

#### Generation of $R26^{CreERT2/+};Taf7^{ff};Supt7l^{-/-}$ mESCs

The RT7#13 mESC were transfected with the plasmid construct containing *Cas9-EGFP* and gRNA (Supplementary Table 4) at about 70% confluence using Lipofectamine 2000 following manufacturer's instruction. Two to three days after, EGFP<sup>+</sup> cells were sorted using the BD FACS Aria TM II (BD Biosciences) instrument. Each fluorescent cell was isolated in a well of 96 well plates and thereafter amplified and frozen. Clones were screened by PCR and by western blot. The  $R26^{CreERT2/+};Taf7^{ff};Supt7l^{-/-}$  clone #72 was used for the experiments (RT7;*Supt7l*<sup>-/-</sup>#72).

#### Genotyping and screening

Mouse tail tips were digested in 300 µL of 200 µg/mL of proteinase K in tail digestion buffer (10 mM Tris pH 7, 200 mM NaCl, 5 µM EDTA, 0.2% SDS). Embryonic yolk sacs and mESC cells were digested in 100 µL of 200 µg/mL of proteinase K in 1X PCR buffer. In all cases, 0.6 µL of lysate was used in 25 µL of PCR reaction (Taq DNA polymerase kit, #4728858001) following the manufacturer's instruction. Primers are listed in Supplementary Table 3.

#### 4-hydroxytamoxifen (4-OHT) treatment

Five  $10^4$  or  $8 \times 10^5$  cells were seeded in one well of 6 well plate or a P100, respectively, coated with 0.1% gelatin/PBS at day -1 (D-1). At D0, cells were treated with 100 nM of 4-OHT in a final volume of 2 mL or 12 mL for a well of 6 well plate or a P100, respectively. As a control, the same volume of the vehicle (Ethanol, 0.1% EtOH final) in the same final volume was used. The different experiments were done at D2, D4 or D6. The 6 well plates were used for phenotypic analyses, and the P100 plate for proteomic and transcriptomic analyses.

#### Cell counting

Cells were collected and centrifugated (150 g) for 4 min at RT. Cell pellets were resuspended in a given volume of PBS. One volume of cell suspension was mixed with 1 volume of 0.4% of Trypan Blue, counted on a Countess II Automated Cell Counter (Thermo Fischer Scientific). Different clones were used: RT7: 5 clones (#5, #7, #8, #13, #15), RT10: 5 clones (#3, #6, #9, #15, #41), RT710: 4 clones (#1, #47, #48, #55), RT7; *Supt7l*<sup>-/-</sup>: 3 clones (#45, #71, #72).

#### Apoptosis assay

Floating and attached cells were collected by trypsinization and stained with the FITC Annexin V Apoptosis Detection Kit with PI (Thermo Fischer Scientific) according the recommendation and analyzed by flow cytometry using a BD FACS Celesta (BD Bioscience). A minimum of 10,000 events were recorded. Different independent clones were used: *RT7*: 4 clones (#5, #7, #8, #13), *RT10*: 6 clones (#3, #6, #9, #15, #19, #41), *RT710*: 4 clones (#1, #47, #48, #55), *RT7*; *Supt7l<sup>-/-</sup>*: 3 clones (#3, #71, #72).

### **Measuring of surface area of colonies**

Cells were rinsed twice with PBS and fixed with 4% PFA 30 min at 4°C. After PBS washing, cells were stained 30 min with 0.1% crystal violet dye and then rinsed with PBS. Pictures were taken with a macroscope M420 (Leica) coupled with a CoolSNAP camera (RS Photometrics). The surface of the clones was measured automatically with ImageJ (Fiji, v 1.53c). Different independent clones were used: *RT7*: 2 clones (#5, #13), *RT10*: 3 clones (#3, #6, #9), *RT710*: 2 clones (#1, #47).

### **Immunofluorescence (IF) on mESC**

Cells plated on gelatinized round glass slides were rinsed, fixed with 4% PFA 10 min at RT, rinsed again in PBS and then permeabilized in 0.1% Triton X-100/PBS 20 min at RT. After two PBS washes, primary antibodies diluted in 10% foetal calf serum (FCS)/PBS were incubated 1 h at RT. Cells were rinsed twice with 0.02% Triton X-100/PBS 5 min and once again 5 min with PBS. Secondary antibodies (Goat anti-mouse or goat anti-rabbit or goat anti-rat, conjugated to Alexa 488 nm or 594 nm, see Sup. Table. S5) were diluted 1/1000 in 1 µg/mL DAPI/10% FCS/PBS and incubated 1 h at RT. Cells were then washed 5 min in 0.02% Triton X-100/PBS and once again 5 min with PBS. The slides were assembled using Vectashield® and sealed with nail polish. They were imaged with an LSM 510 laser-scanning microscope (63x Plan APO objective). The pictures are shown with the LUT “Green Fire Blue” scale to better show the variation in fluorescence intensity. Measures of fluorescence intensity were done automatically using ImageJ software.

### **EU labelling**

Cells plated on gelatinized round glass slide were incubated in 2i+lif medium containing 1 mM of 5-ethynyl-uridine (EU, ThermoFischer Scientific) 1 h at 37°C under 5% CO<sub>2</sub> and then fixed with 4% PFA 15 min at RT. Analysis was realized using the Click-it RNA Imaging Kits (Invitrogen) according the manufacturer's guidelines. As a control, cells were incubated 1 h either with 500 µM of triptolide (Trp, Sigma Aldrich) or 300 µM of flavopiridol (Flavo, Sigma Aldrich) transcription inhibitors followed by incubation 1 hour in 1 mM EU in the presence of the inhibitors. Cells were imaged and EU signal quantified as described above.

### **Nuclear enriched whole cell extracts (NWCE)**

Cells were pelleted, rinsed with PBS, resuspended in 1 mL/g of cells of ice-cold Hypotonic Buffer (10 mM Tris pH8, 1.5 mM MgCl<sub>2</sub>, 10m M KCl supplemented with cOmplete protease inhibitor mix 1x

(Roche). Cells were then lysed by 10 gently strokes with a B pestle in a glass dounce grinder (Kimble). After 10 min centrifugation at 9000g, the pellet, which contains the nuclei, was resuspended in 1 mL/g of cells of ice-cold High Salt Buffer (20 mM Tris pH8, 1.5 mM MgCl<sub>2</sub>, 450 mM NaCl, 0.2 mM EDTA, 25% glycerol, 0.5% NP40, supplemented with cOmplete protease inhibitor mix 1x). Nuclei were lysed by 10 gently strokes and incubated in ice 30 min. After 10 min centrifugation at 9000 g, the supernatant was recovered as nuclear enriched whole cell extract.

### **Immunoprecipitation (IP)**

One mg of NWCE was first incubated 1 h with 120 µL of Protein-G Sepharose beads slurry (GE healthcare) in 1 mL of IP100 buffer (25 mM Tris HCl pH7.9, 10% Glycerol, 0.1% NP40, 5 mM MgCl<sub>2</sub>, 100 mM KCl, 1X cOmplete) at 4°C under gently agitation. Then 10 to 30 µL of antibodies were incubated with the NWCE during 2 hours at 4°C under gently agitation. Finally, the mix NWCE plus antibodies was incubated with fresh 120µL of Protein-G sepharose beads slurry overnight at 4°C under gently agitation. Beads were then washed at 4°C twice with IP500 Buffer (25 mM Tris HCl pH7.9, 10% Glycerol, 0.1% NP40, 5 mM MgCl<sub>2</sub>, 500 mM KCl, 1X cOmplete) under gentle agitation and then three times with IP100 buffer. Immunoprecipitated proteins were eluted in 50 µL Glycine buffer (0.1 M glycine pH2.8) and neutralized with 10 mM Tris HCl pH8.

### **Mass spectrometry**

#### Liquid digestion

Protein mixtures were TCA-precipitated overnight at 4°C. Samples were then centrifuged at 14000 rpm for 30 minutes at 4°C. Pellet were washed twice with 500 µL cold acetone and centrifuged at 14000 rpm for 10 minutes at 4°C. Washed pellet were then urea-denatured with 8 M urea in Tris-HCl 0.1 mM, reduced with 5 mM TCEP for 30 minutes, and then alkylated with 10 mM iodoacetamide 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) at a ratio 1/100 (enzyme/proteins) in 8 M urea for 4h, followed by an overnight modified trypsin digestion (Promega) at a ratio 1/100 (enzyme/proteins) in 2 M urea. Both LysC 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 (Thermo Scientific, San Jose California) coupled in line with a LTQ-Orbitrap ELITE mass spectrometer via a nano-electrospray ionization source (Thermo Scientific, San Jose California).

Peptide mixtures were loaded on a C18 Acclaim PepMap100 trap-column (75 µm ID x 2 cm, 3 µm, 100Å, Thermo Fisher Scientific) for 3.5 minutes at 5 µL/min with 2% ACN, 0.1% FA in H<sub>2</sub>O and then

separated on a C18 Accucore nano-column (75 µm ID x 50 cm, 2.6 µm, 150Å, Thermo Fisher Scientific) with a 100 minutes linear gradient from 5% to 50% buffer B (A: 0.1% FA in H<sub>2</sub>O / B: 99% ACN, 0.1% FA in H<sub>2</sub>O), then a 20 minutes linear gradient from 50% to 70% buffer B, followed with 10 min at 99% B and 10 minutes of regeneration at 5% B. The total duration was set to 140 minutes at a flow rate of 200nL/min. The oven temperature was kept constant at 40°C.

The mass spectrometer was operated in positive ionization mode, in data-dependent mode with survey scans from m/z 300-1600 acquired in the Orbitrap at a resolution of 240,000 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 30 s, repeat count was set to 1 and exclusion mass width was ± 10 ppm.

#### Data Analysis

Proteins were identified by database searching using SequestHT (Thermo Fisher Scientific) with Proteome Discoverer 2.2 software (Thermo Fisher 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 0.5 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.

#### Fold change quantification

Fold change value were quantified from XIC value using custom R scripts available on demand (R software version 4.0.2). Only peptides with at least two XIC values greater than the average of XIC value of the mock IP were conserved. The XIC values of all peptides of a protein were summed for each replicate and each condition. These summed XIC values were then divided by the average of summed XIC values of the bait protein. These normalized XIC values were then averaged between replicates of the same condition. Fold change value for a specific protein corresponds to the Log2 of the averaged XIC values from the mutant (4-OHT) condition divided by the averaged XIC values from the wild-type (EtOH) condition.

#### **Gel Filtration (GF)**

One mg of NWCE were diluted twice in the GF buffer (20 mM Tris pH8, 1.5 mM MgCl<sub>2</sub>, 450 mM NaCl, 0.2 mM EDTA, 10% glycerol, 0.5% NP40). Then they were passed through a Superose 6 GL

10/300 column (Sigma Aldrich) previously equilibrated and run at 0.4 mL/min. About 60 250  $\mu$ L fractions were collected and analyzed by western blot.

#### **Whole cell extract (WCE)**

Cells were pelleted at 150 g, rinsed with PBS. The cell pellet was mixed with one equivalent volume of WCE buffer (50 mM Tris HCl pH7.9, 25% Glycerol, 0.2mM EDTA, 5 mM MgCl<sub>2</sub>, 600 mM KCl, 0.5% NP40, 1 mM DTT, 1X cOmplete) and incubated 30 min on ice. Then, 3 volumes of IP0 (25mM Tris HCl pH7.9, 5% Glycerol, 0.1% NP40, 1 mM DTT, 1X PIC) was added and incubated 30 min on ice. For whole cell extract from organs, organs were lysed in Organ WCE buffer (20 mM Hepes pH7, 10% glycerol, 0.2 mM EDTA, 0.1% Triton, 0.35 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 1X PIC) with a pestle stroke followed by freeze-thaw in N<sub>2</sub> liquid. Proteins present in the supernatant were recovered after high speed centrifugation.

#### **Western blot (WB)**

Twenty  $\mu$ L of input or supernatant, 15  $\mu$ L of eluted proteins from IP, 27  $\mu$ L of fractions from GF, or 20  $\mu$ g of WCE or NWCE were boiled for 5 min in 100 mM Tris HCl pH 6.8, 30% glycerol, 4% SDS, 0.2% Bromophenol Blue, 100 mM DTT, resolved in 10% SDS-polyacrylamide gel and transferred to a nitrocellulose membrane (Protran, Amersham). After blocking in 3% milk/PBS, primary antibodies diluted 1/1000 in 0.3% milk/PBS (Supplementary Table 2) were incubated overnight at 4°C. After 3 washes in 0.05% Tween 20/PBS, HRP-coupled secondary antibodies (Supplementary Table 2) diluted 1/10000 in 0.3% milk/PBS were incubated 2 h at RT followed by ECL detection (ThermoFisher Scientific).

#### **RNA extraction**

RNA was extracted using TRI® Reagent (Molecular Research Center Inc) following the manufacturer's recommendations. RNAs were precipitated with a volume of isopropanol, washed with 75% EtOH, resuspended in RNase-free water and quantified with a Nanodrop (ThermoFisher Scientific). RNA extraction from organs started by stroking then in Trizol followed by high speed centrifugation.

#### **RT-qPCR**

Reverse Transcription (RT) was performed using with 1  $\mu$ g of total RNA using QuantiTect Reverse Transcription Kit (Quiagen) following the manufacturer's instructions. For qPCR, cDNAs were diluted 5 times and amplified using LightCycler 480 SYBR Green 2x PCR Master Mix I and 0.6 mM of forward and reverse primers (Sigma Aldrich, Supplementary Table 2), respectively, in 8  $\mu$ L of reaction volume. qPCR reaction was realized using a LightCycler 480 (Roche). Primers are listed in Sup. Table S3. Normalized values correspond to  $\Delta$ CT.



**Supplementary Table 1: Biological Resources**

Mouse lines	References
Tg(T-Cre)	(Perantoni et al., 2005)
<i>R26<sup>CreERT2</sup></i>	(Ventura et al., 2007)
<i>Taf10<sup>ff</sup></i>	(Mohan II et al., 2003)
<i>Taf7<sup>ff</sup></i>	(Gegonne et al., 2012)
Tg(T-Cre/+); <i>Taf7<sup>ff</sup></i>	This study
Tg(T-Cre/+); <i>Taf7<sup>ff</sup></i> ; <i>Taf10<sup>ff</sup></i>	This study
Tg(T-Cre/+); <i>Taf10<sup>ff</sup></i>	(Bardot et al., 2017)

Cell lines	References
<i>R26<sup>CreERT2/+</sup></i> ; <i>Taf7<sup>ff</sup></i>	This study
<i>R26<sup>CreERT2/+</sup></i> ; <i>Taf10<sup>ff</sup></i>	(Bardot et al., 2017)
<i>R26<sup>CreERT2/+</sup></i> ; <i>Taf7<sup>ff</sup></i> ; <i>Taf10<sup>ff</sup></i>	This study
<i>R26<sup>CreERT2/+</sup></i> ; <i>Taf7<sup>ff</sup></i> ; <i>Sup7l<sup>-/-</sup></i>	This study
Feeder CD1 WT	Routinely generated at IGBMC cell culture platform

**Supplementary Table 2: List of antibodies used in this study**

Name	Company name	Use	Reference
anti-TAF4 antibody (monoclonal mouse)	IGBMC antibody service (32TA-2B9)	WB	(Mohan II et al., 2003)
anti-TAF5 antibody (monoclonal mouse)	IGBMC antibody service (1TA-1C2)	WB	(Jacq et al., 1994)
anti-TAF6 antibody (monoclonal mouse)	IGBMC antibody service (25TA-2G7)	WB	(Langer et al., 2016)
anti-TAF7 antibody (monoclonal mouse)	IGBMC antibody service (19TA-2C7)	WB	(Langer et al., 2016)
anti-TAF7 antibody (polyclonal rabbit)	IGBMC antibody service (3475)	WB + IF	This study
anti-TAF7L antibody (monoclonal mouse)	IGBMC antibody service (46TA-2D5)	WB	(Martianov et al., 2016)
anti-TAF8 antibody (polyclonal rabbit)	IGBMC antibody service (3478)	WB	(Kamenova et al., 2019)
anti-TAF10 antibody (monoclonal mouse)	IGBMC antibody service (6TA-2B11)	WB + IF	(Bardot et al., 2017)
anti-TAF12 antibody (monoclonal mouse)	IGBMC antibody service (22TA-2A1)	WB + IP	(Langer et al., 2016)
anti-TBP antibody (monoclonal mouse)	IGBMC antibody service (3TF1-3G3)	WB + IP	(Brou et al., 1993)
anti-SUPT7L antibody (monoclonal mouse)	Bethyl laboratories (A302-803A)	WB	(Kamenova et al., 2019)
anti-RPB1pSer5 antibody (monoclonal rat)	GmbH antibody service (CTD4-3E8)	IF	(Chapman et al., 2007)
anti-RPB1pSer2 antibody (monoclonal rat)	GmbH antibody service (CTD7-3E10)	IF	(Chapman et al., 2007)
anti-GST antibody (monoclonal mouse)	IGBMC antibody service (15TF2-1D10)	IP	(Kamenova et al., 2019)
anti-Rabbit IgG, Alexa Fluor® 488 conjugate (goat polyclonal)	Life technology (A-11008)	IF	

anti-Mouse IgG Alexa Fluor® 546 conjugate (goat polyclonal)	Life technology (A-11003)	IF	
anti-Rat IgG, Alexa Fluor® 488 conjugate (goat polyclonal)	Life technology (A-11006)	IF	
anti-Rabbit IgG Peroxydase conjugate (goat polyclonal)	Jackson ImmunoResearch (111-035-144)	WB	
anti-Mouse IgG Peroxydase conjugate (goat polyclonal)	Jackson ImmunoResearch (111-036-071)	WB	

WB; western blot, IP; immunoprecipitation, IF; immunofluorescence

**Supplementary Table 3: List of PCR and qPCR primers**

Species	Gene	Forward	Reverse	Usage
mouse	<i>Taf7</i>	GTATGAAAACCTGTGTCCTGG TCTG	GAAGGCAAGTTCTCAATGAA AGGG	PCR
mouse	<i>Taf10</i>	GTAGTGTCCAGCACACCTCT	CAGTCTAACCTGCTCCGAGT	PCR
mouse	<i>Cre</i>	TGATGGACATGTTCAGGGATC	CAGCCACCAGCTTGCATGA	PCR
mouse	<i>Rosa-Wt</i>	AAAGTCGCTCTGAGTTGTTAT	CCTGATCCTGGCAATTTTCG	PCR
mouse	<i>Rosa-Cre</i>	GGAGCGGGAGAAATGGATATG	CCTGATCCTGGCAATTTTCG	PCR
mouse	<i>Supt7l-exon1-2</i>	ATTGTGGCGACTGCTTGATAG	ACCCAGAGAGTGACTTTTACCG	PCR
mouse	<i>Supt7l-5'UTR</i>	GCAGTTCCCACATAAGAAGCA	AGCCGCGTATAACCACTCCT	PCR
mouse	<i>Rn7sk</i>	CCATTGTAGGAGAACGTAGGG TAGTCAAGC	CCACATGCAGCGCCTCATT	qPCR
mouse	<i>Taf7</i>	AATATGCCGCTACGGTGAGG	TCAGGTTGACATGCCAGAC	qPCR
mouse	<i>Taf7L</i>	CCTGAGAAACATCCGCGGTC	AGACAACGTCTCACTGCCTG	qPCR
mouse	<i>Taf7L2</i>	GCAACGGAACGTGTGAAGTG	CCGATGGGAAGTCGTTGTTG	qPCR
mouse	<i>Taf10</i>	GAGGGGGCAATGTCTAACGG	CGCGGTTCAGGTAGTAACCA	qPCR

**Supplementary Table 4: List of sgRNAs used to generate the  $R26^{CreERT2/+};Taf^{7ff};Sup7l^{-/-}$  mouse ES cell lines ( $RT7;Sup7l^{-/-}$ )**

Name	Sequence	PAM	Reference
<i>mSupt7l-3</i>	ACCATCTCCCTCGCCCCG	AGG	(Fischer et al., 2021b)
<i>mSupt7l-4</i>	ACCAGTACGTATTCAGAG	TGG	

**Supplementary Table 5: Reagents**

Reagent	Company name	Catalogue #
4',6-diamidino-2-phenylindole dihydrochloride (DAPI)	Molecular Probes	D1306
4-OHT	Sigma Aldrich	H7904
Acetonitrile MS grade	Sigma Aldrich	1207802
Acrylamide:Bisacrylamide 29:1	Euromedex	EU0077-B
Albumine, Bovine, Fraction V (BSA)	MP Biomedicals	160069
Ammonium peroxydisulfate (APS)	Euromedex	1012010500
BlueTrypan Staining 0,4%	Invitrogen	T10288
Bromophenol Blue	Serva	15375
CHIR99021	axon medchem	1386
Chloroform	Carlo Erba	438603
Click-it RNA Imaging Kits	Invitrogen	2192300
Complete Protease Inhibitor Cocktail (cOmplete), EDTA free	Roche	11873580001
Cristal Violet	Sigma Aldrich	C3886-100g
Cryomatrix	Thermo Scientific	6769006
Dimethy sulfoxid (DMSO)	Sigma Aldrich	1.02931.1000
DMEM (4,5g/L glucose)	ThermoFisher Scientific	41966-029
DTT (Dithiothreitol)	Euromedex	EU0006
EDTA	Euromedex	EU0007
Endoproteinase Lys-C	Wako	125-05061
Ethanol	VWR	83813360
FITC Annexin V Apoptosis Detection Kit with PI	Biolegend	640914
Flavopiridol	Sigma Aldrich	F3055-1mg
Foetal calf serum	Sigma Aldrich	F7524
Foetal calf serum ES-tested	ThermoFisher Scientific	10270-10

Formic acid	Sigma Aldrich	94318
Gelatine solution	PAN BIOTECH	P06-20410
Glycerol	PanReacAppliChem	A0970.5000
Glycine	Sigma Aldrich	G8898
Goat Serum	Sigma Aldrich	G9023
HCl 37%	VWR	20252.290
Hepes pH 6	Sigma Aldrich	H3375-1KG
Iodoacetamide	Sigma Aldrich	I1149
Isopropanol	VWR	20842330
KCl	Sigma Aldrich	P9333-1KG
KOH	VWR	26668.296
L-glutamine	ThermoFisher Scientific	25030-024
LightCycler® 480 SYBR® Green 2x PCR Master Mix I	Roche	4887352001
Lipofectamine2000	Thermo Scientific	11668019
M2 medium	Sigma Aldrich	M7167
MgCl2 hexahydrate	Sigma Aldrich	63068-25G
Nonfat Dry Milk	Régilait	
N,N,N',N'-Tetramethyl-ethylenediamine (TEMED)	Serva	35930
NaCl	Carlo Erba	479687
Nitrocellulose membrane	Protran, Amersham	10600033
Non-essential amino acids	ThermoFisher Scientific	11140-035
NP40 (IGEPAL CA-630)	Sigma Aldrich	I-3021
Paraformaldehyde (PFA)	Electron Microscopy Sciences	15710
PBS	Sigma Aldrich	D5652
PD0325901	Axon Medchem	1408
Penicillin and streptomycin	ThermoFisher Scientific	15140-122
Pierce™ ECL Western Blotting Substrate	ThermoFisher Scientific	32109

Protein G Sepharose	GE healthcare	17-0618-05
Proteinase K	Thermo Scientific	EO0491
QuantiTect Reverse Transcription Kit	Qiagen	205311
RNase-free water	Sigma Aldrich	95284
Sucrose	Euromedex	200-301-B
Taq DNA polymerase, dNTPack	Roche	4728858001
TRI® Reagent (Trizol)	Molecular Research Center Inc.	TR 188
Trichloroacetic acid	Sigma Aldrich	T0699
Triptolide	Sigma Aldrich	T3652-1mg
Tris(2-carboxyethyl)phosphine (TCEP)	X	
Triton X-100	Sigma Aldrich	T8787
Trizma base	Sigma Aldrich	T1503-1Kg
Trypsine	Promega	V5111
Trypsine-EDTA	Invitrogen	25200-072
Tween 20	Sigma Aldrich	P2287
Urea	Sigma Aldrich	U0631
VECTASHIELD® Media without DAPI	Mounting VectorLabs	H-1000
β-mercaptoethanol	ThermoFisher Scientific	31350-010



## II. Validation de l'anticorps anti-TAF12 (22TA-2A1) pour l'étude de l'assemblage de TFIID par expérience d'immunoprécipitation.

L'étude du phénotype des mESC *RT7*, *RT10* et *RT710* a montré que la déplétion de TAF7 et/ou TAF10 induit l'apparition de phénotypes dont la sévérité diffère. La déplétion de TAF10 notamment réduit significativement la croissance et la viabilité de la population cellulaire, contrairement à la déplétion de TAF7. Le double mutant présente le phénotype le plus sévère (Hisler et al.). Ces résultats soulignent l'importance différentielle de TAF7 et TAF10 dans l'assemblage de TFIID. Le module TAF2-8-10 est ainsi recruté au niveau du core-TFIID, composé de TAF4, TAF5, TAF6, TAF9 et TAF12, pour former le complexe 8TAFs (Bieniossek et al., 2013a; Trowitzsch et al., 2015). On suppose que cette étape est requise pour la poursuite de l'assemblage de TFIID avec le recrutement entre autres de TAF7. Ce modèle d'assemblage repose principalement sur la surexpression et l'assemblage des différentes sous-unités de TFIID *in vitro* et sur l'étude par cryo-EM des complexes obtenus (Bieniossek et al., 2013a; Trowitzsch et al., 2015). Il est à noter que TAF10 est également important pour l'assemblage du complexe coactivateur SAGA (Bardot et al., 2017; Herbst et al., 2021).

Dans le but de mieux connaître la capacité d'assemblage de TFIID dans les cellules mutantes *RT7*, *RT10* et *RT710*, j'ai voulu réaliser des IP de TFIID suivi d'analyses par western blot et spectrométrie de masse. J'ai tout d'abord testé des anticorps anti-TAFs afin de réaliser des IPs de TFIID. La grande majorité des anticorps disponible au sein de notre institut a été générée de manière à être dirigée contre les protéines TAF humaines (table 1), ces dernières ne présentant pas 100% d'identité avec les protéines TAF murines. L'anticorps que je cherchais devait répondre à 2 critères :

1) ne pas être un anticorps anti-TAF7 ou anti-TAF10, car ce sont les deux protéines conditionnellement déplétées dans mon modèle cellulaire,

2) permettre d'étudier à quelle étape de l'assemblage de TFIID est bloquée en absence de TAF7 et/ou TAF10. L'anticorps devra donc reconnaître une des sous-unités du core-TFIID. La quantité d'anticorps polyclonaux purifiée à partir de sérum étant limitée et donc afin d'avoir suffisamment d'anticorps tout au long de ma thèse, j'ai focalisé ma recherche sur les anticorps monoclonaux disponibles à l'IGBMC.

Dans un premier temps, j'ai voulu connaître les anticorps qui permettent de révéler par western blot les sous-unités de TFIID. J'ai réalisé pour ce faire une IP mock anti-GST, comme contrôle négatif, et une IP anti-TBP, comme contrôle positif. L'efficacité de l'anticorps 3G3 anti-TBP pour immunoprécipiter TFIID à partir de lysats de cellules de souris a déjà été montré (Bardot et al., 2017) (Figure 41 A, B). Les anticorps anti-TAF4, anti-TAF5, anti-TAF6, anti-TAF7, anti-TAF8, anti-TAF10, anti-TAF12 et anti-TBP, déjà couramment utilisés dans notre laboratoire, ont été validés ce qui m'a permis de connaître le profil de migration des protéines TAF murines (Figure 41 C).

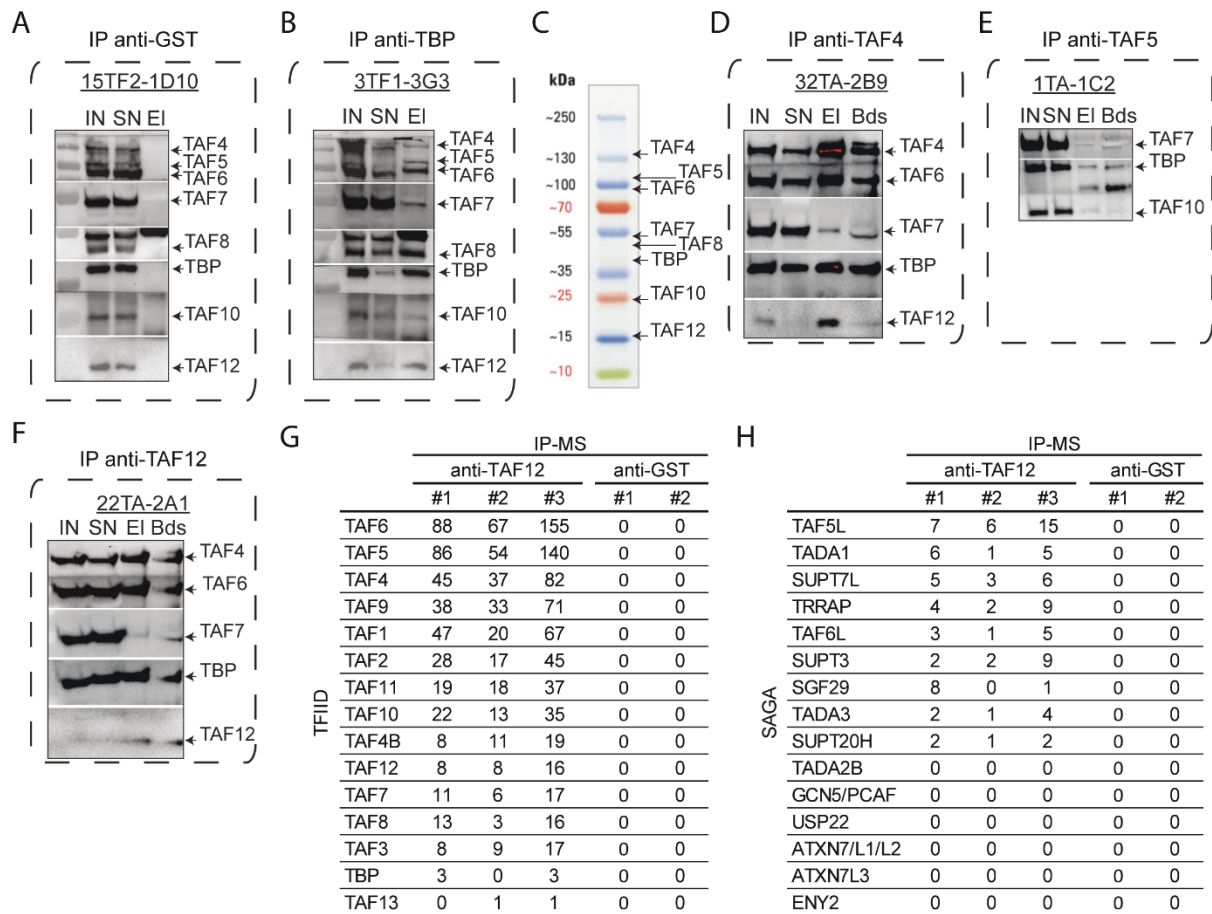


Figure 41 : Validation de l'anticorps anti-TAF12 22TA-2A1 pour immunoprécipiter TFIID. (A, B, E, D, E et F) Analyses par western blot d'expériences d'immunoprécipitation anti-GST (A), anti-TBP (B), anti-TAF7 (D), anti-TAF5 (E) et anti-TAF12 (F). La référence de chaque anticorps utilisé pour l'IP est indiquée au-dessus de chaque membrane. Les TAFs révélés sont indiqués à droite de chaque membrane. IN (input), SN (surnageant), EI (élution), Bds (billes). (G-H) Nombre total des PSM (peptide-spectrum matches) mesurés par spectrométrie de masse à partir des éluions d'IP anti-TAF12 (à gauche) et anti-GST (à droite), pour les sous-unités de TFIID (G) et les sous-unités spécifiques à SAGA (H). Les PSM indiqués sont la somme des PSM détectés dans le tripliqua expérimental.

Parmi les 25 anticorps testés pour immunoprécipiter TFIID (Figure 42 et 43, table 2), seulement 3 respectaient les critères définis ci-dessus : l'anticorps 2B9 anti-TAF4 (Figure 41 D), l'anticorps 1C2 anti-TAF5 (Figure 41 E) et l'anticorps 2A1 anti-TAF12 (Figure 41 F). Néanmoins, l'anticorps TAF5 est moins efficace pour immunoprécipiter TFIID que les anticorps anti-TAF4 et anti-TAF12 (Figure 41, comparer E à D et F). Contrairement à TAF4, TAF12 fait également partie du complexe SAGA, par conséquent j'ai choisi d'utiliser cet anticorps pour les expériences ultérieures en espérant pouvoir également étudier la composition de SAGA en absence de TAF10. Cependant, bien que l'analyse par spectrométrie de masse ait permis de valider sa capacité à immunoprécipiter l'ensemble du complexe TFIID (Figure 41 G), cette analyse a également montré qu'il n'était pas capable d'immunoprécipiter efficacement SAGA (Figure 41 H). Dans le but d'analyser la composition de SAGA en absence de TAF10, l'anticorps polyclonal de lapin anti-SUPT20H, qui avait déjà été validé pour immunoprécipiter SAGA murin (Fischer et al., 2021a), a été utilisé.

**Résultats**

*Validation de l'anticorps anti-TAF12 pour l'étude de l'assemblage de TFIID*

Table 2 : Liste des anticorps utilisés. - : ne fonctionne pas, +/- : Ip peu efficace ou n'IP pas le complexe TFIID entier. + et ++ fonction relativement bien et très bien.

Nom	Cible	Type	Fournisseur	Expérience	Référence
15TF2-1D10	GST	monoclonal de souris	IGBMC	IP	Nagy et al., 2010
3006	SUPT20H	polyclonal de lapin	IGBMC	IP (+/-)	(Krebs et al., 2011)
3TF1-3G3	TBP	monoclonal de souris	IGBMC	IP (++) et WB	(Brou et al., 1993)
1727	TBPL1	polyclonal de lapin	IGBMC	WB	(Martianov et al., 2016)
1TRF-2D5	TBPL1	monoclonal de souris	IGBMC	WB	/
1TRF-2D7	TBPL1	monoclonal de souris	IGBMC	WB	/
3TRF-2A1	TBPL1	monoclonal de souris	IGBMC	WB	(Martianov et al., 2001)
J17	TFIIA- $\alpha$	polyclonal de lapin	cadeau de HG Stunnberg	WB	(Yu et al., 2020a)
42TA-1A2	TAF3	monoclonal de souris	IGBMC	IP (+/-)	/
39TA-2B1	TAF3	monoclonal de souris	IGBMC	IP (-)	/
39TA-2F5	TAF3	monoclonal de souris	IGBMC	IP (-)	/
20TA-1B12	TAF4	monoclonal de souris	IGBMC	IP (-)	/
32TA-2B9	TAF4	monoclonal de souris	IGBMC	IP (+) et WB	(Perletti et al., 2001)
1TA-1C2	TAF5	monoclonal de souris	IGBMC	IP (+/-)	(Jacq et al., 1994)
15TA-2D2	TAF5	monoclonal de souris	IGBMC	IP (-)	/
25TA-2G7	TAF6	monoclonal de souris	IGBMC	IP et WB	(Dantonel et al., 1997)
37TA-1C2	TAF6	monoclonal de souris	IGBMC	IP et WB	Jacq et al., 1994
33TA-1H7	TAF7	monoclonal de souris	IGBMC	IP (-)	/
19TA-2C7	TAF7	monoclonal de souris	IGBMC	IP et WB	(Lavigne et al., 1996)
3475	TAF7	polyclonal de lapin	IGBMC	WB	(Bardot et al., 2017)
37TA-2C12	TAF7	monoclonal de souris	IGBMC	IP (+/-)	/
1FR-1A2	TAF8	monoclonal de souris	IGBMC	IP (-)	(Mohan II et al., 2003)
1FR-1B6	TAF8	monoclonal de souris	IGBMC	IP (-)	Kamenova et al., 2019
3478	TAF8	polyclonal de lapin	IGBMC	WB	Kamenova et al., 2019
53TA-1H1	TAF9	monoclonal de souris	IGBMC	IP (-)	/
49TA-2B6	TAF9B	monoclonal de souris	IGBMC	IP (-)	/
6TA-2B11	TAF10	monoclonal de souris	IGBMC	IP(++) et WB	(Mohan II et al., 2003)

23TA-1H8	TAF10	monoclonal de souris	IGBMC	IP(++)	(Wieczorek et al., 1998)
15TA-2B4	TAF11	monoclonal de souris	IGBMC	IP (-)	(Mengus et al., 1995)
35TA	TAF11	monoclonal de souris	IGBMC	IP (-)	/
22TA-2A1	TAF12	monoclonal de souris	IGBMC	IP (++) et WB	(Jacq et al., 1994)
16TA-3C12	TAF13	monoclonal de souris	IGBMC	IP (-)	(Mengus et al., 1995)

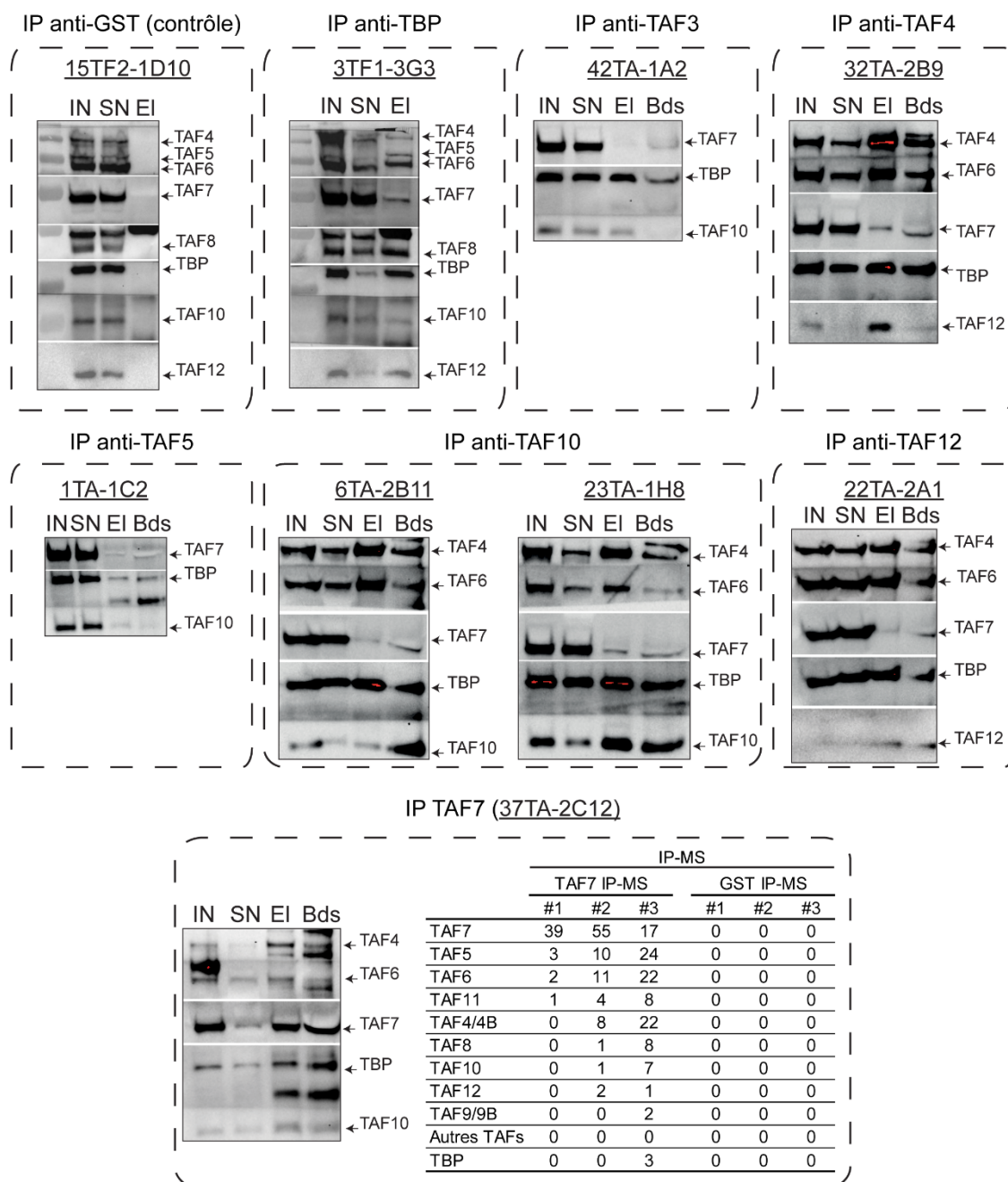


Figure 42 : Validation par western blot d'anticorps ciblant des protéines de TFIID. Au-dessus de chaque panel est indiquée la protéine TAF immunoprécipitée et au-dessus de chaque membrane le nom de l'anticorps utilisé. A gauche de chaque membrane sont indiquées les sous-unités de TFIID révélées. IP GST sert d'IP mock prouvant la spécificité des bandes observées. IN : input, SN, surnageant, EI, élution, Bds : billes. Pour l'IP TAF7 utilisant l'anticorps 37TA-2C12, le nombre total de PSM pour chaque sous-unité TFIID observées dans les mesures de spectrométrie de masse, réalisé en tripliqua, sont données dans le tableau. Pour comparer, le nombre total de PSM obtenu à partir d'une élution d'IP anti-GST est également indiqué.

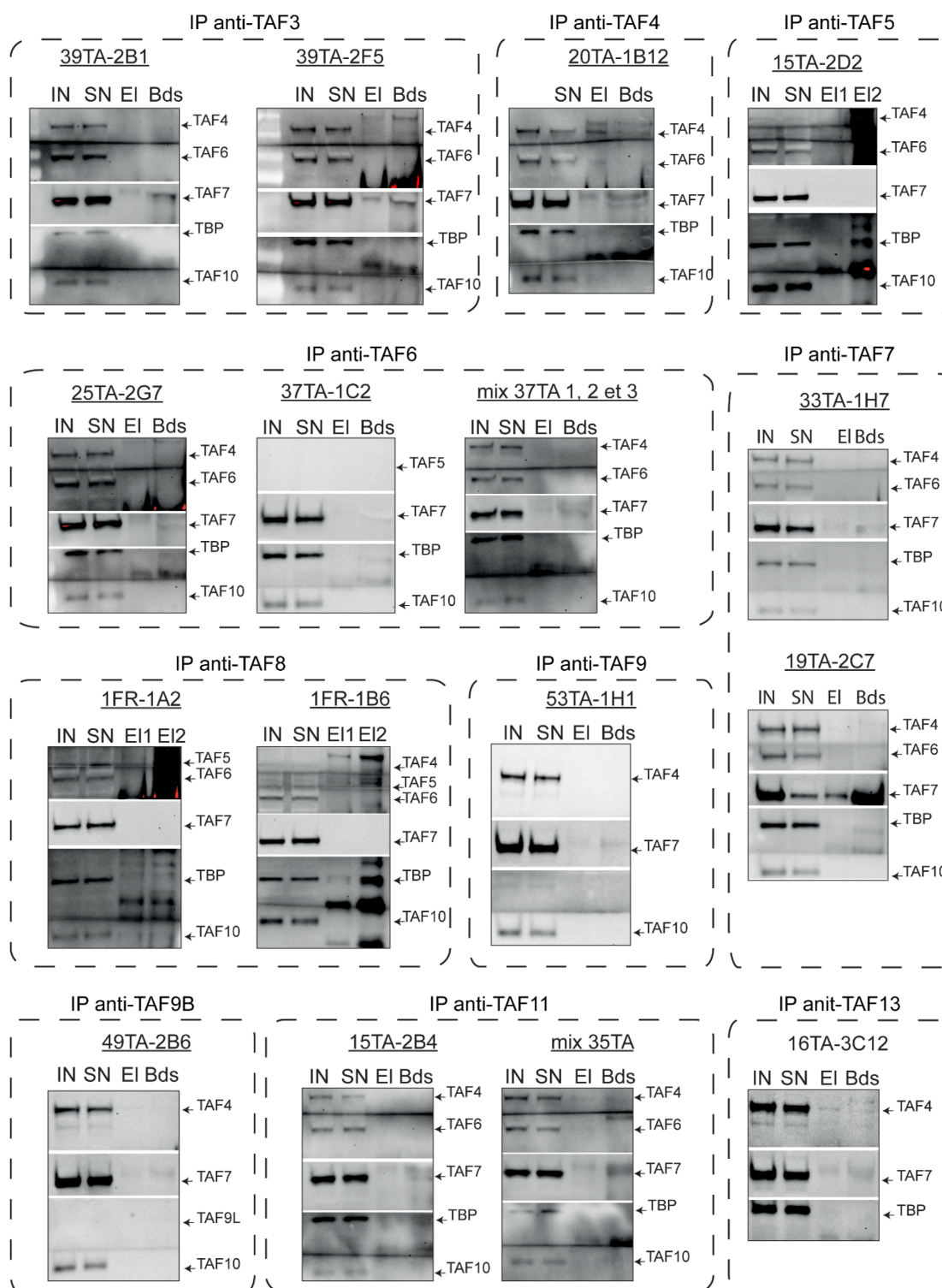


Figure 43 : Anticorps non-validés pour immunoprécipiter TFIID. Au-dessus de chaque panel est indiquée la protéine TAF immunoprécipitée et au-dessus de chaque membrane le nom de l'anticorps utilisé. A gauche de chaque membrane sont indiquées les sous-unités de TFIID révélées. IP anti-GST sert d'IP mock prouvant la spécificité des bandes observées. IN : input, SN, surnageant, EI, élution, Bds : billes.



### III. Etude de l'importance de TAF7 dans l'assemblage de TFIID.

La différence de sévérité des phénotypes observés entre les mutants *RT7*, *RT10* et *RT710* corrélient avec la perturbation plus ou moins importante de l'assemblage de TFIID en absence de TAF7 et/ou TAF10. Ainsi, alors qu'au sein des mutants *RT710* seul le core-TFIID est retrouvé, dans les mutants *RT10*, une partie de ces core-TFIID semble être couplé à TAF1. L'importance de TAF7 dans l'assemblage de TAF1 au complexe TFIID en formation n'est pas connue. D'autre part, on ne sait pas à quelle étape et avec quelle(s) partenaire(s) TAF7 s'assemble au TFIID en formation.

L'analyse par western blot des fractions obtenues par chromatographie d'exclusion à partir de lysats cellulaires enrichis en composants nucléaires (NWCE, *Nuclear enriched Whole Cell Extract*) indique que TAF7, ainsi que TAF10 et TBP, sont présents dans les fractions Q, R, S (Figure 44 Aa, rectangle jaune), en plus des fractions D et E contenant TFIID (Figure 44 Aa, rectangle bleu). Le fait que TAF7, TAF10 et TBP sortent dans les mêmes fractions de chromatographies a déjà été observé dans des extraits de mESC (Langer et al., 2016). Les fractions Q, R et S contiennent des protéines ou complexes protéiques d'un poids supérieur à 158 kDa. TAF7 a un poids d'environ 50 kDa, TAF10 d'environ 25 kDa et TBP d'environ 40 kDa (Figure 41 C). Par conséquent, le fait que ces trois protéines soient présentes dans ces fractions indique qu'elles sont complexées à d'autres partenaires. En absence de TAF7, TAF10 et TBP ne sont plus présents dans ces fractions (Figure 44 Ab, rectangle jaune). En absence de TAF10, TAF7 et TBP sont de nouveau trouvés dans les fractions Q, R, S (Figure 44 Ac, rectangle jaune). Ainsi, ensemble ces résultats suggèrent que TAF7 et TBP sont peut-être présents dans un même complexe avant qu'il soit assemblé à TFIID.

Pour obtenir plus de détails sur cet hypothétique complexe, des IP anti-TAF7 ont été réalisées à partir des fractions Q, R et S provenant de la condition contrôle et de la condition déplétée en TAF10 (Figure 44 B). Dans l'*input*, TAF4, TAF5 et TAF6 ne sont pas présents indiquant que le core-TFIID ne se trouve pas dans ces fractions, en accord avec les données de chromatographie d'exclusion de taille. Bien que TAF7, TAF8, TBP et TAF10 sont détectés à un faible niveau dans l'*input*, seul TAF7 et TBP sont détectés dans l'élution en condition contrôle et déplétée en TAF10 (Figure 44, B). Comme attendu, en condition déplétée en TAF10, TAF8 est également déplété (Figure 44, B). Ainsi, TBP et TAF7 semblent être présents au sein d'un même complexe, comme suggéré par l'analyse des fractions de chromatographie.

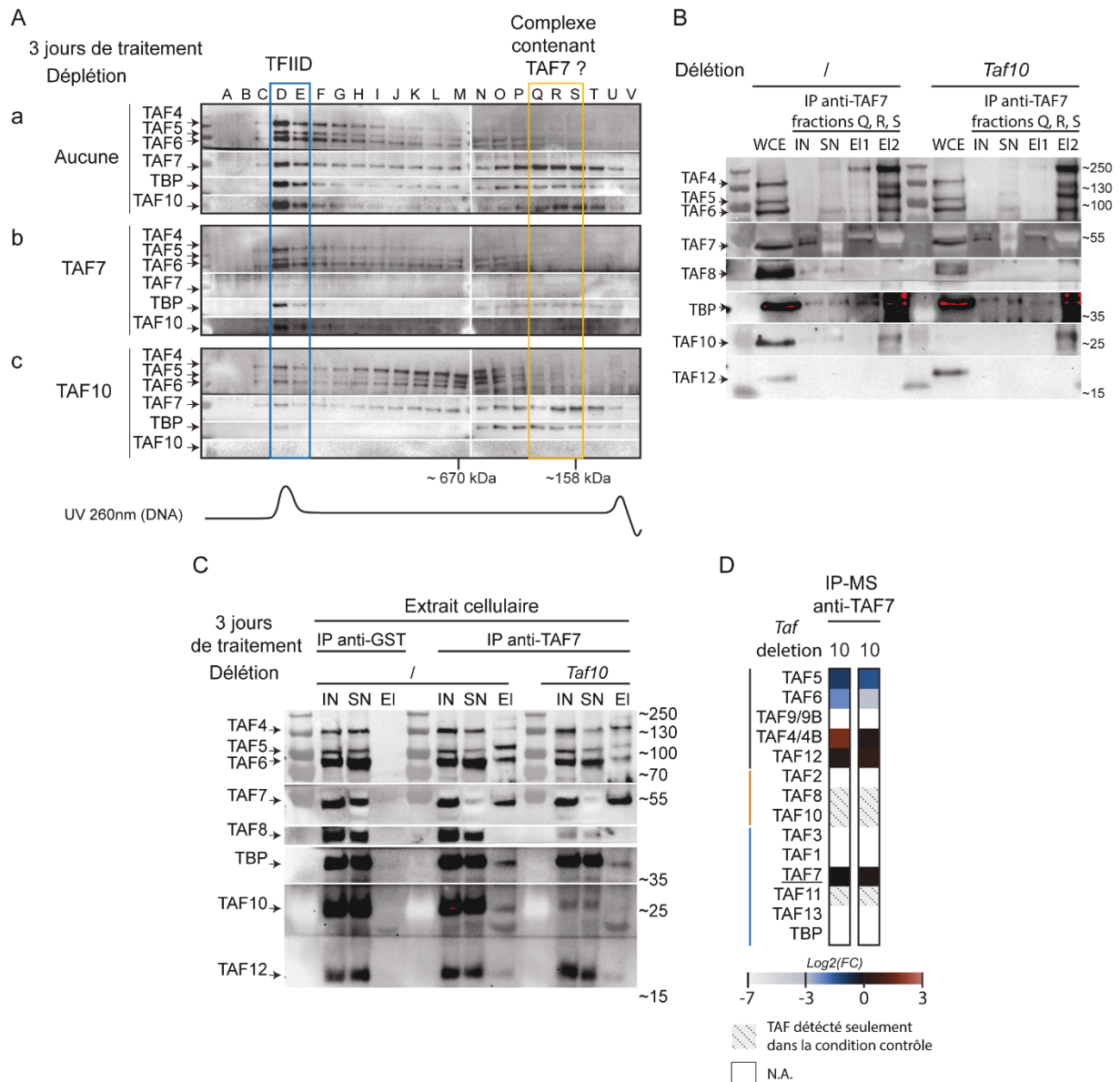


Figure 44 : Recherche d'un complexe contenant TAF7. A) Analyse par western blot des fractions obtenues par chromatographie d'exclusion de taille réalisée à partir de mESC RT10 contrôle (a), mutantes (c) et de mESC RT7 mutantes (b) traités durant 3 jours avec de l'EtOH (a) ou du 4-OHT (b, c). B) Analyse par western blot d'extrait cellulaire (WCE) ainsi que d'input (IN), surnageant (SN) et des éluions (EI) provenant d'IP anti-TAF7 réalisées à partir des fractions Q, R et S de la condition contrôle (Aa) et de la condition déplétée en TAF10 (Ac). C) Analyse par western blot d'IP anti-TAF7 et d'IP mock anti-GST réalisées à partir de NWCE de RT10 contrôle (déplétion : /) ou mutante (déplétion : Taf10) traités comme indiqué dans (A). Pour (A), (B) et (C), les différentes protéines étudiées sont indiquées sur la gauche. D) Valeurs XIC normalisées représentées en heatmap pour deux éluions d'IP TAF7 réalisées comme dans (C). Le FC correspond à la valeur XIC relative obtenue à partir de la condition déplétée en TAF10 sur la valeur XIC relative obtenue à partir de la condition contrôle.

Que ce soit en condition contrôle ou déplétée en TAF10, TAF7 est retrouvée dans toutes les fractions entre les fractions M, N, O contenant le core-TFIID et les fractions D, E contenant de complexe TFIID. J'ai voulu savoir si TAF7 pouvait incorporer le core-TFIID en absence de TAF8 et TAF10. Dans ce but, des IPs anti-TAF7 ont été réalisées sur des NWCE cellules RT10 contrôles et mutantes. En condition contrôle, seul les sous-unités TAF4, TAF5, TAF6, TAF7, TAF8, TAF10, TAF11, TAF12 et TBP ont été détectées (Figure 44, C et D). En condition mutante, TAF4, TAF5 et TAF6 sont détectées dans l'élution, en plus de TBP et TAF7 qui étaient déjà détectées dans les IPs réalisées sur les fractions de

chromatographie (Figure 44, C et D). Ces données préliminaires suggèrent que TAF7 et TBP sont dans un premier temps présents dans un même complexe. Ce complexe par la suite peut s'assembler au core-TFIID bien que le module TAF2-8-10 n'est pas incorporé le core-TFIID induisant son réarrangement. La liaison de TBP et de TAF7 n'est pas directe, mais se fait via TAF1 (Chen et al., 2021a; Patel et al., 2018), TAF1 pourrait faire partie de ce complexe.

#### IV. Etude de l'importance de TAF7 et TAF10 dans le recrutement des sous-unités de TFIID à la chromatine.

L'assemblage de TFIID est affecté différemment par la perte de TAF7 et/ou TAF10 : un complexe *TAF7-less TFIID* pour les mutants *RT7*, un core-TFIID couplé à TAF1, et potentiellement TAF7 et TBP pour les mutants *RT10*, et seulement le core-TFIID pour les mutants *RT710*. Les différences de phénotypes observés entre les mutants *RT7*, *RT10* et *RT710* pourraient être dues à la capacité de ces complexes à être recrutés au promoteur et à initier la transcription.

Dans ce but, j'ai voulu analyser les TAFs associés à la chromatine en absence de TAF7 et/ou TAF10. La chromatine de cellules *RT7*, *RT10* et *RT710* traitées durant 3 jours à l'EtOH (contrôle) ou au 4-OHT (mutant), a été récupérée, lavée, dénaturée 10min à 100°C et les surnageants, contenant les protéines extraites de la chromatine, ont été isolés. L'analyse par spectrométrie de masse d'extraits de chromatine de condition contrôle n'a permis de détecter que 6 TAFs (Figure 45 A). En condition contrôle, TAF4, TAF5, TAF6, TAF7, TAF10 et TBP ont été détectés dans les extraits cytoplasmiques et les NWCE à des niveaux variables vraisemblablement dus à la qualité de l'extraction (Figure 45 B, comparer les lignes 1, 7 et 13 à 2, 8 et 14). Seul TAF4, TAF7 et TAF10 ont été détectés dans l'extrait de chromatine (Figure 45 B, ligne 3, 11, 15), en accord avec les données de spectrométrie de masse. L'absence de la plupart des sous-unités de TFIID en condition contrôle suggère que l'extraction des protéines liées à la chromatine n'a pas été suffisamment efficace. Néanmoins, en condition déplétée pour TAF7, les protéines TAF4 et TAF10 sont retrouvées dans l'extrait de chromatine (Figure 45 B, comparer les lignes 3 et 6). En condition déplétée pour TAF10, couplée ou non à la déplétion de TAF7, TAF4 n'est plus présent dans cet extrait (Figure 45 B, comparer les lignes 11 à 12 et 15 à 18). Que ce soit en présence ou absence de TAF10, TAF7 est toujours présent dans l'extrait de chromatine (Figure 45 B, comparer ligne 11 à 12). Par conséquent, ces données préliminaires semblent indiquer que TAF10 est requis pour la liaison de TAF4 à la chromatine, mais pas de TAF7. A contrario, TAF7 n'est pas requis pour la liaison de TAF4 et TAF10 à la chromatine.



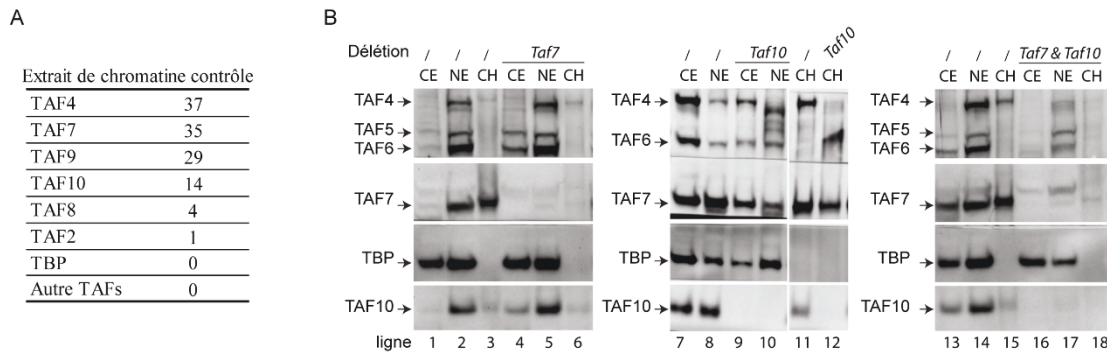


Figure 45 : TAF10 n'est pas requis pour recruter/maintenir TAF7 sur la chromatine et vice versa. A) Tableau récapitulatif du nombre de PSM détectés après l'analyse par spectrométrie de masse d'extrait de chromatine issu de mESC en condition contrôle. Les PSM indiqués sont la somme des PSM détectés dans le tripliqua expérimental. B) Analyse par western blot d'extrait cellulaire (CE), d'extrait enrichi en composant nucléaire (NE) et d'extrait de chromatine (CH). Les cellules RT7, RT10 et RT710 ont été traitées durant trois jours soit avec de l'EtOH (condition contrôle, indiquée « / » pour la délétion), soit avec de 4-OHT (condition mutante) induisant la délétion respectivement de Taf7, Taf10 ou des deux, comme indiquée. Les TAFs observés sont indiqués sur la gauche de chaque western blot.

## V. Etude de l'importance de TAF7 et TAF10 dans la transcription médiée par l'ARN pol II.

Dans le but de mieux comprendre le rôle fonctionnel de TFIID, et de ses complexes partiels, dans la transcription au sein des mESC, j'ai voulu analyser le transcriptome naissant par TT-seq (Schwalb et al., 2016). Cette expérience consiste en un marquage métabolique des ARNs nouvellement synthétisés via le 4-thiouridine (4sU). Après ce marquage, les ARN sont extraits, fragmentés puis les régions marquées sont purifiées (Figure 46). Le jour de traitement à laquelle le marquage au 4sU est réalisé est important. Trop tôt les cellules risquent de ne pas être suffisamment déplétée en TFIID et par conséquent la transcription pourrait ne pas être significativement perturbée. Trop tard, des effets indirects pourraient apparaître et par conséquent l'interprétation des résultats pourrait être biaisée.

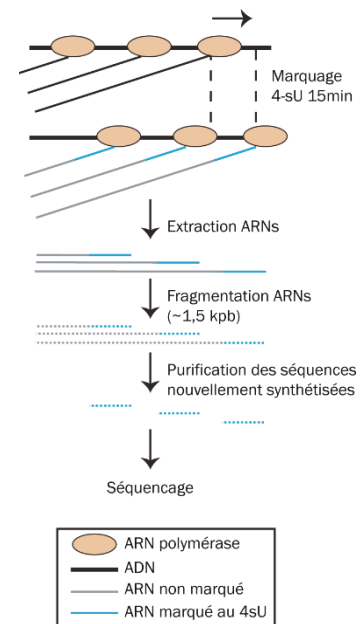


Figure 46 : Stratégie du protocole TT-seq permettant d'étudier le transcriptome naissant. Issu de Schwalb et al., 2016

Les cellules mutantes *RT7*, *RT10* et *RT710* présentent un phénotype dû à la perte de TAF7, TAF10 ou des deux respectivement à partir de quatre jours de traitement (J4) au 4-OHT (Hisler et al.). Bien que dès le jour 2, TAF7 et/ou TAF10, en fonction du mutant, sont significativement déplétés, l'assemblage de TFIID n'est perturbé qu'à partir du jour 3 dans chacun de ces mutants. Néanmoins,

l'ARN pol II est encore en train de transcrire à J3 dans les trois mutants. Ainsi, le jour 3 semble être le jour adéquat pour réaliser le marquage métabolique.

D'après le protocole optimisé par Véronique Fischer (Fischer et al., 2021b), les cellules traitées au 4sU devaient avoir été passées 24 heures avant le marquage pour qu'elles soient isolées les unes des autres et ne forment que de petites colonies. Ceci favoriserait l'entrée du 4sU uniformément dans toutes les cellules. Cependant, les cellules *RT7*, *RT10* et *RT710* sont incapables de proliférer et de former de nouvelles colonies lorsqu'elles sont passées au 4<sup>ème</sup> jour de traitement (Hisler et al.).

Dans le but de marquer les cellules de façon optimale, j'ai passé les cellules au deuxième jour (J2) de traitement et analysé la croissance et la viabilité cellulaire 24 heures après (J2+1) (Figure 47A). A J2+1, TAF7 est indétectable en western blot dans les cellules mutantes *RT7* et *RT710*, comme TAF10 pour les cellules mutantes *RT10* et *RT710* (Figure 47 B). Comme attendu, la déplétion de TAF10 est corrélée aux déplétions de SUPT7L et TAF8 (Bardot et al., 2017; Kamenova et al., 2019; Hisler et al.). Le nombre de cellules vivantes en condition mutante à J2+1, pour les trois lignées cellulaires, est similaire au nombre de cellules en condition contrôle (Figure 47 C, D). Ceci contraste avec l'incapacité de ces mêmes cellules mutantes à former de nouvelles colonies lorsqu'elles sont passées à J4 (Hisler et al.).

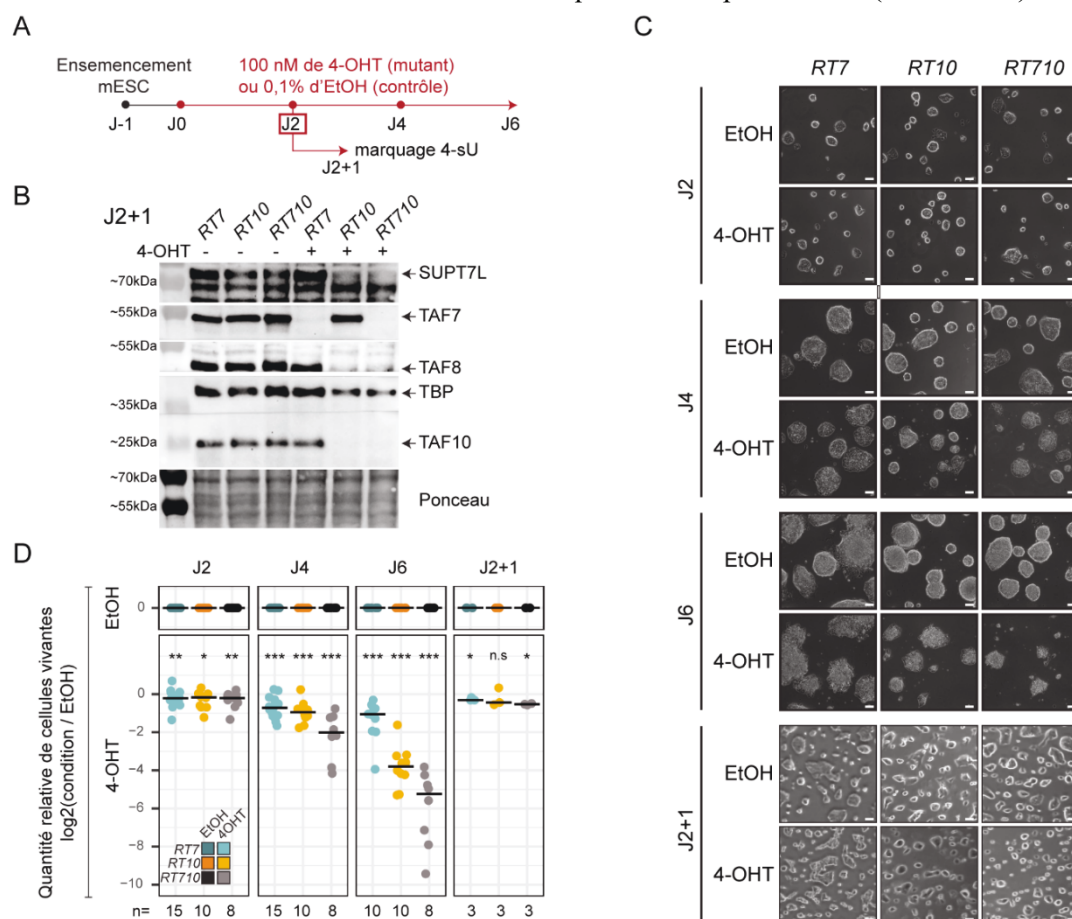


Figure 47 : Après avoir été passées au deuxième jour de traitement, les cellules mutantes se maintiennent en vie et forment des colonies. A) Stratégie pour induire la délétion des protéines d'intérêts au sein des mESC RT7, RT10 et RT710. Au deuxième jour de traitement (J2) les cellules sont passées dans de nouvelles plaques, leur viabilité et leur capacité à former de nouvelles colonies sont analysées un jour plus tard (J2+1). (B) Analyse par western blot de SUPT7L, TAF7, TAF8, TBP et TAF10 à partir de lysats de RT7, RT10 et RT710 contrôle (-) et mutant (+) à J2+1. En dessous se trouve le ponceau. (C) Images des

colonies formées par les cellules RT7, RT10 et RT710 en condition contrôle (EtOH) et mutant (4-OHT). Barre d'échelle : 50  $\mu$ m. (D) Log2 de la quantité relative de cellules vivantes en condition contrôle (EtOH) et mutant (4-OHT) à J2, J4, J6 et J2+1. La ligne noire correspond à la médian des mesures. n indique le nombre de répliquas (4 répliquas biologiques pour J2 à J6, 1 répliquas biologique pour J2+1). Test de Wilcoxon condition mutante par rapport à condition contrôle : \* 0.05, \*\* 0.005, \*\*\* 0.0005.

Le TT-seq repose sur la purification des transcrits nouvellement synthétisés marqués au 4-sU. L'efficacité de cette purification n'est pas constante d'une expérience à l'autre. Par conséquent, 8/10 d'ARNs totaux extraits des mESC sont mixés à 1/10 d'ARNs totaux marqués issus de cellules de drosophile S2, nommé *spike-in*, et 1/10 d'ARNs totaux non marqués de *S.cerevisiae* (Figure 48 A). Après la purification, seuls les ARNs marqués issus des mESC et des S2 devraient être purifiés. Par la suite, durant l'analyse bio-informatique, les ARN marqués purifiés de S2 servent à normaliser les données de séquençage. Par cette normalisation, les variabilités d'efficacité de purification sont gommées permettant de comparer les différents séquençages.

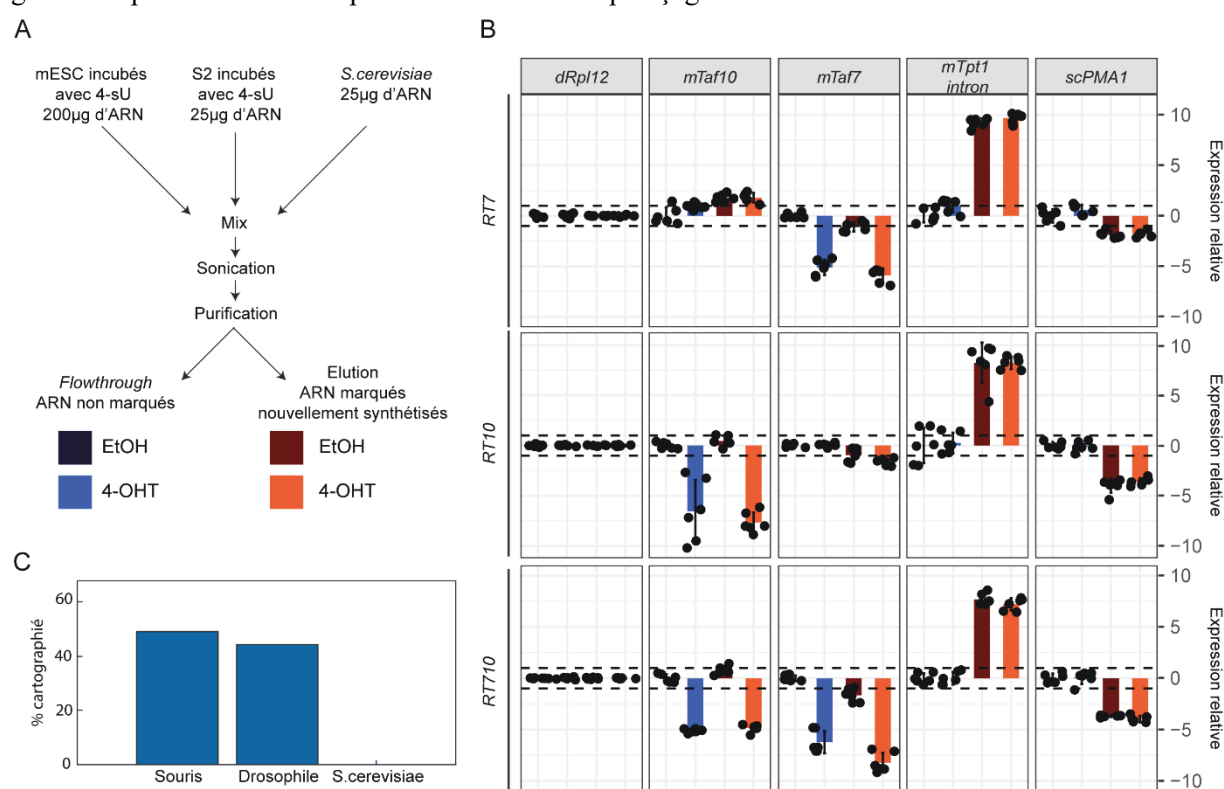


Figure 48 : Les ARNs purifiés après marquage au 4sU et sonication correspondent aux ARNs marqués nouvellement synthétisés. (A) Récapitulatif du protocole pour mixer les ARNs des mESC aux ARNs de drosophile utilisés comme spike-in et aux ARNs de levure non marqués utilisés comme contrôle de purification. Suite au mélange des ARNs, le mix est soniqué puis les ARNs marqués sont purifiés. (B) Analyse par RT-qPCR de la présence des transcrits de dRpl12, mTaf10, mTaf7, mTpt1 et scPMA1 à partir des ARN purifiés (Rouge foncé et rouge) et des ARNs présent dans le flowthrough (bleu foncé et bleu) en conditions contrôle (EtOH, couleur foncé) et mutant (4-OHT, couleur plus clair) à J2+1. Les primers utilisés dans le cas mTpt1 couvrent la jonction de l'exon 3 à l'intron 4 du pre-ARNm Tpt1. Le gène de drosophile dRpl12 est utilisé comme gène de référence. n= triplicata pour le marquage et la purification x duplicata pour l'analyse qPCR. (C) Analyse Fastq-screen permettant de connaître le nombre de read total pour le génome de souris, de drosophile et de levure.

Pour vérifier la bonne purification des ARNs marqués, les ARN obtenus dans la fraction *flowthrough* et les ARNs purifiés ont été analysés par RT-qPCR. Comme attendu, l'expression de *Taf7* par les mutants RT7 et RT710 est fortement réduite, de même que l'expression de *Taf10* pour les mutants RT10 et RT710. Comme espéré, les ARNs purifiés correspondent aux ARNs marqués nouvellement synthétisés. D'une

part, la jonction exon n°3 - intron n°4 du pré-ARNm *Tpt1* est enrichie dans la fraction purifiée par rapport à la fraction *flowthrough* (Figure 48 B). A l'inverse les ARNs de levure sont moins présents dans la fraction purifiée par rapport à la fraction *flowthrough*, comme l'ont confirmé les premières analyses du séquençage (Figure 48 B, C)

Le mélange des ARNm totaux de mESC et des ARNs totaux *spike in* en fonction de leur masse empêche la prise en compte de la quantité de cellules ce qui peut poser problème. Ainsi, si pour l'un des mutants, la transcription est significativement altérée (plus de 50%) alors les ARNs totaux de relativement deux fois plus de cellules seront utilisés pour le mélange. De ce fait, la comparaison de la condition mutante par rapport à la condition contrôle amènerait à une analyse biaisée. Dans le but de contrôler ce paramètre, la quantité totale d'ARNs purifiée à jour 3 a été divisée par le nombre de cellules compté à ce même jour. Bien que la quantité d'ARN par cellule varie d'un rapport de 10 d'une expérience à une autre (Figure 49 A), dû vraisemblablement à l'efficacité variable de l'extraction d'ARN, la normalisation de la quantité d'ARNs par cellule en condition mutante sur cette quantité en condition contrôle indique qu'il n'y a pas de différence significative de quantité d'ARN totaux par cellule entre les conditions mutantes et les conditions contrôles.

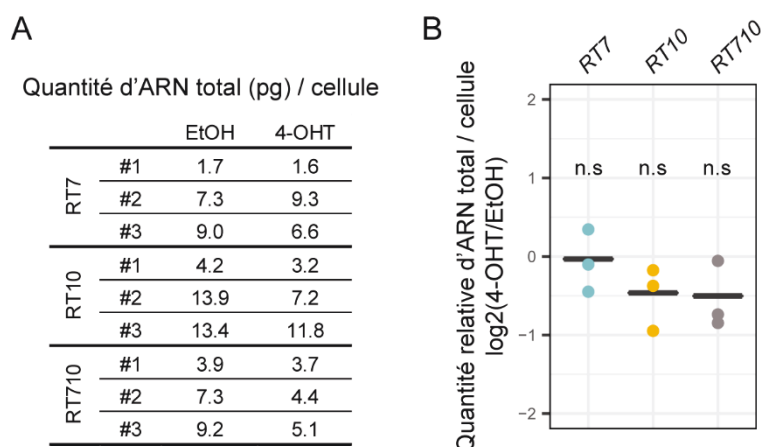


Figure 49 : Les cellules contrôles et mutantes RT7, RT10 et RT710 ont une quantité d'ARN totaux similaire. (A) Récapitulatif de la quantité d'ARN total (pg) par cellule à J2+1 pour des tripliquas expérimentaux en condition contrôle (EtOH) et mutant (4-OHT). (B) Log2 de la quantité relative d'ARN totaux réalisé à partir du tableau en (A). Test de Wilcoxon entre condition mutante et condition contrôle : n.s : non significatif

## VI. Potentiel rôle d'un complexe TAF10 dans le recrutement de TBPL1 aux promoteurs.

Il a récemment été rapporté que TBP n'est pas requis pour la transcription médiée par l'ARN pol II au sein des mESC (Kwan et al., 2021). Il est supposé que TBPL1 pourrait compenser la perte de TBP pendant un certain temps chez ces cellules. En effet, que ce soit en présence ou absence de TBP, TBPL1 est localisé au niveau des promoteurs des gènes transcrits, ayant ou n'ayant pas de boîte TATA. (Kwan et al., 2021). Dans cette étude, la machinerie transcriptionnelle liée à TBPL1 n'a pas été élucidée.

Afin d'avoir une meilleure connaissance du recrutement de TBPL1 au niveau des promoteurs, j'ai voulu savoir si TBPL1 est complexé à TFIID. J'ai tout d'abord cherché à mettre TBPL1 en évidence par western blot à partir d'extraits cellulaires totaux. Quatre anticorps ont été testés dans ce but : l'anticorps polyclonale de lapin 1727, et trois anticorps monoclonaux de souris, 1TRF-2D5, 1TRF-2D7 et 3TRF-2A1. Comme attendu, les anticorps 1727 et 1TRF-2D5 ont donné un signal à environ 20kDa dans l'extrait de testicule et de cerveau, exprimant tout deux TBPL1, mais pas dans l'extrait de foie et le poumon qui ne l'expriment pas (Martianov et al., 2001, 2016) (Figure 50A a-b, flèche). Il est important de noter que ces deux anticorps ont déjà été validés pour leur utilisation en western blot à partir d'extrait de testicule murin (Martianov et al., 2001, 2016).

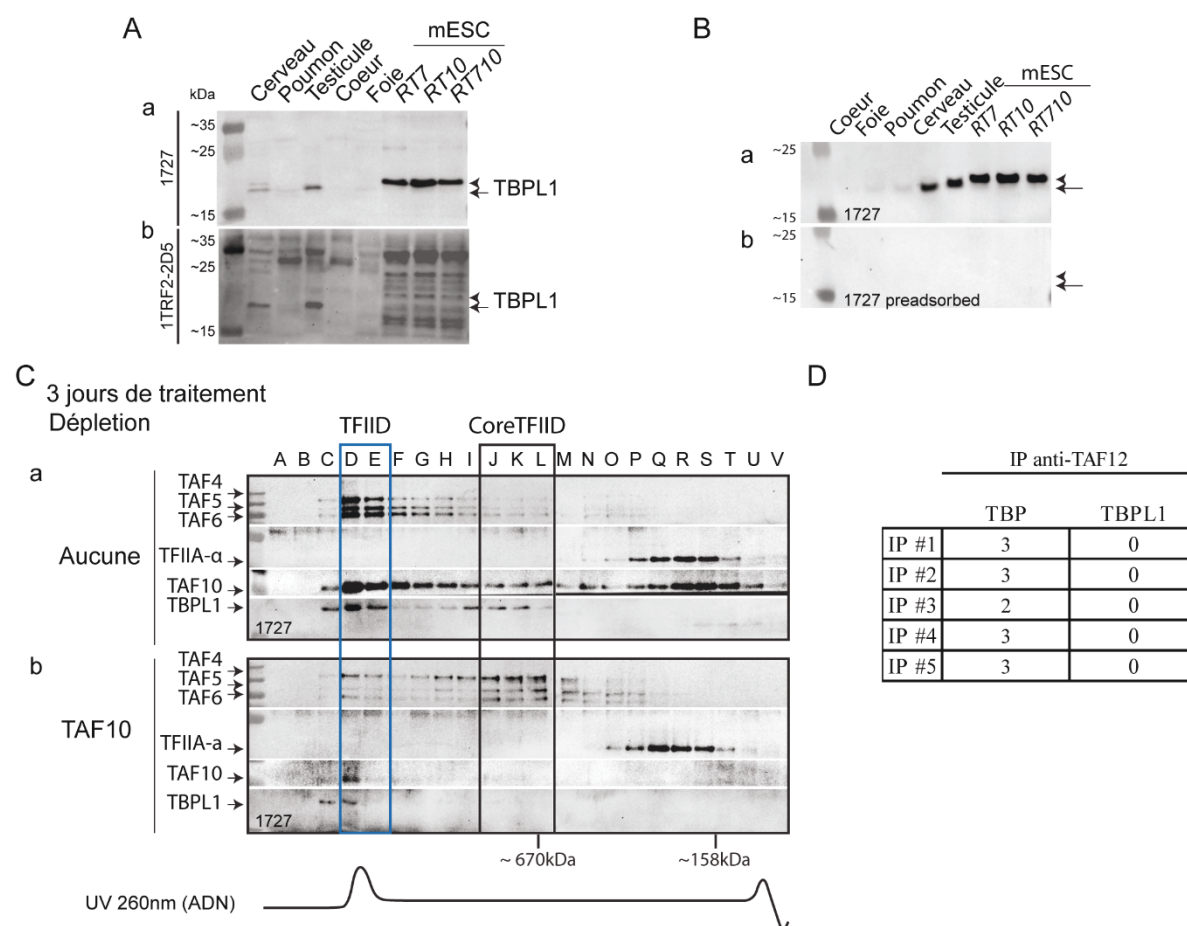


Figure 50 : TBPL1 est présent au sein d'un complexe de 1 à 2 MDa contenant TAF10. A) Analyse de la capacité de l'anticorps 1727 (a) et de l'anticorps 1TRF-2D5 (b) à détecter TBPL1 par western blot. Des extraits de cerveau et de testicule, exprimant TBPL1, ainsi que de foie et poumon, n'exprimant pas TBPL1, sont utilisés comme contrôle. Des extraits de RT7, RT10 et RT710 contrôle ont également été analysés avec ces deux anticorps. B) Analyse par western blot de TBPL1 dans les différents extraits protéiques décrits en (A) avec l'anticorps 1727 non pré absorbé (a) et après avoir été pré absorbé avec le peptide P18 (b) utilisé pour le produire. C) Analyse par western blot des fractions obtenues par chromatographie d'exclusion de taille de NWCE issues de cellules RT10 contrôles (a) et mutantes (b) au troisième jour de traitement. TAF4, TAF5, TAF6, TFIIA- $\alpha$ , TAF10 et TBPL1 ont été révélés. L'anticorps polyclonale de lapin J17 a été utilisé pour mettre en évidence TFIIA- $\alpha$  (Yu et al., 2020a) et l'anticorps 1727 a été utilisé pour révéler TBPL1. D) Tableau récapitulant le nombre de PSM détectés après l'analyse par spectrométrie de masse des éluions d'IP anti-TAF12 réalisées à partir de NWCE de mESC. Ces données sont issues des IP réalisées dans le cadre de l'étude décrite dans le papier (Hisler et al) à partir d'extraits de cellules contrôles RT7 (#1 à #3) et RT10 (#4 et #5) à J2 (#1 et #4), J3 (#2 et #4) et J4 (#5). Les PSM indiqués sont la somme des PSM détectés dans le tripliqua expérimental.



La révélation de TBPL1 avec l'anticorps 1727 montre que TBPL1 issu du testicule est plus grand que celui issu du cerveau (Figure 50 B a, flèche). Dans les extraits de mESC, l'anticorps 1727 fait apparaître une bande à une taille supérieure à la taille de TBPL1 issu du testicule (Figure 50 B a, comparer la flèche et la tête de flèche). La pré absorption de l'anticorps avec le peptide, utilisé pour le produire, entraîne comme attendu la disparition des bandes (Figure 50 B b comparer à B a, flèche et tête de flèche). L'anticorps 1TRF-2D5 fait apparaître dans les extraits de mESC des bandes de taille relativement proche de celles issues de testicules et de cerveaux (Figure 50 A b, comparer les têtes de flèche et la flèche) ainsi que la bande mise en évidence par l'anticorps 1727 (Figure 50 A, comparer les têtes de flèche de a et de b). Ensemble ces résultats confortent l'idée que l'anticorps 1727 détecte spécifiquement TBPL1 par western blot.

Dans le but de savoir si TBPL1 fait partie de TFIID au sein des mESC, les cellules *RT10* ont été traitées 3 jours soit avec de l'EtOH (contrôle) soit avec du 4-OHT (mutant). Des NWCE ont été réalisés puis séparés en différentes fractions par chromatographie d'exclusion de taille. L'analyse par western blot des fractions issues de la condition contrôle montre que TBPL1 se trouve dans les fractions D et E (Figure 50 C a). Ces fractions contiennent TFIID, comme le montre la présence de TAF4, TAF5, TAF6 et TAF10. Contrairement à ce qui a été rapporté dans les testicules (Martianov et al., 2016), TBPL1 n'est pas élué dans les fractions contenant TFIIA, indiquant que dans les mESC, TBPL1 n'est pas complexé avec TFIIA (Figure 50 C a). L'analyse des fractions issues des conditions mutantes montre comme attendu la déplétion quasi-totale de TAF10 (Figure 50 C b). TAF4, TAF5 et TAF6 se trouvent dans les fractions J à L de plus bas poids moléculaire, indiquant la perturbation de l'assemblage de TFIID résultant à la formation uniquement du core-TFIID (Hisler et al.). La présence de TBPL1 est également fortement diminuée dans cette condition mutante. Ces résultats indiquent que TBPL1 fait partie d'un complexe dans lequel TAF10 joue un rôle structural important. Sur la base de cette observation, les données de spectrométrie de masse obtenues à partir de l'analyse des éluions d'IP anti-TAF12 ont été réanalysées (Figure 50 D). Alors que les peptides spécifiques de TBP sont relativement bien détectés dans chaque élution, aucun peptide spécifique à TBPL1 n'a été détecté. Ces données préliminaires semblent indiquer que TBPL1 ne fait pas partie de TFIID mais d'un complexe d'une taille d'environ 1-2MDa pour l'assemblage duquel TAF10 est important.

## VII. Matériels et méthodes.

Les protocoles utilisés dans ces études préliminaires ont été décrits dans (Hisler et al.). Ci-dessous est décrit la réalisation des extraits de chromatine ainsi que le protocole détaillé pour la réalisation du TT-seq.

### 1. Réalisation d'extrait de chromatine

Les cellules ont été pelletées, rincées avec du PBS, remises en suspension dans 1 volume (1 mL de buffer par gramme de cellules) de tampon hypotonique froid (10 mM Tris (pH8), 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 1X cOmplete). Elles ont été lysées par 10 coups délicats dans un broyeur (Kimble, Cat#885301-0002). Après centrifugation à 9000 g pendant 10 min, le culot, qui contient les noyaux, a été remis en suspension dans 1 volume de tampon riche en sel froid (20 mM Tris, 1.5 mM MgCl<sub>2</sub>, 450 mM NaCl, 0,2 mM EDTA, 25% glycérol, 0,5% NP40, 1X cOmplete, 1 mL de buffer par gramme de cellules). Ils ont été lysés dans un broyeur et incubés dans la glace pendant 30 minutes. La chromatine sous forme de méduse a été récupérée et lavée toute la nuit dans du tampon riche en sel sous agitation. Après avoir été lavé dans de l'IP100 (25 mM Tris Hcl pH7,9, 10% Glycerol, 0,1% NP40, 5 mM MgCl<sub>2</sub>, 100 mM KCl, 1X cOmplete), l'ADN est dénaturé à 100°C 10 min dans 1 volumes d'IP100 puis refroidi rapidement sur glace. Le surnageant correspond à l'extrait de chromatine.

### 2. Protocol TT-seq

Le protocole pour la réalisation du TT-seq est décrit par la suite en anglais.

## Newly transcribed RNA labelling and isolation (From Fischer V, optimized and clarified for some points)

### Specific material / solution we need along the protocol:

- 4-Thiouridine (4sU) from Glentham Life Sciences<sup>1</sup> [Ref: GN6085-100MG].  
For 100mM stock solution: 100mg of powder are diluted in 3.84mL of DMSO.
- Tri Reagent (Trizol) from MRC<sup>2</sup> [Ref: TR 118].
- RiboPure – Yeast kit from Invitrogen<sup>1</sup> [Ref: AM1926].
- RNasin (40U/μL) from Promega<sup>2</sup>.
- Turbo DNA-free kit from Invitrogen<sup>2</sup> [Ref: AM1907].
- SnapCap tubes<sup>3</sup> (Covaris).
- μMACS Streptavidin Kit<sup>2</sup> [Ref: 130-074-101].
- EZ-Link™ HPDP-Biotin from Thermo Scientific<sup>1</sup> [Ref: 21341].  
For 1mg/mL stock solution: 50mg of powder are diluted in 50mL of DMSO
- SuperScript IV Reverse Transcriptase from Invitrogen<sup>2</sup> [Ref: 18090050].
- Random hexamer primer from Thermo Scientific<sup>2</sup> [Ref: So142].
- dNTPs (10mM) from Roche<sup>2</sup> [Ref: 42793424].
- LightCycler 480 SYBR Green I Master from Roche<sup>2</sup> [Ref: 04707516001].

<sup>1</sup>Purchased externally

<sup>2</sup>Purchased from the “petit produit” facility

<sup>3</sup>Purchased from the sequencing facility

### Solution :

10X biotinylation buffer	Conc.init (mM)	Conc.final (mM)	Volume (mL)
Tris-HCl (pH7.5)	1000	100	5
EDTA	500	10	1
DEPC-treated water			44
Volume total			50

} Stored at RT°C

RT washing buffer	Conc.init (mM)	Conc.final (mM)	Volume (mL)
Tris-Hcl (pH7.5)	1000	100	20
EDTA	500	10	4
NaCl	5000	1	40
Tween20	100%	0.10%	0.2
DEPC-treated water			135.8
Volume total			200

} Stored at RT°C



### **4sU-labelling in mouse ES E14 cells**

*Material to prepare for day 3 to do the experiment with one clone and two conditions of treatment:*

- 24 mL of Trizol (3mL/P150)
- 24 x 1.5 mL tube well annotated for a one condition
- Cold PBS
- 100mL of pre-warmed (37°C) of 500 µM of 4sU in 2i+lif for

#### *Protocol*

- 1) **Day -1:** The morning of the day before the day on which the treatment begins:
  - For a single condition, 10 P100 plates previously gelatinized are seeded with  $0.8 \times 10^6$  cell per plate. 5 will be treated with 4-OHT and the other one with EtOH.
  - The same day prepare the 2i+lif medium for the day after + gelatinized for 10 P150 plates
- 2) **Day 0:** On the morning of the treatment day:
  - Treat cells with either 1/1000 EtOH or 100nM 4-OHT (1/1000 from 100 µM 4-OHT stock) diluted in 2i+lif.
- 3) **Day 2:** After two days of treatment, during the afternoon
  - $10^7$  cells are trypsinized and seeded again in P150 plates. For each condition (EtOH or 4-OHT), 5 P150 plate have to be plated, so a minimum of  $5 \times 10^7$  cell are needed.
- 4) **Day 3:** On the morning of the third day of treatment, beginning of the morning
  - 8 plates will be treated with 4sU to label newly synthesized RNA.
  - Prepare 100 mL of preheated (37°C) 500µM 4sU/2i+lif to treat 8 plates (4 from 4-OHT condition and 4 from EtOH condition). This corresponds to 250 µL of 100 mM 4sU in 100 mL 2i+lif.
  - Remove culture medium
  - Treat cells with 10 mL of preheated (37°C) 4sU/2i+lif medium per P150
  - Incubate for 15 minutes in incubator
  - Remove medium
  - Wash with ice cold 1x PBS

#### *Under the hood*

- For one condition, add 3 ml of Trizol reagent per 15 cm plate, rinse the plates with the Trizol to collect all cells.
- Distribute the volume obtained in 12 x 1.5 mL tubes and freeze then in liquid nitrogen.

*If other cell / other condition has to be labelled, the 15 mL tube can be store for one hour on ice.*

*Frozen samples can be stored at -80°C for several months.*

- From the two plates left (one EtOH treated, the other one 4-OHT treated), cells are collected, counted and then used to check the depletion of the protein(s) of interest by western blot.

### **S2 cell culturing + 4sU labeling**

- Culture S2 cells in Schneider medium with 10% FCS 98i + 0.5% PenStrep at ~ 27°C (S2 cells don't need special CO<sub>2</sub> levels and can also be grown at room temperature)

*NOTE: S2 cells attach to plate until they get confluent, then they also start to grow in suspension*

- For labelling:
  - Do some up and down with the medium to get cell attached on the plate.
  - Transfer medium containing cells in a 50 mL tube protected from light, measure the volume and add an appropriate volume of 4sU (same procedure as for mES cells)

*NOTE: For instance, from 3 x 75 mm flask, about 43 mL of cell suspension were collected in which 215µL of 100 mM 4sU were added.*

- For the labelling, S2 cells should be kept under cover (aluminium) at RT under toggle for 15min.
- Pellet the cells (centrifugation at 1000 g for 5 min)
- Wash with PBS
- Resuspend in 12 mL Trizol / ~50 mL of cell suspension
- Spread in 12 tubes (1mL/tube) and freeze then in liquid nitrogen.
- Do RNA extraction as done for mESC.

### **S.cerevisiae culturing**

- Yeasts are grown in YPD medium for several hours until the Do<sub>595nm</sub> reaches 1.
- RNAs are then extracted using the kit "RiboPure™ – Yeast"

### **RNA extraction (mESC and S2)**

*Material to prepare to do the experiment with one clone and two conditions of treatment:*

- 24 x 1.5 mL tube to recover aqueous phase (12 for each condition)
  - 6 x 1.5 mL tube to pull resuspended RNA (3 for each condition)
  - Chloroform
  - Isopropanol
  - 75% EtOH (diluted with DEPC-treated water)
  - DEPC-treated water
- 
- ✓ Keep RNA in the dark as much as possible
  - ✓ Use an odd number of tubes, and no more than 24 tubes, because the centrifuge can only hold 24 tubes.

### *Protocol*

- Tube are defreeze by incubating them during 5-10 minutes at 37°C tubes. This permits complete dissociation of the nucleoprotein complex
- Add 0.2 mL of chloroform
- Shake tube vigorously by hand for 15 seconds
- Incubate for 2-3 minutes at RT
- Centrifuge the sample at 14,000 g (small centrifuge) 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*

- Recover ~700 to 900 µL of aqueous phase in a new clean tube.

*NOTE: To do that, use the P200 with filter tips.*

- Add 0.7 mL of 100% isopropanol to the aqueous phase
- Shake tube vigorously by hand for 15 seconds and incubate at RT for 10 minutes
- Centrifuge the sample at 14,000 g (small centrifuge) for 15 minutes at 4°C.
- 
- Pour the supernatant from tube, leaving the RNA pellet
- Wash the pellet with 1 mL of 75% ethanol/DEPC-treated water
- Centrifuge the sample at 14,000 g (small centrifuge) for 5 minutes at 4°C
- Pour the supernatant and dry the wall of the tube with clean paper
- Air dry the RNA pellet for 5 minutes at RT°C, in the dark
- Prepare 1250 µL of DEPC-treated water in which 3 µL of RNase inhibitor will be added.
- Resuspend the RNA pellet in 50 µL RNase-free water.
- Incubate in a heat block set at 55-60°C, 450 rpm, for 10-15 minutes
- Pool several tubes of the same condition in a 1,5 mL tube up to have 200 µL per tube.
- Proceed to DNase treatment or freeze sample in liquid nitrogen and store them at -80°C.

### DNase treatment using TURBO DNA-free Kit

*Material to prepare to do the experiment with one clone and two conditions of treatment:*

- TURBO DNase kit
- 6 x 1.5 mL tubes

- ✓ Use an odd number of tubes, and no more than 24 tubes, because the centrifuge can only hold 24 tubes.
- ✓ Keep RNA in the dark as much as possible

#### *Protocol*

- Add 0.1 volume of 10X TURBO DNase Buffer and 2-3  $\mu$ L TURBO DNase to the RNA, mix gently

*NOTE: Make a mix: For 6 tubes with 200  $\mu$ L of RNA, mix 120  $\mu$ L of 10 X TURBO buffer + 18  $\mu$ L of TURBO DNase and then put 22  $\mu$ L of the mix per tube containing RNA.*

- Incubate at 37°C for 20-30 min in the dark.
- Add 0.1 volumes (~20  $\mu$ L) of resuspended DNase Inactivation Reagent
- Incubate 5 min at RT (24°C), mixing occasionally, flick the tube 2-3 times during the incubation period to redisperse the DNase Inactivation Reagent

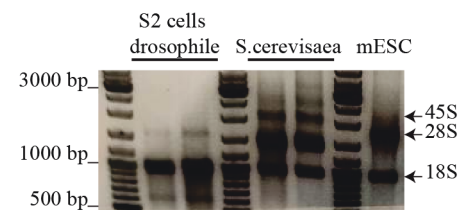
*NOTE: the DNase inactivation Reagent are beads. It's important to flick the tube to maintain the tube in suspension.*

- Centrifuge at 9,000 g (small centrifuge) for 5 min
- Recover the supernatant containing RNA: pool in a same 1.5 mL tube all supernatant from the same condition.

*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*

- Measure by nanodrop the concentration and migrate 1  $\mu$ g of RNA on 1% agarose gel to check the quality of the RNA.

*NOTE: It is really important to migrate RNA in clean TAE, clean migration tank and use clean DNA/RNA loading buffer. Otherwise, RNA could be degraded during the migration. Also keep in mind that the 28S RNA from drosophile is actually a dimer that will be denaturated during the migration. Therefore, a single band will be visible.*



*Figure 51 : Profil of RNA from S2 cells, yeast and mESC analyzed on 1% agarose gel.*

- To check the quality of RNA by bioanalyzer, 20 $\mu$ L at 75ng/ $\mu$ L were given to the sequencing platform.

### Mix spik-in + mESC RNA

Labeled RNA from mESC and S2 as well as yeast unlabeled RNA were measured using Qbit machine and the Quant-it RNA Broad Range kit according the manufacturer's instruction. Note that before measuring the concentration, 2  $\mu$ L of RNA were diluted in 2  $\mu$ L of RNase-free water to ensure that the sample concentration is in the correct measurement interval

For one condition, in a clean tube, mix about 200µg of total labeled mESC RNA with 25 µg of total labeled S2 RNA and 25 µg of total non-labeled RNA.

## **Precipitation**

*Material to prepare:*

- 5M NaCl
  - Isopropanol
  - 75% ethanol/DEPC-treated water
  - RNase-free water
  - RNase inhibitor
- 
- Add 0.1 volume of 5M NaCl
  - Mix
  - Add 1 volume of isopropanol
  - Centrifuge at 14,000 g (small centrifuge) for 15 minutes at 4°C
  - Wash the pellet with 1 mL of 75% ethanol/DEPC-treated water
  - Centrifuge at 14,000 g (small centrifuge) for 5 minutes at 4°C
  - Pour the wash
  - Dry the wall of the tube with clean paper
  - Air dry the RNA pellet for 5 minutes at RT°C, in the dark
  - Resuspend the RNA pellet in 130 µL RNase-free water
  - Incubate the RNA 15 min at 60°C and 750 rpm.

## Fragmentation

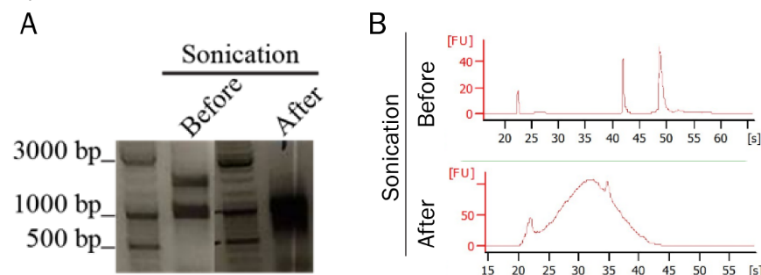
*Material to prepare:*

- Prepare two 1.5 mL tubes per RNA sample
  - Prepare one 2 mL tube per RNA sample
  - Prepare one Snap Cap tubes per RNA sample
  - Prepare 1% agarose gel with clean TAE
  - Clean the migration tank and fill it with clean TAE
  - DNA loading buffer and DNA ladder
- ✓ The fragmentation is carried out with a Covaris E220. The sonication is done inside the bath. I figured out the level of water can influence the efficiency of the sonication. For all my experiments, the level of water was at 10 for the rule “Fill” and 5 on the rule “Run” when the platform containing the tube is in “start position”, meaning on the water. **In any case, fragmentation must be checked each time before starting the purification!**
- ✓ Keep RNA in the dark as much as possible

### Protocol

- In a new 1.5 mL tube, keep 1  $\mu$ L of RNA in 9  $\mu$ L of H<sub>2</sub>O RNase-free water (tube A). Mix, and again collect 1  $\mu$ L from tube A and put it in tube B containing 25  $\mu$ L of RNase-free water. These two tubes could be used to check the quality of the RNA after making the spik-in and the precipitation. The tube A will be used to check the quality on 1% agarose gel and the tube B will be used on the bioanalyzer.
- Open the Snap Cap tubes and put the ~129  $\mu$ L of RNA (containing ~250  $\mu$ g of RNA) in the Snap Cap tubes.
- Covaris E220 settings: 1% duty factor, 100 W, 200 cycles per burst, 80 sec. Don't forget to put "concentrator"
- Open the Snap Cap tubes and recover the RNA in a 2mL tube.
- In a new 1.5 mL tube, proceed as the first bullet point. This will be used to check if the fragmentation worked.
- Migrate in 1% agarose gel RNA

*NOTE: Fragmented RNA is supposed to be in a range between 10 kb and above 200 bp (average of >1.5 kb).*



*Figure 52 : RNA before and after sonication. RNAs were analyzed on 1% agarose gel (A) or with the bioanalyzer (B).*

## Newly transcribed RNA extraction

### *Material to prepare:*

- Biotinylation buffer
- DMSO from cell culture
- DEPC-treated water
- Biotin-HPDP (of 1 mg/mL stock, in DMSO, store at -20°C, *light-sensitive*, wrap in aluminium) → defreeze it at 37°C in water bath
- 0.1M DTT prepared with DEPC-treated water
- Chloroform
- Isopropanol

- ✓ Keep RNA in the dark as much as possible until 4sU-labeled RNA are purified.

### *Protocol*

- Heat RNA for 10 minutes at 60°C and immediately chill it on ice for 2 minutes
- In ~130 µL of fragmented RNA, add in order:
  - 200 µL of 1mg/mL Biotin
  - 100 µL of 10X biotinylation buffer
  - 200 µL of DMSO
  - 370 µL of DEPC-treated water.
- Incubate at RT (= 24°C) and protected from light for 3 hours with gentle agitation (thermomixer, 550 rpm, covered by aluminium)
- ✓ During this time, prepare for one sample
  - ❖ 2 x 2 mL tubes to recover upper phase
  - ❖ 3 x 1.5 mL tubes to recover FT (tube 1, 2 et 3)
  - ❖ 1 x 1.5 mL tube to recover 4-sU RNA purified
- Add approximately 1 mL of chloroform to the tubes

*NOTE: that correspond to add 1 volume of chloroform for 1 volume of RNA*

- Mix vigorously
- Spin at 16,000 g for 5 minutes, at 4°C

*NOTE: This step allows to remove biotin that did not biotinylate the RNA*



- Carefully transfer only upper phase into new 2 mL tubes (already prepared)

*NOTE: Introduce unbound biotin will reduce the efficiency of labelled RNA purification.*

- Add 80  $\mu$ L of 5 M NaCl (1/10 of volume)
- Mix
- Add 800  $\mu$ L of isopropanol (one volume)
- Mix and spin at 16,000 g for at least 15 minutes, at 4°C
- Pour the supernatant and add 1 mL of 75% ethanol
- Spin at 16,000 rpm for 10 minutes, at 4°C
- Pour supernatant, dry the wall of the tube with paper.
- Resuspend RNA in 100  $\mu$ L RNase-free water
- Heat biotinylated RNA for 10 minutes (thermomixer, 65°C and 750rpm)
- Chill the samples on ice for 5 minutes
- Add 100  $\mu$ L of magnetic streptavidin beads ( $\mu$ MACS Streptavidin beads and kit, Miltenyi, stored at +4°C) to biotinylated RNA (final volume of 200  $\mu$ L)
- Incubate with slight shaking for 90 minutes, at RT (thermomixer, 24°C, 550 rpm, in 2 mL tubes)
- ✓ During this time,
  - Put 100% EtOH at -20°C
  - Prepare 0.1M DTT in DEPC-water.
- Place  $\mu$ MACS columns in the magnetic stand (one column/sample, Miltenyi)
- Add 900  $\mu$ L of RT washing buffer to columns (for pre-run and equilibration)
- Apply beads/RNA mix (200  $\mu$ L) to the columns
- Collect flow-through in 1.5 mL tubes (tube 1)  
and apply it to the column again to collect it in tube 2  
and again, to collect it in tube 3. If the beads are bound to the column, the solution collected should be transparent.  
The tube 3 is kept on ice and contains normally unlabeled RNA.
- Wash columns 5 times with 600, 700, 800, 900 and 1000  $\mu$ L of washing buffer
- Elute RNA in new 1.5 mL Eppendorf tubes:
  - First 200 $\mu$ L of 0.1 M DTT
  - Wait 5 min
  - Do a second elution with 200  $\mu$ L of 0.1 M DTT



*Figure 53 : Pictures to show how to use the  $\mu$ MACS columns in the magnetic stand. The top picture is when the columns are washed. The bottom picture is when the flowthrough is collected.*

- To elute purified RNA, add:
  - 2  $\mu$ L of 20 mg/mL glycogen (RNA grade) and mix
  - 40  $\mu$ L of 3 M NaOAc (pH 5.2, 0.1 volume) and mix
  - 1200  $\mu$ L of ice cold 100% ethanol (3 volumes) and mix
- To elute RNA from the flow through, add:
  - 1  $\mu$ L of 20 mg/mL glycogen (RNA grade) and mix
  - 20  $\mu$ L of 3M NaOAc (pH 5.2, 0.1 volume) and mix
  - 600  $\mu$ L of ice cold 100% ethanol (3 volumes) and mix
- Let RNA precipitate overnight, at -20 °C
- Recover RNA by centrifugation (13,500 g for 10-30 minutes, at 4°C)
- Wash with 75% EtOH
- Let dry the pellet
- Resuspend
  - Purified RNA in 15  $\mu$ L DEPC-treated water and incubate 15 min at 60°C under agitation → 2  $\mu$ L are diluted in 12  $\mu$ L of DEPC-treated water and will be used for the RT-qPCR
  - RNA from flowthrough in 150  $\mu$ L DEPC-treated water and incubate 15 min at 60°C under agitation → 7.5  $\mu$ L are diluted in 2.5  $\mu$ L of DEPC-treated water and will be used for the RT-qPCR.

### **RT-qPCR**

The enrichment in labelled newly synthesized RNA can be verified by RT-qPCR. About 10 genes can be checked in duplicate, below the table of primer used for the qPCR.

RT is realized with Superscript IV with the thermomixer and qPCR with 480 Sybr Green I Master.

qPCR program used in the Light Cycler 480 machine.

Pre-incubation : 95°C – 5min – 4.8°C/sec

Amplification : 95°C – 10sec – 4.8°C/sec

60°C – 10sec – 4.8°C/sec

72°C – 10sec – 4.8°C/sec

45X

<i>S.cerevisiae</i>	PMA1	PMA1 - O53F	CTCATCAGCCAACTCAAGAAA	Check that unlabeled RNA are not in purified fraction
		PMA1 - O124R	CGTCATCGTCAGAAGATTCA	
	PRPS3	RPS3 - F	ATTGTTGAACGGTTTGGC	
		RPS3 - R	CCCTTAGCACCAGATTCCATA	
<i>Drosophile</i>	Rpl12	dRpl12_F	AAGGGAACCTGCAAGGAAGT	Check that labeled RNA are enriched in purified fraction
		dRpl12_R	CCCTCGTTCAGTTCGTCAATA	
	Tubuline	alphaTub84B_Foward	GCTTCCTCATCTTCCACTCG	
		alphaTub84B_Reverse	GCTTGGACTTCTTGCCGTAG	
Mouse	Tpt1	mTpt1 Int forward	TTAAGCACATCCTTGCTAATTTCA	To check intronic enrichment
		mTpt1 Int reverse	TGTACGAGACAGCAAACAGACTTT	
	Clf1	mCLF1 Intron_forward	TATGAGACCAAGGAGAGCAAGAA	
		mCLF1 Intron_forward	GTTAAGCTCTGAGAAAGGGAACC	
	mTaf7	qVH054	AATATGCCGCTACGGTGAGG	To check deletion
		qVH055	TCAGGTTGACATGCCCAGAC	
	mTaf10	qVH062	GAGGGGGCAATGTCTAACGG	
		qVH063	CGCGGTTCAGGTAGTAACCA	
		qVH041	CCACATGCAGCGCCTCATT	

---

# Discussion

---

---

# Discussion

---

## I. La réorganisation du core-TFIID par le module TAF2-8-10 n'est pas une étape requise pour le recrutement des autres TAFs.

Le modèle actuel expliquant l'assemblage de TFIID repose sur l'assemblage séquentiel de sous-module (Bieniossek et al., 2013b; Patel et al., 2020; Trowitzsch et al., 2015). Au core-TFIID s'assemble le module TAF2-8-10 permettant de former le complexe 8-TAFs (Trowitzsch et al., 2015). Lors de cette intégration, le core-TFIID perd sa symétrie axiale, le pré-lobes B se balançant vers le pré-lobes A d'environ 40 Å et le pré-lobes A de 14 Å vers le pré-lobes B (Bieniossek et al., 2013b). Cette réorganisation a été suggérée comme étant requise pour le recrutement des sous-unités TAF1, TAF7, TBP ainsi que des hétérodimères TAF3-10 et TAF11-13.

Mes analyses par chromatographie indiquent que TAF7 et TBP sont présents dans un même complexe différent de TFIID en condition normale. Mes données d'IP anti-TAF7 indiquent que TAF7 et TBP sont également capables de se lier au core-TFIID en absence de TAF10. Mes données d'IP-MS anti-TAF10 suggèrent que TAF1 peut se lier au core-TFIID en absence de TAF10. D'autres IP-MS en condition déplétée en TAF10 à jour 2 ont été réalisées par un ancien étudiant en doctorat Paul Bardot (publiées dans sa thèse). Dans les IP-MS anti-TAF7, TBP est présent à une abondance relativement équivalente entre la condition mutante et la condition contrôle. L'abondance en TAF1 est deux fois inférieure en condition mutante qu'en condition contrôle. En guise de comparaisons, l'abondance de l'ensemble des autres TAFs était au moins 3 fois inférieure en condition mutante qu'en condition contrôle. Ainsi, TAF1 et TBP sont les deux sous-unités les plus interactions avec TAF7 en absence de TAF10. Dans TFIID, TAF1 est en contact avec le core-TFIID par l'intermédiaire de ses 3 motifs de liaison à TAF6 (T6BMs, *TAF6-binding motifs*) (Chen et al., 2021a), avec TAF7 par son domaine DU3591 (Bhattacharya et al., 2014; Chen et al., 2021a; Wang et al., 2014) et avec TBP par ses domaines TAND1 et TAND2 (Anandapadamanaban et al., 2013; Liu et al., 1998; Patel et al., 2018).

Il a également été montré qu'en condition *in vitro* la mise en contact de l'hétérodimère TAF11-13 à TBP entraîne la formation d'un complexe TAF11-13-TBP. Dans ce complexe, TAF11-13 interagit dans la partie concave de TBP, lieu de liaison avec l'ADN et le TAND1 de TAF1 (Gupta et al., 2017). Néanmoins, la mise en contact de ce complexe TAF11-13-TBP au TAND1 entraîne la relocalisation efficace de TBP au niveau de TAND1 (Gupta et al., 2017). Ainsi, TBP présente une plus grande affinité pour le TAND1 que pour l'hétérodimère TAF11-13. Au sein de TFIID, TAF11-13 interagit avec un côté d'un lobe de TBP, TAND1 de TAF1 se trouvant à l'intérieur du domaine concave de TBP (Patel et al., 2018). Ainsi, il semble que TAND1 interagit en premier avec la partie concave de TBP résultant de l'interaction dans un deuxième temps de TAF11-13 avec le côté de TBP au sein du TFIID.

Ensemble, ces données suggèrent que TAF1, TAF7 et TBP forment un complexe capable de s'intégrer au core-TFIID, nommé complexe 7-TAFs par la suite (Figure 54 A, a). Néanmoins, on ne peut exclure que ce complexe ne se forme qu'en absence de TAF10. De plus, les IP-MS anti-TAF12 détectent bien moins de sous-unités du complexe 7-TAFs en condition mutante *RT10* qu'en condition contrôle à jour 3. Ceci peut suggérer :

1) soit que le complexe core-TFIID associé à TAF1-7-TBP n'est pas stable

2) soit que le recrutement que TAF1-7-TBP au niveau du core-TFID est moins efficace que son recrutement au niveau du complexe 8-TAFs (Figure 54 A, c).

Pour mieux comprendre l'enchaînement des étapes d'assemblage de TFIID qui suit la formation de core-TFIID, il serait intéressant de savoir si c'est le module TAF1-7-TBP ou le module TAF2-8-10 qui s'assemble en premier au core-TFIID. Pour cela, des tests d'assemblage *in vitro* serait tout d'abord nécessaire pour savoir si le module TAF1-7-TBP ainsi que le complexe 7-TAFs peuvent être assemblés *in vitro*. L'assemblage *in vitro* du module TAF2-8-10 ainsi que du complexe 8-TAFs a déjà été montré comme possible (Bieniossek et al., 2013b; Scheer et al., 2021; Trowitzsch et al., 2015).

Si le module TAF1-7-TBP et le complexe 7-TAFs peuvent être assemblés *in vitro*, la cinétique d'assemblage des complexes 7-TAFs et 8-TAFs, ainsi que le recrutement des autres modules TAFs au sein de ces complexes, pourrait être comparée. Ceci permettra d'avoir une idée de la voie d'assemblage optimale de TFIID après la formation du complexe core-TFIID.

Pour vérifier en condition *in cellulo* la présence des complexes 8-TAFs et 7-TAFs en condition normale, une succession de co-immunoprécipitations pourrait être réalisée à partir de NWCE de mESC. Une première IP anti-TRRAP permettrait d'enlever le complexe SAGA ainsi que les protéines qui se lient de façon non spécifique aux billes et/ou aux anticorps. Suite à cela :

- 1) une IP anti-TBP sur le surnageant d'une IP anti-TAF10 permettrait d'étudier la présence du complexe du complexe 8-TAFs (Figure 54 B).
- 2) une IP anti-TAF10 sur le surnageant d'une IP anti-TBP permettrait d'étudier la présence du complexe 7-TAFs (Figure 54 C).

Des anticorps anti-TRRAP, TAF10 et TBP ont déjà été validés au sein de notre équipe et sont disponibles. Des approches de co-immunoprécipitations successives ont déjà été réalisées au sein de notre équipe pour mettre en évidence des complexes protéiques (Wieczorek et al., 1998; Yu et al., 2020a). Néanmoins, il est possible qu'en condition normale, l'assemblage de TFIID est très rapide et que par conséquent les complexes 8-TAFs et 7-TAFs isolés sont très peu abondants. De ce fait, ils ne seront peut-être pas détectés lors de l'analyser par spectrométrie de masse. De plus, il ne sera pas possible d'exclure que plusieurs complexes aient été élués après la dernière IP anti-TAF10 ou anti-TBP.

*La réorganisation du core-TFIID par le module TAF2-8-10 n'est pas une étape requise pour le recrutement des autres TAFs.*

Par exemple, l'élution de l'IP anti-TBP devrait contenir également SL1 et TFIIB, deux complexes dans lequel il est également présent. Si les expériences fonctionnent, l'ensemble des données *in vitro* et *in vivo* permettront de mieux comprendre l'assemblage de TFIID.

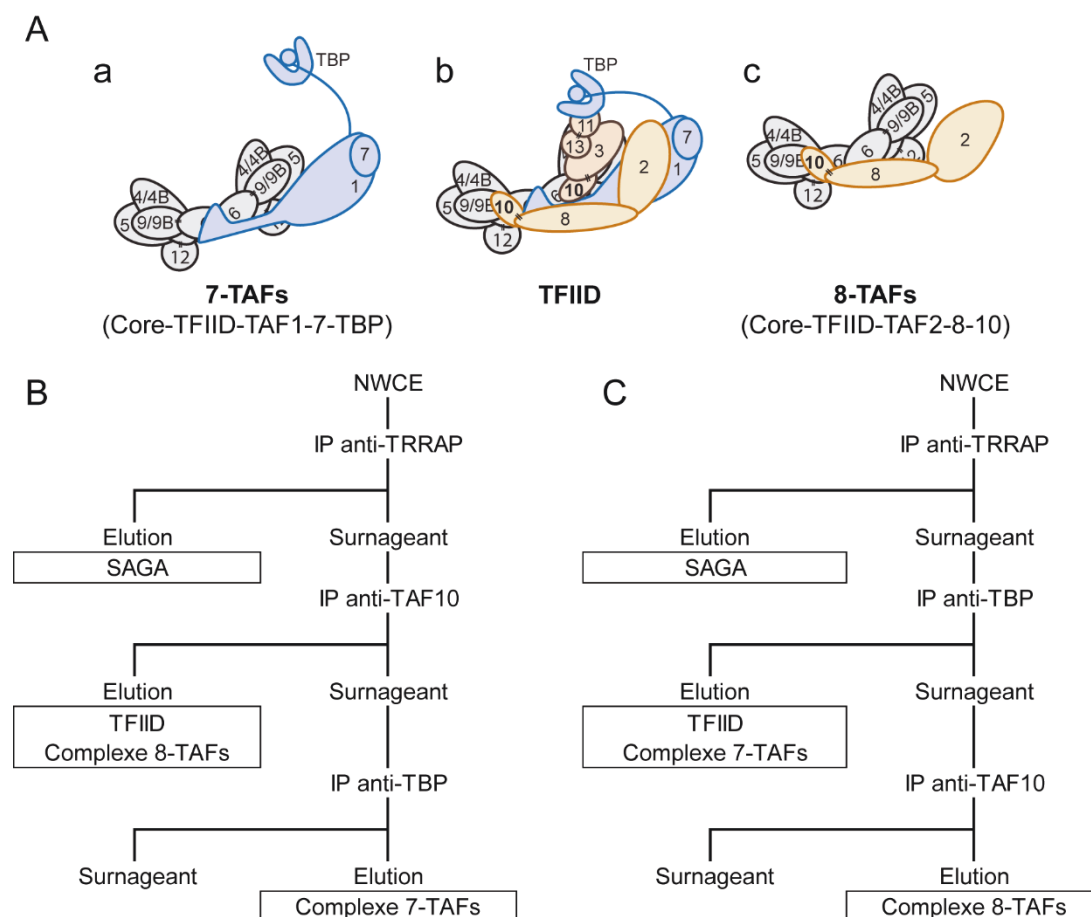


Figure 54 : Approche pour mettre en évidence *in cellulo* les complexes 7-TAFs et 8-TAFs. A) Schéma des complexes 7-TAFs (a), TFIID (b) et 8-TAFs (c). B-C) Approches de co-immunoprécipitation pour mettre en évidence les complexes 7-TAFs (B) et 8-TAFs (C) à partir de NWCE (Nuclear-enriched Whole Cell Extract).

## II. TAF10 joue un rôle important dans l'assemblage de TFIID en stabilisant ses partenaires.

Malgré le fait que d'autres TAFs peuvent intégrer le complexe core-TFIID en absence de TAF10, mes données IP-MS anti-TBP et anti-TAF12 indiquent que l'assemblage de TFIID en est gravement impacté. L'importance de TAF10 s'explique par le fait que la présence de TAF10 est nécessaire à la stabilisation de la protéine TAF8 après sa traduction. En effet, il a été précédemment montré que la déplétion de TAF10 entraîne la déplétion de TAF8 (Bardot et al., 2017; Kamenova et al., 2019). L'assemblage du module TAF2-8-10 est dépendante de TAF8 car il est lié par son HFD à TAF10



et par son extrémité C-terminal à TAF8. Enfin, TAF8 est requis pour l'intégration du module TAF2-8-10 au sein du core-TFIID (Demény et al., 2007b; Scheer et al., 2021; Trowitzsch et al., 2015). Ainsi, TAF10 en maintenant l'intégrité de TAF8 permet de former le core-TFIID. Sachant que TAF8 disparaît en absence de TAF10, il serait intéressant de regarder si TAF2 également disparaît en absence de TAF10. Ceci pourrait suggérer que TAF8 et/ou TAF8-10 est important pour la stabilité de TAF2.

En plus de TAF8, TAF10 a deux autres partenaires auxquels il est lié : la protéine SUPT7L et la protéine TAF3. Les analyses par western blot indiquent qu'en absence de TAF10, SUPT7L est absent tout comme TAF8. De ce fait, il est envisageable de penser que TAF3 disparaît également suite à la déplétion de TAF10. Cette hypothèse se repose sur le fait que TAF3 interagit avec TAF10 par des interactions HFDs tout comme SUPT7L et TAF8 (Herbst et al., 2021; Soutoglou et al., 2005). De plus, TAF3, TAF8 et SUPT7L sont issus d'une même protéine ancestrale (Antonova et al., 2019). De ce fait il est possible que cette caractéristique d'être stabilisé par TAF10 a été conservé au cours de l'évolution. Néanmoins, cette hypothèse demande à être vérifiée. Il a par exemple été rapporté que TAF3 existe non lié à TAF10 dans les hESC (Maston et al., 2012). Dans les mESC, il a été rapporté la présence de complexe TAF3-CTCF (Liu et al., 2011). Ainsi, en absence de TAF10 la présence de CTCF pourrait être suffisant pour maintenir TAF3 dans la cellule. Il serait donc intéressant de purifier et de valider un anticorps pour étudier la présence de TAF3 par western blot dans le but de regarder si en condition déplétée en TAF10 TAF3 disparaît ou non.

Des analyses par spectrométrie de masse sur des complexes TFIID après 'cross-link' ont montré que TAF3 contacte TAF1, TAF4, TAF6, TAF7, TAF9 et TAF12. Ainsi, il se pourrait que le recrutement de l'hétérodimère TAF3-10 au sein de TFIID est réalisé via TAF3. Des essais d'assemblage *in vitro* de TAF3-10, avec des versions de TAF3 mutées pour ces sites, aux complexes core-TFIID, 7-TAFs ou 8-TAFs permettraient de connaître l'importance de TAF3 dans l'assemblage de TAF3-10 à TFIID.

Enfin, au sein de TFIID, TAF10 se trouve proche de l'hétérodimère TAF11-13 (Chen et al., 2021a; Patel et al., 2018). Les analyses par spectrométrie de masse sur des complexes TFIID après 'cross-link' ont également montré des interactions entre TAF10 et TAF13 (Chen et al., 2021a; Scheer et al., 2021). Ainsi il est également possible que l'intégration de TAF11-13 au sein du complexe TFIID nécessite la présence de TAF10. De la même façon, l'assemblage *in vitro* de complexe TFIID avec des versions de TAF10 ou de TAF13 mutées pour ces sites de liaisons permettrait de connaître l'importance de TAF10 dans le recrutement de TAF11-13 à TFIID.

Pour valider ces résultats *in vitro* dans un modèle *in vivo*, il serait intéressant d'utiliser le modèle Drosophile. Cet organisme est peut-être le seul ayant un paralogue à drTAF10, nommé drTAF10B, s'intégrant spécifiquement au sein de drSAGA (voir introduction IV, 2) drTAF10 s'intègre quant-à-lui uniquement dans TFIID. En prenant ce modèle, il serait intéressant d'étudier quelles sont les différences entre drTAF10 et drTAF10B à l'origine de leur ancrage spécifique dans drTFIID et dans drSAGA

respectivement. Il sera tout d'abord nécessaire de vérifier que les séquences d'acides aminés de TAF10 sont conservées entre les mammifères et la drosophile. Si c'est le cas, l'utilisation du modèle drosophile permettra de mieux comprendre l'importance fonctionnelle des interactions entre TAF10 et ses partenaires TAF3, TAF8 dans l'assemblage de TFIID.

### III. TAF7 est impliqué dans le recrutement de TAF1 à TFIID

Contrairement à TAF10, TAF7 semble relativement peu important pour l'assemblage et la stabilité de TFIID. Mes données d'IP-MS indiquent qu'en absence de TAF7 un complexe TFIID déplété en TAF7, nommé *TAF7-less TFIID*, s'assemble. Néanmoins, les données d'IP-MS anti-TAF12 indiquent que TAF1 est deux fois moins détecté que les autres TAFs à jour 3. Il est intéressant de noter que mes données d'IP-MS anti-TAF12 montrent que le core-TFIID des mutants *RT710* et *RT10* se différencie par le fait qu'en présence de TAF7, TAF1 est associé au core-TFIID. Ainsi, il semble que l'intégration de TAF1 au TFIID en formation est facilitée par la présence de TAF7. Il avait déjà été suggéré que le recrutement de TAF7 au domaine DU3591 de TAF1 pourrait être important pour le repliement et la stabilisation de TAF1 (Bhattacharya et al., 2014). Ensemble, ces données suggèrent que TAF7, en participant au repliement et/ou la stabilisation de TAF1, favoriserait son recrutement au sein de TFIID. La mutation par CRISPR-Cas9 du domaine de TAF7 liant le DU3591 au sein des mESC permettrait de mieux comprendre l'importance de cette interaction dans l'assemblage de TAF1 à TFIID.

### IV. TAF1 est requis pour le recrutement optimal de TBP sur TFIID : possible mise en évidence d'un complexe TFTC contenant l'ensemble des sous-unités de TFIID à l'exception de TAF1 et TBP.

Les données d'IP-MS anti-TBP détecte le même niveau de TAF1 dans les mutants et les contrôles *RT7*, alors que l'IP-MS anti-TAF12 en détecte deux fois moins à jour 3 dans les mutants. Ces différences peuvent s'expliquer par le fait qu'en absence de TAF7 :

- 1) deux populations de complexes *TAF7-less TFIID* sont immunoprécipitées par TAF12 : l'une contenant TAF1 et TBP et l'autre ne contenant ni l'un ni l'autre.
- 2) lors de l'IP anti-TBP, uniquement les complexes *TAF7-less TFIID* contenant TBP et TAF1 sont immunoprécipités.
- 3) les complexes *TAF7-less TFIID* contenant TBP et TAF1 dans le mutant *RT7* sont présents à un niveau relativement identique à la présence de TFIID dans les contrôles *RT7*.

La présence de ces deux populations de complexes *TAF7-less TFIID* indique que le recrutement de TBP au niveau de TFIID requière la présence de TAF1. Il est intéressant de noter que l'existence d'un complexe TFIID ne contenant ni TAF1 ni TBP, nommé TFTC, a déjà été suggérée au sein des cellules HeLa. La réalisation d'IP anti-TBP sur l'élution d'une IP anti-TAF10 a permis de mettre en évidence un surnageant contenant TAF4, TAF5, TAF6, TAF7, TAF8, TAF10 et TAF12 ainsi que des sous-unités retrouvées dans SAGA : SUPT7L, SUPT3H, GCN5 et SAP130 (Demény et al., 2007b; Wieczorek et al., 1998). Ainsi, il est possible que le surnageant contenait en réalité deux complexes : d'une part le complexe SAGA et d'autre part un complexe constitué de l'ensemble des TAFs à l'exception de TAF1 et TBP. Je nommerai par la suite ce complexe TCTF.

D'autres approches d'immunoprécipitation ou d'analyse biochimique, seront nécessaire pour valider la présence de ce complexe. En particulier, la présence de TAF7 au sein de TFIID dépourvu en TAF1 reste à être prouvée, sachant que TAF1 est le seul partenaire de TAF7 au sein du complexe TFIID. Néanmoins, les analyses par spectrométrie de masse sur des complexes TFIID après 'cross-link' ont montré des interactions entre TAF7 et d'une part TAF2 et d'autre part TAF3 (Chen et al., 2021a; Scheer et al., 2021). Un complexe TAF7-11-13 a également été découvert dans le cytoplasme des cellules HeLa (Gupta et al., 2017). En reprenant l'approche de co-immunoprécipitation décrite précédemment (Figure 54 C), la réalisation d'une IP anti-TAF7 et anti-TBP sur l'élution de l'IP-TAF10 permettrait de mettre en évidence le complexe TFTC, ainsi que la présence de TAF7 dans ce complexe.

## V. Potentielle mise à jour du modèle d'assemblage de TFIID.

En conclusion, avec l'ensemble des hypothèses formulées ci-dessus, le modèle d'assemblage de TFIID peut être renouvelé de la manière suivante (Figure 55) :

- Après la formation du core-TFIID, le module TAF1-7-TBP ou le module TAF2-8-10 sont recrutés.
- Si le module TAF1-7-TBP s'intègre en premier, alors suit le recrutement du module TAF2-8-10, de l'hétérodimère TAF3-10 puis de TAF11-13 pour former le complexe TFIID. L'intégration de TAF2-8-10 pourrait être important pour stabiliser le complexe core-TFIID associé à TAF1-7-TBP. Il est également possible que le recrutement en premier de TAF1-7-TBP ne soit pas la voie d'assemblage la plus efficace.
- Si le module TAF2-8-10 s'intègre en premier, suit alors 1) soit le recrutement de TAF1-7-TBP, puis des autres sous-unités, formant le complexe TFIID, 2) soit le recrutement de TAF3-10, puis de TAF11-13 et de TAF7 formant le complexe TFTC. Il est possible que sur le complexe TFTC puisse

être recruté par la suite TAF1-TBP. Ceci pourrait expliquer en quoi TAF7 est important pour le recrutement de TAF1 au sein de TFIID. L'assemblage de TAF1 au sein du complexe TFTC déjà formé nécessiterait la présence de TAF7. Ceci suppose que lors de l'intégration de TAF1-TBP au sein de TFTC, TAF7 change de partenaire.

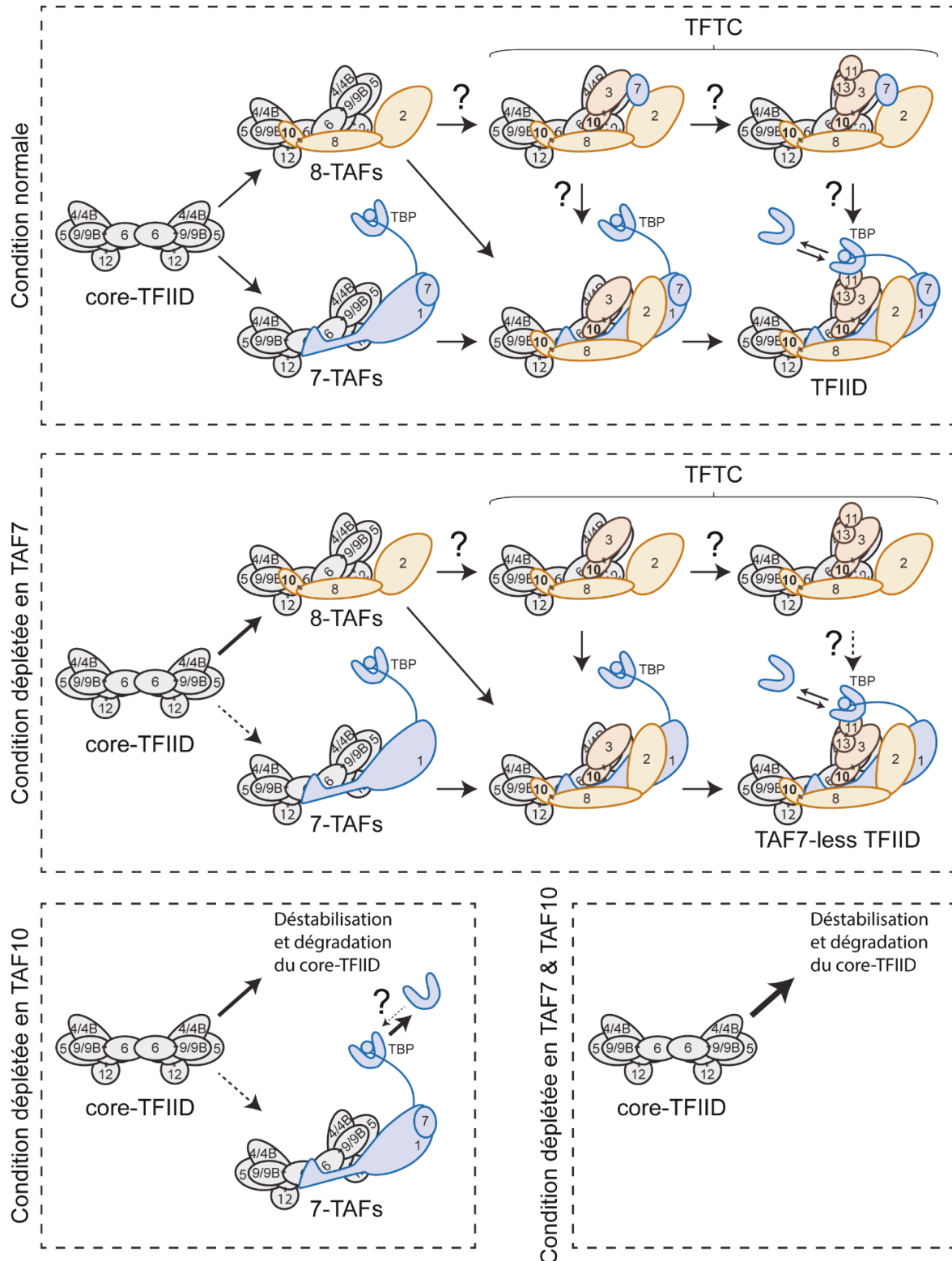


Figure 55 : Possibles différentes voies d'assemblage de TFIID. En haut, la formation du complexe TFIID et TFTC en conditions normales. En dessous est un récapitulatif de l'impact de la perte de TAF7 et/ou TAF10 sur la formation de ces complexes résultant de la formation de complexes partiels. La largeur des flèches indique la cinétique (k) d'assemblage et de désassemblage des sous-unités/modules : plus la flèche est large plus le k est important.

## VI. La transcription dépendante de l'ARN pol II est maintenue en absence de TFIID

Mes données montrent qu'en absence de TFIID, l'activité globale de l'ARN pol II n'est pas altérée. Que ce soit en conditions déplétées en TAF7 ou en TAF10 conduisant à la formation des complexes TFIID partiels, les phosphorylations de RBP1<sup>ser5</sup> et de RBP1<sup>ser2</sup> ne sont pas perturbées. De la même manière, le niveau d'incorporation de EU dans les cellules mutantes n'est pas perturbé par la déplétion de TAF7 ou TAF10. Ce marquage est spécifique de la transcription médiée par l'ARN pol II, car le traitement des cellules au flavopiridole et au triptolide, deux inhibiteurs spécifiques de l'ARN pol II (Bensaude, 2011), réduisant drastiquement le marquage.

A l'heure actuelle, nous ne pouvons pas exclure qu'au troisième et au quatrième jour de traitement, des complexes TFIID résiduels sont encore présents au sein des PICs potentiellement maintenus au niveau des promoteurs. De plus, à jour 3, les IP-MS anti-TBP détectent encore un niveau résiduel de TFIID dans la condition déplétée en TAF10. Ces complexes TFIID pourraient être suffisants pour initier la transcription. Pour vérifier ce point, la localisation sur la chromatine de TAF7 et TAF10 par CUT&TAG ou CUT&RUN sera vérifiée dans les mutants déplétés en TAF7 et en TAF10 respectivement à jour 3. Néanmoins, les analyses préliminaires par western blot d'extrait de chromatine ont montré que TAF7 et TAF10 ne sont pas présents sur la chromatine à partir du jour 3 en condition déplétée en TAF7 et TAF10 respectivement.

## VII. La double déplétion TAF7 et TAF10 affecte la transcription médiée par l'ARN pol II : de possibles effets indirects qui demandent de choisir le bon jour de traitement pour les analyses transcriptomiques.

Contrairement aux simples déplétions, la double déplétion TAF7 et TAF10 perturbe la transcription dépendant de l'ARN pol II à partir du quatrième jour de traitement. Bien que les phosphorylations de RBP1<sup>ser5</sup> et de RBP1<sup>ser2</sup> ne soient pas perturbées, la transcription de l'ARN pol II est limitée à des sous-compartiments nucléaires. L'incorporation de EU sera réalisée à jour 5 pour voir si cette transcription sub-nucléaire est maintenue dans le temps.

La perturbation de la transcription en condition double déplétions pourrait s'expliquer par des effets indirects encore non visible dans les conditions simples mutants. Il est possible que la double déplétion affecte plus rapidement l'assemblage de TFIID que les simples déplétions, résultant à une perte de TFIID plus rapide. L'analyse des IP-MS anti-TAF12 et anti-TBP indique que les sous-unités

du core-TFIID sont moins détectées en condition double déplétion qu'en condition déplétée en TAF10 à jour 3. Bien qu'en absence de TFIID, la transcription globale réalisée par l'ARN pol II est maintenue, on ne peut pas exclure que l'expression de certains gènes soit fortement impactée par l'absence de TFIID. La perturbation de l'expression de ces gènes pourrait par la suite conduire à la perturbation générale de la transcription médiée par l'ARN pol II. Par conséquent, la perte plus rapide de TFIID en condition double déplétion pourrait affecter plus rapidement la transcription de ces quelques gènes entraînant rapidement une perturbation globale de la transcription. Ainsi, l'analyse de l'incorporation de EU à jour 5 chez les simples mutants pourrait également montrer une perturbation générale de la transcription dépendante de l'ARN pol II.

De plus, la perturbation de l'assemblage de TFIID semble également avoir des effets sur la formation des autres machineries de transcription. Ainsi, les IP-MS anti-TBP en condition double déplétion à jour 4 ont montré que TBP est délocalisé de la machinerie de transcription pol I. La perturbation de la transcription pol I entraîne la perturbation de la production de ribosomes qui indirectement peut perturber profondément la transcription de l'ARN pol II.

Pour étudier le plus précisément possible le rôle de TFIID, il est donc important de choisir le bon jour de traitement pour faire les analyses transcriptomiques. A jour 4, le phénotype de l'ensemble des mutants est assez mineur. Néanmoins, la transcription commence à être fortement perturbée par l'absence de TFIID dans les doubles mutants, ceci étant peut-être dû à des effets indirects. A jour 3, l'ensemble des mutants présentent une forte déplétion de TFIID et ne contiennent majoritairement que des complexes partiels. Par conséquent, il a été conclu que pour étudier au mieux la fonction de TFIID et des complexes partiels contenant TAF7 ou TAF10 dans la transcription, les approches transcriptomiques telles que le TT-seq seront réalisées à jour 3.

### VIII. Rôle fonctionnel de TAF7 dans la régulation du couplage des premières étapes de la transcription : mise en place du PIC, initiation et élongation de la transcription.

A jour 6, les mutants *RT10* présentent un phénotype beaucoup plus sévère, alors que le phénotype des mutants *RT7* restent mineur. Ainsi, dans les mESC TAF7 semble être moins requis que TAF10. Néanmoins, après avoir été passées à jour 4, les mESC déplétées en TAF7 présentent une capacité très réduite à reformer des colonies. De plus, les embryons mutants *Taf7*<sup>-/-</sup> ne sont pas capables de se développer normalement. Au sein des mESC, dès le jour 3, TFIID laisse place à un complexe *TAF7-less TFIID* (Figure 55). Ensemble, ces données suggèrent que le complexe *TAF7-less TFIID* n'est pas capable de maintenir l'expression et/ou de mettre en place un programme d'expression génétique



Le complexe *TAF7-less TFIID* contient l'ensemble des TAFs qui permet à TFIID de reconnaître et lier les promoteurs, de charger TBP sur l'ADN et de recruter les autres GTFs (voir introduction, IV.6.). Néanmoins, TAF7 semble réguler différentes étapes de la transcription.

Au sein de TFIID, TAF7 est lié au niveau du domaine DU3591 de TAF1. Ce domaine DU3591 contient un sous-domaine *winged-helix* (WH) qui contacte DPE/MTE des promoteurs (Louder et al., 2016; Wang et al., 2014). TAF1 a également un autre domaine de liaison à l'ADN nommé ZnK (*CCHC zinc knuckle*). TAF1 est important pour l'expression des gènes du cycle cellulaire *cyclin A*, *D1* et *E*. La mutation du domaine WH et ZnK de TAF1 perturbe la liaison de TAF1 au promoteur des gènes *cyclin A* et *D1* conduisant à une réduction de leur niveau d'expression (Curran et al., 2018). Il est intéressant de noter qu'au sein de cellule HeLa, TAF7 inhibe également l'expression de ces deux cyclines : la réduction de la quantité de TAF7 via des approches de *siRNA* entraînent une augmentation du niveau d'expression des cyclines (Kloet et al., 2012). D'autres essais de transcription *in vitro* ont également montré que TAF7 diminue le taux de transcription (Gegonne et al., 2006). Ensemble, ces données suggèrent que TAF7 pourraient réguler la liaison de TAF1 au promoteur, et en conséquence le recrutement de TFIID au promoteur, ce qui *in fine* régule le niveau d'expression des gènes.

D'autres données ont montré que l'initiation de la transcription des gènes est corrélée avec la dissociation de TAF7 de TFIID (Gegonne et al., 2006; Kloet et al., 2012). Cette dissociation de TAF7 est induite par sa phosphorylation (Gegonne et al., 2006; Kloet et al., 2012). La phosphorylation de la sérine 264 de TAF7 perturbe notamment son interaction à TAF1. Ainsi, il a été suggéré que la phosphorylation de TAF7 est un point de contrôle dans la transition de l'assemblage du PIC à l'initiation de la transcription (Gegonne et al., 2006).

TAF7 serait phosphorylé par TAF1 (Gegonne et al., 2001, 2006; Weissman et al., 1998). Initialement, il a été suggéré que TAF1 avait deux modules kinases : le premier à l'extrémité N-terminal et le deuxième à l'extrémité C-terminal (Wassarman and Sauer, 2001). Cependant, il est maintenant connu que l'extrémité N-terminal forme en réalité le domaine TAND1 (Patel et al., 2018) qui n'a pas de fonction kinase (Bhattacharya et al., 2014). A ma connaissance, aucune information n'est disponible sur l'extrémité C-terminal de TAF1 chez les mammifères. Ainsi, la phosphorylation de TAF7 par l'extrémité C-terminal de TAF1 n'est pas à exclure.

Des essais de phosphorylation *in vitro* de CTD de RPB1 ont montré que la présence de TAF7 diminue le taux de phosphorylation de RPB1<sup>Ser5</sup> (Gegonne et al., 2008). Il a été rapporté que TAF7 inhibe l'activité kinase de TFIIH. Cette expérience a été réalisée en utilisant de l'ATP marqué radioactivement permettant de mettre en évidence le CTD phosphorylé. Il est également marqué que TAF7 est visible sur la membrane, suggérant que TAF7 est également phosphorylé dans le mixe réactionnel. Par conséquent, il est possible que la diminution de la phosphorylation de RPB1<sup>Ser5</sup> soit due à une compétition entre la phosphorylation du CTD et de TAF7 par CDK9 de TFIIH. Ensemble, ces données

suggèrent que TAF7 pourrait inhiber l'initiation de la transcription, en inhibant l'interaction complet et la stabilisation de TAF1 au promoteur, avant que TFIID soit présent dans le PIC.

TAF7 a également été montré *in vitro* comme associé au facteur d'élongation p-TEF $\beta$  au sein du complexe d'élongation (Gegonne et al., 2008). La réalisation d'IP-MS anti-TAF7 à partir d'extrait de mESC a permis de détecter p-TEF $\beta$  (Bardot, non publié). *In vitro* TAF7 semble également être phosphorylé par p-TEF $\beta$  (Gegonne et al., 2008). Il a également été montré que la déplétion de TAF4 dans les hépatocytes diminue la pause de l'ARN pol II et favorise l'élongation (Alpern et al., 2014). Le complexe S-TAF, composé de TAF1-7-11-13-TBP, a été montré *in vitro* comme capable de réguler la pause de l'ARN pol II (Fant et al., 2020). Ainsi, TAF7 pourrait également avoir un rôle dans la régulation de la pause de l'ARN pol II. Dans ce cas-ci, le rôle régulateur de TAF7 pourrait également dépendre de son niveau de sa phosphorylation.

Ainsi, dans en condition déplétée en TAF7, le couplage entre la formation du PIC, l'initiation de la transcription et l'élongation pourrait être dérégulé modifiant le taux et/ou l'efficacité de la transcription. L'analyse du rôle fonctionnel de la phosphorylation de TAF7 dans la régulation des différentes étapes de la transcription nécessitera de chercher et de muter spécifiquement les résidus de TAF7 phosphorylés par TFIID et p-TEF $\beta$ , et potentiellement TAF1. Par la suite des analyses de l'initiation de la transcription et la pause de l'ARN pol II pourront être réalisées telles que des expériences de ChIP-seq contre l'ARN pol II. Ces expériences permettront de mieux comprendre l'importance de la phosphorylation de TAF7 dans les différentes étapes de la transcription.

D'après les données présentées ci-dessus, il est étonnant que la transcription ne soit pas perturbée au sein des mutant *RT7*. L'analyse du transcriptome naissant via le TT-seq permettra d'avoir une première indication sur le rôle de TAF7 dans la régulation de la transcription. Il est important de noter que contrairement aux autres TAF, la déplétion partielle de TAF7 au sein des mESC par approche *siRNA* ne perturbe pas la pluripotence des cellules (Pijnappel et al., 2013). Ainsi, il est possible que dans le contexte des mESC, TAF7 joue un rôle mineur dans la transcription. Si c'est le cas il sera nécessaire de développer un autre model pour étudier le rôle de TAF7 dans la transcription.

## IX. Rôle fonctionnel de TFIID dans l'activation et la répression de la transcription

La déplétion de TAF10 perturbe fortement l'assemblage de TFIID entrainant la formation d'un complexe 7-TAFs, constitué du core-TFIID associé à TAF1-7-TBP dès jour 3 (Figure 55). En condition normale, TAF1, TAF2, TAF3, TAF10 et TAF4/4B-TAF12 sont impliqués dans le recrutement de TFIID à la chromatine (voir introduction, IV.6). Le complexe 7-TAFs ne contenant pas TAF2, TAF3 et TAF10, son recrutement au niveau de la chromatine devrait être perturbé. Le complexe 7-TAFs ne contient



également pas l'hétérodimère TAF11-13. Lors du chargement de TBP dans le sillon mineur de l'ADN, TAF11-13 est l'unique point de contact entre TBP et le lobe A de TFIID suite à la dissociation de TAND1 et 2 de TBP (Patel et al., 2018) (voir introduction IV. 6). L'interaction de TAF11-13 à TBP est rompue après que TBP est été chargé dans le sillon mineur (Patel et al., 2018). Ceci peut indiquer que TAF11-13 est peut-être important pour maintenir TBP au niveau de l'ADN jusqu'à ce que ce dernier soit chargé dans le sillon mineur. Ainsi, en absence de TAF11-13, le complexe 7-TAFs ne pourrait pas efficacement charger TBP dans l'ADN. De plus, les données d'IP-MS anti-TBP indiquent que TBP n'est pas recruté ou maintenu au niveau du complexe 7-TAFs. L'ensemble de ces données indique que la transcription devrait être fortement altérée en absence de TAF10 à partir du jour 3.

Néanmoins, l'analyse de la phosphorylation de RPB1<sup>Ser5</sup>, RPB1<sup>Ser2</sup> et de l'incorporation de EU indique que l'activité globale de l'ARN pol II est maintenue à jour 3 et 4. De la même manière, la déplétion conditionnelle de TAF10 au sein du mésoderme embryonnaire n'affecte pas la transcription cyclique requise pour la somitogenèse (Bardot et al., 2017). Chez la souris adulte, la déplétion de TAF10 dans les progéniteurs de kératinocyte adulte ne perturbe pas leur prolifération et leur différenciation (Indra et al., 2005), indiquant là encore que la transcription n'est pas significativement perturbée. Néanmoins, les mESC mutantes *RT10* présentent un phénotype sévère : la croissance de leur population est fortement impactée, les cellules entrent en apoptose et ne sont pas capables de former des colonies après avoir été passées. Par conséquent, on peut supposer que bien que l'activité globale de l'ARN pol II est maintenue, l'expression et/ou la mise en place du programme génétique sont perturbées. En particulier, l'absence de TAF10 perturbe plus l'expression génétique que la perte de TAF7.

On peut supposer que l'incapacité du complexe 7-TAFs à recruter TBP pourrait entraîner la liaison ectopique de TBP au niveau de la chromatine conduisant à la transcription de régions intergéniques et/ou intragéniques. En condition *in vitro*, TBP est suffisant pour initier la transcription à partir de promoteurs contenant une boîte TATA (Donczem and Steven, 2018; Remboutsika et al., 2001). De plus, la présence de nucléosome ne perturbe pas la capacité de TBP à initier la transcription (Koster et al., 2014; Remboutsika et al., 2001). De ce fait, en absence du complexe TFIID, il est possible que TBP se lie à TFIIA, formant un complexe TBP-TFIIA initiant la transcription à partir de séquence TATA, de façon similaire à la transcription initiée par le complexe TBPL2-TFIIA durant la croissance ovocytaire (Yu et al., 2020a). Néanmoins, la liaison ectopique de TBP à l'ADN est normalement réprimée par MOT1 et NC2 (nom chez *S.cerevisiae*, orthologue à BTAF1 et au complexe DR1/DRAP1 respectivement) qui répriment la mise en place de transcriptions intergéniques (Butryn et al., 2015; Gomez-Navarro et al., 2016; Koster et al., 2014). De plus, les IP-MS anti-TBP ne détectent pas TFIIA que ce soit en présence ou absence de TFIIA. Ainsi, la perte de TAF10 n'augmente pas l'affinité de TFIIA pour TBP. Ceci peut s'expliquer par le fait que MOT1 interagit avec TBP aux mêmes zones de contact que TFIIA (Butryn et al., 2015; Ravarani et al., 2020). Enfin, il est probable que certains facteurs spécifiques aux ovocytes permettent le chargement de TBPL2 au niveau des promoteurs. Pour ces

raisons, il se peut que TBP ait un comportement très différent de TBPL2 en absence de TFIID. L'analyse du transcriptome naissant ainsi que l'étude de la localisation de TBP sur la chromatine en absence de TAF10 permettra d'étudier le possible rôle de TBP dans l'initiation d'une transcription aberrante.

Deux études différentes ont délété conditionnellement TAF4 (Alpern et al., 2014) ou TAF10 (Tatarakis et al., 2008) dans les hépatocytes de façon post-partum via le transgène *Alb-Cre*. Dans les deux cas, la transcription des gènes qui sont normalement exprimés après la naissance n'a pas été activée. De l'autre, la transcription des gènes qui sont normalement réprimés après la naissance n'a pas été stoppée. Enfin, la transcription des quelques gènes dont l'expression débute durant le développement fœtal et qui continue à l'être après la naissance n'a pas été perturbée par l'absence de TAF10 (Tatarakis et al., 2008). En conséquence, il a été suggéré que TFIID est principalement important dans l'activation ou la répression de la transcription. Il a également été montré qu'au niveau des gènes dont l'expression est maintenue ou réprimée après la naissance, TBP, TFIIB, le médiateur et l'ARN pol II sont maintenus au niveau des promoteurs en absence de TAF10 (Tatarakis et al., 2008). Au contraire, pour les gènes dont l'expression doit être activée après la naissance, ces différents facteurs n'étaient pas retrouvés aux promoteurs des gènes en absence de TAF4 ou de TAF10 (Alpern et al., 2014; Tatarakis et al., 2008). Ainsi, la capacité de TFIID à activer ou réprimer la transcription semble se reposer sur sa capacité à mettre en place ou à dissocier le PIC au niveau des promoteurs. Il a été rapporté également que le recrutement de TFIID au niveau des gènes dont l'expression est activée après la naissance entraîne le recrutement du facteur de transcription HNF4A qui est un régulateur majeur de l'expression génétique des hépatocytes (Alpern et al., 2014). Ainsi, en plus de mettre en place le PIC, TFIID pourrait permettre de recruter un ensemble de facteurs de transcription qui permettent de maintenir l'expression de programme génétique.

Ce modèle pourrait coïncider avec mes données. Les cellules mutantes déplétées en TAF10 sont incapables de former de nouvelles colonies après avoir été passées. Ce phénotype pourrait indiquer que le programme d'expression génétique ne peut être modifié face à un stress ou un changement de condition de culture en absence de TAF10. Ce modèle pourrait également expliquer pourquoi la perte de TAF10 affecte la formation du bourgeon des membres mais pas la somitogenèse. Dans le modèle *Tg(T-Cre/+); Taf10<sup>fl/fl</sup>*, l'expression de la CRE est initiée peu après 6 jours de développement (E6) (Perantoni et al., 2005), menant à la perte de TAF10 vers E8,5 (Bardot et al., 2017). Le début de l'expression de la CRE peut être supposé par le fait que la notochorde n'est que partiellement déplétée en TAF10 (Bardot et al., 2017). La notochorde est produite à partir d'un ensemble de progéniteur commun au mésoderme, mais il se différencie de celui-ci vers E6,25 (Balmer et al., 2016). Si on assume que la perturbation de l'assemblage de TFIID par l'absence de TAF10 au sein du mésoderme suit la même cinétique que la perte de TFIID au sein des mutants mESC *RT10*, on peut supposer que TFIID est significativement perdu et remplacé par le complexe 7-TAFs dans les cellules mésodermique à partir de E9.

La morphogenèse des membres antérieurs est initiée qu'à partir de E9 (Bruneau et al., 1999; Tickle, 2015). Elle débute par l'expression de *Tbx5* vers E9 qui va induire par la suite *Fgf10* et *Fgf8*, tous deux mettant en place l'AER (*Apical Ectodermal Ridge*). L'AER est important pour la croissance des bourgeons de membre et leur développement par la suite (Tickle, 2015). Dans les embryons mutants conditionnels *Taf10*<sup>-/-</sup>, *Tbx5* est faiblement exprimé à E9 mais pas à E9,5, indiquant que son expression n'est pas maintenue. Ni *Fgf10* ni *Fgf8* sont exprimés à E9,5 (résultat master, (Bardot et al., 2017)). Ainsi, ces données préliminaires suggèrent que TFIID est requis pour la mise en place du programme génétique nécessaire à la morphogenèse des membres antérieurs.

La somitogenèse est quand-à-elle initié vers E7 (Guibentif et al., 2021). Dans les embryons mutants conditionnels *Taf10*<sup>-/-</sup>, la somitogenèse est maintenu à E9,5 (Bardot et al., 2017). Il est possible qu'à E7, TFIID soit encore présent dans les cellules mésodermiques et serait capable de mettre en place le programme d'expression génétique nécessaire à la somitogenèse. Le maintien de la somitogenèse à E9.5 en absence de TAF10 (Bardot et al., 2017) indiquerait que TFIID n'est pas requis pour maintenir l'expression de ce programme génétique.

Il est important de noter qu'après avoir été mise en place, une prolifération importante est impliquée dans la somitogenèse. Par conséquent, la reprise de l'expression du programme génétique après la division cellulaire suggère que certains facteurs permettent de réassembler le PIC en absence de TFIID dans les cellules filles.

## X. L'induction de la dégradation des sous-unités de TFIID régule la mise en place des programmes d'expressions génétiques.

Le rôle de régulateur de TFIID dans la mise en place d'un programme génétique a également été mis en lumière par le fait la dégradation de certaines TAFs et TBP sont impliqués dans certains processus cellulaires. Ces dégradations sont orchestrées par l'activation de voie de signalisation. Ainsi, l'activation de la voie de signalisation des TGF- $\beta$  induit l'ubiquitination de TAF7 par TRIM26 entraînant sa dégradation dans les cellules de glande mammaires murines (Nakagawa et al., 2017). Dans ce modèle cellulaire, la réduction de TAF7 permet d'inhiber la prolifération cellulaire et de favoriser l'induction de l'apoptose (Nakagawa et al., 2017). La détermination et la différenciation cellulaire font également appel la régulation de la quantité de certaines TAFs. La différenciation des cellules musculaires murines C2C12 entraine la dégradation de TAF4 et de TBP (Perletti et al., 2001). Dans ce contexte, la différenciation optimale des cellules nécessite que TBP soit ubiquitiné par *Huwei1* qui induit sa dégradation (Li et al., 2015).

La différenciation des cellules F9 en endoderme primitif par traitement à l'acide rétinoïque, entraine également la dégradation de TAF4 et de TBP (Perletti et al., 2001). Le maintien ectopique de TAF4

perturbe le processus de différenciation (Perletti et al., 2001). *A contrario*, TAF3 est requis pour la différenciation en endoderme primitif et définitif (Liu et al., 2011). Il est intéressant de noter que TAF3 et TAF4 ont été indiqués comme nécessaires pour le maintien de la pluripotence des mESC (Pijnappel et al., 2013). Ensemble, ces résultats indiquent que l'induction de la différenciation cellulaire demande la dégradation de certaines sous-unités de TFIID. Ces dégradations pourraient entraîner à la formation de complexes TFIID partiels impliqués dans la régulation de l'expression génétique.

Dans cette idée, il a été montré que durant la différenciation des mESC en progéniteurs neuronaux, LOXL2 oxyde TAF10 méthylé entraînant sa dégradation (Iturbide et al., 2015). La perte de TAF10 est importante pour réprimer l'expression des gènes de pluripotence et activer de façon optimale le programme d'expression génétique nécessaire à la différenciation (Iturbide et al., 2015).

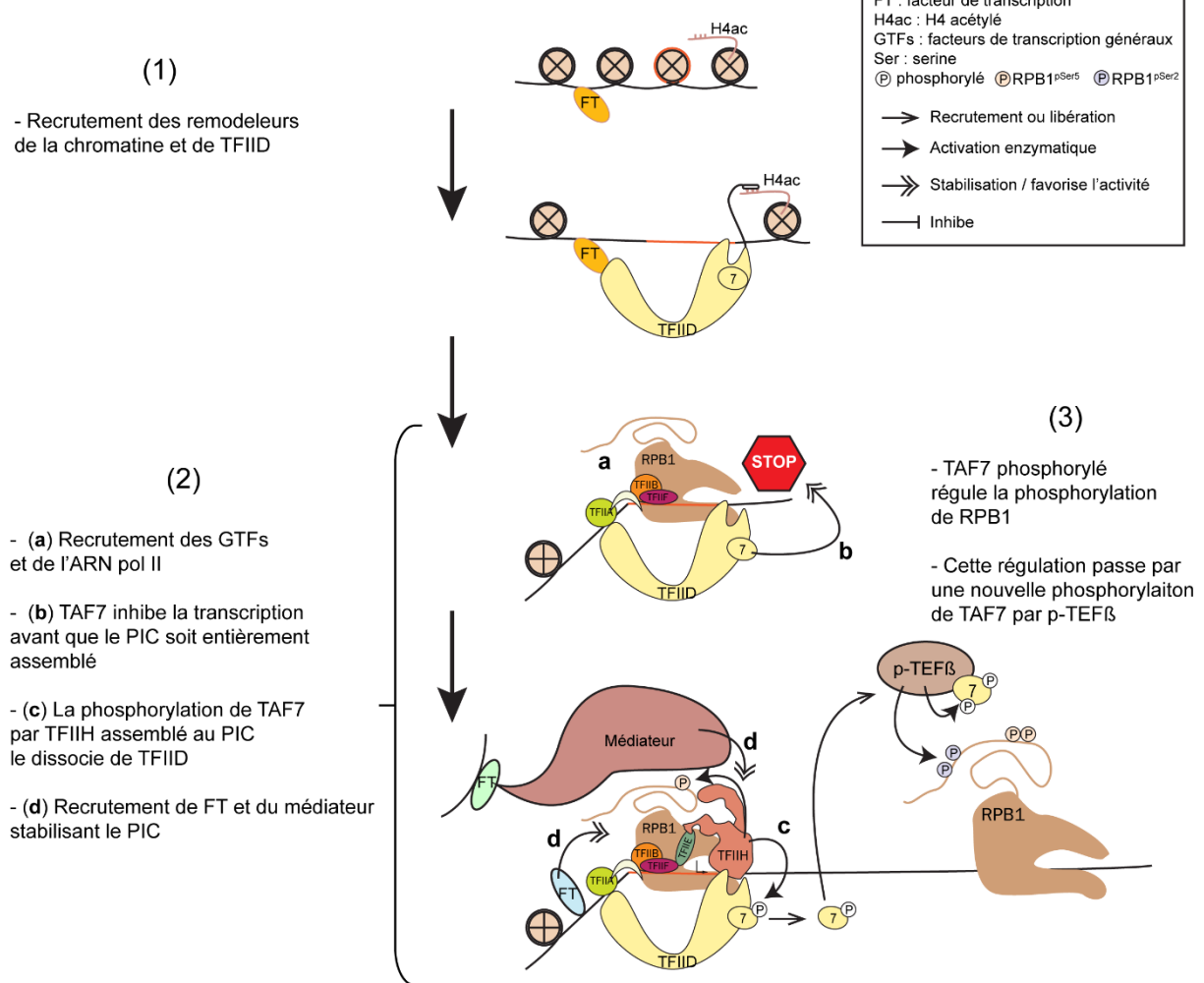
## XI. Potentielle mise à jour du modèle expliquant l'importance de TFIID dans la transcription.

Il est généralement admis que TFIID est requis pour la transcription, permettant le recrutement de l'ARN pol II au niveau des gènes. L'ensemble des nouvelles découvertes décrites ci-dessus permettent de détailler son rôle (Figure 56).

- 1) Dans un premier temps, le recrutement de certains facteurs de transcription et la présence de marques épigénétiques (H3K9me3, H4ac) permettent de recruter TFIID au niveau des promoteurs.
- 2) Lors de l'activation de la transcription, TFIID recrute les GTFs et l'ARN pol II au niveau du promoteur conduisant à l'initiation de la transcription. La mise en place du PIC pourrait être régulée par la modification post-traductionnelle des TAFs, comme expliqué dans le chapitre de l'introduction IV.6 avec l'exemple de la méthylation de TAF10.  
TFIID recrute également au niveau du PIC des facteurs de transcription, le médiateur et/ou des coactivateurs, permettant de mettre en place l'épigénome et la reprise de la transcription après la division cellulaire.
- 3) TFIID permet de réguler le couplage entre mise en place du PIC, initiation et élongation de la transcription. En particulier, TAF7 jouerait un rôle important dans ce processus. D'une part il inhiberait l'initiation de la transcription avant que le PIC soit complètement formé, et d'autre part régulerait la pause de l'ARN pol II. La régulation de ces différentes étapes par TAF7 se ferait par son niveau de phosphorylation catalysé par TFIIH et p-TEFβ.

- 4) L'activation de voies de signalisation, dans le but de réprimer l'expression de certains gènes, mène au recrutement de certains facteurs au niveau de TFIID entraînant sa modification post-traductionnelle, sa recomposition ou sa restructuration. D'une part, les TFIID modifiés induisent alors le désassemblage du PIC et empêchent le recrutement de d'autres TFIID et PIC fonctionnels. D'autre part, les TFIID modifiés peuvent recruter, en fonction des stimuli, certains coactivateurs. Dans le premier cas, ces coactivateurs réactivent la transcription des gènes si besoin, en retirant entre autres les TFIID modifiés des promoteurs. Dans un deuxième cas, les coactivateurs remodelent la chromatine et posent des marques épigénétiques héritées par les cellules filles et réprimant la transcription sur le long terme.

### Activation de l'expression génétique



### Répression de l'expression génétique

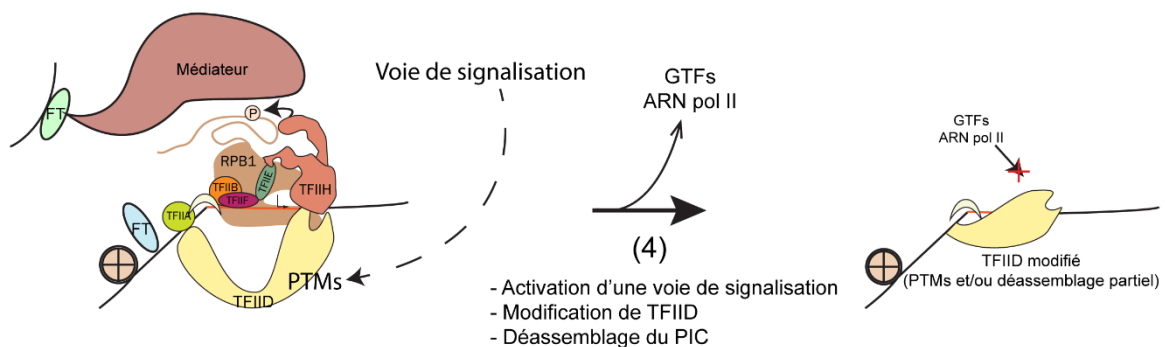


Figure 56 : Modèle du rôle de TFIID dans l'activation et la répression de la transcription ainsi que dans la régulation du couplage entre assemblage du PIC, initiation et élongation de la transcription.

## XII. Expériences pour l'étude de l'importance fonctionnelle de TFIID, et de ses complexes partiels, dans la mise en place et l'expression d'un programme d'expression génétique particulier

Le modèle hypothétique décrit ci-dessus (Figure 56) indique que TFIID serait important pour mettre en place le programme d'expression génétique. Dans le but de mieux comprendre l'importance fonctionnelle de TFIID, il serait intéressant de dépléter TAF7 et/ou TAF10 au cours de la différenciation des mESC. Dans mon modèle cellulaire, 3 jours sont nécessaires entre l'induction de la délétion et la perte de TFIID. De ce fait, des protocoles pourraient être établis pour coordonner la perte de TFIID en fonction des étapes de différenciation. Ainsi, il sera possible d'étudier plus en détail l'importance de TFIID, et des complexes partiels formés en absence de TAF7 ou de TAF10, lors de la mise en place des programmes d'expression génétique requis pour la différenciation.

Une autre solution serait de créer des lignées de mESC dans lesquels les protéines TAFs d'intérêts sont conditionnellement dégradées via une approche degron (Nishimura et al., 2009). Cette approche nécessite d'éditer le génome d'un modèle cellulaire pour que 1) la protéine d'intérêt soit fusionnée au domaine AID (*auxin-inducible degron*), 2) les cellules expriment le transgène *Tir1*. Le traitement de ces cellules avec l'hormone de plante auxine induit l'ubiquitination de l'AID par TIR1. Cette ubiquitination conduit à la dégradation de la protéine par le protéasome. Cette dégradation à l'intérêt d'être rapide car les protéines fusionnées à l'AID ne sont généralement plus détectables après 24 heures de traitement. Sachant que la différenciation est généralement un processus rapide qui se déroule en quelques jours, la possibilité de dépléter la quasi-totalité des protéines en 24 heures permettrait de plus facilement étudier le rôle de TFIID dans l'établissement des programmes d'expression génétique. Cette stratégie a également l'intérêt de pouvoir stopper la dégradation de la protéine juste par l'arrêt du traitement à l'auxine.

Cependant, le système degron fonctionne en ubiquitinant et en dégradant les protéines d'intérêts. Le recrutement et la dégradation de la protéine d'intérêt ubiquitinée au protéasome devrait également entraîner le recrutement et la dégradation de l'ensemble du complexe. Le système Cre-Lox quant-à-lui entraîne une diminution progressive de l'expression de la protéine d'intérêt ce qui *in fine* perturbe l'assemblage du complexe. Par conséquent, les deux techniques ne sont pas équivalentes en tout point et pourrait donner des résultats très différents. Par conséquent il sera nécessaire de réétudier la cinétique de la perte de TFIID et la composition des potentiels complexes partiels formés suite à l'induction du degron avant tout autres expériences. Par la suite, l'étude du transcriptome naissant et de la localisation des TAFs et des GTFs sur le génome au cours de la différenciation après avoir activé le degron permettrait de mieux comprendre le rôle de TFIID dans l'établissement des programmes d'expression génétique.



Une question importante est : si TFIID n'est pas requis pour la transcription après que celle-ci est été activée, comment se refait l'initiation de la transcription en son absence ? Il a été montré que certains GTFs, dont TFIID, reste au promoteur des gènes après que l'ARN pol II soit entrée en élongation (Joo et al., 2017). Il a été suggéré que leur présence favoriserait la ré-initiation de la transcription, c'est-à-dire le recrutement d'une nouvelle ARN pol II après que la précédente soit entrée en élongation. L'assemblage du PIC permet de recruter le médiateur (Sun et al., 2021) ainsi que des facteurs de transcription (Alpern et al., 2014) au niveau des promoteurs. Néanmoins TFIID ne serait pas requis pour maintenir la transcription après que celle-ci ait été activée. Ensemble ces données suggèrent que TFIID permettrait de mettre en place un ensemble de complexes au niveau des promoteurs, sous la forme d'un supra-complexe de ré-initiation, qui permettrait de réinitier la transcription sans sa présence (Tatarakis et al., 2008).

Pour étudier cette hypothèse, il serait intéressant de trouver lors d'un processus de différenciation *in vitro* trois catégories de gènes similaires à celles trouvées dans les études (Alpern et al., 2014; Tatarakis et al., 2008) : une nécessitant d'être réprimée par TFIID, une deuxième nécessitant d'être activée par TFIID et une troisième dont la perte de TFIID n'affecte pas leur transcription. Ces gènes seraient découverts par les analyses transcriptomiques réalisées en absence de TFIID lors de la différenciation cellulaire. A partir de cette découverte, l'analyse des facteurs présents au niveau des promoteurs de ces gènes par des expériences de TurboID (Branon et al., 2018) permettrait de mettre en évidence un ensemble de facteurs qui serait recruté aux promoteurs durant l'activation de la transcription et qui y serait maintenu après la perte de TFIID. La mise en évidence de ces facteurs permettrait de mieux comprendre le rôle de TFIID dans la mise en place d'un complexe de ré-initiation au niveau des promoteurs des gènes déjà en transcription.

### XIII. Les sous-unités TAFs sont partagées avec d'autres complexes que TFIID ou peuvent avoir des fonctions indépendantes à TFIID, complexifiant l'étude de la fonction TFIID

#### 1. La perte de TAF10 perturbe l'assemblage de SAGA : quelle est l'impact de la perte de SAGA, par rapport à la perte de TFIID, sur la transcription ?

L'étude de la fonction de TFIID est complexe dû au fait que certaines de ses sous-unités sont partagées avec d'autres complexes et peuvent être impliquées dans des fonctions indépendantes à TFIID. TBP est par exemple présent dans les trois machineries de transcription (voir introduction, III.3). Chez les mammifères, TAF9/9B, TAF10 et TAF12 sont également présents dans le complexe SAGA (voir introduction, IV.5). TAF10 en stabilisant SUPT7L est important dans l'assemblage de SAGA. SUPT7L est requis pour intégrer TAF10 au sein de SAGA (Fischer et al., 2021b). Mes données d'IP-MS



SUPT20H indiquent que SUPT7L est important pour l'intégration de SUPT20H au sein du corps SAGA. SUPT20H est lui-même impliqué dans le recrutement de TRRAP, du module SPL et DUB au sein de SAGA (Herbst et al., 2021; Nagy et al., 2009). Les résultats préliminaires ont montré que la perte de SAGA au sein des mESC n'affecte pas le niveau global d'acétylation et d'ubiquitination des histones (Bardot, P. non publiée). De la même manière, la perte de SUPT7L n'affecte pas le niveau d'acétylation des histones (Fischer et al., 2021b). Cependant, le module HAT de SAGA est partagé avec le complexe ATAC (Helmlinger and Tora, 2017). D'autres complexes ayant une activité de débubiquitination fonctionnent de façon redondante au module DUB de SAGA, tels que les complexes composés de USP51 et USP27x (Helmlinger and Tora, 2017). Par conséquent ces complexes pourraient compenser la perte de SAGA.

SAGA et ATAC sont impliqués dans l'expression du programme génétique de pluripotence des mESC. SAGA est important pour la mise en place du programme pendant que ATAC est important pour son maintien (Fischer et al., 2021b). L'étude fonctionnelle de SAGA et ATAC a été complexifiée par le fait que la présence ou non des deux inhibiteurs Chir et PD (2i) dans le milieu 2i+lif+serum modifie le phénotype des mESC après la déplétion des sous-unités d'intérêts (voir introduction, IV.5.ii). De ce fait, SAGA et ATAC sont deux complexes dont la fonction est sensible aux stimuli de milieu extérieur. L'ensemble de mon étude a été réalisé en cultivant mes cellules dans le milieu 2i+lif+serum, milieu dans lequel la perte de SAGA n'a pas ou peu d'effet sur le phénotype des mESC (Fischer et al., 2021a; Fischer et al., 2021b, Hisler et al.). Il sera cependant important de croiser les données du transcriptome naissant des mESC *Supt7l*<sup>-/-</sup> aux données générées à partir des cellules déplétées en TAF10. Ceci permettra de connaître quel est la part de la perte de SAGA et de TFIID dans les différences transcriptionnelles visibles.

## 2. TAF10 est potentiellement dans un autre complexe contenant TBPL1

Mes analyses de chromatographie suggèrent qu'au sein des mESC TAF10 est présent dans un complexe différent de TFIID dans lequel TBPL1 est présent. TBPL1 est recruté au niveau de l'ensemble des promoteurs des gènes exprimés dans les mESC (Kwan et al., 2021). La perte de TBP n'affecte pas la transcription médiée par l'ARN pol II dans les mESC (Kwan et al., 2021). La perte de TAF10 pourrait de ce fait perturber le recrutement de TBPL1 au niveau des promoteurs. S'il est avéré que TBPL1 est impliqué dans la transcription dans les mESC et que TAF10 est requis pour le recrutement de TBPL1, alors la perte de TAF10 devrait perturber la transcription dépendante de TBPL1. La caractérisation de la composition et de la fonction du complexe contenant TBPL1 pourraient être étudiées par la réalisation d'IP-MS anti-TBPL1 et l'étude d'une lignée mESC déletée pour *Tbpl1*. Ceci permettra également de comprendre la fonction de TBP et de TBPL1 dans la transcription dans les mESC.

### 3. Possible présence et rôle du complexe TFTC dans la transcription

Le complexe TFTC est également un complexe composé de nombreuses TAFs, dont TAF10. Il a été initialement mis en évidence par une suite d'immunoprécipitation (Wieczorek et al., 1998). Néanmoins, il s'est révélé par la suite que la fraction purifiée contenait également le complexe SAGA, qui était alors non connu (Brand et al., 1999; Demény et al., 2007b; Wieczorek et al., 1998). Cette fraction est capable d'induire une transcription à partir de promoteurs contenant ou non une boîte TATA en condition *in vitro* (Wieczorek et al., 1998). S'il est avéré que TFTC permet d'initier ou de réguler directement ou indirectement la transcription, la déplétion de TAF10 affecte également la présence et par conséquent la fonction de TFTC dans la cellule. Il sera nécessaire de mieux caractériser le complexe TFTC avant d'hypothétiser son potentiel rôle dans la transcription. De nouvelles études de sa composition par la réalisation d'IP permettra de mieux caractériser sa composition.

### 4. TAF7 est impliqué dans des fonctions indépendantes à TFIID

TAF7 est également impliqué dans des complexes indépendants de TFIID. TAF7 régule l'activité acétyltransférase de CIITA impliquée dans l'expression des protéines des complexes majeurs d'histocompatibilités (MHC, *Major Histocompatibility Complex*) (Devaiah et al., 2010). L'expression de ces gènes a été montrée comme indépendante de TAF1 (Howcroft et al., 2003; Raval et al., 2001), indiquant que TFIID n'est pas impliqué dans l'expression de ces gènes. Ainsi la déplétion de TAF7 pourrait également affecter une transcription initiée de façon indépendante à TFIID.

## XIV. Conclusion

Ma thèse s'est concentrée sur la caractérisation *in cellulo* de l'assemblage de TFIID et de sa fonction dans la transcription médiée par l'ARN pol II. Dans un premier temps, ce travail a permis de valider dans un modèle *in cellulo* le modèle d'assemblage de TFIID. En particulier, le rôle essentiel de TAF10 dans l'assemblage de TFIID à partir du core-TFIID a été validé. Mon travail a permis également de proposer une mise à jour de ce modèle d'assemblage. En particulier, la restructuration du core-TFIID imposée par l'intégration du module TAF2-8-10 n'est pas requis pour le recrutement des autres sous-unités. De ce fait, l'intégration du module TAF1-7-TBP au core-TFIID conduit à la formation d'un complexe 7-TAFs. Néanmoins, l'intégration du module TAF2-8-10 semble stabiliser le complexe TFIID en formation et favoriser le recrutement des autres sous-unités. Enfin, TAF7 favorise l'intégration du module TAF1-7-TBP au sein de TFIID, vraisemblablement en participant au repliement et/ou en maintenant l'intégrité de TAF1. Enfin, mes données suggèrent également qu'un complexe TFTC, constitué du core-TFIID associé au moins au module TAF2-8-10 et TAF7 mais ne contenant pas TAF1 ni TBP, existe au sein des mESC.

Dans un deuxième temps, mon travail a permis de montrer que TFIID n'est pas requis pour maintenir l'activité globale de l'ARN pol II. Les données présentes dans la littérature suggèrent que TFIID a plutôt un rôle pour l'activation de la transcription en induisant la primo-transcription. Par la suite, TFIID ne serait plus requis pour la transcription, mais resterait néanmoins au sein du PIC. L'activation de certaines voies de signalisation induirait la mise en place de PTMs sur certains TAFs, pouvant induire son désassemblage partiel, ce qui conduirait au désassemblage du PIC et l'inhibition de la transcription. Par conséquent, TFIID serait essentiel pour réguler la mise en place des programmes d'expression génétique en fonction des signaux intra et extra cellulaire. Enfin, TFIID, via TAF7, permettrait de coupler finement la mise en place du PIC, l'initiation de la transcription et la pause de l'ARN pol II.

Certaines études pointent le rôle des protéines TAFs dans certains cancers. Par exemple, la modification de la stœchiométrie des TAFs au sein des cellules cancéreuses ainsi que leur interaction avec des oncogènes favorisent le développement et l'évolution de cancers. (Le Gallo et al., 2017; Kalogeropoulou et al., 2010; Voulgari et al., 2008; Wardell et al., 2013; Xu et al., 2019; Zeng et al., 2018). De plus, la mutation des sous-unités de TFIID est impliquée dans le développement de maladies congénitales et des syndromes neuro-dégénératives (Cheng et al., 2020; El-Saafin et al., 2018). Par exemple, notre équipe a récemment étudié la maladie d'une petite fille présentant un retard global dans son développement ainsi qu'une déficience intellectuelle grave (El-Saafin et al., 2018). Cette jeune fille présente une mutation du gène *Taf8* faisant apparaître un autre site d'épissage conduisant à un décalage du cadre de lecture durant la traduction. Les fibroblastes issus de cette patiente présentent une quasi absence de protéines TAF8. Cette découverte fut surprenante car la délétion zygotique de *Taf8* chez la souris est létale à l'état blastocyste (Voss et al., 2000). Comme attendu, l'absence de TAF8 perturbe l'assemblage de TFIID. Néanmoins, l'activité globale de l'ARN pol II n'est pas perturbée dans les fibroblastes de la patiente. Ce dernier point ressemble aux phénotypes de mes propres mESC mutantes dans lesquelles la déplétion TAF10, conduisant à la perte de TAF8 entre autres, empêche l'assemblage de TFIID mais n'affecte pas l'activité globale de l'ARN pol II. Par conséquent, l'étude du rôle de TFIID aura des conséquences aussi bien sur notre connaissance générale de l'expression des gènes que sur la compréhension de certaines maladies

---

# Références

---

---

# Références

---

- Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* *13*, 720–731.
- Agnelli, L., Bisognin, A., Todoerti, K., Manzoni, M., Taiana, E., Galletti, S., Cutrona, G., Gaffo, E., Bortoluzzi, S., and Neri, A. (2019). Expanding the repertoire of miRNAs and miRNA-offset RNAs expressed in multiple myeloma by small RNA deep sequencing. *Blood Cancer J.* *9*.
- Akhtar, W., and Veenstra, G.J.C. (2009). TBP2 is a substitute for TBP in *Xenopus* oocyte transcription. *BMC Biol.* *7*, 1–11.
- Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: A central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* *16*, 155–166.
- Alpern, D., Langer, D., Ballester, B., Le Gras, S., Romier, C., Mengus, G., and Davidson, I. (2014). TAF4, a subunit of transcription factor II D, directs promoter occupancy of nuclear receptor HNF4A during post-natal hepatocyte differentiation. *Elife* *3*, e03613.
- Anandapadamanaban, M., Andresen, C., Helander, S., Ohyama, Y., Siponen, M.I., Lundström, P., Kokubo, T., Ikura, M., Moche, M., and Sunnerhagen, M. (2013). High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nat. Struct. Mol. Biol.* *20*, 1008–1014.
- Antonova, S. V., Haffke, M., Corradini, E., Mikuciunas, M., Low, T.Y., Signor, L., van Es, R.M., Gupta, K., Scheer, E., Vos, H.R., et al. (2018). Chaperonin CCT checkpoint function in basal transcription factor TFIID assembly. *Nat. Struct. Mol. Biol.* *25*, 1119–1127.
- Antonova, S. V., Boeren, J., Timmers, H.T.M., and Snel, B. (2019). Epigenetics and transcription regulation during eukaryotic diversification: The saga of TFIID. *Genes Dev.* *33*, 888–902.
- Atanassov, B.S., Mohan, R.D., Lan, X., Kuang, X., Lu, Y., Lin, K., McIvor, E., Li, W., Zhang, Y., Florens, L., et al. (2016). ATXN7L3 and ENY2 Coordinate Activity of Multiple H2B Deubiquitinases Important for Cellular Proliferation and Tumor Growth. *Mol. Cell* *62*, 558–571.
- Balmer, S., Nowotschin, S., and Hadjantonakis, A.K. (2016). Notochord morphogenesis in mice: Current understanding and open questions. *Dev. Dyn.* *245*, 547–557.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* *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. *Mol. Cell* *68*, 130–143.e5.
- Bardot, P., Vincent, S.D., Fournier, M., Hubaud, A., Joint, M., Tora, L., and Pourquié, O. (2017). The TAF10-containing TFIID and SAGA transcriptional complexes are dispensable for early somitogenesis in the mouse embryo. *Development* *144*, 3808–3818.
- Barisic, D., Stadler, M.B., Iurlaro, M., and Schübeler, D. (2019). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* *569*, 136–140.
- Bartfai, R., Balduf, C., Hilton, T., Rathmann, Y., Hadzhiev, Y., Tora, L., Orban, L., and Müller, F. (2004). TBP2, a Vertebrate-Specific Member of the TBP Family, Is Required in Embryonic Development of Zebrafish. *Curr. Biol.* *14*, 593–598.
- Baumann, D.G., and Gilmour, D.S. (2017). A sequence-specific core promoter-binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes. 1–11.
- Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C., et al. (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol. Cell* *66*, 384–397.e8.
- Behe, M., and Felsenfeld, G. (1981). Effects of methylation on a synthetic polynucleotide: the B→Z transition in poly(dG-m5dC).poly(dG-m5dC). *Proc. Natl. Acad. Sci. U. S. A.* *78*, 1619–1623.
- Ben-Ari, Y., Brody, Y., Kinor, N., Mor, A., Tsukamoto, T., Spector, D.L., Singer, R.H., and Shav-Tal, Y. (2010). The life of an mRNA in space and time. *J. Cell Sci.* *123*, 1761–1774.

- Bensaude, O. (2011). Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* 2, 103–108.
- Bhattacharya, S., Lou, X., Hwang, P., Rajashankar, K.R., Wang, X., Gustafsson, J.Å., Fletterick, R.J., Jacobson, R.H., and Webb, P. (2014). Structural and functional insight into TAF1-TAF7, a subcomplex of transcription factor II D. *Proc. Natl. Acad. Sci. U. S. A.* 111, 9103–9108.
- Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013a). The architecture of human general transcription factor TFIID core complex. *Nature* 493, 699–702.
- Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013b). The architecture of human general transcription factor TFIID core complex. *Nature* 493, 699–702.
- 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 Dev.* 28, 1999–2012.
- Bose, D.A., and Berger, S.L. (2017). eRNA binding produces tailored CBP activity profiles to regulate gene expression. *RNA Biol.* 14, 1655–1659.
- Brand, M., Leurent, C., Mallouh, V., Tora, L., and Schultz, P. (1999). Three-dimensional structures of the TAF(II)-containing complexes TFIID and TFTC. *Science* (80-. ). 286, 2151–2153.
- Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N., and Ting, A.Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* 36, 880–898.
- Brooks, T.A., and Hurley, L.H. (2009). The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer* 9, 849–861.
- Brou, C., Chaudhary, S., Davidson, I., Lutz, Y., Wu, J., Egly, J.M., Tora, L., and Chambon, P. (1993). Distinct TFIID complexes mediate the effect of different transcriptional activators. *EMBO J.* 12, 489–499.
- 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* (80-. ). 292, 2333–2337.
- Bruneau, B.G., Logan, M., Davis, N., Levi, T., Tabin, C.J., Seidman, J.G., and Seidman, C.E. (1999). Chamber-specific cardiac expression of Tbx5 and heart defects in Holt- Oram syndrome. *Dev. Biol.* 211, 100–108.
- Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural Basis of Transcription: An RNA Polymerase II-TFIIB Cocystal at 4.5 Angstroms. *Science* (80-. ). 303, 983–988.
- Butryn, A., Schuller, J.M., Stoehr, G., Runge-Wollmann, P., Förster, F., Auble, D.T., and Hopfner, K.P. (2015). Structural basis for recognition and remodeling of the TBP:DNA:NC2 complex by Mot1. *Elife* 4, 1–22.
- Castillo, J., Lopez-Rodas, G., and Franco, L. (2017). Histone Post-Translational Modifications and Nucleosome Organisation in Transcriptional Regulation: Some Open Questions. *Atassi M. Protein Rev. Adv. Exp. Med. Biol.* 966.
- Cavalieri, V. (2021). The Expanding Constellation of Histone Post-Translational Modifications in the Epigenetic Landscape. *Gene* 12, 1–22.
- 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. *EMBO J.* 18, 4835–4845.
- Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA Polymerase II Is Phosphorylated at CTD Residue Serine-7. *Science* (80-. ). 318, 1780–1783.



- Chen, K., Wang, L., Yang, M., Liu, J., Xin, C., Hu, S., and Yu, J. (2010). Sequence signatures of nucleosome positioning in *caenorhabditis elegans*. *Genomics, Proteomics Bioinforma.* 8, 92–102.
- Chen, X., Qi, Y., Wu, Z., Wang, X., Li, J., Zhao, D., Hou, H., Li, Y., Yu, Z., Liu, W., et al. (2021a). Structural insights into preinitiation complex assembly on core promoters. *Science* (80-. ). 372.
- Chen, X., Yin, X., Li, J., Wu, Z., Qi, Y., Wang, X., Liu, W., and Xu, Y. (2021b). Structures of the human Mediator and Mediator-bound preinitiation complex. *Science* (80-. ). 372.
- Chen, X., Chen, X., Qi, Y., Wu, Z., Wang, X., Li, J., Zhao, D., Hou, H., and Li, Y. (2021c). Structural insights into preinitiation complex assembly on core promoters. *8490*, 49–51.
- Cheng, H., Capponi, S., Wakeling, E., Marchi, E., Li, Q., Zhao, M., Weng, C., Stefan, P.G., Ahlfors, H., Kleyner, R., et al. (2020). Missense variants in TAF1 and developmental phenotypes: Challenges of determining pathogenicity. *Hum. Mutat.* 41, 449–464.
- Cheung, A.C.M., and Cramer, P. (2012). A movie of RNA polymerase II transcription. *Cell* 149, 1431–1437.
- Chiu, T.P., Rao, S., Mann, R.S., Honig, B., and Rohs, R. (2017). Genome-wide prediction of minor-groove electrostatic potential enables biophysical modeling of protein-DNA binding. *Nucleic Acids Res.* 45, 12565–12576.
- Chong, J.A., Moran, M.M., Teichmann, M., Kaczmarek, J.S., Roeder, R., and Clapham, D.E. (2005). TATA-Binding Protein (TBP)-Like Factor (TLF) Is a Functional Regulator of Transcription: Reciprocal Regulation of the Neurofibromatosis Type 1 and c-fos Genes by TLF/TRF2 and TBP. *Mol. Cell. Biol.* 25, 2632–2643.
- Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* 78, 273–304.
- Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* 18, 407–422.
- Coleman, R.A., Qiao, Z., Singh, S.K., Peng, C.S., Cianfrocco, M., Zhang, Z., Piasecka, A., Aldeborgh, H., Basishvili, G., and Liu, W.L. (2017). p53 Dynamically Directs TFIID Assembly on Target Gene Promoters. *Mol. Cell. Biol.* 37, 1–19.
- Cormack, B.P., and Struhl, K. (1992). The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell* 69, 685–696.
- Cox, M.M., Doudna, J.A., and O'Donnel, M. (2015). *Molecular Biology : Principles and Practice - Second edition*.
- Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I., and Fakan, S. (2006). Chromosome territories - a functional nuclear landscape. *Curr. Opin. Cell Biol.* 18, 307–316.
- Curran, E.C., Wang, H., Thomas, R.H., Zheng, N., and Wang, E.H. (2018). Zinc knuckle of TAF1 is a DNA binding module critical for TFIID promoter occupancy. 1–13.
- Dana, H., Chalbatani, G.M., Mahmoodzadeh, H., Karimloo, R., Rezaiean, O., Moradzadeh, A., Mehmandoost, N., Moazzen, F., Mazraeh, A., Marmari, V., et al. (2017). Molecular Mechanisms and Biological Functions of siRNA. *Int. J. Biomed. Sci.* 13, 48–57.
- Dantanel, J.C., Murthy, K.G.K., Manjey, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* 389, 399–402.
- Dantanel, J.C., Quintin, S., Lakatos, L., Labouesse, M., and Tora, L. (2000). TBP-like factor is required for embryonic RNA polymerase II transcription in *C. Elegans*. *Mol. Cell* 6, 715–722.
- Darzacq, X., Shav-Tal, Y., De Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 14, 796–806.
- Demény, M.A., Soutoglou, E., Nagy, Z., Scheer, E., Jánosàzi, À., Richardot, M., Argentini, M., Kessler, P., and Tora, L. (2007a). Identification of a small TAF complex and its role in the assembly of TAF-containing complexes. *PLoS One* 2.



- Demény, M.A., Soutoglou, E., Nagy, Z., Scheer, E., Jánoszázi, À., Richardot, M., Argentini, M., Kessler, P., and Tora, L. (2007b). Identification of a small TAF complex and its role in the assembly of TAF-containing complexes. *PLoS One* 2, 25–28.
- Dergai, O., and Hernandez, N. (2019). How to Recruit the Correct RNA Polymerase? Lessons from snRNA Genes. *Trends Genet.* 35, 457–469.
- Dergai, O., Cousin, P., Gouge, J., Satia, K., Praz, V., Kuhlman, T., Lhôte, P., Vannini, A., and Hernandez, N. (2018). Mechanism of selective recruitment of RNA polymerases II and III to snRNA gene promoters. *Genes Dev.* 32, 711–722.
- Devaiah, B.N., Lu, H., Gegonne, A., Sercan, Z., Zhang, H., Clifford, R.J., Lee, M.P., and Singer, D.S. (2010). Novel functions for TAF7, a regulator of TAF1-independent transcription. *J. Biol. Chem.* 285, 38772–38780.
- Dieci, G., Fiorino, G., Castelnovo, M., Teichmann, M., and Pagano, A. (2007). The expanding RNA polymerase III transcriptome. *Trends Genet.* 23, 614–622.
- Donczem, R., and Steven, H. (2018). Mechanistic Differences in Transcription Initiation at TATA-less and TATA-containing promoters. *Mol. Cell. Biol.* 38, 1–14.
- 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 Interdiscip. Rev. RNA* 10, 1–18.
- Dynlacht, B.D., Hoey, T., and Tjian, R. (1991). Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* 66, 563–576.
- El-Saafin, F., Curry, C., Ye, T., Garnier, J.M., Kolb-Cheynel, I., Stierle, M., Downer, N.L., Dixon, M.P., Negroni, L., Berger, I., et al. (2018). Homozygous TAF8 mutation in a patient with intellectual disability results in undetectable TAF8 protein, but preserved RNA polymerase II transcription. *Hum. Mol. Genet.* 27, 2171–2186.
- Emre, N.C.T., Ingvarsdottir, K., Wyce, A., Wood, A., Krogan, N.J., Henry, K.W., Li, K., Marmorstein, R., Greenblatt, J.F., Shilatifard, A., et al. (2005). Maintenance of low histone ubiquitylation by Ubp10 correlates with telomere-proximal Sir2 association and gene silencing. *Mol. Cell* 17, 585–594.
- Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.
- Falender, A.E., Freiman, R.N., Geles, K.G., Lo, K.C., Hwang, K.S., Lamb, D.J., Morris, P.L., Tjian, R., and Richards, J.A.S. (2005a). Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID. *Genes Dev.* 19, 794–803.
- Falender, A.E., Shimada, M., Lo, Y.K., and Richards, J.A.S. (2005b). TAF4b, a TBP associated factor, is required for oocyte development and function. *Dev. Biol.* 288, 405–419.
- Fan, Y., Nikitina, T., Morin-Kensicki, E.M., Zhao, J., Magnuson, T.R., Woodcock, C.L., and Skoultschi, A.I. (2003). H1 Linker Histones Are Essential for Mouse Development and Affect Nucleosome Spacing In Vivo. *Mol. Cell. Biol.* 23, 4559–4572.
- 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., et al. (2020). TFIID Enables RNA Polymerase II Promoter-Proximal Pausing. *Mol. Cell* 78, 785–793.e8.
- Federico, A., Julian, Z., Jan, A., Jernej, U., and Luscombe, N.M. (2020). Intergenic RNA mainly derives from nascent transcripts of known genes. *BioRxiv* 1–19.
- Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R., and Bickmore, W.A. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.* 4.
- Fischer, V., Schumacher, K., Tora, L., and Devys, D. (2019). Global role for coactivator complexes in RNA polymerase II transcription. *Transcription* 10, 29–36.
- Fischer, V., Scheer, E., Lata, E., Morlet, B., Plassard, D., Vincent, S.D., Helmlinger, D., Devys, D., and

- Tora, L. (2021a). SUPT3H-less SAGA coactivator can assemble and function without significantly perturbing RNA polymerase II transcription in mammalian cells. *Biorxiv* 1–45.
- Fischer, V., Plassard, D., Ye, T., San Martin, B.R., Stierle, M., Tora, L., and Devys, D. (2021b). Two Related Coactivator Complexes SAGA and ATAC Control Embryonic Stem Cell Self-Renewal Through Two Different Acetyltransferase-Independent Mechanisms. *SSRN Electron. J.* 36, 109598.
- Fischle, W., Boo, S.T., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* 438, 1116–1122.
- Flores, O., Lu, H., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J. Biol. Chem.* 267, 2786–2793.
- Forero-Quintero, L.S., Raymond, W., Handa, T., Saxton, M.N., Morisaki, T., Kimura, H., Bertrand, E., Munsky, B., and Stasevich, T.J. (2021). Live-cell imaging reveals the spatiotemporal organization of endogenous RNA polymerase II phosphorylation at a single gene. *Nat. Commun.* 12, 1–16.
- Franklin, R.E. (1953). Molecular configuration in sodium thymonucleate. *Nature* 171, 740–741.
- 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. *Mol. Cell. Biol.* 25, 4638–4649.
- Fyodorov, D. V., Zhou, B.R., Skoultschi, A.I., and Bai, Y. (2018). Emerging roles of linker histones in regulating chromatin structure and function. *Nat. Rev. Mol. Cell Biol.* 19, 192–206.
- Le Gallo, M., Rudd, M.L., Urick, M.E., Hansen, N.F., Zhang, S., Lozy, F., Sgroi, D.C., Vidal Bel, A., Matias-Guiu, X., Broaddus, R.R., et al. (2017). Somatic mutation profiles of clear cell endometrial tumors revealed by whole exome and targeted gene sequencing. *Cancer* 123, 3261–3268.
- Gamarra, N., and Narlikar, G.J. (2021). Collaboration through chromatin: motors of transcription and chromatin structure. *J. Mol. Biol.* 433, 166876.
- Gazdag, E., Rajkovic, A., Torres-Padilla, M.E., and Tora, L. (2007). Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction* 134, 51–62.
- Gazdag, E., Santenard, A., Ziegler-Birling, C., Altobelli, G., Poch, O., Tora, L., and Torres-Padilla, M.E. (2009). TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes Dev.* 23, 2210–2223.
- Gazit, K., Moshonov, S., Elfakess, R., Sharon, M., Mengus, G., Davidson, I., and Dikstein, R. (2009). TAF4/4b·TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes. *J. Biol. Chem.* 284, 26286–26296.
- Gegonne, a, Weissman, J.D., and Singer, D.S. (2001). TAFII55 binding to TAFII250 inhibits its acetyltransferase activity. *Proc. Natl. Acad. Sci. U. S. A.* 98, 12432–12437.
- Gegonne, A., Weissman, J.D., Zhou, M., Brady, J.N., and Singer, D.S. (2006). TAF7: A possible transcription initiation check-point regulator. *Proc. Natl. Acad. Sci. U. S. A.* 103, 602–607.
- Gegonne, A., Weissman, J.D., Lu, H.X., Zhou, M.S., Dasgupta, A., Ribble, R., Brady, J.N., and Singer, D.S. (2008). TFIID component TAF7 functionally interacts with both TFIIF and P-TFIIb. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5367–5372.
- Gegonne, A., Tai, X., Zhang, J., Wu, G., Zhu, J., Yoshimoto, A., Hanson, J., Cultraro, C., Chen, Q.-R.Q.-R., Guinter, T., et al. (2012). The general transcription factor TAF7 is essential for embryonic development but not essential for the survival or differentiation of mature T cells. *Mol. Cell. Biol.* 32, 1984–1997.
- Ghiraldini, F.G., Filipescu, D., and Bernstein, E. (2021). Solid tumours hijack the histone variant network. *Nat. Rev. Cancer* 21, 257–275.
- G mez-Navarro, N., Jordán-Pla, A., Estruch, F., and Pérez-Ortín, J.E. (2016). Defects in the NC2 repressor affect both canonical and non-coding RNA polymerase II transcription initiation in yeast.

BMC Genomics 17, 1–9.

Gonzalez-Sandoval, A., Towbin, B.D., Kalck, V., Cabianca, D.S., Gaidatzis, D., Hauer, M.H., Geng, L., Wang, L., Yang, T., Wang, X., et al. (2015). Perinuclear Anchoring of H3K9-Methylated Chromatin Stabilizes Induced Cell Fate in *C. elegans* Embryos. *Cell* 163, 1333–1347.

Greber, B.J., Toso, D.B., Fang, J., and Nogales, E. (2019). The complete structure of the human TFIID core complex. *Elife* 8, 1–29.

Griesenbeck, J., Tschochner, H., and Grohmann, D. (2017). Structure and function of RNA polymerases and the transcription machineries.

Guibentif, C., Griffiths, J.A., Imaz-Rosshandler, I., Ghazanfar, S., Nichols, J., Wilson, V., Göttgens, B., and Marioni, J.C. (2021). Diverse Routes toward Early Somites in the Mouse Embryo. *Dev. Cell* 56, 141–153.e6.

Gupta, K., Watson, A.A., Baptista, T., Scheer, E., Chambers, A.L., Koehler, C., Zou, J., Obong-Ebong, I., Kandiah, E., Temblador, A., et al. (2017). Architecture of TAF11 / TAF13 / TBP complex suggests novel regulation properties of general transcription factor. *Elife* 6, 1–31.

Gustafson, E.A., Seymour, K.A., Sigrist, K., Rooij, D.G.D.E., and Freiman, R.N. (2020). ZFP628 Is a TAF4b-Interacting Transcription Factor Required for Mouse Spermiogenesis. *Mol. Cell. Biol.* 40.

Guzikevich-guerstein, G., and Shakked, Z. (1996). A novel form of the DNA double helix imposed on the TATA-box by the TATA- binding protein. *Nature* 3, 32–37.

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

Hantsche, M., and Cramer, P. (2017). Conserved RNA polymerase II initiation complex structure. *Curr. Opin. Struct. Biol.* 47, 17–22.

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

Harteis, S., and Schneider, S. (2014). Making the bend: DNA tertiary structure and protein-DNA interactions. *Int. J. Mol. Sci.* 15, 12335–12363.

Hartenian, E., and Glaunsinger, B.A. (2019). Feedback to the central dogma: cytoplasmic mRNA decay and transcription are interdependent processes. *Crit. Rev. Biochem. Mol. Biol.* 54, 385–398.

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

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.

Helmlinger, D., and Tora, L. (2017). Sharing the SAGA. *Trends Biochem. Sci.* 42, 850–861.

Herbst, D.A., Esbin, M.N., Louder, R.K., Dugast-Darzacq, C., Dailey, G.M., Fang, Q., Darzacq, X., Tjian, R., and Nogales, E. (2021). Structure of the human SAGA coactivator complex. *BioRxiv* 2021.02.08.430339.

Herceg, Z., Hulla, W., Gell, D., Cuenin, C., Leonart, M., Jackson, S., and Wang, Z.Q. (2001). Disruption of Trapp causes early embryonic lethality and defects in cell cycle progression. *Nat. Genet.* 29, 206–211.

Herrera, F.J., Yamaguchi, T., Roelink, H., and Tjian, R. (2014). Core promoter factor TAF9B regulates neuronal gene expression. *Elife* 3, 1–17.

Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. *Nature* 463, 474–484.

Hodges, A.J., Hudson, N.O., and Buck-Koehntop, B.A. (2020). Cys2His2 Zinc Finger Methyl-CpG Binding Proteins: Getting a Handle on Methylated DNA. *J. Mol. Biol.* 432, 1640–1660.

Høiby, T., Zhou, H., Mitsiou, D.J., and Stunnenberg, H.G. (2007). A facelift for the general transcription factor TFIID. *Biochim. Biophys. Acta - Gene Struct. Expr.* 1769, 429–436.

- Holla, S., Dhakshnamoorthy, J., Folco, H.D., Balachandran, V., Xiao, H., Sun, L. ling, Wheeler, D., Zofall, M., and Grewal, S.I.S. (2020). Positioning Heterochromatin at the Nuclear Periphery Suppresses Histone Turnover to Promote Epigenetic Inheritance. *Cell* 180, 150-164.e15.
- Howcroft, T.K., Raval, A., Weissman, J.D., Gegonne, A., and Singer, D.S. (2003). Distinct Transcriptional Pathways Regulate Basal and Activated Major Histocompatibility Complex Class I Expression. *Mol. Cell. Biol.* 23, 3377–3391.
- Huisinga, K.L. (2005). Global Regulation of Gene Expression in *Saccharomyces Cerevisiae* Via TATA Binding Protein Regulatory Factors. *13*, 573–585.
- Indra, A.K., Mohan, W.S., Frontini, M., Scheer, E., Messaddeq, N., Metzger, D., and Tora, L. (2005). TAF10 is required for the establishment of skin barrier function in foetal, but not in adult mouse epidermis. *Dev. Biol.* 285, 28–37.
- van Ingen, H., van Schaik, F.M.A., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A.C., Kruijzer, J.A.W., Liskamp, R.M.J., Timmers, H.T.M., and Boelens, R. (2008). Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3. *Structure* 16, 1245–1256.
- Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., and Tjian, R. (2007). Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* 21, 2936–2949.
- Iturbide, A., Pascual-Reguant, L., Fargas, L., Cebrià, J.P., Alsina, B., De Herreros, A.G., and Peir, S. (2015). LOXL2 Oxidizes Methylated TAF10 and Controls TFIID-Dependent Genes during Neural Progenitor Differentiation. *Mol. Cell* 58, 755–766.
- Izban, M.G., and Luse, D.S. (1992). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3' → 5' direction in the presence of elongation factor SII. *Genes Dev.* 6, 1342–1356.
- Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAF(II)250 double bromodomain module. *Science* (80-. ). 288, 1422–1425.
- Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., and Tora, L. (1994). Human TAF1130 Is Present in a Distinct TFIID Complex and Is Required for Transcriptional Activation by the Estrogen Receptor. *Cell* 79, 107–117.
- Jaeger, M.G., Schwalb, B., Mackowiak, S.D., Velychko, T., Hanzl, A., Imrichova, H., Brand, M., Agerer, B., Chorn, S., Nabet, B., et al. (2020). Selective Mediator dependence of cell-type-specifying transcription. *Nat. Genet.* 52, 719–727.
- Jallow, Z., Jacobi, U.G., Weeks, D.L., Dawid, I.B., and Veenstra, G.J.C. (2004). Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in *Xenopus*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 13525–13530.
- 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. *EMBO J.* 30, 249–262.
- Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* 2014, 1–25.
- Joo, Y.J., Ficarro, S.B., Soares, L.M., Chun, Y., Marto, J.A., and Buratowski, S. (2017). Downstream promoter interactions of TFIID TAFs facilitate transcription reinitiation. *Genes Dev.* 31, 2162–2174.
- Kalogeropoulou, M., Voulgari, A., Kostourou, V., Sandaltzopoulos, R., Dikstein, R., Davidson, I., Tora, L., and Pintzas, A. (2010). TAF4b and Jun/activating protein-1 collaborate to regulate the expression of integrin  $\alpha 6$  and cancer cell migration properties. *Mol. Cancer Res.* 8, 554–568.
- Kaltenbach, L., Horner, M.A., Rothman, J.H., and Mango, S.E. (2000). The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. *Mol. Cell* 6, 705–713.
- Kamenova, I., Mukherjee, P., Conic, S., Mueller, F., El-saa, F., Bardot, P., Garnier, J., Dembele, D., Capponi, S., Timmers, H.T.M., et al. (2019). Co-translational assembly of mammalian nuclear multisubunit complexes. *Nat. C* 15, 1–13.

- Kao, C.C., Lieberman, P.M., Schmidt, M.C., Zhou, Q., Pei, R., and Berk, A.J. (1990). Cloning of a transcriptionally active Human TATA binding factor. *Science* (80-. ). 248, 1646–1650.
- Kapranov, P., Willingham, A.T., and Gingeras, T.R. (2007). Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* 8, 413–423.
- Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365, 520–527.
- Kim, S., Yu, N.K., and Kaang, B.K. (2015). CTCF as a multifunctional protein in genome regulation and gene expression. *Exp. Mol. Med.* 47, e166.
- Klejman, M.P., Pereira, L.A., van Zeeburg, H.J.T., Gilfillan, S., Meisterernst, M., and Timmers, H.T.M. (2004). NC2 $\alpha$  Interacts with BTAF1 and Stimulates Its ATP-Dependent Association with TATA-Binding Protein. *Mol. Cell. Biol.* 24, 10072–10082.
- Kloet, S.L., Whiting, J.L., Gafken, P., Ranish, J., and Wang, E.H. (2012). Phosphorylation-Dependent Regulation of Cyclin D1 and Cyclin A Gene Transcription by TFIID Subunits TAF1 and TAF7. *Mol. Cell. Biol.* 32, 3358–3369.
- Kobayashi, W., and Kurumizaka, H. (2019). Structural transition of the nucleosome during chromatin remodeling and transcription. *Curr. Opin. Struct. Biol.* 59, 107–114.
- Kojima, M.L., De Rooij, D.G., and Page, D.C. (2019). Amplification of a broad transcriptional program by a common factor triggers the meiotic cell cycle in mice. *Elife* 8, 1–32.
- Kokic, G., Chernev, A., Tegunov, D., Dienemann, C., Urlaub, H., and Cramer, P. (2019). Structural basis of TFIIF activation for nucleotide excision repair. *Nat. Commun.* 10, 1–9.
- Kolesnikova, O., Ben-Shem, A., Luo, J., Ranish, J., Schultz, P., and Papai, G. (2018). Molecular structure of promoter-bound yeast TFIID. *Nat. Commun.* 9.
- Kopytova, D. V., Krasnov, A.N., Kopantceva, M.R., Nabirochkina, E.N., Nikolenko, J. V., Maksimenko, O., Kurshakova, M.M., Lebedeva, L.A., Yerokhin, M.M., Simonova, O.B., et al. (2006). Two Isoforms of *Drosophila* TRF2 Are Involved in Embryonic Development, Premeiotic Chromatin Condensation, and Proper Differentiation of Germ Cells of Both Sexes . *Mol. Cell. Biol.* 26, 7492–7505.
- Koster, M.J.E., Yildirim, A.D., Weil, P.A., Holstege, F.C.P., and Timmers, H.T.M. (2014). Suppression of intragenic transcription requires the MOT1 and NC2 regulators of TATA-binding protein. *Nucleic Acids Res.* 42, 4220–4229.
- Kostova, N.N., Srebrev, L., Markov, D. V., Sarg, B., Lindner, H.H., and Rundquist, I. (2013). Histone H5-chromatin interactions in situ are strongly modulated by H5 C-terminal phosphorylation. *Cytom. Part A* 83 A, 273–279.
- Koudelka, G.B., Mauro, S.A., and Ciubotaru, M. (2006). Indirect Readout of DNA Sequence by Proteins: The Roles of DNA Sequence-Dependent Intrinsic and Extrinsic Forces. *Prog. Nucleic Acid Res. Mol. Biol.* 81, 143–177.
- Kouskouti, A., Scheer, E., Staub, A., Tora, L., and Talianidis, I. (2004). Gene-specific modulation of TAF10 function by SET9-mediated methylation. *Mol. Cell* 14, 175–182.
- Kowalski, A., and Pałyga, J. (2011). Chromatin compaction in terminally differentiated avian blood cells: The role of linker histone H5 and non-histone protein MENT. *Chromosom. Res.* 19, 579–590.
- Kraemer, S.M., Ranallo, R.T., Ogg, R.C., and Stargell, L.A. (2001). TFIIA Interacts with TFIID via Association with TATA-Binding Protein and TAF40. *Mol. Cell. Biol.* 21, 1737–1746.
- 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. *Mol. Cell* 44, 410–423.
- Kribelbauer, J.F., Lu, X.J., Rohs, R., Mann, R.S., and Bussemaker, H.J. (2020). Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors. *J. Mol. Biol.* 432, 1801–1815.
- Kuhlman, T.C., Cho, H., Reinberg, D., and Hernandez, N. (1999). The General Transcription Factors



- IIA, IIB, IIF, and IIE Are Required for RNA Polymerase II Transcription from the Human U1 Small Nuclear RNA Promoter. *Mol. Cell. Biol.* 19, 2130–2141.
- Kwan, J.Z.J., Nguyen, T.F., Budzyński, M.A., Cui, J., Price, R.M., and Teves, S.S. (2021). A TBP-independent mechanism for RNA Polymerase II transcription. *BioRxiv* 2021.03.28.437425.
- Langer, D., Martianov, I., Alpern, D., Rhinn, M., Keime, C., Dollé, P., Mengus, G., and Davidson, I. (2016). Essential role of the TFIID subunit TAF4 in murine embryogenesis and embryonic stem cell differentiation. *Nat. Commun.* 7, 11063.
- Lannes, L., Young, P., Richter, C., Morgner, N., and Schwalbe, H. (2017). Interaction of the N-Terminal Tandem Domains of hnRNP LL with the BCL2 Promoter i-Motif DNA Sequence. *ChemBioChem* 18, 2033–2044.
- Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 152, 1021–1036.
- Lavigne, A., Mengus, G., May, M., Dubrovskaya, V., Tora, L., Chambon, P., and Davidson, I. (1996). Multiple Interactions between hTAF II 55 and Other TFIID Subunits. *J. Biol. Chem.* 271, 19774–19780.
- Learned, R.M., Cordes, S., and Tjian, R. (1985). Purification and characterization of a transcription factor that confers promoter specificity to human RNA polymerase I. *Mol. Cell. Biol.* 5, 1358–1369.
- Lee, D.-H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinber, D., and Lewis, B.A. (2005). Functional Characterization of Core Promoter Elements: the Downstream Core Element Is Recognized by TAF1. *Mol. Cell. Biol.* 25, 9674–9686.
- Lemieux, K., and Gaudreau, L. (2004). Targeting of Swi/Snf to the yeast GAL1 UASG requires the Mediator, TAFII, and RNA polymerase II. *EMBO J.* 23, 4040–4050.
- Li, L., Martinez, S.S., Hu, W., Liu, Z., and Tjian, R. (2015). A specific E3 ligase/deubiquitinase pair modulates TBP protein levels during muscle differentiation. *Elife* 4, 1–19.
- Lin, J.J., Lehmann, L.W., Bonora, G., Sridharan, R., Vashisht, A.A., Tran, N., Plath, K., Wohlschlegel, J.A., and Carey, M. (2011). Mediator coordinates PIC assembly with recruitment of CHD1. *Genes Dev.* 25, 2198–2209.
- 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-TAF(II)230 complex: Protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell* 94, 573–583.
- Liu, E., Li, L., and Cheng, L. (2018). Gene regulatory network review. *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.* 1–3, 155–164.
- Liu, X., Bushnell, D.A., Wang, D., Calero, G., and Kornberg, R.D. (2010). Structure of an RNA Polymerase II-TFIIB Complex and the Transcription Initiation Mechanism. *Science* (80-. ). 327, 206–209.
- Liu, Z., Scannell, D.R., Eisen, M.B., and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* 146, 720–731.
- Lobo, S.M., and Hernandez, N. (1989). A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. *Cell* 58, 55–67.
- Locati, M.D., Pagano, J.F.B., Abdullah, F., Ensink, W.A., Van Olst, M., Van Leeuwen, S., Nehrdich, U., Spaink, H.P., Rauwerda, H., Jonker, M.J., et al. (2018). Identifying small RNAs derived from maternal-and somatic-type rRNAs in zebrafish development. *Genome* 61, 371–378.
- Lorch, Y., and Kornberg, R.D. (2017). Chromatin-remodeling for transcription. *Q. Rev. Biophys.* 50, 1–15.
- Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* 49, 203–210.
- Louder, R.K., He, Y., Lopez-Blanco, J.R., Fang, J., Chacón, P., and Nogales, E. (2016). Structure of

- promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* 531, 604–609.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Machida, S., Takizawa, Y., Ishimaru, M., Sugita, Y., Sekine, S., Nakayama, J. ichi, Wolf, M., and Kurumizaka, H. (2018). Structural Basis of Heterochromatin Formation by Human HP1. *Mol. Cell* 69, 385–397.e8.
- Malecová, B., Caputo, V.S., Lee, D.F., Hsieh, J.J., and Oelgeschläger, T. (2015). Taspase1 processing alters TFIIA cofactor properties in the regulation of TFIID. *Transcription* 6, 21–32.
- Martello, G., and Smith, A. (2014). The Nature of Embryonic Stem Cells. *Annu. Rev. Cell Dev. Biol.* 30, 647–675.
- Martianov, I., Fimia, G.M., Dierich, A., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2001). Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Mol. Cell* 7, 509–515.
- Martianov, I., Viville, S., and Davidson, I. (2002a). RNA polymerase II transcription in murine cells lacking the TATA binding protein. *Science* (80-. ). 298, 1036–1039.
- Martianov, I., Brancorsini, S., Gansmuller, A., Parvinen, M., Davidson, I., and Sassone-Corsi, P. (2002b). Distinct functions of TBP and TLF/TRF2 during spermatogenesis: Requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. *Development* 129, 945–955.
- Martianov, I., Velt, A., Davidson, G., Choukrallah, M.A., and Davidson, I. (2016). TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. *Sci. Rep.* 6, 1–13.
- Martire, S., and Banaszynski, L.A. (2020). The roles of histone variants in fine-tuning chromatin organization and function. *Nat. Rev. Mol. Cell Biol.* 21, 522–541.
- Martinez-Fernández, V., and Navarro, F. (2018). Rpb5, a subunit shared by eukaryotic RNA polymerases, cooperates with prefoldin-like Bud27/URI. *AIMS Genet.* 05, 063–074.
- Maston, G.A., Zhu, L.J., Chamberlain, L., Lin, L., Fang, M., and Green, M.R. (2012). Non-canonical TAF complexes regulate active promoters in human embryonic stem cells. *Elife* 2012.
- Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem.* 255, 11992–11996.
- 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.
- Mencía, M., and Struhl, K. (2001). Region of Yeast TAF 130 Required for TFIID To Associate with Promoters. *Mol. Cell. Biol.* 21, 1145–1154.
- Mengus, G., May, M., Jacq, X., Staub, A., Tora, L., Chambon, P., and Davidson, I. (1995). Cloning and characterization of hTAF(II)18, hTAF(II)20 and hTAF(II)28: Three subunits of the human transcription factor TFIID. *EMBO J.* 14, 1520–1531.
- Metzger, D., Scheer, E., Soldatov, A., and Tora, L. (1999). Mammalian TAF II 30 is required for cell cycle progression and specific cellular differentiation programmes. *Embo* 18, 4823–4834.
- Mishina, Y., Financsek, I., Kominami, R., and Muramatsu, M. (1982). Fractionation and reconstitution of factors required for accurate transcription of mammalian ribosomal RNA genes: identification of a species-dependent initiation factor. *Nucleic Acids Res.* 10, 6659–6670.
- Mohan II, W.S., Scheer, E., Wendling, O., and Metzger, D. (2003). TAF10 ( TAF II 30 ) Is Necessary for TFIID Stability and Early Embryogenesis in Mice. *Mol. Cell. Biol.* 23, 4307–4318.
- Müller, F., Lakatos, L., Dantonel, J.C., Strähle, U., and Tora, L. (2001). TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. *Curr. Biol.* 11, 282–287.
- Müller, F., Zaucker, A., and Tora, L. (2010). Developmental regulation of transcription initiation: More

- than just changing the actors. *Curr. Opin. Genet. Dev.* 20, 533–540.
- 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. *Mol. Cell. Biol.* 29, 1649–1660.
- Nakagawa, T., Hosogane, M., Nakagawa, M., Morohoshi, A., Funayama, R., and Nakayama, K. (2017). Transforming Growth Factor  $\beta$ -Induced Proliferative Arrest Mediated by TRIM26-Dependent TAF7 Degradation and Its Antagonism by MYC. *Mol. Cell. Biol.* 38, 1–19.
- Nikolova, E.N., Zhou, H., Gottardo, F.L., Alvey, H.S., Kimsey, I.J., and Al-Hashimi, H.M. (2013). A Historical Account of Hoogsteen Base-pairs in Duplex DNA *Evgenia*. *Biopolymers* 99, 1–7.
- Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* 6, 917–922.
- Niu, K., Zhang, X., Deng, H., Wu, F., Ren, Y., Xiang, H., Zheng, S., Liu, L., Huang, L., Zeng, B., et al. (2018). BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in *Bombyx mori*. *Nucleic Acids Res.* 46, 1710–1723.
- O’Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front. Endocrinol. (Lausanne)* 9, 1–12.
- Ohbayashi, T., Makino, Y., and Tamura, T.A. (1999). Identification of a mouse TBP-like protein (TLP) distantly related to the *Drosophila* TBP-related factor. *Nucleic Acids Res.* 27, 750–755.
- Ohbayashi, T., Shimada, M., Nakadai, T., and Tamura, T. aki (2001). TBP-like protein (TLP/TLF/TRF2) artificially recruited to a promoter stimulates basal transcription in vivo. *Biochem. Biophys. Res. Commun.* 285, 616–622.
- Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev.* 10, 2657–2683.
- Oyama, T., Sasagawa, S., Takeda, S., Hess, R.A., Lieberman, P.M., Cheng, E.H., and Hsieh, J.J. (2013). Cleavage of TFIIA by caspase 1 activates TRF2-specified mammalian male germ cell programs. *Dev. Cell* 27, 188–200.
- Ozata, D.M., Gainetdinov, I., Zoch, A., O’Carroll, D., and Zamore, P.D. (2019). PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* 20, 89–108.
- Pahi, Z., Borsos, B.N., Vedelek, B., Shidlovskii, Y. V., Georgieva, S.G., Boros, I.M., and Pankotai, T. (2017). TAF10 and TAF10b partially redundant roles during *Drosophila melanogaster* morphogenesis. *Transcription* 8, 297–306.
- Papadopoulos, P., Gutiérrez, L., Demmers, J., Scheer, E., Pourfarzad, F., Papageorgiou, D.N., Karkoulia, E., Strouboulis, J., van de Werken, H.J.G., van der Linden, R., et al. (2015). TAF10 interacts with GATA1 transcription factor and controls mouse erythropoiesis. *Mol. Cell. Biol.* 35, MCB.01370-14.
- 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.
- Park, J.L., Lee, Y.S., Kunkeaw, N., Kim, S.Y., Kim, I.H., and Lee, Y.S. (2017). Epigenetic regulation of noncoding RNA transcription by mammalian RNA polymerase III. *Epigenomics* 9, 171–187.
- Parra-Marín, O., Lopez-Pacheco, K., Hernández, R., and Lopez-Villaseñor, I. (2020). The highly diverse TATA box-binding proteins among protists: A review. *Mol. Biochem. Parasitol.* 239, 111312.
- 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 (80-. )* 362.
- Patel, A.B., Greber, B.J., and Nogales, E. (2020). Recent insights into the structure of TFIID, its assembly, and its binding to core promoter. *Curr. Opin. Struct. Biol.* 61, 17–24.
- Patikoglou, G.A., Kim, J.L., Sun, L., Yang, S.H., Kodadek, T., and Burley, S.K. (1999). TATA element



- recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* *13*, 3217–3230.
- Pederson, T. (2011). The nucleolus. *Cold Spring Harb. Perspect. Biol.* *3*, 1–15.
- Perantoni, A.O., Timofeeva, O., Naillat, F., Richman, C., Pajni-Underwood, S., Wilson, C., Vainio, S., Dove, L.F., and Lewandoski, M. (2005). Inactivation of FGF8 in early mesoderm reveals an essential role in kidney development. *Development* *132*, 3859–3871.
- 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.
- Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Mol. Cell* *38*, 603–613.
- Perletti, L., Kopf, E., Carré, L., and Davidson, I. (2001). Coordinate regulation of RARgamma2, TBP, and TAFII135 by targeted proteolysis during retinoic acid-induced differentiation of F9 embryonal carcinoma cells. *BMC Mol. Biol.* *2*.
- Peterson, M.G., Tanese, N., Franklin Pugh, B., and Tjian, R. (1990). Functional domains and upstream activation properties of cloned human TATA binding protein. *Science* (80-. ). *248*, 1625–1630.
- Di Pietro, C., Ragusa, M., Duro, L., Guglielmino, M.R., Barbagallo, D., Carnemolla, A., Laganà, A., Buffa, P., Angelica, R., Rinaldi, A., et al. (2007). Genomics, evolution, and expression of TBPL2, a member of the TBP family. *DNA Cell Biol.* *26*, 369–385.
- Pijnappel, W.W.M.P., Esch, D., Baltissen, M.P.A., Wu, G., Mischerikow, N., Bergsma, A.J., Van Der Wal, E., Han, D.W., Bruch, H. Vom, Moritz, S., et al. (2013). A central role for TFIID in the pluripotent transcription circuitry. *Nature* *495*, 516–519.
- Pimmett, V.L., Dejean, M., Fernandez, C., Trullo, A., Bertrand, E., Radulescu, O., and Lagha, M. (2021). Quantitative imaging of transcription in living *Drosophila* embryos reveals the impact of core promoter motifs on promoter state dynamics. *Nat. Commun.* *12*.
- Poggi, L., and Richard, G.-F. (2021). Alternative DNA Structures In Vivo : Molecular Evidence and Remaining Questions . *Microbiol. Mol. Biol. Rev.* *85*.
- Pointud, J.-C. (2003). The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID subunit TAF7, is developmentally regulated during male germ-cell differentiation. *J. Cell Sci.* *116*, 1847–1858.
- Poleshko, A., Shah, P.P., Gupta, M., Badu, A., Morley, M., Manderfield, L.J., Ifkovits, J.L., and Jain, R. (2017). Genome-nuclear lamina interactions regulate cardiac stem cell lineage restriction.
- Poleshko, A., Smith, C.L., Nguyen, S.C., Sivaramakrishnan, P., Wong, K.G., Murray, J.I., Lakadamyali, M., Joyce, E.F., Jain, R., and Epstein, J.A. (2019). H3k9me2 orchestrates inheritance of spatial positioning of peripheral heterochromatin through mitosis. *Elife* *8*, 1–24.
- Portin, P., and Wilkins, A. (2017). The evolving definition of the term “Gene.” *Genetics* *205*, 1353–1364.
- Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. *Crit. Rev. Biochem. Mol. Biol.* *48*, 575–608.
- Pueschel, R., Coraggio, F., and Meister, P. (2016). From single genes to entire genomes: the search for a function of nuclear organization. *Development* *143*, 910–923.
- Ransohoff, J.D., Wei, Y., and Khavari, P.A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol.* *19*, 143–157.
- Rao, S., Chiu, T.P., Kribelbauer, J.F., Mann, R.S., Bussemaker, H.J., and Rohs, R. (2018). Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein-DNA binding. *Epigenetics and Chromatin* *11*, 1–11.

- Raval, A., Howcroft, T.K., Weissman, J.D., Kirshner, S., Zhu, X.S., Yokoyama, K., Ting, J., and Singer, D.S. (2001). Transcriptional coactivator, CIITA, is an acetyltransferase that bypasses a promoter requirement for TAFII250. *Mol. Cell* 7, 105–115.
- Ravarani, C.N.J., Flock, T., Chavali, S., Anandapadamanaban, M., Babu, M.M., and Balaji, S. (2020). Molecular determinants underlying functional innovations of TBP and their impact on transcription initiation. *Nat. Commun.* 11, 1–16.
- Remboutsika, E., Jacq, X., and Tora, L. (2001). Chromatin Is Permissive to TATA-binding Protein (TBP)-mediated Transcription Initiation. *J. Biol. Chem.* 276, 12781–12784.
- Roeder, R.G., and Rutter, W.J. (1969). Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms. *Nature* 224, 234–237.
- Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* 79, 233–269.
- Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. *Cell* 169, 1187–1200.
- Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 16, 129–143.
- Sawan, C., Hernandez-Vargas, H., Murr, R., Lopez, F., Vaissière, T., Ghantous, A.Y., Cuenin, C., Imbert, J., Wang, Z.Q., Ren, B., et al. (2013). Histone acetyltransferase cofactor Trapp maintains self-renewal and restricts differentiation of embryonic stem cells. *Stem Cells* 31, 979–991.
- Scheer, E., Luo, J., Bernardini, A., Ruffenach, F., Garnier, J.-M., Kolb-Cheynel, I., Gupta, K., Berger, I., Ranish, J., and Tora, L. (2021). TAF8 regions important for TFIID lobe B assembly or for TAF2 interactions are required for embryonic stem cell survival. *J. Biol. Chem.* 101288.
- Schier, A.C., and Taatjes, D.J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* 34, 465–488.
- Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* 16, 2593–2620.
- Schwalb, B., Michel, M., Zacher, B., Hauf, K.F., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science* (80-. ). 352, 1225–1228.
- See, K., Lan, Y., Rhoades, J., Jain, R., Smith, C.L., and Epstein, J.A. (2019). Lineage-specific reorganization of nuclear peripheral heterochromatin and H3K9Me2 domains. *Dev.* 146.
- Segall, J., Matsui, T., and Roeder, R.G. (1980). Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. *J. Biol. Chem.* 255, 11986–11991.
- Sharifi, S., and Bierhoff, H. (2018). Regulation of RNA Polymerase i Transcription in Development, Disease, and Aging. *Annu. Rev. Biochem.* 87, 51–73.
- Shin, S.I., Ham, S., Park, J., Seo, S.H., Lim, C.H., Jeon, H., Huh, J., and Roh, T.Y. (2016). Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome. *DNA Res.* 23, 477–486.
- Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 16, 1128–1133.
- Sorscher, D.H., Yang, B., Bhaumik, D., Trangas, T., Philips, A. V., Chancellor, K.E., and Coleman, M.S. (1994). Initiation of Transcription at the Human Terminal Deoxynucleotidyl Transferase Gene Promoter: A Novel Role for the TATA Binding Protein. *Biochemistry* 33, 11025–11032.
- Sosińska, P., Mikuła-Pietrasik, J., and Ksiazek, K. (2015). The double-edged sword of long non-coding RNA: The role of human brain-specific BC200 RNA in translational control, neurodegenerative diseases, and cancer. *Mutat. Res. - Rev. Mutat. Res.* 766, 58–67.
- Soutoglou, E., Demény, M.A., Scheer, E., Fienga, G., Sassone-Corsi, P., and Tora, L. (2005). The Nuclear Import of TAF10 Is Regulated by One of Its Three Histone Fold Domain-Containing Interaction

- Partners. *Mol. Cell. Biol.* 25, 4092–4104.
- Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* 19, 262–274.
- Sun, F., Sun, T., Kronenberg, M., Tan, X., Huang, C., and Carey, M.F. (2021). The Pol II preinitiation complex (PIC) influences Mediator binding but not promoter-enhancer looping. *Genes Dev.* 35, 1175–1189.
- Sun, M., Schwalb, B., Schulz, D., Pirkel, 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 Res.* 22, 1350–1359.
- Sutherland, C., Cui, Y., Mao, H., and Hurley, L.H. (2016). A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1. *J. Am. Chem. Soc.* 138, 14138–14151.
- Taatjes, D.J. (2017). The Continuing SAGA of TFIID and RNA Polymerase II Transcription. *Mol. Cell* 68, 1–2.
- Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J.S. (2009). Small RNAs derived from snoRNAs. *Rna* 15, 1233–1240.
- Talbert, P.B., and Henikoff, S. (2017). Histone variants on the move: Substrates for chromatin dynamics. *Nat. Rev. Mol. Cell Biol.* 18, 115–126.
- Tang, W., Niu, K., Yu, G., Jin, Y., Zhang, X., Peng, Y., Chen, S., Deng, H., Li, S., Wang, J., et al. (2020). In vivo visualization of the i-motif DNA secondary structure in the Bombyx mori testis. *Epigenetics and Chromatin* 13, 1–12.
- Tang, X., Ren, H., Guo, M., Qian, J., Yang, Y., and Gu, C. (2021). Review on circular RNAs and new insights into their roles in cancer. *Comput. Struct. Biotechnol. J.* 19, 910–928.
- Tatarakis, A., Margaritis, T., Martinez-Jimenez, C.P., Kouskouti, A., Mohan, W.S., Haroniti, A., Kafetzopoulos, D., Tora, L., and Talianidis, I. (2008). Dominant and Redundant Functions of TFIID Involved in the Regulation of Hepatic Genes. *Mol. Cell* 31, 531–543.
- Teichmann, M., Wang, Z., Martinez, E., Tjernberg, A., Zhang, D., Vollmer, F., Chait, B.T., and Roeder, R.G. (1999). Human TATA-binding protein-related factor-2 (hTRF2) stably associates with HTFIIa in HeLa cells. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13720–13725.
- Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* 41, 105–178.
- Tickle, C. (2015). How the embryo makes a limb: Determination, polarity and identity. *J. Anat.* 227, 418–430.
- Timmers, H.T.M., and Tora, L. (2018). Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation. *Mol. Cell* 72, 10–17.
- Tolsma, T.O., and Hansen, J.C. (2019). Post-translational modifications and chromatin dynamics. *Essays Biochem.* 63, 89–96.
- Tomecki, R., Sikorski, P.J., and Zakrzewska-Placzek, M. (2017). Comparison of preribosomal RNA processing pathways in yeast, plant and human cells – focus on coordinated action of endo- and exoribonucleases. *FEBS Lett.* 591, 1801–1850.
- Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* 16, 673–675.
- Tourigny, J.P., Schumacher, K., Saleh, M.M., Devys, D., and Zentner, G.E. (2021). Architectural Mediator subunits are differentially essential for global transcription in *Saccharomyces cerevisiae*. *Genetics* 217.
- Towbin, B.D., González-Aguilera, C., Sack, R., Gaidatzis, D., Kalek, V., Meister, P., Askjaer, P., and Gasser, S.M. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear

periphery. *Cell* 150, 934–947.

Travers, A. (1999). The location of the linker histone on the nucleosome. *Trends Biochem. Sci.* 24, 4–7.

Trowitzsch, S., Viola, C., Scheer, E., Conic, S., Chavant, V., Fournier, M., Papai, G., Ebong, I.-O., Schaffitzel, C., Zou, J., et al. (2015). Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat. Commun.* 6, 6011.

Tudek, A., Candelli, T., and Libri, D. (2015). Non-coding transcription by RNA polymerase II in yeast: Hasard or nécessité? *Biochimie* 117, 28–36.

Upadhyaya, A.B., Khan, M., Mou, T.C., Junker, M., Gray, D.M., and Dejong, J. (2002). The germ cell-specific transcription factor ALF. Structural properties and stabilization of the TATA-binding protein (TBP)-DNA complex. *J. Biol. Chem.* 277, 34208–34216.

Vannini, A. (2013). A structural perspective on RNA polymerase I and RNA polymerase III transcription machineries. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1829, 258–264.

Varier, R.A., Outchkourov, N.S., De Graaf, P., Van Schaik, F.M.A., Ensing, H.J.L., Wang, F., Higgins, J.M.G., Kops, G.J.P.L., and Timmers, H.M. (2010). A phospho/methyl switch at histone H3 regulates TFIID association with mitotic chromosomes. *EMBO J.* 29, 3967–3978.

Veenstra, G.J.C., Weeks, D.L., and Wolffe, A.P. (2000). Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in *Xenopus*. *Science* (80-. ). 290, 2312–2315.

Ventura, A., Kirsch, D.G., McLaughlin, M.E., Tuveson, D.A., Grimm, J., Lintault, L., Newman, J., Reczek, E.E., Weissleder, R., and Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. *Nature* 445, 661–665.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., and Timmers, H.T.M. (2007). Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* 131, 58–69.

Voss, A.K., Thomas, T., Petrou, P., Anastassiadis, K., Schöler, H., and Gruss, P. (2000). Taube nuss is a novel gene essential for the survival of pluripotent cells of early mouse embryos. *Development* 127, 5449–5461.

Voulgari, A., Voskou, S., Tora, L., Davidson, I., Sasazuki, T., Shirasawa, S., and Pintzas, A. (2008). TATA box-binding protein-associated factor 12 is important for RAS-induced transformation properties of colorectal cancer cells. *Mol. Cancer Res.* 6, 1071–1083.

Wada, Y., Ohta, Y., Xu, M., Tsutsumi, S., Minami, T., Inoue, K., Komura, D., Kitakami, J., Oshida, N., Papantonis, A., et al. (2009). A wave of nascent transcription on activated human genes. *Proc. Natl. Acad. Sci. U. S. A.* 106, 18357–18361.

Wang, F., El-Saafin, F., Ye, T., Stierle, M., Negroni, L., Durik, M., Fischer, V., Devys, D., Vincent, S.D., and Tora, L. (2021). Histone H2Bub1 deubiquitylation is essential for mouse development, but does not regulate global RNA polymerase II transcription. *Cell Death Differ.* 28, 2385–2403.

Wang, H., Curran, E.C., Hinds, T.R., Wang, E.H., and Zheng, N. (2014). Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. *Cell Res.* 24, 1433–1444.

Wardell, J.R., Hodgkinson, K.M., Binder, A.K., Seymour, K.A., Korach, K.S., Vanderhyden, B.C., and Freiman, R.N. (2013). Estrogen Responsiveness of the TFIID Subunit TAF4B in the Normal Mouse Ovary and in Ovarian Tumors1. *Biol. Reprod.* 89, 1–9.

Warrier, S., Nuwayhid, S., Sabatino, J.A., Sugrue, K.F., and Zohn, I.E. (2017). Supt20 is required for development of the axial skeleton. *Dev. Biol.* 421, 245–257.

Wassarman, D.A., and Sauer, F. (2001). Tafi250 : a transcription toolbox. *J. Cell Sci.* 114, 2895–2902.

Waston, J.D., and Crick, F.H. (1953). The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* 18, 123–131.

- Weil, P.A., Luse, D.S., and Roeder, R.G. (1979). Selective and Accurate Initiation of Transcription at the Ad2 Major Late Promotor in a Soluble System Dependent on Purified RNA Polymerase II and DNA. *Cell* 18, 469–484.
- Weissman, J.D., Brown, J.A., Howcroft, T.K., Hwang, J., Chawla, A., Roche, P.A., Schiltz, L., Nakatani, Y., and Singer, D.S. (1998). HIV-1 Tat binds TAFII250 and represses TAFII250-dependent transcription of major histocompatibility class I genes. *Proc. Natl. Acad. Sci. U. S. A.* 95, 11601–11606.
- Wieczorek, E., Brand, M., Jacq, X., and Tora, L. (1998). Function of TAFII-containing complex without TBP in transcription by RNA polymerase II. *Nature* 393, 187–191.
- Winnebeck, E.C., Millar, C.D., and Warman, G.R. (2010). Why does insect RNA look degraded? *J. Insect Sci.* 10, 1–7.
- Wissink, E.M., Vihervaara, A., Tippens, N.D., and Lis, J.T. (2019). Nascent RNA analyses: tracking transcription and its regulation. *Nat. Rev. Genet.* 20, 705–723.
- Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nat. Rev. Genet.* 18, 517–534.
- Xiao, L.J., Kim, M.J., and DeJong, J. (2006). Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice. *Gene Expr. Patterns* 6, 409–419.
- Xiong, J., Ye, T.T., Ma, C.J., Cheng, Q.Y., Yuan, B.F., and Feng, Y.Q. (2019). N6-Hydroxymethyladenine: A hydroxylation derivative of N6-methyladenine in genomic DNA of mammals. *Nucleic Acids Res.* 47, 1268–1277.
- 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, 229–232.
- Xu, Y., Man, N., Karl, D., Martinez, C., Liu, F., Sun, J., Martinez, C.J., Martin, G.M., Beckedorff, F., Lai, F., et al. (2019). TAF1 plays a critical role in AML1-ETO driven leukemogenesis. *Nat. Commun.* 10, 1–15.
- 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 & DNA motifs enriched in TATA-less core promoters\_Gene\_2007.pdf. *Gene* 1, 52–65.
- Yao, R.W., Wang, Y., and Chen, L.L. (2019). Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* 21, 542–551.
- Yeganeh, M., and Hernandez, N. (2020). RNA polymerase III transcription as a disease factor. *Genes Dev.* 34, 865–882.
- Yu, C., Cveticic, N., Hisler, V., Gupta, K., Ye, T., Gazdag, E., Negroni, L., Hajkova, P., Berger, I., Lenhard, B., et al. (2020a). TBPL2/TFIIA complex establishes the maternal transcriptome through oocyte-specific promoter usage. *Nat. Commun.* 11, 1–13.
- Yu, X., Xie, Y., Zhang, S., Song, X., Xiao, B., and Yan, Z. (2020b). tRNA-derived fragments: Mechanisms underlying their regulation of gene expression and potential applications as therapeutic targets in cancers and virus infections. *Theranostics* 11, 461–469.
- Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Molecular biology: Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* (80-). 309, 626–630.
- Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408, 225–229.
- Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355–364.
- Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. *Cell Res.* 21, 564–578.



- Zeng, P., Wang, Y., Zheng, Y., Song, X., and Yin, Y. (2018). Cancer-testis antigen HCA587/MAGEC2 interacts with the general transcription coactivator TAF9 in cancer cells. *Mol. Med. Rep.* 17, 3226–3231.
- Zenk, F., Zhan, Y., Kos, P., Löser, E., Atinbayeva, N., Schächtle, M., Tiana, G., Giorgetti, L., and Iovino, N. (2021). HP1 drives de novo 3D genome reorganization in early *Drosophila* embryos. *Nature* 593, 289–293.
- Zeraati, M., Langley, D.B., Schofield, P., Moye, A.L., Rouet, R., Hughes, W.E., Bryan, T.M., Dinger, M.E., and Christ, D. (2018). I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* 10, 631–637.
- Zhang, D., Penttilä, T.L., Morris, P.L., and Roeder, R.G. (2001). Cell- and stage-specific high-level expression of TBP-related factor 2 (TRF2) during mouse spermatogenesis. *Mech. Dev.* 106, 203–205.
- 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. *Mol. Cell* 29, 102–111.
- Zhou, H., Spicuglia, S., Hsieh, J.J.-D., Mitsiou, D.J., Høiby, T., Veenstra, G.J.C., Korsmeyer, S.J., and Stunnenberg, H.G. (2006). Uncleaved TFIIA Is a Substrate for Taspase 1 and Active in Transcription. *Mol. Cell. Biol.* 26, 2728–2735.
- Zhou, H., Kaplan, T., Li, Y., Grubisic, I., Zhang, Z., Wang, P.J., Eisen, M.B., and Tjian, R. (2013). Dual functions of TAF7L in adipocyte differentiation. *Elife* 2, 1–22.
- 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.





---

# Annexes

---

---

# Annexes

---

**TBPL2/TFIIA complex establishes the maternal transcriptome  
through oocyte-specific promoter usage.**

Changwei Yu, Nevena Cvetic, **Vincent Hisler**, [...] Stéphane D. Vincent & Laszlo Tora.

Cette étude réalisée par Changwei Yu avait pour but de comprendre l'importance fonctionnelle de TBPL2 au sein des ovocytes durant la croissance ovocytaire chez la souris. Au moment de l'initiation de la croissance ovocytaire, TBP est remplacé par son paralogue TBPL2 qui est alors requis pour la transcription médiée par l'ARN pol II. J'ai validé les résultats d'une analyse RNA seq par RT-qPCR. En particulier, j'ai validé plusieurs transcrits qui sont surexprimés en l'absence de TBPL2. Cette observation était un peu surprenante puisque TBPL2 est responsable de la transcription ARN pol II en l'absence de TBP. Mes données ont permis de conclure que la surexpression d'un nombre important de transcrits n'est pas un artefact de normalisation du RNA-seq mais plutôt la conséquence d'une stabilisation de transcrits normalement dégradés au cours de la croissance ovocytaire.

Une des questions posées était de savoir si un complexe de type TFIID contenant TBPL2 est fonctionnel au cours de cette période. Dans ce but, une délétion conditionnelle du gène *Taf7* au cours de la croissance ovocytaire a été réalisée. De manière remarquable, la croissance et la maturation ovocytaires sont normales chez ces souris mutantes conditionnelles. J'ai participé au travail de révision du manuscrit en analysant l'expression de TAF7 dans les ovaires contrôles et mutants. J'avais déjà validé un anticorps anti-TAF7 sur mes cellules ES. J'ai utilisé cet anticorps pour étudier l'expression de TAF7 au cours de la croissance ovocytaire et pour valider la délétion conditionnelle par immunofluorescence. Mes travaux ont montré que l'expression de TAF7 est faible dans les ovocytes contrôles en croissance et restreinte au cytoplasme. Cette expression faible disparaît complètement dans les ovocytes mutants, confirmant l'efficacité de la délétion.

Pour finir, j'ai contribué à la rédaction du papier en écrivant la partie du matériel et méthode concernant mes expériences.

**SUPT3H-less SAGA coactivator can assemble and function without significantly perturbing  
RNA polymerase II transcription in mammalian cells.**

Veronique Fischer, Elisabeth Scheer, Elisabeth Lata, **Vincent Hisler** [...], Didier Devys & Laszlo  
Tora.

SAGA est un complexe coactivateur de la transcription qui peut à la fois débiquitiner et acétyler les nucléosomes. Chez la levure, SAGA est capable d'interagir avec TBP via deux sous-unités SPT3 et SPT8. Seul SPT3 a un orthologue chez les mammifères ; SUPT3H. La comparaison de la structure de SAGA entre la levure et les mammifères montre que SPT3 et SUPT3H n'ont pas exactement la même conformation et il n'est pas clair si chez les mammifères, SAGA interagit avec TBP. L'étude conduit par Véronique Fischer a pour but de mieux comprendre l'importance de SUPT3H dans l'assemblage de SAGA, dans la transcription médiée par l'ARN pol II ainsi que dans la déposition de TBP sur la chromatine.

Des cellules mESCs portant une mutation dans le gène *Supt3h* ont été générées. Une différence de phénotype a été observée en fonction des conditions de culture. Alors que la perte de Supt3h n'a pas d'effet notable lorsque les cellules sont cultivées en présence de sérum, de LIF et des deux inhibiteurs (CHIR99021 et PD0325901) (milieu sérum+LIF+2i), elle affecte le maintien de la pluripotence lorsque les cellules sont cultivées en présence de sérum et de LIF (milieu sérum+LIF). Une analyse du transcriptome naissant a été réalisée dans les conditions sérum+LIF+2i et a montré que la perte de *Supt3h* n'a pas d'effet majeur sur la transcription. Les reviewers nous ont demandé de faire une analyse similaire lorsque les cellules sont cultivées en présence de sérum+LIF. J'ai mis à profit mon expérience dans la réalisation des transcriptomes naissants pour réaliser cette expérience. Cette expérience est en cours.













ARTICLE



<https://doi.org/10.1038/s41467-020-20239-4>

OPEN

# TBPL2/TFIIA complex establishes the maternal transcriptome through oocyte-specific promoter usage

Changwei Yu <sup>1,2,3,4</sup>, Nevena Cvetesic <sup>5</sup>, Vincent Hisler <sup>1,2,3,4</sup>, Kapil Gupta <sup>6</sup>, Tao Ye <sup>1,2,3,4</sup>, Emese Gazdag <sup>1,2,3,4</sup>, Luc Negroni <sup>1,2,3,4</sup>, Petra Hajkova<sup>5</sup>, Imre Berger <sup>6</sup>, Boris Lenhard <sup>5</sup>, Ferenc Müller <sup>7</sup>, Stéphane D. Vincent <sup>1,2,3,4,8</sup>✉ & László Tora <sup>1,2,3,4,8</sup>✉

During oocyte growth, transcription is required to create RNA and protein reserves to achieve maternal competence. During this period, the general transcription factor TATA binding protein (TBP) is replaced by its paralogue, TBPL2 (TBP2 or TRF3), which is essential for RNA polymerase II transcription. We show that in oocytes TBPL2 does not assemble into a canonical TFIID complex. Our transcript analyses demonstrate that TBPL2 mediates transcription of oocyte-expressed genes, including mRNA survey genes, as well as specific endogenous retroviral elements. Transcription start site (TSS) mapping indicates that TBPL2 has a strong preference for TATA-like motif in core promoters driving sharp TSS selection, in contrast with canonical TBP/TFIID-driven TATA-less promoters that have broader TSS architecture. Thus, we show a role for the TBPL2/TFIIA complex in the establishment of the oocyte transcriptome by using a specific TSS recognition code.

<sup>1</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France. <sup>2</sup>Centre National de la Recherche Scientifique (CNRS), UMR7104, 67404 Illkirch, France. <sup>3</sup>Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, 67404 Illkirch, France. <sup>4</sup>Université de Strasbourg, 67404 Illkirch, France. <sup>5</sup>Imperial College London, South Kensington Campus, London SW7 2AZ, UK. <sup>6</sup>Max Planck Bristol Centre for Minimal Biology, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK. <sup>7</sup>Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK. <sup>8</sup>These authors contributed equally: Stéphane D. Vincent, László Tora. ✉email: [vincent@igbmc.fr](mailto:vincent@igbmc.fr); [laszlo@igbmc.fr](mailto:laszlo@igbmc.fr)

**R**egulation of transcription initiation by RNA polymerase II (Pol II) is central to all developmental processes. Pol II transcription requires the stepwise assembly of multi-protein complexes called general transcription factors (GTFs) and Pol II<sup>1</sup>. The evolutionary conserved TFIID complex plays a major role in transcription initiation as it is the first GTF to initiate the assembly of the pre-initiation complex (PIC) by recognising the core promoter<sup>2</sup>. TFIID is a large multiprotein complex composed of the TATA box-binding protein (TBP) and 13 TBP-associated factors (TAFs) in metazoa<sup>3</sup>. The model suggesting that transcription is always regulated by the same transcription complexes has been challenged in metazoans by the discovery of cell-type-specific complexes containing specialised GTF-, TBP- or TAF-paralogs<sup>4</sup>. Two TBP paralogues have been described in vertebrates: TBPL1 (TBP-like factor; TLF, also known as TRF2) has been identified in all metazoan species<sup>5–10</sup>, while TBPL2 (also known as TRF3 or TBP2) has only been described in vertebrates<sup>11,12</sup>. Remarkably, while *Tbpl1* and *Tbpl2* mutants display embryonic phenotypes in non-mammalian species<sup>7–10,12,13</sup>, *Tbpl1* and *Tbpl2* loss of function in mouse results in male and female sterility, respectively<sup>14–16</sup>, suggesting that in mammals, these two TBP-like proteins are involved in cell-specific transcription. While TBPL2 shares a high degree of identity (92%) within the conserved saddle-shaped C-terminal DNA-binding core domain of TBP<sup>17</sup>, the C-terminus of TBPL1 is more distant with only 42% identity<sup>12</sup>. As a consequence TBPL2, but not TBPL1, is able to bind canonical TATA box sequences in vitro<sup>5,12,18</sup>. The N-terminal domains of the three vertebrate TBP-related factors do not show any conservation. All three vertebrate TBP-related factors can interact with the GTFs TFIIA and TFIIB, and can mediate Pol II transcription initiation in vitro<sup>12,13,18–20</sup>. However, how alternative initiation complexes form, how they regulate cell-type-specific transcription and how they recognise promoter sequences remain unknown.

Mapping of the transcription start sites (TSSs), at single nucleotide by Cap Analysis of Gene Expression (CAGE) revealed two main modes for transcription start site (TSS) usage<sup>21</sup>. Transcription initiation within a narrow region, called “sharp” (or focused) TSS-type, is common in highly active, tissue-specific gene promoters containing TATA boxes. While transcription initiation with multiple initiation positions within an about 100 bp region, called “broad” TSS promoter architecture<sup>21</sup>, is more characteristic to ubiquitously expressed and developmentally regulated genes (reviewed in ref. <sup>22</sup>). During zebrafish maternal to zygotic transition, it was described that two TSS-defining grammars coexist, in core promoters of constitutively expressed genes to enable their expression in the two regulatory environments<sup>23</sup>. Maternally active promoters in zebrafish tend to be sharp, with TATA-like, AT-rich (W-box) upstream elements guiding TSS selection, while embryonically active broad promoter architectures of the same genes appear to be regulated by nucleosome positioning. Although a number of germ cell-specific, as well as somatic transcriptional regulators, have been well characterised during folliculogenesis (reviewed in ref. <sup>24</sup>), the exact actors and mechanisms required for setting up the oocyte-specific transcriptome have not yet been identified in vertebrates.

Female germ cells develop during oogenesis leading to the formation of a highly differentiated and specialised cell, the oocyte. In females, oocytes enter meiosis during embryonic life. Quiescent primordial follicles composed of meiotically arrested oocytes at the late diplotene stage and surrounded by granulosa cells are formed perinatally in mice (reviewed in ref. <sup>24</sup>). Shortly after birth, some primordial follicles enter folliculogenesis and undertake a growth phase during which a specific oocyte-specific transcriptome is set up, and oocytes increase their size until the pre-antral follicular stage<sup>25</sup>. A remarkable feature of oocytes is the

very high expression of retrotransposons driven by Pol II transcription. These elements are interspersed with repetitive elements that can be mobile in the genome. One of the three major classes of retrotransposons in mammals is the long terminal repeat (LTR) retrotransposons derived from retroviruses, also known as endogenous retroviruses (ERVs) that is subdivided in three sub-classes: ERV1, ERVK and endogenous retrovirus-like ERVL-MaLR (mammalian apparent LTR retrotransposons) (reviewed in ref. <sup>26</sup>). Transcription of mobile elements in specific cell types depends on the presence of a competent promoter recognition transcription machinery and/or the epigenetic status of the loci where these elements have been incorporated. Remarkably, MaLRs encode no known proteins, but MaLR-dependent transcription is key in initiating synchronous developmentally regulated transcription to reprogramme the oocyte genome during growth<sup>27</sup>.

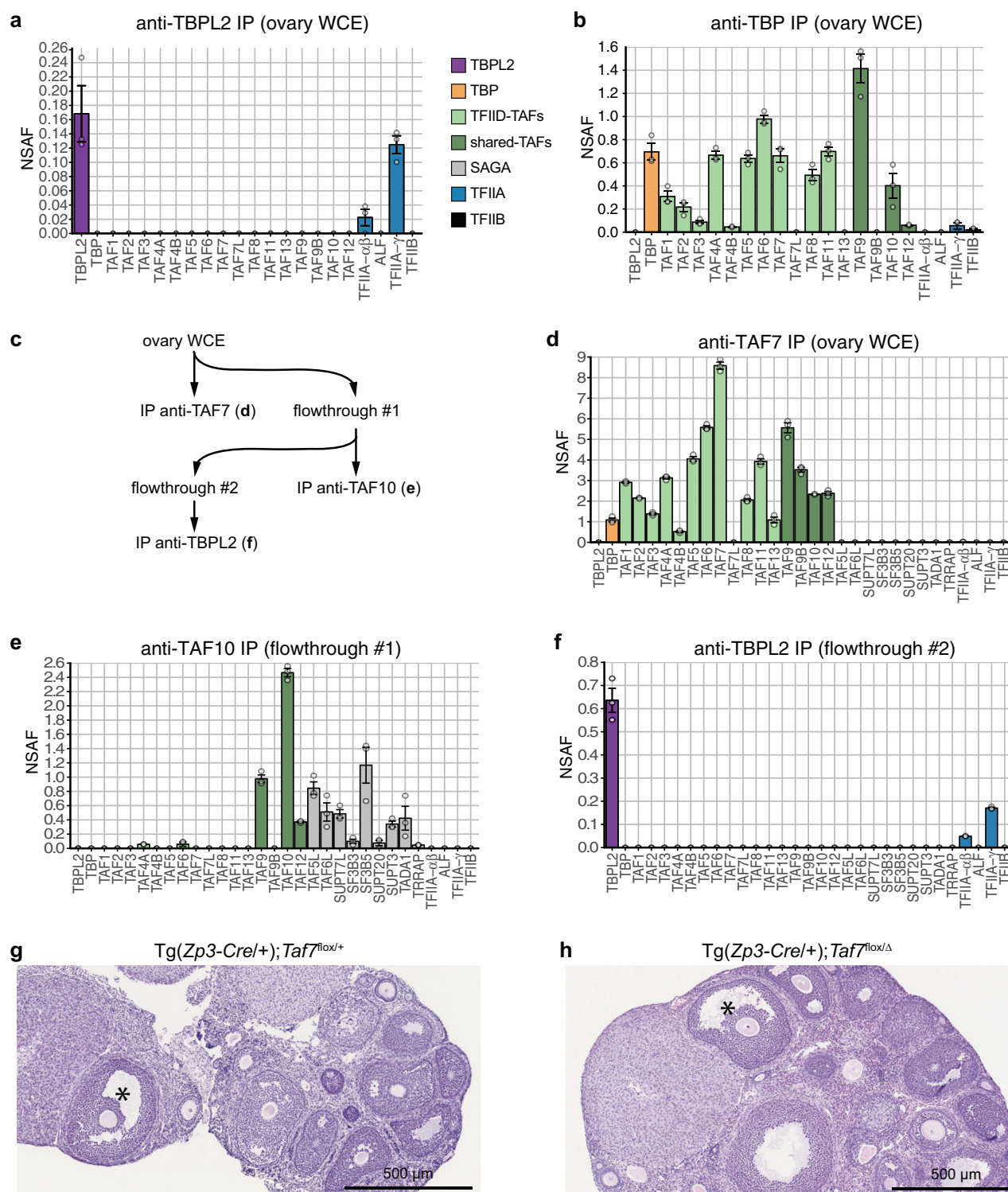
Remarkably, during oocyte growth, TBP protein is absent and replaced by TBPL2<sup>28</sup>. Indeed, TBP is only expressed up to the primordial follicular oocytes and becomes undetectable at all subsequent stages during oocyte growth. In contrast, TBPL2 is highly expressed in the growing oocytes, suggesting that TBPL2 is replacing TBP for its transcription initiating functions during folliculogenesis<sup>28</sup>. In agreement with its oocyte-specific expression, a crucial role of TBPL2 for oogenesis was demonstrated in *Tbpl2*<sup>−/−</sup> females, which show sterility due to defect in secondary follicle production<sup>16,29</sup>. In the absence of TBPL2, immunofluorescent staining experiments showed that elongating Pol II and histone H3K4me3 methylation signals were abolished between the primary and secondary follicle stage oocytes, suggesting that Pol II transcription was impaired<sup>16</sup>. Initially, TBPL2/TRF3 was suggested to be expressed during muscle differentiation<sup>30</sup>, but this observation was later invalidated<sup>16,29</sup>. Altogether, the available data suggested that TBPL2 is playing a specialised role during mouse oocyte development. However, how does TBPL2 regulate oocyte-specific transcription and what is the composition of the associated transcription machinery, remained unknown.

Here, we demonstrate that in oocytes TBPL2 does not assemble into a canonical TFIID complex, while it stably associates with TFIIA. The observation that the oocyte-specific deletion of *Taf7*, a TFIID-specific TAF, does not influence oocyte growth and maturation, corroborates the lack of TFIID in growing oocytes. Our transcriptomics analyses in wild-type and *Tbpl2*<sup>−/−</sup> oocytes show that TBPL2 mediates transcription of oocyte-expressed genes, including mRNA destabilisation factor genes, as well as MaLR ERVs. Our transcription start site (TSS) mapping from wild-type and *Tbpl2*<sup>−/−</sup> growing oocytes demonstrates that TBPL2 has a strong preference for TATA-like motif in gene core promoters driving specific sharp TSS selection. This is in marked contrast with TBP/TFIID-driven TATA-less gene promoters in preceding stages that have broad TSS architecture. Our results show a role for the TBPL2-TFIIA transcription machinery in a major transition of the oocyte transcriptome mirroring the maternal to zygotic transition that occurs after fertilisation, completing a full germline cycle.

## Results

### Oocyte-specific TBPL2/TFIIA complex distinct from TFIID.

To characterise TBPL2-containing transcription complexes, we prepared whole-cell extracts (WCE) from 14 days postnatal (P14) mouse ovaries and analysed TBPL2-associated proteins by anti-TBPL2 immunoprecipitation (IP) coupled to label-free mass spectrometry (Fig. 1a and Supplementary Fig. 1a, b). To determine the stoichiometry of the composition of the immunoprecipitated complexes, normalised spectral abundance factor



**Fig. 1 TBPL2 does not assemble in a TFIID-like complex in growing oocytes.** **a** Anti-TBPL2 immunoprecipitation followed by mass spectrometry (IP-MS) analysis from three biological replicates of mouse ovarian whole-cell extracts (WCE). The colour code for the different proteins or complexes is indicated on the right. NSAF; normalised spectral abundance factor. **b** Anti-TBP IP-MS from ovarian WCE (three technical replicates). The colour code is the same as in **(a)**. **c-f** Sequential IP-MS experiment from ovarian WCE (three technical replicates). The strategy of the sequential immunoprecipitation **(c)**, anti-TAF7 IP-MS **(d)**, followed by an anti-TAF10 IP-MS **(e)** and then an anti-TBPL2 IP-MS **(f)**. The colour code is the same as in **(a)**. **g, h** Representative views of haematoxylin and eosin-stained ovaries sections from control **(g)** and oocyte-specific *Taf7* mutant (*Tg(Zp3-Cre/+); Taf7<sup>fllox/Δ</sup>*, **h**) ovaries (analyses of three sections from three biological replicates). The presence of antral follicles is indicated by an asterisk. Scale bars: 500 μm. In **(a, b, d-f)** grey dots indicate replicates and error bars, +/– standard error of the mean.



(NSAF) values were calculated<sup>31</sup>. In the anti-TBPL2 IPs, we identified TFIIA- $\alpha\beta$  and TFIIA- $\gamma$  subunits as unique GTF subunits associated with TBPL2 (Fig. 1a and Supplementary Data 1). As ovaries contain many other non-oocyte cell types that express TBP, in parallel from the same extracts we carried out an anti-TBP IP. The mass spectrometry of the anti-TBP IP indicated that TBP assembles into the canonical TFIID complex in non-oocyte cells (Fig. 1b and Supplementary Data 2). As growing oocytes represent only a tiny minority of ovary cells, we further tested the TBPL2-TFIIA interaction by a triple IP strategy (Fig. 1c): first, we depleted TAF7-containing TFIID complexes with an anti-TAF7 IP; second, the remaining TFIID and SAGA complexes, which contain also shared TAFs<sup>32</sup>, were depleted with an anti-TAF10 IP using the anti-TAF7 IP flow-through as input; third, we performed an anti-TBPL2 IP on the anti-TAF7/anti-TAF10 flow-through fraction (Fig. 1d–f and Supplementary Data 3). The analysis of this third consecutive IP further demonstrated that TBPL2 forms a unique complex with TFIIA- $\alpha\beta$ , and TFIIFA- $\gamma$ , but without any TFIID subunits.

To further analyse the requirement of TFIID during oocyte growth, we carried out a conditional depletion of the TFIID-specific *Taf7* gene during oocyte growth using the *Zp3-Cre* transgenic line<sup>33</sup> (Supplementary Fig. 1c–g). Remarkably, TAF7 is only detected in the cytoplasm of growing oocytes (Supplementary Fig. 1c). The oocyte-specific deletion of *Taf7* did not affect the presence of secondary and antral follicles and the numbers of collected mature oocytes after superovulation (Fig. 1g, h and Supplementary Fig. 1f). The lack of phenotype is not due to an inefficient deletion of *Taf7*, as TAF7 immunolocalization is impaired (Supplementary Fig. 1d, e), and as oocyte-specific *Taf7* mutant females are severely hypofertile (Supplementary Fig. 1g). The observations that TBP is not expressed in growing oocytes, and that the oocyte-specific deletion of *Taf7* abolishes the cytoplasmic localisation of TAF7, but does not influence oocyte growth, show that canonical TFIID does not assemble in the nuclei of growing oocytes. Thus, our results together demonstrate that during oocyte growth a stable TBPL2-TFIIA complex forms, and may function differently from TBP/TFIID.

In order to further characterise the composition of the TBPL2-TFIIA complex, we took advantage of NIH3T3 cells artificially overexpressing TBPL2 (NIH3T3-II10 cells<sup>28</sup>). In this context where TBP and TAFs are present, TFIID is efficiently pulled down by an anti-TBP IP, but no interaction with TFIIA could be detected (Fig. 2b). Interestingly, the anti-TBPL2 IP showed that the artificially expressed TBPL2 can incorporate in TFIID-like complexes as TAFs were co-IP-ed (Fig. 2a); however, with much lower stoichiometry (NSAF values) than that of TBP (Fig. 2b). In contrast, strong interaction with TFIIA- $\alpha\beta$  and TFIIFA- $\gamma$  were detected, suggesting that the TBPL2-TFIIA complex can be formed in the NIH3T3-II10 cells and that TBPL2, to the contrary to TBP has the intrinsic ability to interact with TFIIA. Remarkably, in spite of the high similarity between the core domains of TBP and TBPL2, no interaction with Pol I-associated SL1 (TAF1A–D) and Pol III-associated TFIIB (BRF1) complexes<sup>34</sup> could be detected in the anti-TBPL2 IPs either in NIH3T3-II10 cells or in ovary WCEs (Fig. 2a and Supplementary Data 1). In contrary, in the same extracts TBP associates with these Pol I and Pol III complexes (Fig. 2b and Supplementary Data 2), suggesting that TBPL2 is not involved in Pol I and Pol III transcription initiation in the growing oocytes.

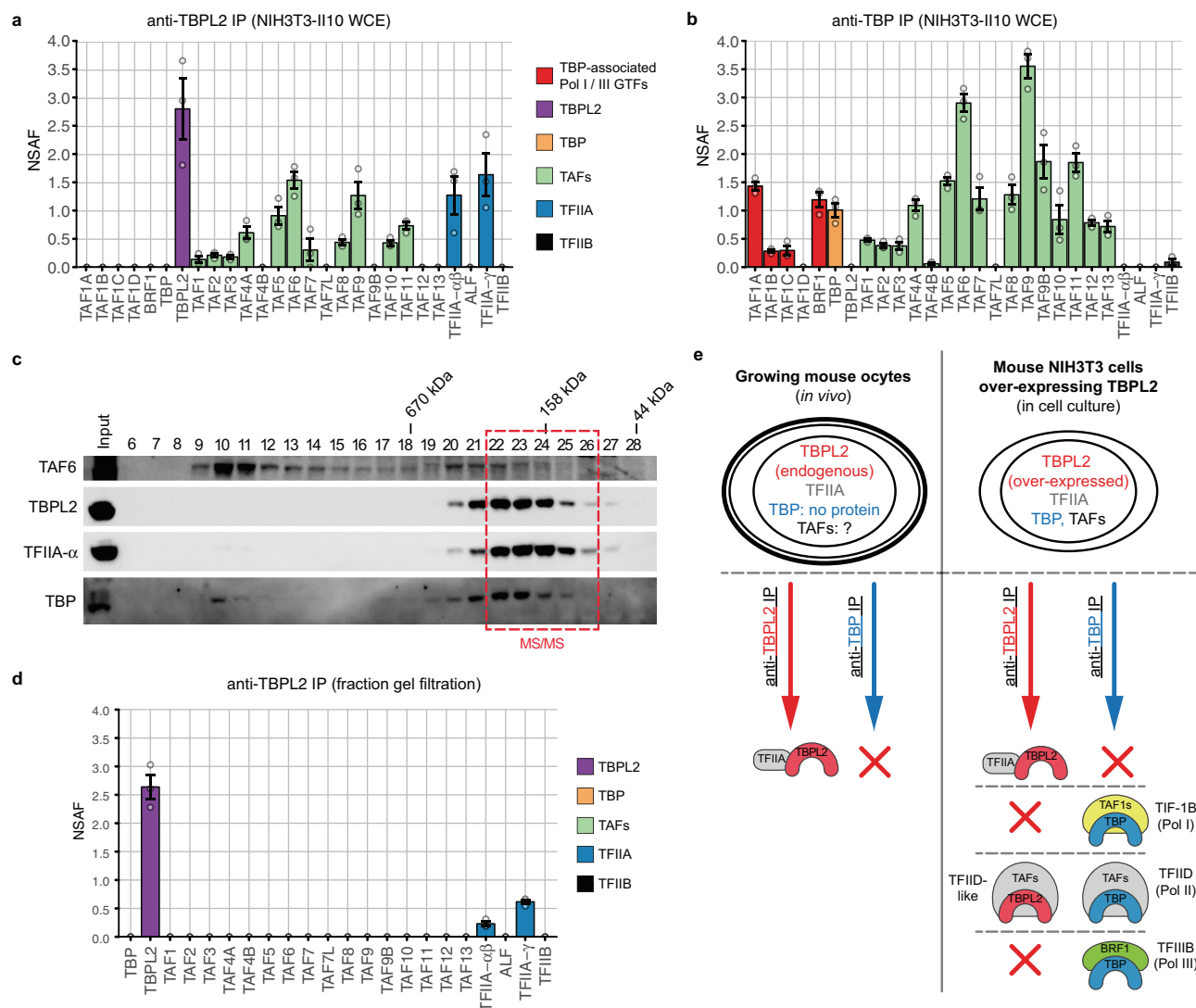
To analyse whether TBPL2 associates with TFIID TAFs and TFIIA in the same complex, we performed a gel filtration analysis of NIH3T3-II10 WCE. The profile indicated that most of the TBPL2 and TFIIA could be found in the same fractions (22–26) eluting around 150–200 kDa, while TBPL2 protein was below the detection threshold of the western blot assay in the TAF6-

containing fractions 9–15 (Fig. 2c). To verify that TBPL2 and TFIIA are part of the same complex in fractions 22–26, we IP-ed TBPL2 from these pooled fractions and subjected them to mass spectrometric analysis. Our data confirmed that in these fractions eluting around 170 kDa, TBPL2 and TFIIA form a stable complex that does not contain any TAFs (Fig. 2d and Supplementary Data 4). Thus, all these experiments together demonstrate that TBPL2/TFIIA form a stable complex in oocytes, where TBP is not expressed and TBPL2/TFIIA is the only promoter recognising transcription complex that could direct Pol II transcription initiation (see the summary of all the IPs in Fig. 2e).

**TBPL2-dependent oocyte transcriptome.** To characterise the growing oocyte-specific transcriptome and its dependence on TBPL2, we have performed a transcriptomic analysis of wild-type (WT) and *Tbpl2*<sup>-/-</sup> oocytes isolated from primary (P7) and secondary (P14) follicles (Figs. 3, 4 and Supplementary Fig. 2, Supplementary Data 5). We observed the downregulation of a high number of oocyte-specific genes, out of which *Bmp15* and *Gdf9* served as internal controls<sup>35,36</sup>, as they were already described to be regulated by TBPL2<sup>16</sup> (Fig. 3a, b and Supplementary Fig. 2a). The principal component analysis showed that the four distinct RNA samples clustered in individual groups and that the main explanation for the variance is the genotype, and then the stage (Supplementary Fig. 2b). Comparison of the RNA-level fold changes between mutant and WT oocytes showed that in *Tbpl2*<sup>-/-</sup>, there is a massive downregulation of the most highly expressed transcripts, both at P7 and P14 (Supplementary Fig. 2c). The Pearson correlation between the P7 and P14 fold change datasets for transcripts expressed above 100 normalised reads was close to 0.8 (Supplementary Fig. 2c), indicating that *Tbpl2* loss of function similarly altered RNA levels at P7 and P14 stages. We, therefore, focused on the P14 stage for the rest of the study.

In WT P14 oocytes transcripts corresponding to 10791 genes were detected. Importantly, many of these detected transcripts have been transcribed at earlier stages and are stored in growing oocytes<sup>37</sup>. As there is no Pol II transcription in *Tbpl2*<sup>-/-</sup> growing oocytes<sup>16</sup>, RNAs detected in the *Tbpl2*<sup>-/-</sup> mutant oocytes represent mRNAs transcribed by a TBP/TFIID-dependent mechanism and deposited into the growing oocytes independently of TBPL2 activity at earlier stages, i.e., at the primordial follicular stage, where TBP is still expressed. The proportion of genes (1396) upregulated following *Tbpl2* deletion (Fig. 3c) can be explained by two mutually not exclusive ways: (i) the consequence of the normalisation to the library size resulting in a slight overestimation of upregulated transcripts, and underestimation of downregulated transcripts and/or (ii) by transcript buffering mechanisms due to mRNA stabilisation<sup>38</sup>. Validation of the upregulation of some candidate transcripts levels (Supplementary Fig. 2d, e) strongly supports the latter hypothesis (but see also the next paragraph).

Nevertheless, we detected 1802 significantly downregulated transcripts in the *Tbpl2*<sup>-/-</sup> oocytes (Fig. 3c). The downregulation of key genes known to be expressed during oocyte growth, such as *Bmp15*, *Eloc*, *Fgf8*, *Gdf9* and *Zar1*<sup>35,36,39</sup>, were confirmed by RT-qPCR (Supplementary Fig. 2f, g). These results suggest that TBPL2 has an important role in gene expression in the growing oocytes. Gene Ontology (GO) analyses of the biological process of the identified downregulated categories of genes (Supplementary Data 6) indicated that many genes, involved in meiosis II and distinct cell cycle processes, were significantly downregulated (Supplementary Fig. 2h). The most enriched molecular function GO category was “poly(A)-specific ribonuclease activity” containing many genes coding for factors or subunits of complexes



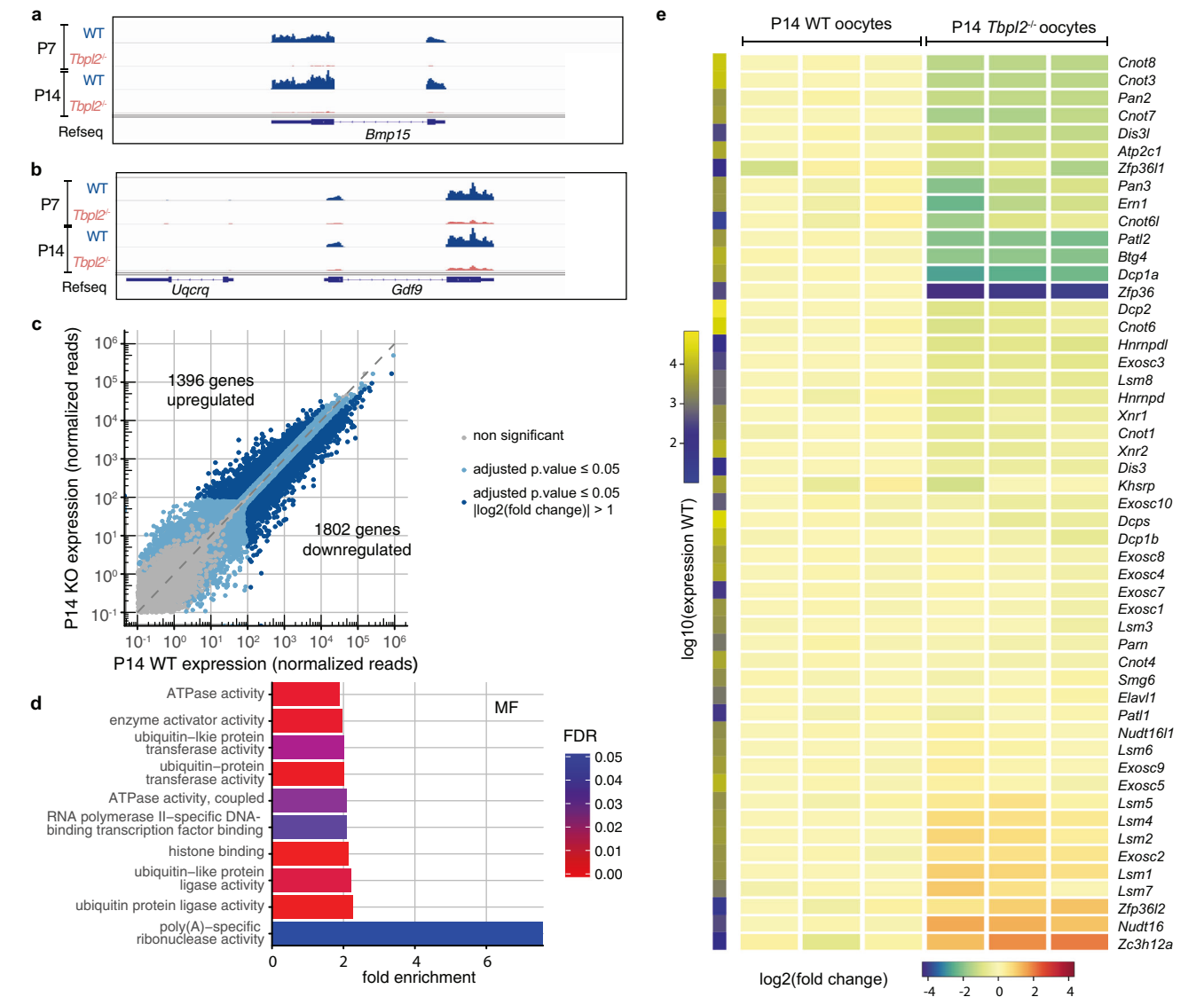
**Fig. 2** TBPL2 assembles into a TBPL2/TFIIA and TFIID-like complexes in NIH3T3 cells overexpressing TBPL2. **a** Anti-TBPL2 immunoprecipitation followed by mass spectrometry (IP-MS) analysis (three technical replicates) of NIH3T3 overexpressing TBPL2 (NIH3T3-II10) whole-cell extracts (WCE). The colour code for the different proteins or complexes is indicated on the right. **b** Anti-TBP IP-MS analysis (three technical replicates) of NIH3T3-II10 WCE. Colour legend for the different proteins is the same as in **(a)**. **c** Western blot of a Superose 6 gel filtration analysis of NIH3T3-II10 WCE probed with anti-TAF6 (top panel), anti-TBPL2 and anti-TFIIA- $\alpha$  (middle panels) and anti-TBP (bottom panel) antibodies. Fraction numbers are shown above each lane, and the elution of known molecular mass markers is indicated above the panels. The pooled fractions used for mass spectrometry analysis are indicated in red. **d** Anti-TBPL2 IP-MS analysis (three technical replicates) of the gel filtration fraction indicated in **(c)**. The colour code for the different proteins or complexes is indicated on the right. NSAF normalised spectral abundance factor. **e** Schematic representation of the fundamental differences existing between TBPL2- and TBP-containing complexes in growing oocytes and NIH3T3-II10 cells. In **(a, b, and d)** grey dots indicate replicates and error bars,  $\pm$  standard error of the mean.

contributing to deadenylation/decapping/decay activity in eukaryotes (Fig. 3d) (i.e., CCR4-NOT, PAN2/PAN3<sup>40</sup>, DCP1A/DCP2<sup>41</sup> or BTG4<sup>39</sup>). In good agreement with the transcriptome analyses, transcripts coding for these “poly(A)-specific ribonuclease activity” factors were significantly downregulated in *Tbpl2*<sup>-/-</sup> mutant P14 oocytes when tested by RT-qPCR (Fig. 3e and Supplementary Fig. 2i). Thus, in P14 oocytes TBPL2 is regulating the transcription of many genes coding for factors, which are in turn crucial in regulating the stability and translation of the mRNA stock deposited during early oogenesis, as well as transcription of meiosis II- and cell cycle-related genes to prepare the growing oocytes for the upcoming meiotic cell division.

A remarkable feature of oocytes is the very high expression of retrotransposons driven by Pol II transcription (see “Introduction”). As expected, in WT P7 and P14 oocytes, the expression of

ERVs was found to be the most abundant<sup>27,42</sup> (Supplementary Fig. 3a–c). Importantly, the transcription of the vast majority of MaLR elements was the most affected in *Tbpl2*<sup>-/-</sup> mutant oocytes at P7 and P14 (Fig. 4). Among them, three highly expressed members, *MT-int*, *MTA\_Mm* and *MTA\_Mm-int*, were dramatically downregulated in P7 and P14 *Tbpl2*<sup>-/-</sup> mutant oocytes (Supplementary Fig. 3d, e). As in P14 oocytes, TBPL2 depletion is reducing transcription more than fourfold from MaLR ERVs, which often serve as promoters for neighbouring genes<sup>27,42</sup>, TBPL2 could seriously deregulate oocyte-specific transcription and consequent genome activation.

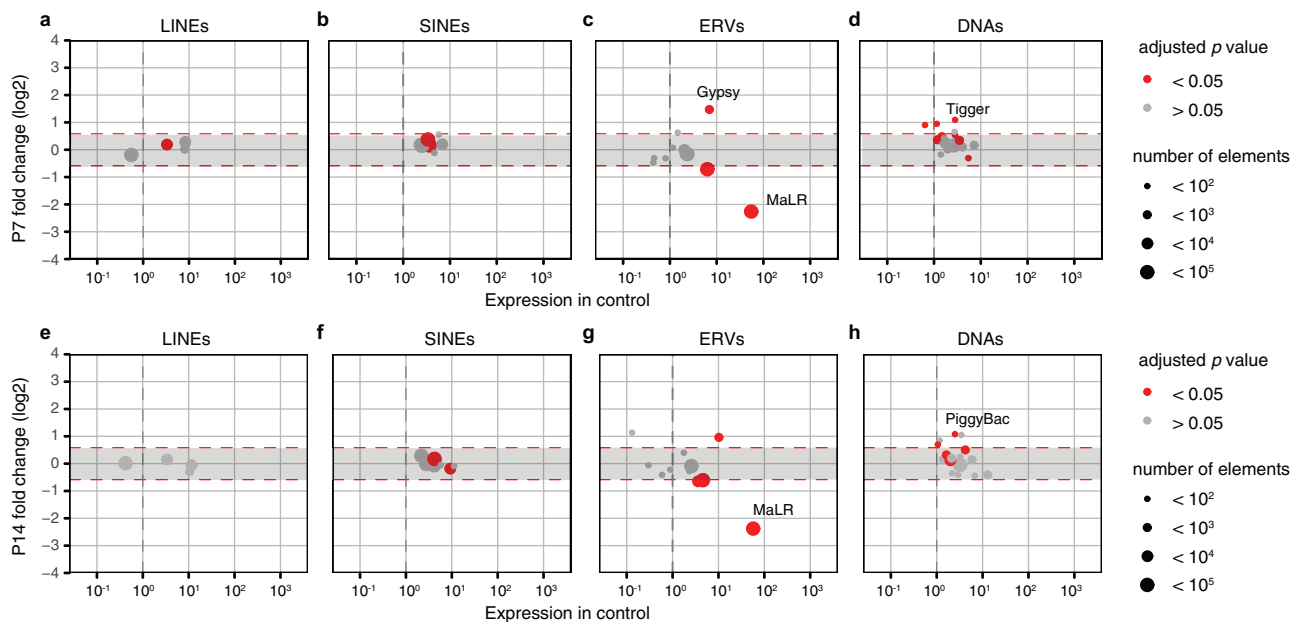
This demonstrates that TBPL2 is orchestrating the de novo restructuring of the maternal transcriptome and that TBPL2 is crucial for indirectly silencing the translation of the earlier deposited TBP-dependent transcripts.



**Fig. 3 Expression of genes related to the mRNA deadenylation/decapping/decay pathways in growing *Tbp12*<sup>-/-</sup> mutant oocytes.** **a, b** Normalised Integrative Genomic Viewer (IGV) snapshots of *Bmp15* (**a**) and *Gdf9* (**b**) loci. Exons and introns are indicated. **c** Expression comparison between wild-type (WT) and *Tbp12*<sup>-/-</sup> mutant postnatal day 14 (P14) oocytes (biological triplicates). Expression was normalised to the median size of the transcripts in kb. Grey dots correspond to non-significant genes and genes with high Cook's distance, light-blue dots to significant genes for an adjusted *P* value ≤ 0.05 and dark-blue dots to significant genes for an adjusted *P* value ≤ 0.05 and an absolute log<sub>2</sub> fold change > 1, after two-sided Wald test and Benjamini-Hochberg correction for multiple comparisons. The number of up- or downregulated genes is indicated on the graph. **d** Downregulated genes GO category analyses for the molecular functions (MF). The top ten most enriched significant GO categories for a FDR ≤ 0.05 are represented. **e** Heatmap of selected genes involved in mRNA decay, decapping or deadenylation pathways. Expression levels in fold change (compared to the mean of WT) of three biological replicates of P14 WT and P14 *Tbp12*<sup>-/-</sup> mutant oocytes are indicated. The fold change colour legend is indicated at the bottom. The first column on the left corresponds to the log<sub>10</sub> of expression, scale on the left.

**TBPL2-driven promoters contain TATA box and are sharp.** The promoter usage changes during zebrafish maternal to zygotic transition revealing different rules of transcriptional initiation in oocyte and in embryo, driven by independent and often overlapping sets of promoter “codes”<sup>23</sup>. Importantly, this switch has not yet been demonstrated in mammals and the role of TBPL2 in this switch during oogenesis remained to be investigated. To this end, we mapped the TSS usage by carrying out super-low input carrier-CAGE (SLIC-CAGE)<sup>43</sup> from WT and *Tbp12*<sup>-/-</sup> P14 oocytes. To characterise only the TBPL2-driven promoters, we removed the CAGE tags present in the *Tbp12*<sup>-/-</sup> dataset from the WT P14 dataset, to eliminate transcripts that have been deposited at earlier stages (hereafter

called “TBPL2-dependent”). Conversely, the *Tbp12*<sup>-/-</sup> dataset corresponds to the TBP/TFIID-dependent, or TBPL2-independent TSSs (hereafter called “TBPL2-independent”). Next, we analysed the genome-wide enrichment of T- and/or A-rich (WW) dinucleotide motifs within the -250/+250 region centred on the dominant TSSs of the TBPL2-dependent and TBPL2-independent oocyte TSS clusters (Fig. 5a, b). TBPL2-dependent TSS clusters are strongly enriched in a well-defined WW motif around their -30 bp region (Fig. 5a, red arrowhead)<sup>44</sup>. In contrast, only about 1/3rd of the TBPL2-independent TSS clusters contained WW-enriched motifs at a similar position (Fig. 5b, red arrowhead), as would be expected from promoters that lack maternal promoter code determinants<sup>23,44</sup>. As canonical



**Fig. 4 Expression of the mammalian apparent LTR retrotransposons (MaLR) endogenous retroviral elements (ERV) are downregulated in growing *Tbp12*<sup>-/-</sup> mutant oocytes.** **a–h** Differential expression between wild-type and *Tbp12*<sup>-/-</sup> mutant postnatal day (P) 7 (**a–d**) and P14 (**e–h**) oocytes of the different transposon classes; RNA transposon classes [LINEs (**a, e**), SINEs (**b, f**) and ERVs (**c, g**)] and DNA transposons (DNAs (**d, h**)). The ERV sub-class III mammalian apparent LTR retrotransposon (MaLR) family is the most severely affected in *Tbp12*<sup>-/-</sup> mutant oocytes at P7 and P14. Dot size and colour are explained on the right side, adjusted *P* value after two-sided Wald test and Benjamini-Hochberg correction for multiple comparisons.

TATA boxes are often associated with tissue-specific gene promoters, we investigated whether the above observed WW motif densities correspond to TATA boxes using the TBP position weight matrix (PWM) from the JASPAR database as a reference. To this end, the presence of TATA boxes was analysed in the TSS clusters of the two datasets and revealed that TBPL2-dependent TSS clusters were enriched in high-quality TATA boxes, including a clear increase in the proportion of canonical TATA boxes, when compared to TBPL2-independent TSS clusters (Fig. 5c). Genome browser view snapshots indicate that TSS clusters in P14 WT oocytes tend to be sharp and are associated with TATA-like motifs (Supplementary Fig. 4a, b). Analysis of the global distribution of the number of TSSs and of the width of the TSS clusters in the above-defined two categories confirmed that TBPL2-dependent TSS are sharper compared to the TBPL2-independent TSS clusters (Supplementary Fig. 4c, d).

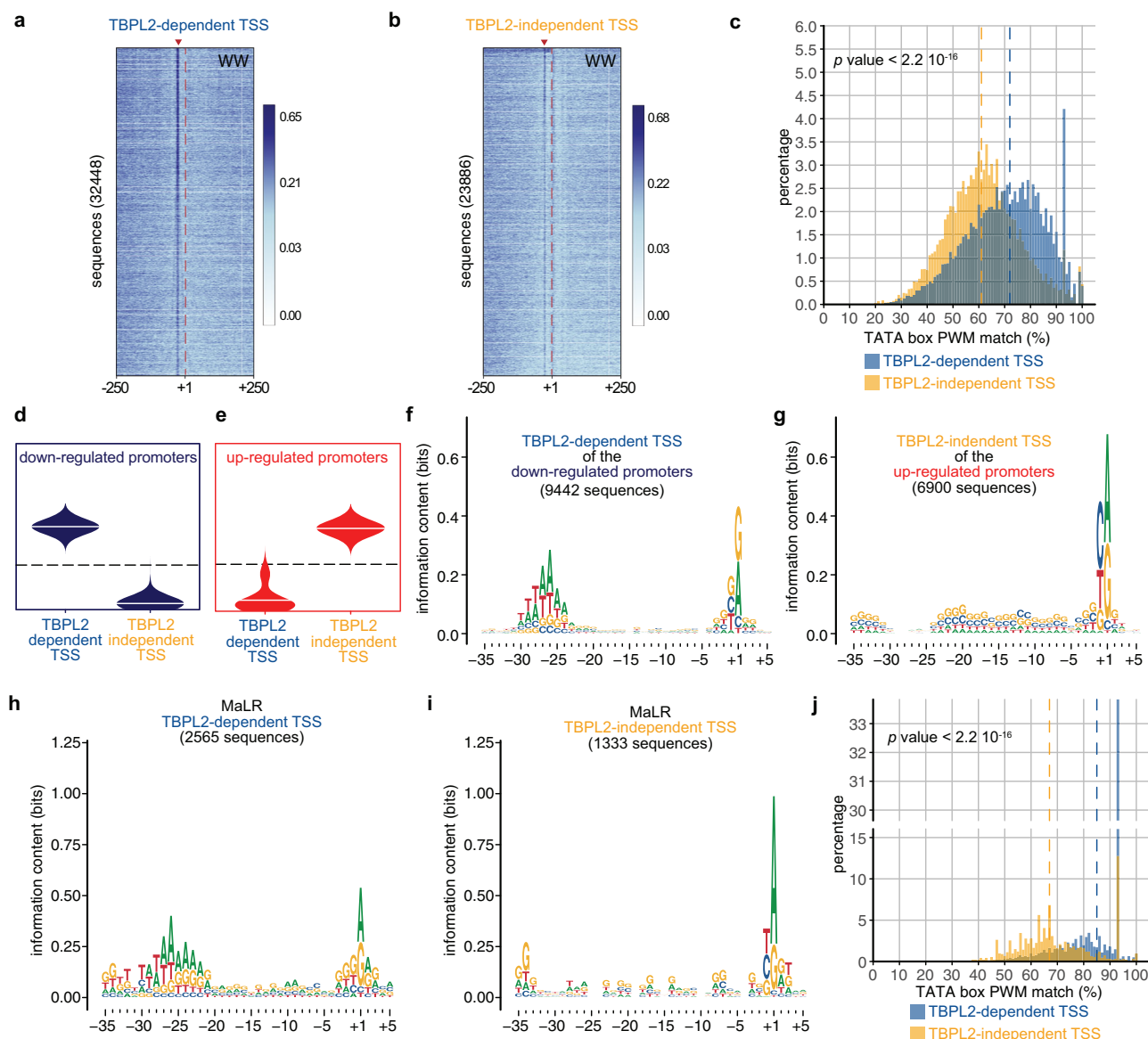
In order to test whether TBPL2 controls transcription initiation from maternal promoter code determinants, we grouped the expression profiles corresponding to each consensus TSS clusters, to characterise promoter activity profiles among datasets by performing self-organising maps (SOMs)<sup>45</sup> (Supplementary Fig. 4e). We then focussed on the two most distinct SOM groups: the downregulated promoters (blue group, containing 9442 consensus TSS clusters) (Fig. 5d) and the upregulated promoters (red group, with 6900 consensus TSS clusters) (Fig. 5e). Motif analyses of these two categories of promoters in their -35/+5 regions relative to the different dominant TSSs indicated that only the core promoters associated with TBPL2-dependent dominant TSSs belonging to the downregulated gene promoters contain a well-defined 7 bp long TATA box-like motif (W-box) in their -31 to -24 regions (Fig. 5f, g and Supplementary Fig. 4f–i). Importantly, W-box-associated TSSs architecture usage distribution for these TBPL2-dependent dominant TSSs was sharp (Supplementary Fig. 4j, l), as expected for motif-dependent transcriptional initiation<sup>23,44</sup>. In contrast, TBPL2-independent TSSs belonging to the upregulated promoters exert a much

broader TSS pattern (Supplementary Fig. 4k, m). Interestingly, GO analyses of the genes associated with the downregulated promoters revealed a strong association with deadenylation/decapping/decay activity (Supplementary Fig. 4n–p, Supplementary Data 7), further confirming our initial RNA-seq analysis observations (Fig. 3).

Importantly, TSS architecture analyses of the TBPL2-dependent MaLR ERV TSSs indicated that the majority of MaLR core promoters contain high-quality TATA box motif (median of the TATA box PWM match is 85%, Fig. 5h–j). These observations together demonstrate that the TBPL2/TFIIA complex drives transcription initiation primarily from core promoters that contain a TATA box-like motif in their core promoter and directs sharp transcription initiation from the corresponding promoter regions to overhaul the growing oocyte transcriptome.

In addition, we observed that TSS usage can shift within the promoter of individual genes depending on the genetic background (Supplementary Fig. 4b). To get more insights into these promoter architecture differences, we identified genome-wide 6429 shifting promoters by comparing either TBPL2-dependent to TBPL2-independent TSS data. These results are consistent with TSS shifts between TBP/TFIID-dependent somatic-like and maternal promoter codes occurring either in 5' or 3' directions (Fig. 6a and Supplementary Fig. 4q)<sup>44</sup>. WW motif analysis indicated that on each shifting promoter, TBPL2-dependent dominant TSSs are associated with WW motifs, while TBPL2-independent dominant TSSs are not (Fig. 6b). In addition, the TATA box PWM match analyses indicated that these WW motifs are enriched in TATA box-like elements compared to the corresponding TBPL2-independent shifting TSSs (Fig. 6c). Thus, our experiments provide a direct demonstration that TBP/TFIID and TBPL2/TFIIA machineries recognise two distinct sequences co-existing in promoters of the same genes with TBPL2 directing a stronger WW/TATA box-dependent sharp TSS selection in them.



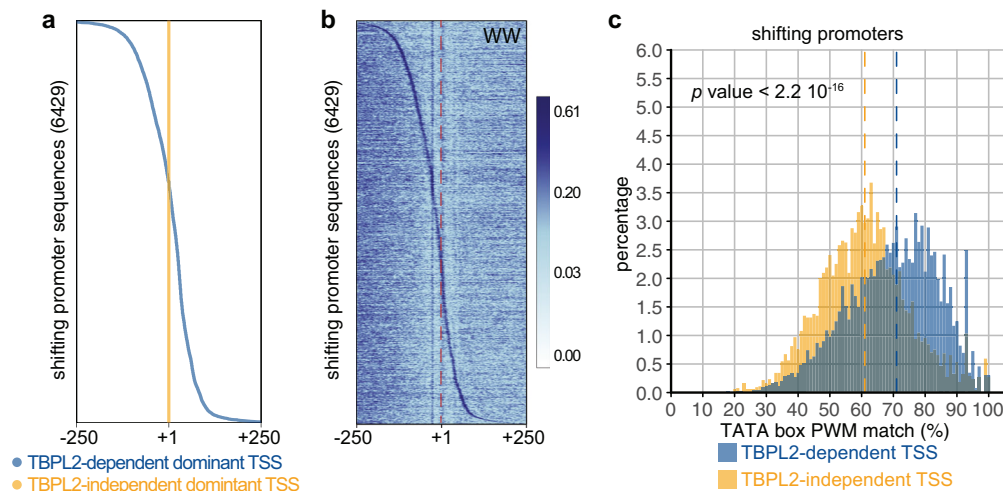


**Fig. 5 Core promoter regions of TBPL2-specific transcription units in postnatal day 14 oocytes are enriched in TATA-like elements and are sharp.** **a, b** Genome-wide A/T-rich dinucleotide (WW) motif analyses of  $-250/+250$  sequences centred on the dominant transcription start sites (TSS, position +1, dashed red line) of TBPL2-dependent (**a**  $n = 32,448$ ) and TBPL2-independent (**b**  $n = 23,886$ ) TSS clusters. Sequences have been ordered by increasing size of the interquartile width of each. The red arrowheads indicate the WW enrichment at position  $-30$  in the TBPL2-dependent TSS clusters (**a**) and the equivalent position in the TBPL2-independent TSS clusters (**b**). The number of TSS clusters is indicated in brackets. **c** Distribution of the best TATA box position weight matrix (PWM) matches within a  $-35$  to  $-20$  region upstream of the dominant TSSs (+1) of TBPL2-dependent (blue) compared to the TBPL2-independent (orange) TSS clusters. The dashed lines indicate the median of the TATA box PWM matches for the TBPL2-dependent (blue) and the TBPL2-independent (orange) TSS clusters ( $P$  value after a two-tailed Wilcoxon rank-sum test). **d, e** Two selected self-organising map (SOM) groups of the consensus TSS clusters: the downregulated promoters (blue, **d**) and the upregulated promoters (red, **e**) groups. **f, g** Sequence logos of the  $-35/+5$  sequence of the TBPL2-dependent dominant TSSs from the downregulated promoters (**f**) and of the TBPL2-independent dominant TSSs from the upregulated promoters (**g**). **h, i** Sequence logo of the  $-35/+5$  sequence of the MaLR ERVs TBPL2-dependent (**h**) and TBPL2-independent (**i**) dominant TSSs. **j** Distribution of the best TATA box PWM matches within a  $-35$  to  $-20$  region upstream of the TBPL2-dependent (blue) and TBPL2-independent (orange) mammalian apparent LTR retrotransposons (MaLR) endogenous retroviral elements (ERV) dominant TSS. The dashed lines indicate the median of the TATA box PWM matches for the TBPL2-dependent (blue) and the TBPL2-independent (orange) TSS clusters ( $P$  value after a two-tailed Wilcoxon rank-sum test).

## Discussion

In this study, we show that a unique basal transcription machinery composed of TBPL2 associated with TFIIA is controlling transcription initiation during oocyte growth, orchestrating a transcriptome change prior to fertilisation using an oocyte-specific TTS usage.

TBPL2 expression in mice is limited to the oocytes and in its absence, oocytes fail to grow and *Tbpl2*<sup>-/-</sup> mouse females are sterile<sup>16,28</sup>. In a mirroring situation, TBPL1 (TRF2) expression is enriched during spermatogenesis, and male germ cells lacking TBPL1 are blocked between the transition from late-round spermatids to early elongating spermatids<sup>14,15</sup>. An interesting



**Fig. 6 Shifting promoters.** **a, b** Analysis of the TBPL2-dependent versus TBPL2-independent shifting promoters within a  $-250/+250$  region centred on the position of the TBPL2-independent transcription start sites (TSS) clusters (position +1 in **(a, b)** and red dashed line in **(b)**). Position of the TBPL2-dependent (blue) and of the TBPL2-independent (orange) dominant TSSs for each shifting promoter sequence **(a)** and WW dinucleotide enrichment heatmap **(b)** from the same set of sequences ordered as in **(b)**. **c** Distribution of the best TATA box PWM matches within a  $-35$  to  $-20$  region upstream of the TBPL2-dependent (blue) and TBPL2-independent (orange) dominant TSS of the shifting promoters. The dashed lines indicate the median of the TATA box PWM matches for the TBPL2-dependent (blue) and the TBPL2-independent (orange) shifting TSS clusters ( $P$  value after a two-tailed Wilcoxon rank-sum test).

parallel between TBPL2 and TBPL1 is that both TBP-type factors form endogenous stable complexes with TFIIA. The beginning of TBPL2 accumulation in the oocyte nuclei or TBPL1 accumulation in male germ cell nuclei coincides with the phase of meiosis<sup>15,28,46</sup>. It is thus conceivable that TBPL2-TFIIA in oocytes or TBPL1-TFIIA during spermatogenesis are involved in the control of gene expression in a meiotic context to set up the corresponding transcriptome. Interestingly, both transcription complexes seem to function in a compacted chromatin environment in which TBP/TFIID probably cannot. However, while TBPL2 and TBP show contrasting expression patterns in the oocytes<sup>28</sup>, TBPL1 and TBP are co-expressed in spermatids<sup>46,47</sup> and it has been suggested that TBPL1 is a testis-specific subunit of TFIIA that is recruited to PIC containing TFIID and might not primarily act independently of TFIID/TBP to control gene expression in round spermatids<sup>48</sup>. While TBPL1 forms a complex also with the TFIIA- $\alpha$ B paralogue, ALF, in testis<sup>48–50</sup>, TBPL2 does not stably associate with ALF, in spite of the fact that it is expressed in oocytes<sup>50</sup>.

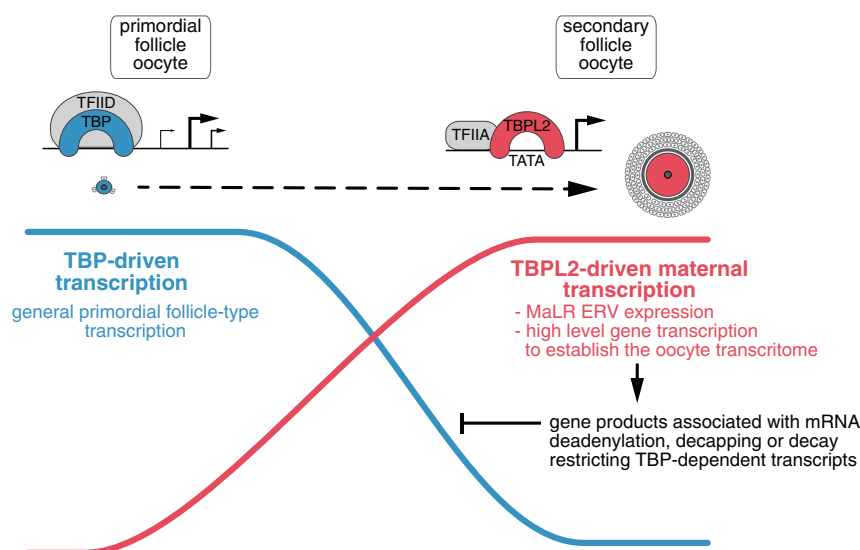
TBP-like factors are bipartite proteins with variable N-terminal domains and relatively well-conserved shared C-terminal domains (core domains) forming a saddle-like structure with a concave surface that is known to bind to DNA<sup>17</sup>. Interestingly, TBPL1 has a very short N-terminal domain<sup>5,18</sup>, suggesting that it lost some abilities to interact with partners. Our data suggest that despite their very high similarity (92% identity between the core domains of TBP and TBPL2; reviewed in ref. <sup>51</sup>), TBP and TBPL2 display different properties as they seem to recognise different DNA sequences to regulate gene promoters with different promoter architectures. Our IP-MS analyses from ovary WCE indicate that contrary to TBP, TBPL2 does not interact with TAFs in growing oocytes. Our analyses in the NIH3T3-II10 cells that overexpress TBPL2 showed that TBPL2 can interact with TAFs in this artificial situation, albeit with less affinity compared to TFIIA, or TBP-TAFs interactions. Our transcriptomic data indicate that all *Taf* mRNAs, except *Taf7l*, are detected in growing oocytes (Supplementary Data 5). However, whether they are also expressed in oocytes at the protein level is not yet known, except for TAF4B that has been detected in female neonate oocytes<sup>52</sup>.

Nevertheless, our data suggest that TAF7 is expressed, but localised to the cytoplasm. It is conceivable that, similarly to *Tbp* mRNA that is transcribed, but not translated in oocytes<sup>53</sup>, *Taf* mRNA translations (other than *Taf7*) are also inhibited and thus, the canonical TFIID, or its building blocks, cannot be assembled, and as a result, the canonical TFIID is not present in the nuclei of growing oocytes. Another reason why TBPL2 does not interact with TAFs or ALF, but rather interacts with TFIIA could be its N-terminal domain that is very different from that of TBP (only 23% identity<sup>51</sup>).

TBPL2 proteins from different vertebrates show a high degree of similarity in their C-terminal core domains amongst themselves, but display very little conservation in their N-terminal domains<sup>12</sup>. It is interesting to note that TBPL2 deficiency leads to embryonic phenotypes in *Xenopus*<sup>13</sup> and zebrafish<sup>12</sup>, because, contrary to the mouse, TBPL2 is still present in the embryo after fertilisation and thus may act in parallel with TBP in the transcription of specific embryonic genes<sup>10,54</sup>. The molecular mechanism by which TBPL2 controls the transcription of these specific sets of genes in frogs and in fish has not been studied. On the contrary, TBPL2 in mammals is only expressed in growing oocytes and the only phenotype that can be observed in mammals is female sterility<sup>16,29</sup>.

LTR retrotransposons, also known as ERVs, constitute ~10% of the mouse genome (reviewed in ref. <sup>55</sup>). While their expression is generally suppressed by DNA methylation and/or repressive histone modifications, a subset of ERV subfamilies retains transcriptional activity in specific cell types<sup>56</sup>. ERVs are especially active in germ cells and early embryos (reviewed in ref. <sup>26</sup>). Indeed, many genome-wide transcripts are initiated in LTRs, such as for example of MaLRs in mouse oocytes, which constitute ~5% of the genome<sup>57</sup>. Members of the MT subfamily of MaLRs are particularly active in oocytes and hundreds of MT LTRs have been co-opted as oocyte-specific gene promoters<sup>27,58</sup>. As LTR-initiated transcription units shape also the oocyte methylome, it will be important to analyse also how TBPL2 influences DNA methylation in oocytes.

Oocytes display remarkable post-transcriptional regulatory mechanisms that control mRNA stability and translation. During



**Fig. 7 Transcriptome overhaul controlled by TBPL2/TFIIA during oocyte growth.** At the beginning of oocyte growth, the transcriptome in primordial and early primary follicles (blue cell) depends on TFIID/TBP (blue complex) transcription from broad promoters (blue line). As TBP protein disappears, Pol II transcription initiation is mediated (red line) only by the oocyte-specific TFIIA/TBPL2 complex (red complex) from sharp promoters. At the growing oocyte stage (red cell) TFIIA/TBPL2 complex is responsible for initiating high levels of Pol II transcription, such as mammalian apparent LTR retrotransposons (MaLR) endogenous retroviral elements (ERV) expression and also the expression of genes of which the gene products survey/limit the stability and/or translatability of transcripts previously deposited by TFIID/TBP-dependent Pol II transcription.

oogenesis, the oocyte genome is transcriptionally active, and the newly synthesised maternal mRNAs are either translated or stored in a dormant form (reviewed in ref. <sup>37</sup>). The newly synthesised transcripts receive a long poly(A) tail and subsequently undergo poly(A) shortening in the oocyte cytoplasm, preventing translation. Until resumption of meiosis, mRNAs with a short poly(A) tail are stored in the cytoplasm in a dormant form (for a review, see ref. <sup>59</sup>). Thus, poly(A) tail deadenylation, amongst other activities, coordinates post-transcriptional regulation of the oocyte mRNA pool. Interestingly, TBPL2 is regulating the activity of several deadenylation/decapping/decay complexes and in the absence of TBPL2, we observed apparent stabilisation of a significant number of transcripts, suggesting that in wild-type oocytes TBPL2 is indirectly inhibiting the translation of mRNAs, and/or inducing the degradation of the mRNAs, previously transcribed by TFIID/TBP-driven Pol II and deposited in the primordial follicular oocytes (Fig. 7). To put in place the growing oocyte-specific maternal transcriptome TBPL2 is controlling the production of new mRNAs using a maternal-specific TSS grammar, as most of these transcripts will remain in the oocyte after transcriptional quiescence. Remarkably, as TBPL2 does not interact with Pol I and Pol III transcription machineries in the growing oocytes, this strongly suggests that rRNA and tRNA are deposited very early during oogenesis in amounts sufficient for the initiation of development.

Therefore, it seems that TBPL2 contributes to establish a novel TBPL2-dependent growing oocyte transcriptome and consequent proteome required for further development and oocyte competence for fertilisation (Fig. 7). The indirect regulation of previously deposited mRNAs by a global transcription regulator resembles the well-characterised maternal to zygotic transition (MZT), during which clearance of inherited transcriptome is mediated by de novo gene products generated by newly activated transcription machinery (reviewed in ref. <sup>59</sup>). At hundreds of gene promoters, two distinct TSS-defining “grammars” coexist in close proximity genome-wide and are differentially utilised either by TBPL2/TFIIA in primary/secondary follicular oocytes, or by TBP/TFIID in primordial follicular oocytes or in the fertilised

embryo. This again shows a striking parallel to MZT<sup>23</sup>, where multiple layers of information are embedded in the same promoter sequence, each representing a different type of regulatory grammar interpreted by dedicated transcription machinery depending on the cellular environment.

## Methods

**Cell lines and cell culture.** The NIH3T3-II10 line overexpressing TBPL2 and the control NIH3T3-K2 have already been described<sup>28</sup> and were maintained in high glucose DMEM supplemented with 10% of new-born calf serum at 37 °C in 5% CO<sub>2</sub>.

**Whole-cell extracts.** NIH3T3-II10 and NIH3T3-K2 cells cultured in 15-cm dish were washed twice with 1× PBS, subsequently harvested by scrapping on ice. Harvested cells were centrifuged 1000 × g at 4 °C for 5 min and then resuspended in one packed cell volume of whole-cell extraction buffer (20 mM Tris-HCl pH 7.5, 2 mM DTT, 20% glycerol, 400 mM KCl, 1× protease inhibitor cocktail (PIC, Roche)). Cell lysates were frozen in liquid nitrogen and thawed on ice three times, followed by centrifugation at 20,817 × g, at 4 °C for 15 min. The supernatant was collected, and protein concentration was measured by Bradford protein assay (Bio-Rad). The cell extracts were used directly for immunoprecipitation and western blot, or stored at –80 °C.

Ovaries collected from postnatal day 14 (P14) CD1 and C57BL/6N female mice were homogenised in whole-cell extraction buffer [20 mM Tris-HCl pH 7.5, 2 mM DTT, 20% glycerol, 400 mM KCl, 5× PIC (Roche)]. Cell lysates were frozen in liquid nitrogen and thawed on ice for three times, followed by centrifugation at 20,817 × g, at 4 °C for 15 min. The supernatant extracts were used directly for immunoprecipitation.

**Antibodies and antibody purification.** The antibodies are listed in Supplementary Table 1. The IGBMC antibody facility raised the anti-TBPL2 polyclonal 3024 serum against the CPDEHGSELNLSNNSPPDPQ peptide (amino acids 111–129) coupled to ovalbumin and injected into one 2-month-old female New-Zeland rabbit. The resulting serum was affinity purified by using the Sulfolink Coupling Gel (Pierce) following the manufacturer’s recommendations.

**Immunoprecipitation.** Ovary extract were incubated with anti-GST (10 µg per IP), anti-TBP (10 µg per IP), anti-TBPL2 (3024, 12 µg (36 µg for gel filtration) per IP), anti-TAF7 (10 µg per IP), or anti-TAF10 (10 µg per IP)-coated Dynabeads (Invitrogen) at 4 °C overnight. After incubation, beads were washed 3 × 5 min at 4 °C with 500 mM KCl buffer [25 mM Tris-HCl (pH 7.9), 5 mM MgCl<sub>2</sub>, 10% glycerol, 0.1% NP40, 2 mM DTT, 500 mM KCl and 1× PIC (Roche)], then washed 3 × 5 min at 4 °C with 100 mM KCl buffer (25 mM Tris-HCl pH 7.9, 5 mM MgCl<sub>2</sub>, 10% glycerol, 0.1% NP40, 2 mM DTT, 100 mM KCl and 1×). Immunoprecipitated



proteins were eluted with 0.1 M glycine pH 2.8 and neutralised with 1.5 M Tris-HCl pH 8.8.

Immunoprecipitation performed from whole-cell extracts of NIH3T3-III0 and NIH3T3-K2 cells were following the same procedures with protein G Sepharose beads (GE Healthcare): 18 µg of rabbit anti-TBPL2 (3024) and 15 µg anti-TBP per IP.

**Western blot analyses.** Protein samples (15–25 µg of cell extracts or 15 µL of IP elution) were mixed with 1/4th volume of loading buffer (100 mM Tris-HCl pH 6.8, 30% glycerol, 4% SDS, 0.2% bromophenol blue and freshly added 100 mM DTT) and boiled for 10 min. Samples were then resolved on a 10% SDS-PAGE and transferred to a nitrocellulose membrane (Protran, Amersham). Membranes were blocked in 3% non-fat milk in 1× PBS at room temperature (RT) for 30 min, and subsequently incubated with the primary antibody overnight at 4 °C (dilution 1/1000). Membranes were washed three times (10 min each) with 1× PBS—0.05% Tween-20. Membranes were then incubated with HRP-coupled goat anti-mouse Ig (Jackson ImmunoResearch, #115-036-071, dilution 1/10,000) or HRP-coupled goat anti-rabbit Ig (Jackson ImmunoResearch, #115-035-144, dilution 1/10,000) for 1 h at RT, followed by ECL detection (Thermo Fisher). The signal was acquired with the ChemiDoc imaging system (Bio-Rad).

**Mass spectrometry analyses and NSAF calculations.** Samples were TCA precipitated, reduced, alkylated, and digested with LysC and Trypsin at 37 °C overnight. After C18 desalting, samples were analysed using an Ultimate 3000 nano-RSLC (Thermo Scientific, San Jose, CA) coupled in line with a linear trap Quadrupole (LTQ)-Orbitrap ELITE mass spectrometer via a nano-electrospray ionisation source (Thermo Scientific). Peptide mixtures were loaded on a C18 Acclaim PepMap100 trap column (75-µm inner diameter × 2 cm, 3 µm, 100 Å; Thermo Fisher Scientific) for 3.5 min at 5 µL/min with 2% acetonitrile (ACN), 0.1% formic acid in H<sub>2</sub>O and then separated on a C18 Accucore nano-column (75-µm inner diameter × 50 cm, 2.6 µm, 150 Å; Thermo Fisher Scientific) with a 240-min linear gradient from 5% to 50% buffer B (A: 0.1% FA in H<sub>2</sub>O/B: 80% ACN, 0.08% FA in H<sub>2</sub>O) followed with 10 min at 99% B. The total duration was set to 280 min at a flow rate of 200 nL/min.

Proteins were identified by database searching using SequestHT with Proteome Discoverer 1.4 software (Thermo Fisher Scientific) a combined *Mus musculus* database generated using Uniprot [<https://www.uniprot.org/uniprot/?query=proteome:UP000000589&sort=score>] (Swissprot, release 2015\_11, 16730 entries) where five interesting proteins sequences (TrEMBL entries: TAF4, ATXN7L2, TADA2B, BTAF1 and SUPT3) were added. Precursor and fragment mass tolerances were set at 7 ppm and 0.5 Da, respectively, and up to two missed cleavages were allowed. Oxidation (M) was set as variable modification and Carbamidomethylation<sup>®</sup> as fixed modification. Peptides were filtered with a false discovery rate (FDR) at 5%, rank 1 and proteins were identified with one unique peptide. Normalised spectral abundance factor (NSAF)<sup>31</sup> were calculated using custom R scripts (R software version 3.5.3). Only proteins detected in at least two out of three of the technical or biological replicates were considered for further analyses.

**Gel filtration.** A Superose 6 (10/300) column was equilibrated with buffer consisting of 25 mM Tris-HCl pH 7.9, 5 mM MgCl<sub>2</sub>, 150 mM KCl, 5% Glycerol, 1 mM DTT and 1× PIC (Roche). Five hundred µL of whole-cell extracts containing ~5 mg of protein were injected in an ÄKTA avant chromatography system (Cytiva) and run at 0.4 mL/min. Protein detection was performed by absorbance at 280 nm and 260 nm. Five hundred µL fractions were collected and analysed by western blot and IP-MS.

**Animal experimentation.** Animal experimentations were carried out according to animal welfare regulations and guidelines of the French Ministry of Agriculture, and procedures were approved by the French Ministry for Higher Education and Research ethical committee C2EA-17 (project n°2018031209153651). The *Tg(Zp3-Cre)*, *Taf7<sup>fllox</sup>* and *Tbpl2<sup>-/-</sup>* mouse lines have already been described<sup>16,33,60</sup>.

**Histology analyses of ovaries.** Ovaries were collected from 6-week-old *Tg(Zp3-Cre/+)*, *Taf7<sup>fllox/+</sup>* and *Tg(Zp3-Cre/+);Taf7<sup>fllox/Δ</sup>* oocyte-specific mutant females, fixed in 4% paraformaldehyde (Electron Microscopy Sciences) overnight at 4 °C, washed three times in PBS at room temperature and embedded in paraffin. Five micrometre-thick sections were stained with haematoxylin and eosin, and images were acquired using a slide scanner Nanozoomer 2.0HT (Hamamatsu Photonics).

**Immunolocalization of TAF7 in the oocytes.** Ovaries were dissected in PBS, fixed overnight in 4% PFA/PBS at 4 °C, rinsed three times in PBS, equilibrated in 30% sucrose/PBS, and embedded in Cryomatrix (Thermo Scientific) in liquid nitrogen vapour. Fifteen micrometre-thick sections were obtained on a Leica cryostat and stored at -80 °C. Sections were rehydrated in TBS (50 mM Tris, 150 mM NaCl, pH 7.5), and permeabilized with 0.5% Triton-X-100 (Sigma) and rinsed twice again in TBS before blocking in 3% BSA, 1% goat serum, 0.1% Tween-20 (Sigma). Immunolabeling was then performed using M.O.M.<sup>®</sup> Immunodetection Kit, Basic (Vector Laboratories, BMK-2202). Purified anti-TAF7 rabbit polyclonal antibody

(dilution 1/300) was revealed using an Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen #A-11108, dilution 1/1000). Sections were counterstained with DAPI (4',6-diamidino-2-phenylindole dihydrochloride; Molecular Probes). Pictures were taken using a TCS SP5 Inverted confocal (Leica) with a ×40 Plan APO objective (CX PL APO 40x/1.25-0.75 OIL CS) and analysed using Fiji 2.0.

**Superovulation.** Five units of pregnant mare serum (PMS) was injected intraperitoneally in 4-week-old female mice between 2 and 4 pm. After 44–46 h, GV oocytes were collected from the ovaries by puncturing with needles.

**Oocytes collection.** After dissection, ovaries were freed from adhering tissues in 1× PBS. Series of six ovaries were digested in 500 µL of 2 mg/mL Collagenase (SIGMA), 0.025% trypsin (SIGMA) and 0.5 mg/mL type IV-S hyaluronidase (SIGMA), on a ThermoMixer (Eppendorf) with gentle agitation for 20 min. The digestion was then stopped by the addition of 1 mL of 37 °C pre-warmed αMEM—5% FBS. The oocytes were then size-selected under a binocular.

**RNA preparation.** Pool of 100–200 oocytes collected were washed through several M2 drops, and total RNA was isolated using NucleoSpin RNAXS kit (Macherey-Nagel) according to the user manual. RNA quality and quantity were evaluated using a Bioanalyzer. Between 5 and 10 ng of RNA was obtained from each pool of oocytes.

**RNA-seq analyses.** PolyA+ RNA-seq libraries were prepared using the SMART-Seq v4 UltraLow Input RNA kit (Clontech) followed by the Nextera XT DNA library Prep kit (Illumina) according to the manufacturer recommendations from three biological replicates for each condition (P7 wild-type (WT), P7 *Tbpl2<sup>-/-</sup>* mutant, P14 WT and P14 *Tbpl2<sup>-/-</sup>* mutant oocytes) and sequenced 50 pb single end using an Illumina HiSeq 4000 (GenomEast platform, IGBMC).

Reads were preprocessed in order to remove the adapter, poly(A) and low-quality sequences (Phred quality score below 20). After this preprocessing, reads shorter than 40 bases were discarded for further analysis. These preprocessing steps were performed using cutadapt version 1.10<sup>61</sup>. Reads were mapped to spike sequences using bowtie version 2.2.8<sup>62</sup>, and reads mapping to spike sequences were removed for further analysis. Reads were then mapped onto the mm10 assembly of *Mus musculus* genome using STAR version 2.7.0f<sup>63</sup>. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.9.1<sup>64</sup>, with annotations from Ensembl version 96 and “union” mode. Read counts were normalised across samples with the median-of-ratios method to make these counts comparable between samples, and differential gene analysis was performed using the DESeq2 version 1.22.2<sup>65</sup>. All the figures were generated using R software version 3.5.3.

**RT-qPCR.** Complementary DNA was prepared using random hexamer oligonucleotides and SuperScript IV Reverse Transcriptase (Invitrogen) and amplified using LightCycler<sup>®</sup> 480 SYBR Green I Master (Roche) on a LightCycler<sup>®</sup> 480 II (Roche). Primers used for qPCR analysis are listed in Supplementary Table 2.

**Repeat element analyses.** Data were processed as already described<sup>66</sup> using Bowtie1 (version 1.2.2)<sup>67</sup> instead of Maq. The RepeatMasker annotation was used to identify the different types of repeat elements (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013-2015 <http://www.repeatmasker.org>). Differential expression analyses were performed using DESeq2 (version 1.22.2)<sup>65</sup>. All the figures were generated using R custom scripts (version 3.5.3).

**SLIC-CAGE analyses.** Twenty-eight and 13 ng of total RNA isolated from P14 oocytes (biological replicate 1 and replicate 2, ~500–1000 oocytes pooled for each replicate) and 15 ng of the total RNA isolated from P14 *Tbpl2<sup>-/-</sup>* mutant oocytes (approximately 550 pooled oocytes) were used for SLIC-CAGE TSS mapping<sup>43</sup>. Briefly, 5 µg of the carrier RNA mix were added to each sample prior to reverse transcription, followed by the cap-trapping steps designed to isolate capped RNA polymerase II transcripts. The carrier was degraded from the final library prior to sequencing using homing endonucleases. The target library derived from the oocyte RNA polymerase II transcripts was PCR-amplified (15 cycles for P14 WT, 16 cycles for P14 *Tbpl2<sup>-/-</sup>* mutant) and purified using AMPure beads (Beckman Coulter) to remove short PCR artifacts (<200 bp, size selection using 0.8× AMPure beads to sample ratio). The libraries were sequenced using HiSeq2500 Illumina platform in single-end, 50 bp mode (Genomics Facility, MRC, LMS).

Sequenced SLIC-CAGE reads were mapped to the reference *M. musculus* genome (mm10 assembly) using Bowtie2<sup>62</sup> with parameters that allow zero mismatches per seed sequence (22 nucleotides). Uniquely mapped reads were kept for downstream analyses using CAGER Bioconductor package (version 1.20.0)<sup>68</sup> and custom R/Bioconductor scripts. Bam files were imported into R using the CAGER package, where the mismatching additional G, if added through the template-free activity of the reverse transcriptase, was removed. Same samples sequenced on different lanes and biological replicates were merged prior to final analyses.

**Promoter analyses.** In order to consider only the CAGE TSS dependent only on TBPL2, we removed all the P14 WT CAGE tags at the position where CAGE tags were also present in the P14 *Tbpl2*<sup>-/-</sup> mutant CAGE tags dataset: for the rest of the analysis, this dataset was called “TBPL2-dependent” and we compared it to the P14 *Tbpl2*<sup>-/-</sup> mutant CAGE data (hereafter called “TBPL2-independent”). Briefly, a CAGE set object was created from the TBPL2-dependent and TBPL2-independent CTSS files using CAGER Bioconductor package (version 1.20.0)<sup>68</sup>, data were normalised using normalizeTagCount (fitlnRange = c(5,1000), alpha = 1.53, T = 1e6) and the powerLaw option. Cluster of CTSS were collected using clusterCTSS (threshold = 1, thresholdIsTpm = TRUE, nrPassThreshold = 1, method = “distclu”, maxDist = 20, removeSingletons = TRUE, keepSingletonsAbove = 5). Width of the TSS regions was calculated using cumulativeCTSSdistribution and quantilePositions (clusters = “tagClusters”, qLow = 0.1, qUp = 0.9): interquantile width corresponds to the 10th–90th percentile of the total tag cluster signal. In order to compare the different samples, consensus promoters were computed using aggregateTagCluster (tpmThreshold = 3, qLow = 0.1, qUp = 0.9, maxDist = 100). Self-organising map (SOM) expression profiling was performed using getExpressionProfiles using a tpmThreshold of 3, the method “som”, xDim = 3 and yDim = 2. Shifting TSS were obtained after calculation of the cumulative distribution along the consensus clusters using cumulativeCTSSdistribution and calculation of the shift score using scoreShift with the Kolmogorov–Smirnov test. Shifting promoters were extracted using getShiftingPromoters (tpmThreshold = 3, scoreThreshold = -Inf, fdrThreshold = 0.01).

TSSs corresponding to the MaLR ERVs were identified after annotation using HOMER (version 4.10)<sup>69</sup>.

Sequence analyses were performed using Bioconductor R seqPattern (version 1.14) and R custom scripts. WW dinucleotides enrichment was computed using plotPatternDensityMap on -250/+250 regions centred on the dominant TSSs. TATA box position weight matrix (PWM) matches analyses were performed using the MotifScanScores function applied on the -35/-20 sequences centred on the dominant TSSs, using the TBP PWM provided in the SeqPattern package (derived from the JASPAR database). The distribution of the best match for each sequence was then plotted. Sequence Logo was created using Bioconductor R package SeqLogo.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The datasets generated during the current study are available in different repositories: proteomic data; ProteomeXchange PRIDE database with accession [PXD0316347](https://proteomecentral.proteomexchange.org/protein/PXD0316347), RNA-seq data; Gene Expression Omnibus database [GSE140090](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140090) and SLIC-CAGE data; ArrayExpress [E-MTAB-8866](https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8866). Source data are provided with this paper.

## Code availability

RNA-seq data were analysed using Bioconductor package DESeq2, SLIC-CAGE data were analysed using Bioconductor package CAGER. All custom codes are available upon request.

Received: 22 September 2020; Accepted: 20 November 2020;

Published online: 22 December 2020

## References

- Roeder, R. G. 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat. Struct. Mol. Biol.* **26**, 783–791 (2019).
- Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).
- Tora, L. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* **16**, 673–675 (2002).
- Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13–25 (2014).
- Ohbayashi, T., Makino, Y. & Tamura, T. A. Identification of a mouse TBP-like protein (TLP) distantly related to the drosophila TBP-related factor. *Nucleic Acids Res.* **27**, 750–755 (1999).
- Teichmann, M. et al. Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. *Proc. Natl Acad. Sci. USA* **96**, 13720–13725 (1999).
- Dantoni, J. C., Quintin, S., Lakatos, L., Labouesse, M. & Tora, L. TBP-like factor is required for embryonic RNA polymerase II transcription in *C. elegans*. *Mol. Cell* **6**, 715–722 (2000).
- Kaltenbach, L., Horner, M. A., Rothman, J. H. & Mango, S. E. The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. *Mol. Cell* **6**, 705–713 (2000).
- Veenstra, G. J. C. Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in *Xenopus*. *Science* **290**, 2312–2315 (2000).
- Müller, F., Lakatos, L., Dantoni, J., Strähle, U. & Tora, L. TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. *Curr. Biol.* **11**, 282–287 (2001).
- Persengiev, S. P. et al. TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proc. Natl Acad. Sci. USA* **100**, 14887–14891 (2003).
- Bárfai, R. et al. TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Curr. Biol.* **14**, 593–598 (2004).
- Jallow, Z., Jacobi, U. G., Weeks, D. L., Dawid, I. B. & Veenstra, G. J. C. Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in *Xenopus*. *Proc. Natl Acad. Sci. USA* **101**, 13525–13530 (2004).
- Zhang, D., Penttilä, T. L., Morris, P. L., Teichmann, M. & Roeder, R. G. Spermiogenesis deficiency in mice lacking the Trf2 gene. *Science* **292**, 1153–1155 (2001).
- Martianov, I. et al. Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Mol. Cell* **7**, 509–515 (2001).
- Gazdag, E. et al. TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes Dev.* **23**, 2210–2223 (2009).
- Hernandez, N. TBP, a universal eukaryotic transcription factor? *Genes Dev.* **7**, 1291–1308 (1993).
- Rabenstein, M. D., Zhou, S., Lis, J. T. & Tjian, R. TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proc. Natl Acad. Sci. USA* **96**, 4791–4796 (1999).
- Hansen, S. K., Takada, S., Jacobson, R. H., Lis, J. T. & Tjian, R. Transcription properties of a cell type-specific TATA-binding protein, TRF. *Cell* **91**, 71–83 (1997).
- Ohbayashi, T. et al. Vertebrate TBP-like protein (TLP/TRF2/TLF) stimulates TATA-less terminal deoxynucleotidyl transferase promoters in a transient reporter assay, and TFIIA-binding capacity of TLP is required for this function. *Nucleic Acids Res.* **31**, 2127–2133 (2003).
- Carninci, P. et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).
- Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**, 233–245 (2012).
- Haberle, V. et al. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* **507**, 381–385 (2014).
- Jagarlamudi, K. & Rajkovic, A. Oogenesis: transcriptional regulators and mouse models. *Mol. Cell Endocrinol.* **356**, 31–39 (2012).
- Pedersen, T. & Peters, H. Proposal for a classification of oocytes and follicles in the mouse ovary. *J. Reprod. Fertil.* **17**, 555–557 (1968).
- Thompson, P. J., Macfarlan, T. S. & Lorincz, M. C. Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* **62**, 766–776 (2016).
- Peaston, A. E. et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
- Gazdag, E., Rajkovic, A., Torres-Padilla, M.-E. & Tora, L. Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction* **134**, 51–62 (2007).
- Malecova, B. et al. TBP/TFIID-dependent activation of MyoD target genes in skeletal muscle cells. *eLife* **5**, 23 (2016).
- Deato, M. D. E. et al. MyoD targets TAF3/TRF3 to activate myogenin transcription. *Mol. Cell* **32**, 96–105 (2008).
- Zybailov, B. et al. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5**, 2339–2347 (2006).
- Helmlinger, D. & Tora, L. Sharing the SAGA. *Trends Biochem. Sci.* **42**, 850–861 (2017).
- Lewandoski, M., Wassarman, K. M. & Martin, G. R. Zp3-cre, a transgenic mouse line for the activation or inactivation of loxP-flanked target genes specifically in the female germ line. *Curr. Biol.* **7**, 148–151 (1997).
- Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* **45**, 439–446 (2012).
- Dong, J. et al. Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature* **383**, 531–535 (1996).
- Galloway, S. M. et al. Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nat. Genet.* **25**, 279–283 (2000).
- Bettgeowda, A. Mechanisms of maternal mRNA regulation: implications for mammalian early embryonic development. *Front. Biosci.* **12**, 3713 (2007).
- Timmers, H. T. M. & Tora, L. Transcript buffering: a balancing act between mRNA synthesis and mRNA degradation. *Mol. Cell* **72**, 10–17 (2018).

39. Yu, C. et al. BTG4 is a meiotic cell cycle-coupled maternal-zygotic-transition licensing factor in oocytes. *Nat. Struct. Mol. Biol.* **23**, 387–394 (2016).
40. Wolf, J. & Passmore, L. A. mRNA deadenylation by Pan2–Pan3. *Biochem. Soc. Trans.* **42**, 184–187 (2014).
41. Ma, J., Flemr, M., Strnad, H., Svoboda, P. & Schultz, R. M. Maternally recruited DCP1A and DCP2 contribute to messenger RNA degradation during oocyte maturation and genome activation in mouse. *Biol. Reprod.* **88**, 11 (2013).
42. Veselovska, L. et al. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol.* **16**, 209 (2015).
43. Cvetesic, N. et al. SLIC-CAGE: high-resolution transcription start site mapping using nanogram-levels of total RNA. *Genome Res.* **28**, 1943–1956 (2018).
44. Cvetesic, N. et al. Global regulatory transitions at core promoters demarcate the mammalian germline cycle. Preprint at <https://www.biorxiv.org/content/10.1101/2020.10.30.361865v2> (2020).
45. Törönen, P., Kolehmainen, M., Wong, G. & Castrén, E. Analysis of gene expression data using self-organizing maps. *FEBS Lett.* **451**, 142–146 (1999).
46. Zhang, D., Penttilä, T. L., Morris, P. L. & Roeder, R. G. Cell- and stage-specific high-level expression of TBP-related factor 2 (TRF2) during mouse spermatogenesis. *Mech. Dev.* **106**, 203–205 (2001).
47. Martianov, I. et al. Distinct functions of TBP and TLF/TRF2 during spermatogenesis: requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. *Development* **129**, 945–955 (2002).
48. Martianov, I., Velt, A., Davidson, G., Choukrallah, M.-A. & Davidson, I. TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. *Sci. Rep.* **6**, 32069 (2016).
49. Upadhyaya, A. B., Lee, S. H. & DeJong, J. Identification of a general transcription factor TFIIA $\alpha$ /beta homolog selectively expressed in testis. *J. Biol. Chem.* **274**, 18040–18048 (1999).
50. Xiao, L., Kim, M. & DeJong, J. Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice. *Gene Expr. Patterns* **6**, 409–419 (2006).
51. Müller, F., Zaucker, A. & Tora, L. Developmental regulation of transcription initiation: more than just changing the actors. *Curr. Opin. Genet. Dev.* **20**, 533–540 (2010).
52. Falender, A. E., Shimada, M., Lo, Y. K. & Richards, J. S. TAF4b, a TBP associated factor, is required for oocyte development and function. *Dev. Biol.* **288**, 405–419 (2005).
53. Akhtar, W. & Veenstra, G. J. C. TBP2 is a substitute for TBP in Xenopus oocyte transcription. *BMC Biol.* **7**, 45 (2009).
54. Jacobi, U. G. et al. TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation. *EMBO J.* **26**, 3900–3909 (2007).
55. Crichton, J. H., Dunican, D. S., MacLennan, M., Meehan, R. R. & Adams, I. R. Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. *Cell Mol. Life Sci.* **71**, 1581–1605 (2014).
56. Faulkner, G. J. et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **41**, 563–571 (2009).
57. McCarthy, E. M. & McDonald, J. F. Long terminal repeat retrotransposons of *Mus musculus*. *Genome Biol.* **5**, R14 (2004).
58. Franke, V. et al. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. *Genome Res.* **27**, 1384–1394 (2017).
59. Vastenhouw, N. L., Cao, W. X. & Lipshitz, H. D. The maternal-to-zygotic transition revisited. *Development* **146**, dev161471 (2019).
60. Gegonne, A. et al. The general transcription factor TAF7 is essential for embryonic development but not essential for the survival or differentiation of mature T cells. *Mol. Cell Biol.* **32**, 1984–1997 (2012).
61. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
62. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
63. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
64. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
65. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 31 (2014).
66. Fadloun, A. et al. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* **20**, 332–338 (2013).
67. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
68. Haberle, V., Forrest, A. R. R., Hayashizaki, Y., Carninci, P. & Lenhard, B. CAGER: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res.* **43**, e51–e51 (2015).
69. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

## Acknowledgements

We thank D. Singer and A. Gegonne for the gift of the *Taf7<sup>fllox</sup>* mouse line and H. Stunnenberg for TFIIA antibodies. We would also like to thank D. Devys for critically reading the manuscript, all members of the Tora lab for thoughtful discussions and suggestions throughout the course of the work. We are grateful to I. Kukhtevich, M. Borsos, M.E. Torres Padilla, T. Gupta, L. Casini, G. Barzaghi and A. Krebs for help in preliminary experiments, and A.H.F.M. Peters for suggestions on the analysis of the retrotransposon data. We thank C. Hérouard and M. Jung from the GenomEAST platform for library preparation and preliminary analyses, P. Eberling for peptide synthesis, F. Ruffenach for proteomic analyses, G. Duval for polyclonal antibody generation, the histology platform, the IGBMC cell culture facility and S. Falcone, M. Poirot and F. Memedov of the IGBMC animal facility for animal care taking. This work was supported by funds from CNRS, INSERM, and Strasbourg University. This study was also supported by the European Research Council (ERC) Advanced grant (ERC-2013-340551, Birtocaction) (to LT) and grant ANR-10-LABX-0030-INRT and a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02 (to IGBMC). IB and FM acknowledge support by Wellcome Trust Senior Investigator awards (106115/Z/14/Z and 106955/Z/15/Z, respectively).

## Author contributions

C.Y., S.D.V. and L.T. designed the study; S.D.V. and L.T. supervised the project; C.Y. performed all molecular lab and mouse experiments; V.H. performed immunolocalizations and helped for the histological analysis of ovaries; E.G. generated the anti-TBPL2 polyclonal antibodies; L.N. carried out the proteomic analyses; K.G. and I.B. carried out preliminary analyses; P.H. organised the SLIC-CAGE; N.C. carried out SLIC-CAGE analyses; N.C. and B.L. analysed the SLIC-CAGE data and S.D.V. analysed the proteomic, RNA-seq and SLIC-CAGE data. F.M. oriented the promoter analyses. C.Y., F.M., S.D.V. and L.T. wrote the paper with contributions to the text and figure legends from all authors. All authors gave final approval for publication.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-20239-4>.

**Correspondence** and requests for materials should be addressed to S.D.V. or L.T.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

**SUPT3H-less SAGA coactivator can assemble and function without significantly perturbing RNA polymerase II transcription in mammalian cells**

**Veronique Fischer<sup>1,2,3,4,\*</sup>, Elisabeth Scheer<sup>1,2,3,4,\*</sup>, Elisabeth Lata<sup>1,2,3,4,∞</sup>, Vincent Hisler<sup>1,2,3,4</sup>, Bastian Morlet<sup>1,2,3,4</sup>, Damien Plassard<sup>1,2,3,4,5</sup>, Stéphane D. Vincent<sup>1,2,3,4</sup>, Dominique Helmlinger<sup>6</sup>, Didier Devys<sup>1,2,3,4</sup> and László Tora<sup>1,2,3,4,#</sup>**

<sup>1</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France;

<sup>2</sup>Centre National de la Recherche Scientifique (CNRS), UMR7104, 67404 Illkirch, France;

<sup>3</sup>Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, 67404 Illkirch, France;

<sup>4</sup>Université de Strasbourg, 67404 Illkirch, France;

<sup>5</sup>Plateforme GenomEast, infrastructure France Génomique; 67404 Illkirch, France.

<sup>6</sup>CRBM, University of Montpellier, CNRS, Montpellier, France

\*These authors contributed equally to this work

∞Present address: Transcription and Tumor Laboratory, University of Bordeaux, INSERM U1212, CNRS UMR 5320, 33076 Bordeaux, France;

#Corresponding author: 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; Tel +33 3 88 65 34 44, Fax: +33 3 88 65 32 01; e-mail: laszlo@igbmc.fr

**Running Title:** SUPT3H-lacking SAGA can assemble and function

**Key words:** SAGA (Spt-Ada-Gcn5 acetyltransferase) complex, Spt3, RNA polymerase II, transcription, knock-out, osteosarcoma, mESC, 4sU RNA-seq, histone fold, U2OS cells, mouse embryonic stem cells, proteomic, co-activator, gene regulation.



## **Abstract**

Coactivator complexes regulate chromatin accessibility and transcription. SAGA (Spt-Ada-Gcn5 Acetyltransferase) is an evolutionary conserved coactivator complex. The core module scaffolds the entire SAGA complex and adopts a histone octamer-like structure, which consists of six histone fold domain (HFD)-containing proteins forming three histone fold (HF) pairs, to which the double HFD-containing SUPT3H adds an HF pair. Spt3, the yeast ortholog of SUPT3H, interacts genetically and biochemically with the TATA binding protein (TBP) and contributes to global RNA polymerase II (Pol II) transcription. Here we demonstrate that i) SAGA purified from human U2OS or mouse embryonic stem cells (mESC) can assemble without SUPT3H; ii) SUPT3H is not essential for mESC survival, iii) SUPT3H is required for mESC growth and self-renewal, and iv) the loss of SUPT3H from mammalian cells affects the transcription of only a specific subset of genes. Accordingly, in the absence of SUPT3H no major change in TBP accumulation at gene promoters was observed. Thus, SUPT3H is not required for the assembly of SAGA, TBP recruitment, or overall Pol II transcription, but plays a role in mESC growth and self-renewal. Our data further suggest that yeast and mammalian SAGA complexes contribute to transcription regulation by distinct mechanisms.

## Introduction

Formation of the transcription preinitiation complex (PIC), containing RNA polymerase II (Pol II) and six general transcription factors (GTFs), is a major regulatory step in eukaryotic gene expression (1,2). PIC formation is mediated by TFIID and the accessibility of the transcription machinery to template DNA is controlled by co-activator complexes. The SAGA complex (Spt-Ada-Gcn5 acetyltransferase) is an evolutionary conserved, multifunctional co-activator composed of about 18-20 subunits (3,4). Early genetic studies, predominantly performed in yeast, have established that SAGA is organized into distinct functional and structural entities, now called modules. These modules comprise a structural core, histone acetyltransferase (HAT), histone deubiquitinase (DUB), TBP-binding, and transcription factor (TF)-binding activities (5).

*Saccharomyces cerevisiae* (y) *SPT3* was isolated as an allele-specific suppressor of mutation in *TBP* coding for the TATA binding protein (TBP), suggesting that Spt3 may play a role in recruiting yeast SAGA (ySAGA) to promoters (6-8). The TBP-binding module of ySAGA comprises Spt3 and Spt8, which both interact with TBP directly (9,10). Recent high resolution cryo-electron microscopy (cryo-EM) structures of ySAGA complexes indicated that the core structural module contains a histone octamer like structure consisting of seven histone fold domain (HFD)-containing proteins, including Ada1/Taf12, Taf6/Taf9 and Taf10/Spt7, which form three HF pairs, and Spt3, which harbours two HF domains forming an intramolecular HF pair (10-12). Interestingly, Spt3 is homologous to the corresponding Taf11/Taf13 HF heterodimer in TFIID, and Spt3 of ySAGA binds the same side of TBP as does the Taf11/Taf13 HF pair (12-16). Purified recombinant TBP binds to the histone octamer-like structure of ySAGA through Spt3 and Spt8 (10,14), further arguing that yeast SAGA contributes to TBP delivery to core promoters. Indeed, structural and biochemical evidence suggest

a model by which Spt3 prevents spurious TBP binding to DNA through steric hindrance, which is relieved by the synergistic binding of TFIIA and a cognate TATA element to TBP (10). In agreement, transcription of stress-inducible genes containing a TATA box in their promoters is more sensitive to SAGA subunit mutations, than to TFIID subunit mutations (17). However, recent analyses in yeast based on quantification of newly synthesized mRNA, demonstrated that both TFIID and SAGA are required for the transcription of almost all Pol II genes, although each subunit may act through different mechanisms and thus to different extent (18-20).

The modular organization of mammalian SAGA complexes is very similar to that of their yeast counterparts (4). For example, their histone octamer-like structure is built up by three homologous histone fold pairs, TADA1/TAF12, TAF6L/TAF9/9b, and TAF10/SUPT7L, and SUPT3H, the mammalian homologue of ySpt3 (21,22). However, metazoan SAGA complexes are lacking a Spt8 homologue (4), suggesting that TBP interactions with mammalian SAGA complexes may be different from that of yeast SAGA. In addition, fly and mammalian SAGA complexes contain also a splicing module (4,22,23).

Here we investigated the role of SUPT3H in SAGA complex assembly, Pol II transcription regulation and TBP recruitment to target gene promoters in human U2OS and mouse embryonic stem cells (ESCs). Our results indicate that, in both cell types, SAGA subunit composition is not affected by the absence of SUPT3H, similar to yeast (9,24,25). In contrast, we observed a striking divergence between ySpt3 and mammalian SUPT3H orthologues in their regulatory roles. While we previously showed that yeast Spt3 contributes to global Pol II transcription (18), we show here that SUPT3H is required only for the expression of a limited number of genes in both cell types. TBP recruitment experiments further show that absence of SUPT3H does not



affect TBP binding to selected gene promoters. Despite these limited regulatory roles, we found that *Supt3* gene inactivation in mESCs affects growth and self-renewal. Together these data suggest that the role of SUPT3H in gene transcription diverged substantially between yeast and mammals, suggesting that mammalian SAGA complexes contribute to transcription regulation by distinct mechanism, or that SAGA-dependent TBP delivery to promoters requires other subunits.

## Materials and Methods

### Reagents

Reagents are described in Supplementary Table S1.

### Biological Resources

Biological resources are described in Supplementary Table S2, and see below:

#### a) Generation of stable Flag-SUPT3H overexpressing U2OS cell lines

The coding region (CDS) of the human *SUPT3H* gene was PCR amplified from the pREV-*SUPT3H* vector (26) using HA 13662980/1 primers described in Supplementary Table S3, and Phusion™ High Fidelity polymerase (Thermo Fisher Scientific, Cat# F-530), following manufacturer's instruction. This first PCR product was purified on gel using NucleoSpin Gel and PCR Clean-up, Mini kit for gel extraction and PCR clean up (Machery-Nagel, Cat# 740609.50). A second round of PCR was performed to amplify this first obtained PCR product and eliminate pREV-*SUPT3H* matrix, using the same primers. Final product was purified on gel. This cDNA was cloned into pSG5-puro A frame expression vector (27). First the *SUPT3H* PCR product was digested with Xho I and Sma I type II restriction enzymes from New England Biolabs, and purified using Nucleospin® PCR clean-up kit (Machery Nagel Cat# 740609.50). Then ligated into the XhoI and SmaI fragment of the pSG5-puro expression vector from an SV40 promoter, digested and purified same way as described above, using T4 DNA ligase (Biolabs, Cat# M0202). This product was transformed in competent DH5α and ampicillin resistant colonies were selected after. Final plasmid constructs were verified by sequencing using.

To obtain a U2OS cell line stably expressing the SUPT3H protein, U2OS cells were transfected with 5 μg of circular pSG5puro-*hSUPT3H* plasmid, using Lipofectamine2000 following manufacturer's instruction (ThermoFisher Scientific, Cat#

11668019). One single cell per well was seeded in 96-well plates using the BD Biosciences FACS Aria II (BD Biosciences) apparatus. After a 10-days puromycin selection, positive cell clones were selected by western blot analysis using the anti-Flag and anti-SUPT3H antibody.

#### **b) Generation of *Supt3*<sup>-/-</sup> mutant mESC lines**

Mouse E14 ESCs were transfected with a plasmid construct encoding for two sgRNAs (sequences in Supplementary Table S4) as well as a Cas9-GFP fusion protein at a confluency of 70-80% using Lipofectamine2000 (ThermoFisher Scientific, Cat# 11668019) following manufacturer's instruction. Two days after transfection, cells were selected for expression of the Cas9-GFP fusion protein by fluorescence activated sorting (FACS). Five 96-well plates were seeded with one GFP positive cell per well using the BD FACS Aria<sup>TM</sup> II (BD Biosciences) apparatus. Clones were amplified, genomic DNA extracted and knockout clones were selected by PCR, using the Phire direct PCR kit (Thermo Scientific, Cat# F-1265) following manufacturer's instructions. Primer sequences for PCR are shown in Supplementary Table S3. Homozygous mutant clones were sequenced using primers spanning the deletion site (Supplementary Table S3).

#### **Cell culture conditions**

Human U2OS osteosarcoma cells (HTB-96; ATCC) and U2OS-FI-SUPT3H cells were cultured using DMEM medium supplemented with 10% foetal calf serum (Sigma Aldrich, Cat# F7524) and 40 µg/ml gentamycin (KALYS, Cat# G0124-25). Cells were grown at 37°C with 5% CO<sub>2</sub> levels.

Mouse ES E14 cells were cultured on plates coated with 0.1% gelatin 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%  $\beta$ -mercaptoethanol (ThermoFisher Scientific, Cat# 31350-010), 100 U/ml penicillin and 100  $\mu$ g/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 medium, 3  $\mu$ M CHIR99021 (Axon Medchem, Cat# 1386) and 1  $\mu$ M PD0325901 (Axon Medchem, Cat# 1408) were added freshly to the medium. Cells were grown at 37°C with 5% CO<sub>2</sub> levels.

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

### **Clonal assays of mouse ESCs**

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% solution) for 30 minutes followed by two washes with 1x PBS. To assess the alkaline phosphatase (AP) activity of mouse ESC colonies, Alkaline Phosphatase Kit (Vector Laboratories, Cat# SK-5100) was used following manufacturer's instructions. Colonies were stained with AP for 5-10 minutes. For clonal assay analyses in FCS+LIF medium, an additional staining with crystal violet was performed after AP staining. Colonies were stained with 0.1% crystal violet solution for at least 30 minutes.

### **Quantification of clonal assays**

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, the colonies were counted manually using the ImageJ interface. 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. The total number of colonies as assessed by crystal violet staining was used for normalization between replicates.

### **Metabolic labelling**

Metabolic labelling of newly synthesized RNA was adapted from previously described protocols (28-30). In brief, the nucleoside analogue 4-thiouridine (4sU) (Glenthane Life Sciences, Cat# GN6085) was added to the cell culture medium at a final concentration of 500  $\mu$ M for a 20-minute pulse for mouse ES E14, human U2OS cells or Schneider S2 cells. After the labelling period, the medium containing 4sU was removed, the cells were washed with ice cold 1x PBS and immediately lysed using TRI<sup>®</sup> Reagent (Molecular Research Center Inc., Cat# TR 188).

*S. pombe* cultures were grown to an OD600 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.

### **Total RNA extraction**

Total RNA was extracted following TRI<sup>®</sup> Reagent (Molecular Research Center Inc., Cat# TR 188) manufacturer's instruction. To remove any potential genomic DNA contamination from the total RNA extracts, the TURBO DNA-free<sup>™</sup> Kit (ThermoFisher Scientific, Cat# AM1907) was used following manufacturer's instructions for rigorous DNase treatment.

For total RNA extraction of yeast cells, the RiboPure<sup>™</sup> RNA Purification Kit (ThermoFisher Scientific, Cat# AM1926) was used following manufacturer's instruction.

## **Reverse Transcription**

Reverse Transcription (RT) was performed with 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.

## **qPCR**

For qPCR, the cDNA samples were amplified using LightCycler® 480 SYBR® Green 2x PCR Master Mix I (Roche, Cat# 04887352001) and 0.3 or 0.6 µM of forward and reverse primer respectively. The primer pairs used for qPCR are listed in Supplementary Table S3. The qPCR was conducted using a LightCycler® 480 (Roche). For the assessment of mRNA levels, the obtained threshold-values were used to calculate the relative gene expression using the  $2^{-\Delta\Delta CT}$  method and considering the individual primer pair efficiencies (31). For TBP ChIP-qPCR, the percentage of input was calculated.

## **Newly synthesized RNA purification**

The purification of newly synthesized RNA was based on previously described protocols (28-30). As spike-in, labelled *S. pombe* total RNA was added to labelled mouse ESC total RNA preparations in a ratio 1:10; or labelled *S. pombe* total RNA was added to labelled U2OS total RNA preparations in a ratio of 1:25 respectively, to a final amount of 200 µg of total RNA prior to newly synthesized RNA purification. The RNA was precipitated and resuspended in 130 µL RNase-free water (Sigma Aldrich, Cat# 95284) and sonicated using the following settings 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). Biotinylation was carried out for 3 hours at 24°C in the dark and with gentle agitations. After incubation, excess biotin was removed by adding an equal volume of chloroform and centrifugation at 16,000 g 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 g 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 either using RNeasy MinElute Cleanup Kit (Qiagen, Cat# 74204) following manufacturer's instruction or by precipitation. Libraries were prepared and sequenced with 1x 50 base pairs on a HiSeq4000 machine (Illumina).

### **Library preparation of 4sU RNA-seq**

For U2OS cells, 4sU RNA-Seq libraries were generated from 15 ng of purified, newly synthesized RNA using Illumina Stranded Total RNA Prep, Ligation with RiboZero Plus kit and IDT for Illumina RNA UD Indexes, Ligation (Illumina, San Diego, USA,



Cat# 20040525 and 20040553/4, respectively), according to manufacturer's instructions. Briefly, abundant ribosomal RNAs were depleted by hybridization to specific DNA probes and enzymatic digestion. The depleted RNAs were purified and fragmented using divalent cations at 94°C for 2 minutes. After random hexamers annealing, fragmented RNAs were then reverse transcribed into first strand complementary DNA (cDNA). Second strand cDNA synthesis further generated blunt-ended double-stranded cDNA and incorporated dTTP in place of dUTP to achieve strand specificity by quenching the second strand during amplification. Following A-tailing of DNA fragments and ligation of pre-index anchors, PCR amplification was used to add indexes and primer sequences and to enrich DNA libraries (30 sec at 98°C; [10 sec at 98°C, 30 sec at 60°C, 30 sec at 72°C] x 12 cycles; 5 min at 72°C). Surplus PCR primers were further removed by purification using SPRIselect beads (Beckman-Coulter, Villepinte, France, Cat# B23319) and the libraries were sequenced with 1x 50 base pairs on a HiSeq4000 machine (Illumina).

For mESCs, 4sU RNA-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, Cat# RS-122-2301) 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 libraries were sequenced with 1x 50 base pairs on a HiSeq4000 machine (Illumina).

#### **Data analysis of 4sU RNA-seq**

Reads were preprocessed using CUTADAPT 1.10 (32) in order to remove adaptors and low-quality sequences and reads shorter than 40 bp were removed for further analysis. rRNA sequences were removed for further analysis. For mESC samples VQFR25, VQFR26, VQFR29, VQFR30, remaining reads were aligned to a hybrid genome composed of mm10 and ASM294v2 assemblies of *M. musculus* and *S. pombe* genomes respectively with STAR 2.5.3a (33). For samples VQFR188, VQFR189, VQFR191, VQFR192, the hybrid genome was composed of hg38 and ASM294v2 assemblies of *H. sapiens* and *S. pombe* genomes respectively. Gene quantification was performed with htseq-count 0.6.1p1 (34), using “union” mode and Ensembl 93 annotations for all organisms except for *S. 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 (35) Bioconductor R package on *H. sapiens* or *M. musculus* counts normalized with size factors computed by the median-of-ratios method proposed by Anders and Huber (36) based on the spike-in counts (using the following options: `cooksCutoff=TRUE`, `independentFiltering=TRUE`, `alpha=0.05`). P-values were adjusted for multiple testing using the Benjamini and Hochberg method (37). For subsequent data analyses and visualization, only protein-coding genes were considered. Further, a threshold of 100 or 50 reads was used to define expressed genes in the U2OS or mESC datasets respectively.

For TATA-less and TATA-box promoter analysis (violin plots in Fig. 2D and Fig. 5D), TATA-box containing promoters were extracted for the human (hg38) and mouse (mm10) genomes using the Eukaryotic Promoter Database (EPD, <https://epd.epfl.ch//index.php>, (38)).

### **Whole cell extract preparation from U2OS cells**

The required number of cells were trypsinized, transferred into 1.5 ml Eppendorf tubes, centrifuged 600 g 4°C for 2 min, and washed twice with 1ml 1x PBS. Pellets were resuspended in 1 packed cells volume (PCV) extraction buffer (400 mM KCl, 20 mM Tris-HCl pH 7.5, 20% glycerol, 2 mM DTT, 1x protease inhibitor cocktail). Tubes were frozen and thawed subsequently 4 times (from liquid nitrogen to ice) and centrifuged at 14,000 g 4°C for 10 min. The supernatant was stored at -80°C.

### **Human cell nuclear extract preparation**

To enrich extracts for nuclear proteins, cells were harvested and washed twice with 1x PBS. Cell pellets were resuspended in 4 times PCV of hypotonic buffer (50 mM Tris pH 7.9, 1 mM EDTA, 1 mM DTT and 1x protein inhibitor cocktail), left to swell for 30 minutes on ice, then dounced 10 times using a B dounce homogenizer to break cytoplasmic membrane. After a 10 minutes centrifugation at 1,000-1,800 g, 4°C, supernatant was removed and the pellet resuspended in a high salt buffer (50 mM Tris pH 7.9, 25 % glycerol, 500 mM NaCl, 0.5 mM EDTA, 1 mM DTT and 1x protein inhibitor cocktail). To break the nuclear membranes, suspension was homogenized by douncing 20 times using a B dounce, then incubated at 4°C (under stirring), for 30 minutes prior to centrifugation at 10,000 g for 20 minutes at 4°C. The supernatant was dialyzed o/n at 4°C against an isotonic salt buffer (50 mM Tris pH 7.9, 20 % glycerol, 5 mM MgCl<sub>2</sub>, 100 mM KCl, 1 mM DTT and 1x protein inhibitor cocktail). The dialyzed fraction was kept as nuclear extract.

### **Mouse ESC nuclear extract preparation**

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<sub>2</sub>, 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 g 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<sub>2</sub>, 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 10,000 g for 10 minutes at 4°C. The supernatant was kept at -80°C as nuclear extract.

### **Antibodies**

The list of antibodies used in this study is shown in Supplementary Table S5. Anti-TADA2B (3122) polyclonal antibody (pAb) was obtained by immunization of rabbits with the C-terminal region (amino acids 221-420) of hTADA2B (Q86TJ2). For this, the *TADA2b* cDNA fragment was PCR amplified and cloned in pET15b vector (Novagen) using Nde I and Bam HI sites. The 6xHis-TADA2B protein fragment was expressed in *E. coli* (BL21), centrifuged, lysed in a buffer L, containing 50 mM Tris-HCl pH 7.5, 400 mM NaCl, 10% glycerol, 1 mM DTT and 1x protein inhibitor cocktail. Recombinant proteins were in inclusion bodies, which were solubilized in buffer L containing 8 M urea. Proteins were purified under denaturing conditions using a Ni<sup>2+</sup>-NTA column, eluted with 300 mM imidazole and dialyzed in 1x PBS. Rabbits were immunized with the purified proteins for 6 weeks as described in (39). The obtained crude rabbit sera were then purified on an Affigel column on which the purified protein fragment was immobilized. Then the column was extensively washed with 1x PBS, pAb3122 eluted with 0.1 M glycine pH 2.5 buffer and immediately neutralized with 2 M Tris pH 8.8.

### **Immunoprecipitation protocol from human cell extracts**

Prior to immunoprecipitation, protein-A sepharose and ANTI-FLAG® M2 Affinity Gel beads (Sigma Aldrich, Cat# A2220) were washed three times with 1x PBS and two times with IP100 buffer (25 mM Tris-HCl 7.5, 5 mM MgCl<sub>2</sub>, 10 % glycerol, 0,1 % NP40, 100 mM KCl, 2 mM DTT and 1x protein inhibitor cocktail), prior to use. Nuclear extracts

were pre-cleared with 1/10 volume of packed bead volume for 1 hours at 4°C with agitation. For antibody binding, packed bead volume corresponding to 1/10 volume of input extract was incubated with the respective antibodies for 1 hours at room temperature, with agitation. Protein-A antibody bound sepharose beads were washed twice with IP500 (25 mM Tris-HCl 7.5, 5 mM MgCl<sub>2</sub>, 10% glycerol, 0,1% NP40, 500 mM KCl, 2 mM DTT and 1x protein inhibitor cocktail) buffer and then with IP100 buffer. Pre-cleared extracts were then incubated overnight with washed protein-A antibody bound sepharose beads at 4°C. Protein complex bound beads were washed twice with IP500 buffer and twice with IP100 buffer. Complexes were subsequently eluted twice with one bead volume of 0.1 M glycine pH 2.8 buffer at room temperature and with agitation. Eluates were immediately neutralized to pH7.5 by adding the required quantity of Tris-HCl pH 8.8 buffer. Eluates were characterized by western blot or mass spectrometry analyses.

### **Immunoprecipitation protocol for mESC nuclear extracts**

Prior to immunoprecipitations, 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<sub>2</sub>, 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 agitation. For antibody binding, the 50% bead slurry was incubated with 5-8 µg of the respective antibodies for 2 hours at 4°C with agitation. After incubation, beads were washed three times with IP500 buffer 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 agitation. After incubation, resins were washed three times with IP500 buffer and twice with IP100 buffer. Complexes were eluted from the beads using two subsequent 0.1 M

glycine pH 2.8 elutions at room temperature and with agitation. Importantly, prior to anti-TAF10 IPs in mESC, nuclear extracts were depleted for TFIID by overnight incubation with beads coated with antibodies targeting the TFIID-specific subunit TAF7. This allowed to increase the purification efficiency for SAGA in anti-TAF10 IPs given that TAF10 is shared between the SAGA and TFIID complexes. All other IPs have been performed without pre-depletion. Eluates were then characterized by western blot or mass spectrometry analyses.

### **Mass spectrometry**

Protein mixtures were Trichloroacetic acid (Sigma Aldrich, Cat# T0699) - precipitated overnight at 4°C. Samples were then centrifuged at 14000 rpm for 30 minutes at 4°C. Pellet were washed twice with 1 ml cold acetone and centrifuged at 16000 g rpm 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 (tris(2-carboxyethyl)phosphine) 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 4h, followed by an overnight modified trypsin digestion (Promega, Cat# V5111) 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. Samples were analyzed using an Ultimate 3000 nano-RSLC (Thermo Scientific, San Jose California) coupled in line with a LTQ-Orbitrap ELITE mass spectrometer via a nano-electrospray ionization source (Thermo Scientific, San Jose California). Peptide mixtures were loaded on a C18 Acclaim PepMap100 trap-column (75  $\mu$ m ID x 2 cm, 3

$\mu\text{m}$ , 100Å, Thermo Fisher Scientific) for 3.5 minutes at 5  $\mu\text{L}/\text{min}$  with 2% Acetonitrile MS grade (Sigma Aldrich, Cat# 1207802), 0.1% formic acid (FA, Sigma Aldrich, Cat# 94318) in  $\text{H}_2\text{O}$  and then separated on a C18 Accucore nano-column (75  $\mu\text{m}$  ID x 50 cm, 2.6  $\mu\text{m}$ , 150Å, Thermo Fisher Scientific) with a 90 minutes linear gradient from 5% to 35% buffer B (A: 0.1% FA in  $\text{H}_2\text{O}$  / B: 99% Acetonitrile MS grade, 0.1% FA in  $\text{H}_2\text{O}$ ), 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 220 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  $\pm 10$  ppm. Proteins were identified by database searching using SequestHT (Thermo Fisher Scientific) with Proteome Discoverer 2.4 software (PD2.4, Thermo Fisher Scientific) on *Homo sapiens* database (UniProt, reviewed, release 2020\_11\_27, 20309 entries) and *Mus musculus* database (UniProt, non-reviewed, release 2020\_07\_13, 55428 entries). Precursor and fragment mass tolerances were set at 10 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. Normalized spectral abundance factors (NSAF) were calculated for each protein as described earlier (40,41). First to obtain spectral abundance factors (SAF), spectral counts identifying a protein were divided by the protein length represented by the number of amino acids. Then to calculate NSAF values, the SAF values of each protein were divided by the sum of SAF values of all detected proteins.

### **Anti-TBP chromatin immunoprecipitation (ChIP)**

Cells were washed with 1x PBS, fixed at room temperature with 1% paraformaldehyde in 1x PBS for 20 minutes. Fixation was stopped by adding glycine to a final concentration of 125 mM for 10 minutes at room temperature. Cells were washed twice with ice-cold 1x PBS and collected by scrapping. Cells were centrifuged at 2000 g for 5 minutes at 4°C and washed once with ice-cold 1x PBS. Cells were lysed in L1 buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.1% NP40, 10% glycerol and protease inhibitory cocktail) and incubated on ice for 10 minutes before centrifugation at 2000g for 10 minutes at 4°C. Nuclei were resuspended in L2 buffer (0.5% SDS, 10 mM EDTA pH 8.0, 50 mM Tris-HCl pH 8.0 and protease inhibitory cocktail) and sonicated using a Covaris E210 sonicator to on average 300 bp fragments. Protein-G sepharose beads were washed twice with TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA) and blocked for three hours with 1  $\mu$ g/ $\mu$ L denatured yeast tRNA and 1,5% fish-skin gelatine. Beads were washed twice with TE buffer and stored at 4°C. For ChIP, 50  $\mu$ g of chromatin was diluted with chromatin dilution buffer (0.01% SDS, 1.1% Triton-X 100, 1.2 mM EDTA, 167 mM NaCl, 16.7 mM Tris-HCl pH 8.0 and protease inhibitory cocktail) to a volume of 800  $\mu$ L and incubated with 30  $\mu$ L of blocked bead slurry for 1 hour at 4°C for preclearing. Five  $\mu$ g of anti-TBP antibody (Abcam, cat # ab51841) was added to precleared chromatin and incubated overnight at 4°C. 50  $\mu$ L of bead slurry

was added to samples and incubated for two to four hours at 4°C with overhead shaking. After centrifugation for 2 minutes at 1000 g, beads were washed two times with Low salt washing buffer (0.1% SDS, 1% Triton-X 100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris HCl pH 8.0 and protease inhibitory cocktail) for 10 minutes, two times with high salt washing buffer (0.1% SDS, 1% Triton-X 100, 2 mM EDTA, 500 mM NaCl, 20 mM Tris HCl pH 8.0 and protease inhibitory cocktail) for 10 minutes, two times with LiCl washing buffer (250 mM LiCl, 1% NP40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0 and protease inhibitory cocktail) for 10 minutes and two times with TE buffer for 10 minutes. Bound fragments were eluted using with freshly prepared elution buffer (1% SDS and 0.1 M sodium bicarbonate) at room temperature. Chromatin was reverse crosslinked by addition of 0.2 M NaCl and 50 µg/mL RNase A. Samples were incubated at 37°C for one hour. 20 µg of Proteinase K was added and samples were incubated at 65°C overnight in a thermomixer. DNA was isolated by Phenol-chloroform purification and resuspended in water.

### **Statistical analysis**

All statistical tests were performed using either R (version 3.6.0). Two-sided Welch's t-test were performed to compare colony areas after quantification with ImageJ between WT and mutant mESCs. Two-sided Wilcoxon rank-sum test was used to compare two means in all the other indicated comparisons. The statistical details for individual experiments can be found in the figure legends or Results section. This includes number of replicates and statistical tests performed. The statistical threshold was defined with a *p*-value of less than 0.05.

### **Data availability**

The 4sU-seq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (42) and are accessible through GEO Series accession

numbers GSE175901. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (43) partner repository with the dataset identifier PXD026991.

## **Results**

### **Human U2OS cells do not express the SUPT3H subunit of the SAGA complex**

When analyzing RNA-sequencing data from the Human Protein Atlas (HPA) for mRNAs expressing different SAGA subunits in a panel of 64 human cell lines, we noticed that *SUPT3H* is not expressed in U2OS cells, a human osteosarcoma cell line, unlike the majority of other cell lines tested (Supplementary Figure S1A). To verify this observation, we designed primer pairs to amplify the whole coding sequence (CDS) of *SUPT3H* from cDNA prepared from U2OS and HeLa cells (Figure 1A, Supplementary Figure S1B). As an additional control, we generated an U2OS cell line stably expressing Flag-tagged wild type (WT) SUPT3H, which we refer to as U2OS-FI-SUPT3H. In agreement with the HPA data, we could not detect the *SUPT3H* mRNA in U2OS cells, while it was clearly detectable in HeLa and U2OS-FI-SUPT3H cells. Similar results were obtained when amplifying smaller fragments of the *SUPT3H* mRNA (Supplementary Figure S1D, scheme of primer positions Supplementary Figure S1B). Together these data indicate that the human U2OS osteosarcoma cell line does not express the SAGA subunit SUPT3H. Importantly, we could not detect major differences in the promoter sequence of the *SUPT3H* gene when comparing U2OS and HeLa cells, suggesting that the loss of *SUPT3H* expression is not caused by changes in the tested promoter sequence (Suppl. Fig. S1E). In addition, growth curve

analyses revealed that U2OS-FI-SUPT3H cells proliferate at a comparable rate than the parental U2OS cells (Supplementary Figure S1C).

### **Absence of SUPT3H does not affect human SAGA complex composition**

As SUPT3H is a subunit of the core module of the SAGA complex interacting with several subunits of the complex, we were wondering if, and to which extent, SAGA is assembled in U2OS osteosarcoma cells lacking SUPT3H. To this end we prepared nuclear extracts from U2OS cells and carried out immunoprecipitation (IP) coupled to mass spectrometry using antibodies raised against SUPT20H (a core module subunit), TADA2B and TADA3 (two HAT module subunits). As a positive control we used human HeLa cell nuclear extracts (NEs) for IPs with the same antibodies. Quantitative mass spectrometry analyses of anti-SUPT20H, -TADA2B and -TADA3 affinity purifications showed that the three antibodies purified all SAGA subunits except SUPT3H from U2OS NEs, and all SAGA subunits from HeLa NEs (Figure 1B and 1C). Thus, the absence of SUPT3H does not affect the overall subunit composition of human SAGA. Importantly, anti-SUPT20H and anti-Flag IPs performed using nuclear extracts from U2OS-FI-SUPT3H cells revealed that SUPT3H, when exogenously expressed, can incorporate in the SAGA complex in the U2OS nucleus (Figure 1C, 1D and Supplementary Figure S1F), indicating that the other subunits of the module are able to incorporate SUPT3H. To conclude, the absence of SUPT3H expression in U2OS cells allowed us to demonstrate that this subunit is dispensable for SAGA assembly.

### **Restoring SUPT3H expression modifies RNA polymerase II transcription at a subset of genes in U2OS cells**

The ability of SUPT3H to reconstitute an intact SAGA complex in U2OS-FI-SUPT3H allowed us to test the role of SUPT3H in Pol II transcription in human cells. For this, we performed 4-thiouridine (4sU) labelling of nascent RNA coupled with

sequencing of the labelled RNA (4sU-seq) (28-30), comparing two independent U2OS-FI-SUPT3H clones with parental U2OS control cells (Supplementary Figure 2A). We observed a very similar enrichment in intronic reads in all samples (Figure 2A), confirming a reproducible enrichment of newly synthesized RNAs genome-wide. Differential expression analyses revealed that restoring SUPT3H expression has only a limited effect on nascent transcript levels (Figure 2C, 2D, Supplementary Figure S2B). Specifically, 176 genes were significantly deregulated in WT U2OS cells lacking SUPT3H compared to U2OS-FI-SUPT3H cells at a threshold of 0.5 absolute log<sub>2</sub> fold change and an adjusted *p*-value of 0.05, including 80 genes downregulated and 96 genes upregulated (Figure 2C, Supplementary Figure S2B). Only 33 and 37 genes are down- and upregulated, respectively, more than two-fold. Overall, no GO categories were significantly enriched amongst genes deregulated by *SUPT3H* overexpression and only very few genes are responsive to the re-expression of SUPT3H in U2OS cells.

Earlier studies suggested that ySAGA predominantly regulates TATA-box containing genes, presumably through Spt3 (7,44,45). We therefore compared expression changes between U2OS and U2OS-FI-SUPT3H for genes with either a TATA-less or a TATA-box in their promoters, as defined by the Eukaryotic Promoter Database (Material and Methods) (Figure 2D). We observed no difference between the two gene classes, suggesting that TATA-box containing gene promoters do not show a stronger sensitivity to SUPT3H re-expression than TATA-less promoters in human U2OS cells. In conclusion, SUPT3H appears to regulate the transcription of only a few genes in osteosarcoma U2OS cells, independently of the presence a TATA-box element in their promoter. This observation is in marked contrast with the strong, global decrease of nascent transcript levels observed upon deletion of *SPT3* in *S. cerevisiae*.

## **Loss of mouse *Supt3* does not affect SAGA subunit composition in mouse embryonic stem cells**

Our results so far indicate that the function of ySpt3 in RNA Pol II transcription has diverged substantially between yeast and humans. To explore this further, we next turned to non-cancerous diploid mammalian cells and used CRISPR/Cas9 genome editing to inactivate the *Supt3* gene in mouse E14 ESCs. We obtained two individual clones with a homozygous deletion of exon 3 from the *Supt3* gene (Figure 3A), resulting in an out of frame stop codon. RT-qPCR analyses confirmed the deletion of the exon and the resulting decrease of remaining *Supt3* mRNA, presumably through non-sense decay (Figure 3B). Western blot analyses confirmed the deletion as SUPT3H is not detectable in extracts from knock-out mESCs (Figure 3C). To determine the role of SUPT3H on mouse SAGA assembly, we purified SAGA using anti-SUPT20H or anti-TAF10 antibodies from WT and *Supt3*<sup>-/-</sup> ESCs nuclear-enriched cell extracts and analyzed their composition by mass spectrometry and western blot analyses. Both immune-purifications coupled to mass spectrometry analyses from *Supt3*<sup>-/-</sup> cells identified all core, HAT and TF-binding SAGA subunits, with the exception of SUPT3H, indicating that mouse SAGA can also assemble without SUPT3H (Figure 3D). We note that the association of the DUB complex with SAGA was weaker in both anti-SUPT20H or anti-TAF10 IPs, but independently of the genotype of mESCs. These results confirm that, as observed in human U2OS cells, mouse SUPT3H has no major role in SAGA integrity in ESCs. Thus, SAGA central core can assemble without the intramolecular HF pair of SUPT3H, suggesting the SAGA can assemble around an hexameric core structure.

## ***Supt3* is required for mouse ESC growth and self-renewal**

The obtention of two independent *Supt3*<sup>-/-</sup> mouse ESC lines indicates that *Supt3* is not essential for mouse ESC viability when cultured in medium containing foetal calf serum (FCS), leukemia inhibitory factor (LIF) and two small molecules that maintain efficient mouse ESC self-renewal (hereafter called FCS+LIF+2i medium). However, we observed in clonal assays that *Supt3*<sup>-/-</sup> cells formed significantly smaller colonies compared to wildtype cells revealing that SUPT3H loss impairs mouse ESC growth (Figure 4A). In agreement, growth curve analyses revealed decreased proliferation of *Supt3*<sup>-/-</sup> cells as compared to wildtype ESCs (Figure 4B).

To assess the importance of SUPT3H for the self-renewal capacities of ESCs, we performed clonal assays in medium without 2i (hereafter referred to as FCS+LIF medium) and observed about a 4-fold decrease in colony numbers in *Supt3*<sup>-/-</sup> ESCs grown in FCS+LIF medium, as compared to WT cells (Figure 4C). As a read-out of the self-renewal capacities of ESCs, we measured the proportion of alkaline phosphatase positive (AP+) colonies. When compared to WT ESCs, the *Supt3*<sup>-/-</sup> cells produced about 3-fold less AP+ colonies (Figure 4D), demonstrating that SUPT3H loss has a major impact on the self-renewal capacities of mouse ESCs.

### **The loss of SUPT3H in mouse ESCs has no major effect on Pol II transcription**

To determine the impact of SAGA lacking SUPT3H on Pol II transcription in mouse ESCs, we analyzed nascent RNA by 4sU-seq in FCS+LIF+2i medium in which *Supt3*<sup>-/-</sup> ESCs can maintain their self-renewal capacities (as assessed by AP staining, Figure 4A). We could observe a similar enrichment of intronic reads in all samples (Fig. 5A). We observed that only a few genes, about 100, are differentially expressed between *Supt3*<sup>-/-</sup> and WT ESCs (Fig. 5B, 5C and Supplementary Figure S2B). Specifically, the newly synthesized levels of about 40 transcripts were significantly



decreased when applying a threshold of 0.5 log<sub>2</sub> fold change and an adjusted *p*-value of 0.05. Conversely, the nascent levels of 60 transcripts increased in *Supt3*<sup>-/-</sup> cells using the same thresholds (Supplementary Table 6). We could not find specific GO categories to be affected among the differentially expressed genes, however, the deregulation of some of these genes may explain the growth and self-renewal defects observed in *Supt3*<sup>-/-</sup> mESCs.

Finally, we compared transcription changes between genes possessing TATA-less and TATA-box containing promoters (Fig. 5D). We observed no difference between the two gene classes suggesting that TATA-box containing promoters do not show a stronger sensitivity to the loss of *Supt3* than TATA-less promoters in mouse ESCs (Fig. 2D). These data together suggest that mouse SAGA complexes lacking Supt3, as observed in human U2OS cells, affect the expression of a small subset of genes in FCS+LIF+2i medium.

#### **Loss of SUPT3H in human U2OS cells or mouse ESCs does not significantly affect TBP binding**

Genetic, biochemical and cryo-EM studies of ySAGA indicate that the Spt3 and Spt8 subunits bind to TBP and are involved in delivering TBP at specific promoters (6,7,10,11,46,47). We therefore wanted to assess whether SUPT3H has a role in recruiting TBP to mammalian gene promoters. Therefore, we performed TBP chromatin immunoprecipitation (ChIP)-coupled-qPCR experiments at selected gene promoters, in both U2OS and mESC cell lines (Figure 6A and 6B). We observed no difference in TBP occupancy at the tested gene promoters when comparing cells with and without SUPT3H. This suggests that TBP binding at these promoters does not require SUPT3H in human and mouse cells.

## Discussion

Previous SAGA subunit deletion and purification studies in *S. cerevisiae* demonstrated that Spt3 is not essential for viability and that ySAGA can assemble in the absence of Spt3, and in the absence of the Spt3/Spt8 TBP interacting module (9,24,25). However, constitutive deletion of *SPT3* has a global effect on Pol II transcription, which decreases the expression of the majority of genes (18).

In our present study we show that SUPT3H is not required for human U2OS cell survival, that mammalian SAGA complexes lacking SUPT3H can stably assemble, and that TBP recruitment at several TATA-containing and TATA-less promoters is not influenced by the absence of SUPT3H in mammalian SAGA. Nevertheless, mouse ESCs lacking SUPT3H show impaired self-renewal capacities and growth potential. Our experiments further show that constitutive loss of SUPT3H in both human U2OS and mouse ESCs grown in FCS+LIF+2i medium affects only a small subset of genes.

Some genes transcriptionally deregulated may play important roles in the cellular homeostasis of either U2OS or ESCs lacking SUPT3H: i.e. several histone genes of the *HIST1* cluster, *PAX7* and *JAK3* in U2OS cells), or *Sal1* and *Pclaf* in mESCs (Supplementary Table 6). Further studies will be necessary to identify which effects are directly caused by the loss of SUPT3H and how these changes contribute for example to the growth and self-renewal defects observed in mESCs.

Importantly, *Supt3* deletion in mice indicated that SUPT3H is important for mouse embryogenesis, but not for early development, as embryos die between E9.5 and E14.5 (48). This observation suggests that SUPT3H is not essential for early mouse development, similarly to other SAGA subunits, such as GCN5, PCAF, SUPT20H, USP22 and ATXN7L3 (49-54).

Cryo-EM structural studies suggested that yeast Spt3, or human SUPT3H, maybe an important structural subunit of SAGA, as both double HF-containing proteins assemble with three pairs of other HFD-containing SAGA subunits to form a deformed histone octamer-like structure (10,11,22). Importantly, it has been described in ySAGA that while the six histone-fold pairs in the histone-fold octamer-like structure are oriented similarly as in the canonical nucleosome, the intramolecular HF pair of Spt3 is tilted by 20 degrees compared with its analogous nucleosome histone H2A–H2B histone pair (10). This tilt therefore could almost completely free Spt3 from its association with the histone-fold octamer. However, ySAGA subunit composition is not affected by the loss of Spt3 (9,24,25), which we confirmed in both human and mouse cells. Interestingly, plants are probably also lacking a Spt3 orthologue (5,55). Thus, we speculate that the flexibility of Spt3 may be important for TBP binding and release and perhaps to couple its delivery with other SAGA regulatory activities through an allosteric mechanism.

Moreover, no ortholog of ySpt8 was detected in mammalian SAGA (5), suggesting that the TBP-interacting module described in yeast SAGA, would be affected in mammalian cells lacking SUPT3H. We were therefore surprised that the absence of SUPT3H has a minimal impact on Pol II transcription and TBP recruitment. It is thus conceivable that in mammalian cells the TBP promoter delivery mechanism is less, or not at all, dependent on SUPT3H. Alternatively, other SAGA subunits or TFIID may be able to compensate the constitutive loss of SUPT3H, masking the true *in vivo* contribution of SAGA to TBP delivery. Finally, it is conceivable that SUPT3H may be important for TBP delivery and transcription of genes that are induced in response to specific developmental and/or environmental signals.

## **Authors' contribution**

V.F. and E.S. performed most experiments, E.L. generated U2OS-FI-SUPT3H cell lines, B.M. performed mass spectrometry analysis, D.P. performed the bioinformatics analyses of 4sU-seq experiments. D.H. provided conceptual help, V.F. and L.T. wrote manuscript with support from E.S., B.M., D.P., S.D.V., D.H., and D.D. L.T. conceived and supervised the study.

## **Acknowledgements**

We thank all members of the Tora lab for protocols, thoughtful discussions and suggestions, A. Ben Shem for comments and critically reading the manuscript, the GenomEast platform [France Génomique consortium (ANR-10-INBS-0009)] for library preparation and sequencing; C. Ebel and M. Philipps for help with FACS sorting, B. Reina San Martin and N. Jung for help with the CRIPR/Cas9 *Supt3* KO experiment, designing and establishing the corresponding vectors, and the IGBMC cell culture facility for help with U2OS, HeLa and E14 ES cells.

## **Funding**

This study was supported by grants from Agence Nationale de la Recherche (ANR) ANR-19-CE11-0003-02, ANR-PRCI-19-CE12-0029-01, ANR-20-CE12-0017-03, NIH 5R01GM131626-02 and NSF (Award Number:1933344) grants (to LT), ANR-18-CE12-0026 (to DD) and ANR-20-CE12-0014-02 grants (to DH and DD); and the IdEx-University of Strasbourg PhD program and by the 'Fondation pour la Recherche Médicale' (FRM) association (FDT201904008368) fellowships (to VF); and an ANR-10-LABX-0030-INRT grant, under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02.

## References

1. Roeder, R.G. (2019) 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat Struct Mol Biol*, **26**, 783-791.
2. Chen, X., Qi, Y., Wu, Z., Wang, X., Li, J., Zhao, D., Hou, H., Li, Y., Yu, Z., Liu, W. *et al.* (2021) Structural insights into preinitiation complex assembly on core promoters. *Science*.
3. Spedale, G., Timmers, H.T. and Pijnappel, W.W. (2012) ATAC-king the complexity of SAGA during evolution. *Gene Dev*, **26**, 527-541.
4. Helmlinger, D. and Tora, L. (2017) Sharing the SAGA. *Trends Biochem Sci*, **42**, 850-861.
5. Helmlinger, D., Papai, G., Devys, D. and Tora, L. (2021) What do the structures of GCN5-containing complexes teach us about their function? *Biochim Biophys Acta Gene Regul Mech*, **1864**, 194614.
6. 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 Dev*, **6**, 1319-1331.
7. Laprade, L., Rose, D. and Winston, F. (2007) Characterization of new Spt3 and TATA-binding protein mutants of *Saccharomyces cerevisiae*: Spt3 TBP allele-specific interactions and bypass of Spt8. *Genetics*, **177**, 2007-2017.
8. Grant, P.A., Winston, F. and Berger, S.L. (2021) The biochemical and genetic discovery of the SAGA complex. *Biochim Biophys Acta Gene Regul Mech*, **1864**, 194669.
9. Sermwittayawong, D. and Tan, S. (2006) SAGA binds TBP via its Spt8 subunit in competition with DNA: implications for TBP recruitment. *Embo J*, **25**, 3791-3800.
10. 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.
11. Wang, H., Dienemann, C., Stutzer, A., Urlaub, H., Cheung, A.C.M. and Cramer, P. (2020) Structure of the transcription coactivator SAGA. *Nature*, **577**, 717-720.
12. 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.
13. Anandapadamanaban, M., Andresen, C., Helander, S., Ohyama, Y., Siponen, M.I., Lundstrom, P., Kokubo, T., Ikura, M., Moche, M. and Sunnerhagen, M. (2013) High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nat Struct Mol Biol*, **20**, 1008-1014.
14. Han, Y., Luo, J., Ranish, J. and Hahn, S. (2014) Architecture of the *Saccharomyces cerevisiae* SAGA transcription coactivator complex. *EMBO J*, **33**, 2534-2546.
15. Patel, A.B., Louder, R.K., Greber, B.J., Grunberg, 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*, **362**.

16. Gupta, K., Watson, A.A., Baptista, T., Scheer, E., Chambers, A.L., Koehler, C., Zou, J., Obong-Ebong, I., Kandiah, E., Temblador, A. *et al.* (2017) Architecture of TAF11/TAF13/TBP complex suggests novel regulation properties of general transcription factor TFIID. *eLife*, **6**.
17. 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*. *Mol Cell*, **13**, 573-585.
18. Baptista, T., Grunberg, 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 e135.
19. 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. *Mol Cell*, **68**, 118-129 e115.
20. 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**.
21. Yu, J., Madison, J.M., Mundlos, S., Winston, F. and Olsen, B.R. (1998) Characterization of a human homologue of the *Saccharomyces cerevisiae* transcription factor spt3 (SUPT3H). *Genomics*, **53**, 90-96.
22. Herbst, D.A.E., M.N. Louder, R.K. Dugast-Darzacq, C. Dailey, G.M. Fang, Q. Darzacq, X. Tjian, R. Nogales, E. (2021), Structure of the human SAGA coactivator complex: The divergent architecture of human SAGA allows modular coordination of transcription activation and co-transcriptional splicing, *bioRxiv*. <https://doi.org/10.1101/2021.02.08.430339>
23. Stegeman, R., Spreacker, P.J., Swanson, S.K., Stephenson, R., Florens, L., Washburn, M.P. and Weake, V.M. (2016) The Spliceosomal Protein SF3B5 is a Novel Component of *Drosophila* SAGA that Functions in Gene Expression Independent of Splicing. *J Mol Biol*, **428**, 3632-3649.
24. Wu, P.Y. and Winston, F. (2002) Analysis of Spt7 function in the *Saccharomyces cerevisiae* SAGA coactivator complex. *Mol Cell Biol*, **22**, 5367-5379.
25. Lee, K.K., Sardi, 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.
26. Umlauf, D., Bonnet, J., Waharte, F., Fournier, M., Stierle, M., Fischer, B., Brino, L., Devys, D. and Tora, L. (2013) The human TREX-2 complex is stably associated with the nuclear pore basket. *Journal of cell science*, **126**, 2656-2667.
27. Green, S., Issemann, I. and Sheer, E. (1988) A versatile in vivo and in vitro eukaryotic expression vector for protein engineering. *Nucleic Acids Res.*, **16**, 369.
28. 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. *Nat Biotechnol*, **29**, 436-442.
29. Radle, B., Rutkowski, A.J., Ruzsics, Z., Friedel, C.C., Koszinowski, U.H. and Dolken, L. (2013) Metabolic labeling of newly transcribed RNA for high

- resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. *J Vis Exp*.
30. Schwalb, B., Michel, M., Zacher, B., Fruhauf, K., Demel, C., Tresch, A., Gagneur, J. and Cramer, P. (2016) TT-seq maps the human transient transcriptome. *Science*, **352**, 1225-1228.
  31. Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*, **29**, e45.
  32. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10-12.
  33. 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.
  34. Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166-169.
  35. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
  36. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol*, **11**, R106.
  37. Benjamini, Y., Hochberg, Y. (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statist. Soc. B* **57**, 289-300.
  38. Dreos, R., Ambrosini, G., Perier, R.C. and Bucher, P. (2015) The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res*, **43**, D92-96.
  39. Muratoglu, S., Georgieva, S., Papai, G., Scheer, E., Enunlu, I., Komonyi, O., Cserpan, I., Lebedeva, L., Nabirochkina, E., Udvardy, A. *et al.* (2003) Two different Drosophila ADA2 homologues are present in distinct GCN5 histone acetyltransferase-containing complexes. *Mol Cell Biol*, **23**, 306-321.
  40. Zybaylov, B., Mosley, A.L., Sardu, M.E., Coleman, M.K., Florens, L. and Washburn, M.P. (2006) Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *Journal of proteome research*, **5**, 2339-2347.
  41. Zhang, Y., Wen, Z., Washburn, M.P. and Florens, L. (2010) Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Analytical chemistry*, **82**, 2272-2281.
  42. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*, **30**, 207-210.
  43. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M. *et al.* (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*, **47**, D442-D450.
  44. 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.
  45. de Jonge, W.J., O'Duibhir, E., Lijnzaad, P., van Leenen, D., Groot Koerkamp, M.J., Kemmeren, P. and Holstege, F.C. (2017) Molecular mechanisms that distinguish TFIID housekeeping from regulatable SAGA promoters. *EMBO J*, **36**, 274-290.



46. Eisenmann, D.M., Chapon, C., Roberts, S.M., Dollard, C. and Winston, F. (1994) The *Saccharomyces cerevisiae* SPT8 gene encodes a very acidic protein that is functionally related to SPT3 and TATA-binding protein. *Genetics*, **137**, 647-657.
47. 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. *Mol Cell Biol*, **20**, 634-647.
48. 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.
49. 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. *Mol Cell Biol*, **27**, 3405-3416.
50. 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.
51. 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. *Proc Natl Acad Sci U S A*, **97**, 11303-11306.
52. Atanassov, B.S., Mohan, R.D., Lan, X., Kuang, X., Lu, Y., Lin, K., McIvor, E., Li, W., Zhang, Y., Florens, L. *et al.* (2016) ATXN7L3 and ENY2 Coordinate Activity of Multiple H2B Deubiquitinases Important for Cellular Proliferation and Tumor Growth. *Mol Cell*, **62**, 558-571.
53. Koutelou, E., Wang, L., Schibler, A.C., 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**.
54. Wang, F., El-Saafin, F., Ye, T., Stierle, M., Negroni, L., Durik, M., Fischer, V., Devys, D., Vincent, S.D. and Tora, L. (2021) Histone H2Bub1 deubiquitylation is essential for mouse development, but does not regulate global RNA polymerase II transcription. *Cell Death Differ*.
55. Antonova, S.V., Boeren, J., Timmers, H.T.M. and Snel, B. (2019) Epigenetics and transcription regulation during eukaryotic diversification: the saga of TFIID. *Genes Dev*, **33**, 888-902.

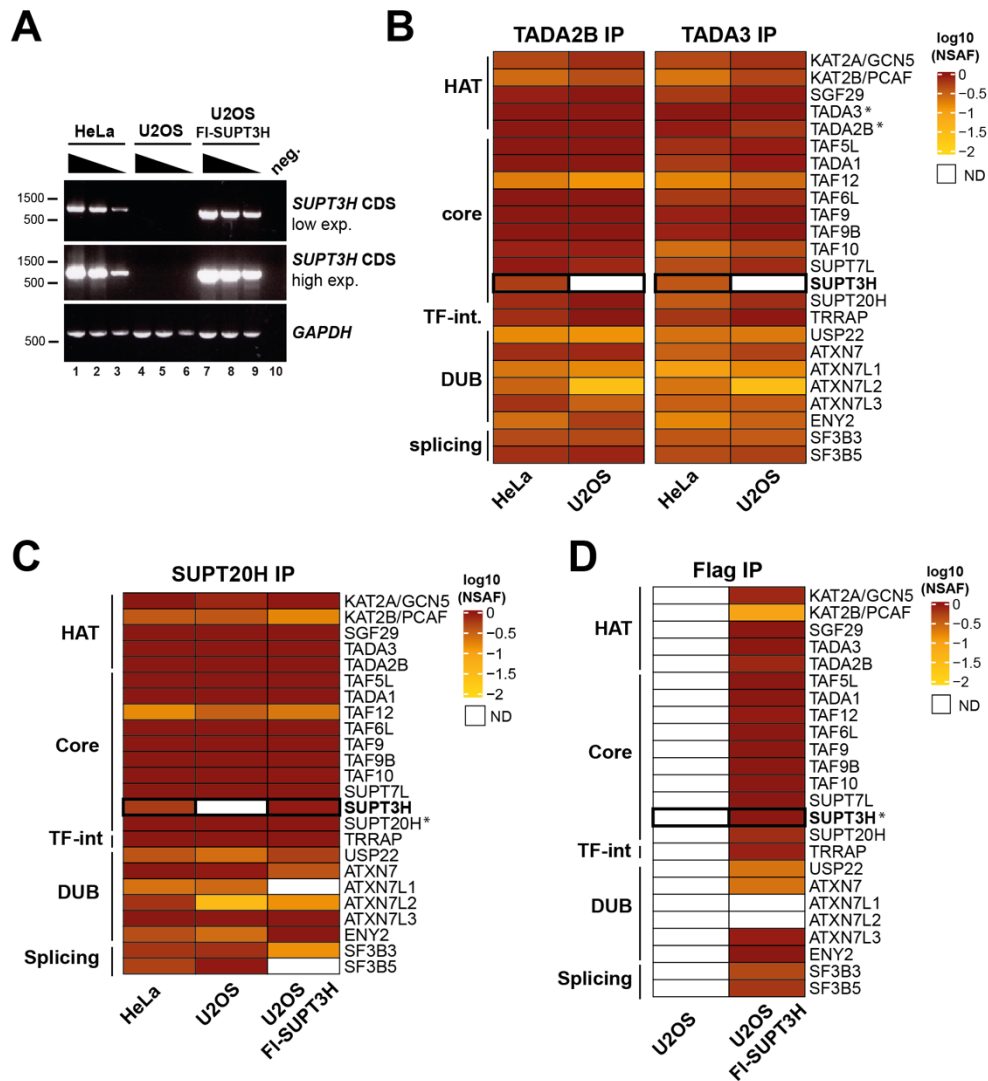
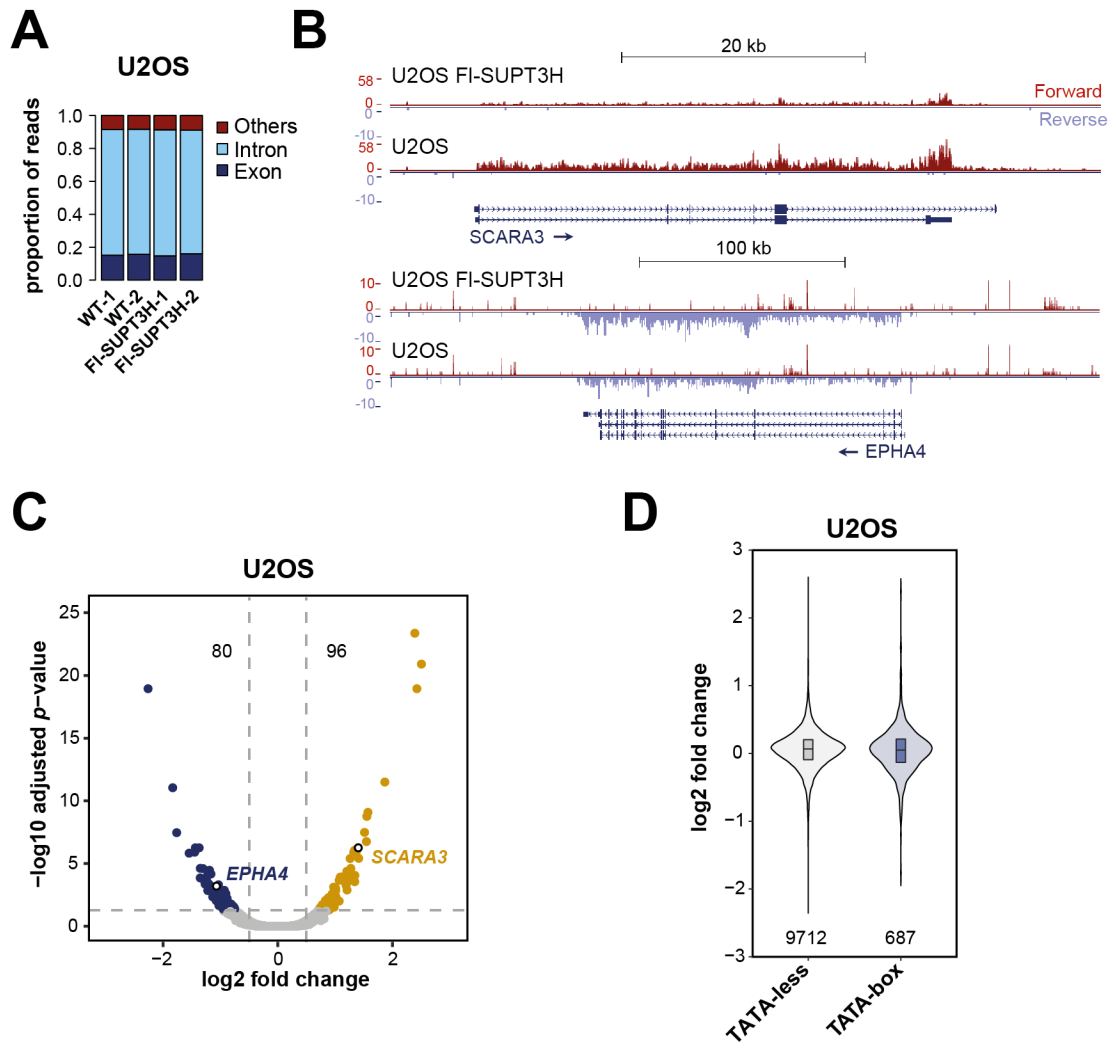


Figure 1

**Figure 1: Absence of the SAGA subunit SUPT3H in human U2OS cells does not hinder formation of the SAGA complex. (A)** PCR of the coding sequence (CDS) of *SUPT3H* using cDNA obtained from HeLa cells, U2OS cells or U2OS cells re-expressing Flag-SUPT3H (FI-SUPT3H). Low and high exposures (exp.) are shown. *Gapdh* expression serves as loading control. n = 3 technical replicates. **(B)** Log10-transformed bait-normalized NSAF (Normalized Spectral Abundance Factor) values of mass spectrometry results from anti-TADA2B and anti-TADA3 immunoprecipitations (IP) of the SAGA complex from HeLa and U2OS nuclear extracts. n = 3 technical replicates in each IPs. **(C)** Log10-transformed bait-normalized NSAF values of mass spectrometry results from SUPT20H immunoprecipitations (IP) of the SAGA complex from HeLa, U2OS, and U2OS FI-SUPT3H nuclear extracts. n = 3 technical replicates in each IP. **(D)** Log10-transformed bait-normalized NSAF values of mass spectrometry results from Flag IP of the SAGA complex from U2OS and U2OS FI-SUPT3H nuclear extracts. n = 3 technical replicates in each IP.

from HeLa, U2OS and U2OS FI-SUPT3H nuclear extracts. n = 3 technical replicates.

**(D)** Log10-transformed bait-normalized NSAF values of mass spectrometry results from Flag IPs shown in (Supplementary Figure S1F). n = 3 technical replicates. In (B, C, D) star (\*) indicates the bait proteins. The distinct SAGA modules are indicated as HAT = histone acetyltransferase; TF-int = transcription factor-interacting; DUB = deubiquitylation. ND = not detected.



**Figure 2**

**Figure 2: Only a small subset of Pol II-transcribed genes is responsive to the re-expression of SUPT3H in human U2OS cells. (A)** Proportion of reads per genomic element for 4sU-seq experiments for two independent U2OS and FI-SUPT3H clones. Exon = reads aligning to exons; Intron = reads aligning to introns, Others = reads aligning to intergenic regions, exon-intergenic junctions and exon-intron junctions. **(B)** UCSC genome browser views of 4sU-seq experiments on two differentially expressed genes (*SCARA3* and *EPHA4*) between U2OS FI-SUPT3H and U2OS cells. Forward and reverse strands are shown. Arrows indicate direction of

transcription. **(C)** Volcano plot representation of 4sU-seq with significantly down- or upregulated genes shown in dark blue and dark yellow respectively with numbers indicated on the top, using two independent U2OS and FI-SUPT3H clones. A threshold of 100 reads was used to define expressed genes. The position of the two genes shown in (B) are indicated by circles. **(D)** Violin plot representation comparing the distribution of the log2 fold changes between TATA-less and TATA-box containing gene classes. The number of genes per category are indicated at the bottom. The statistical test performed is two-sided Welch's t-test. Statistical significance ( $p < 0.05$ ) was not reached.

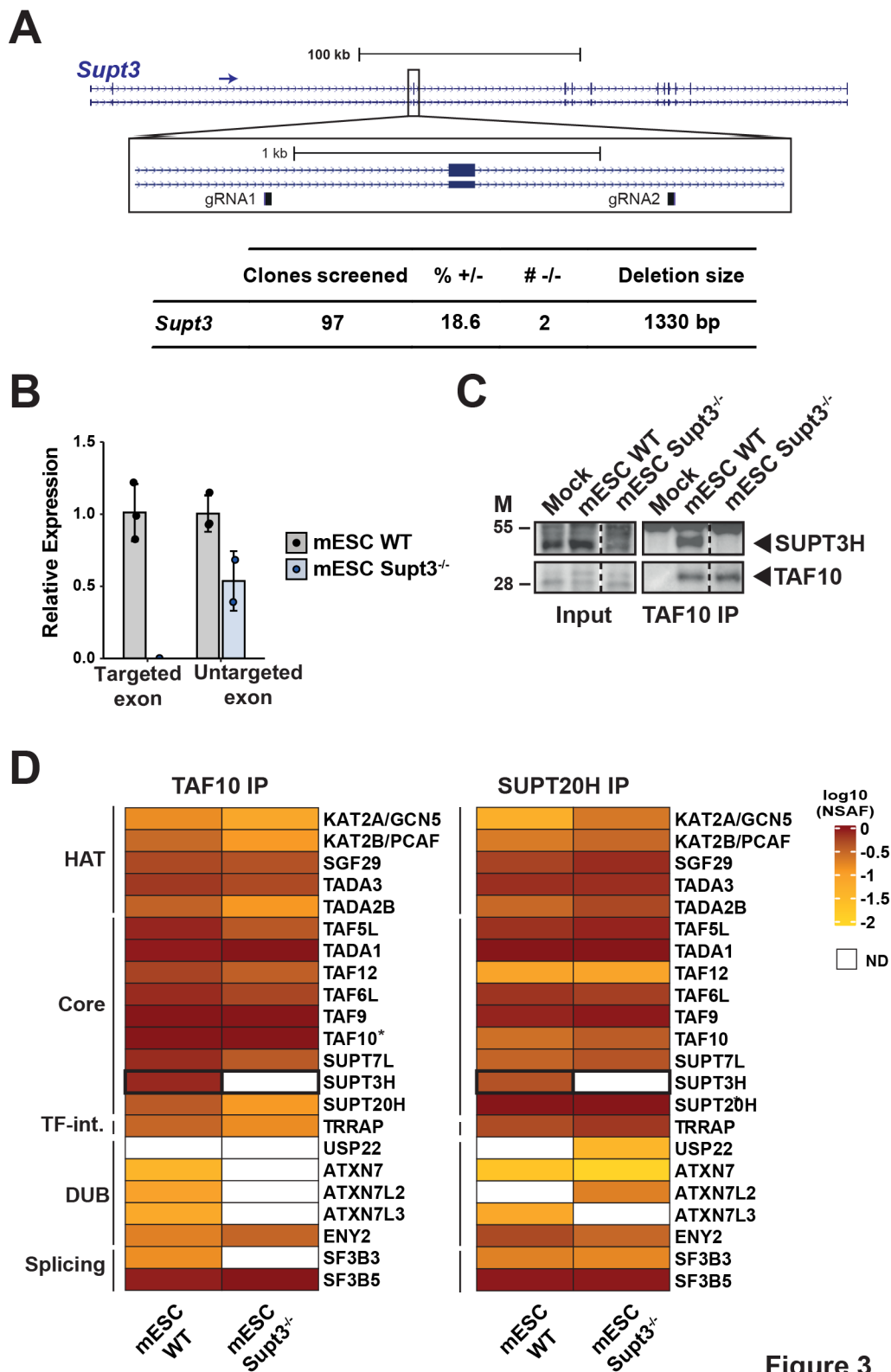
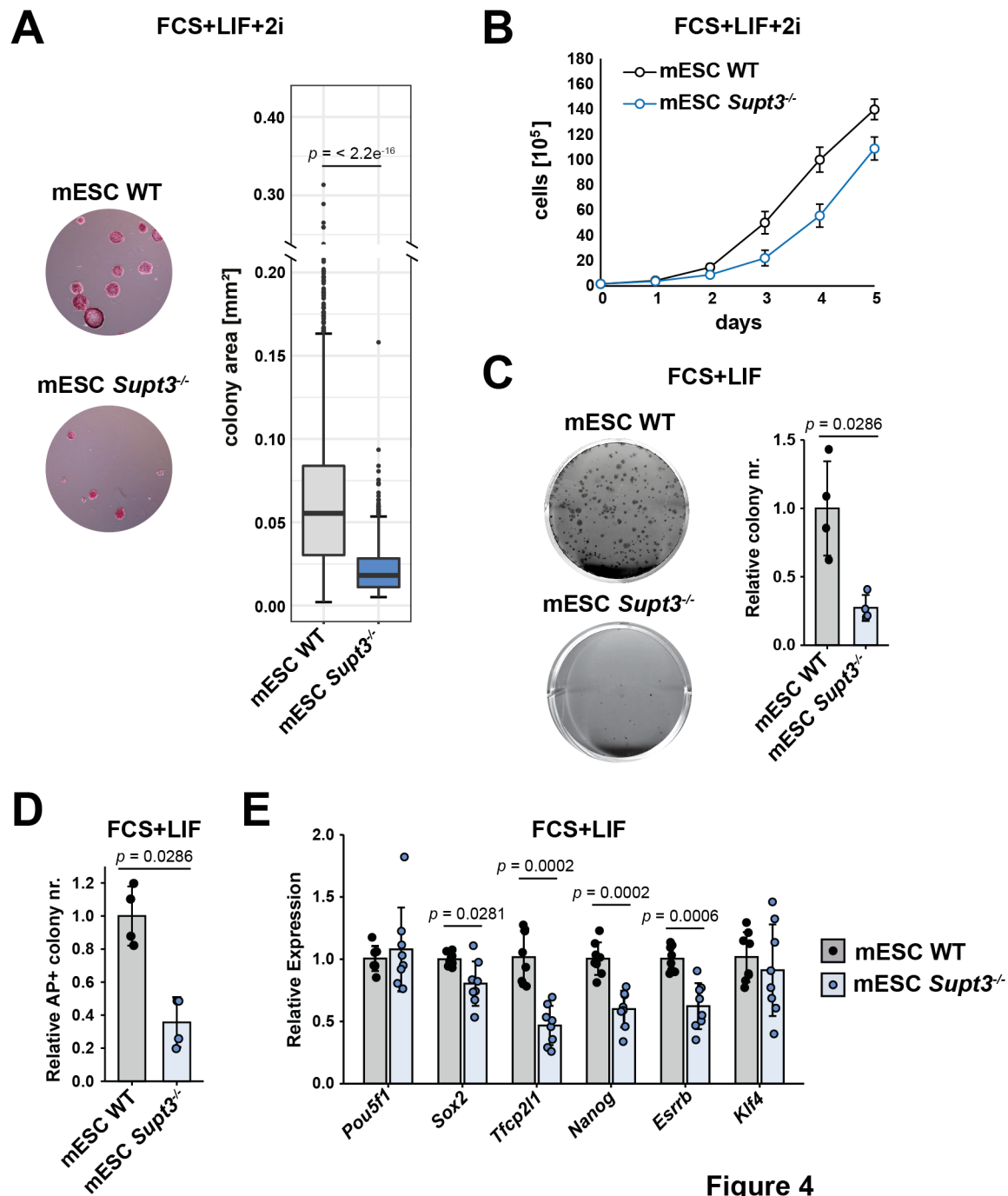


Figure 3

**Figure 3: Loss of *Supt3* does not impair mouse ESC survival or formation of the SAGA complex. (A)** Top, schematic representation of the mouse *Supt3* locus. The insert shows the position of the two gRNAs used to generate the *Supt3*<sup>-/-</sup> cell lines.

Bottom, table showing the number of clones screened, the percentage of heterozygous clones (+/-) and the number of homozygous (-/-) clones obtained as well as the deletion size. **(B)** Validation of *Supt3*<sup>-/-</sup> cell lines by RT-qPCR revealing the absence of the targeted, out-of-frame exon and overall reduced levels of the exon-deleted *Supt3* mRNA. Error bars show mean  $\pm$  standard deviation (SD) of two independent clones, with each assessment of the cell lines being based on the mean of three technical RT-qPCR replicates. **(C)** Western blot analysis of input and elution fractions of TAF10 immunoprecipitations (IP) from wildtype (WT mESCs) and *Supt3*<sup>-/-</sup> mESCs. M = molecular weight markers (in kDa). **(D)** Log10-transformed bait-normalized NSAF values of mass spectrometry results from TAF10 IPs shown in (C) and SUPT20H IPs. n = three technical replicates each. Star (\*) indicates the bait proteins. The distinct SAGA modules are indicated as HAT = histone acetyltransferase; TF-int = transcription factor-interacting; DUB = deubiquitylation. ND = not detected.

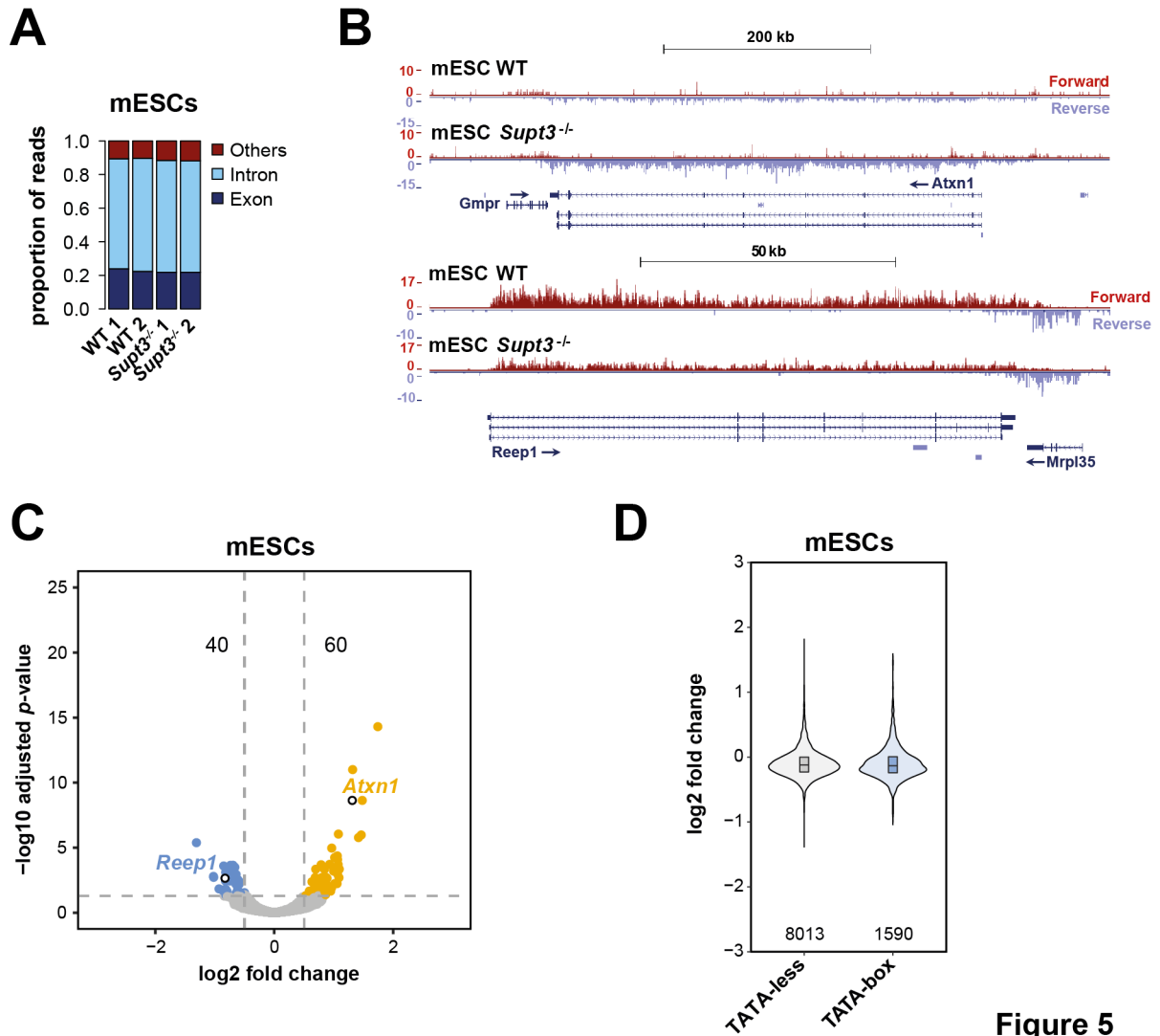




**Figure 4**

**Figure 4: SUPT3H is required for mouse ESC growth and efficient self-renewal.** (A) Clonal assays in FCS+LIF+2i medium comparing WT and *Supt3*<sup>-/-</sup> mouse ESCs. Colonies were stained with alkaline phosphatase. Left, representative images; Right, quantification of colony areas using ImageJ. The statistical test performed is two-sided Welch's t-test. Error bars show mean  $\pm$  standard deviation (SD) of four replicates from two biological replicates, with two independent clones each. p value is

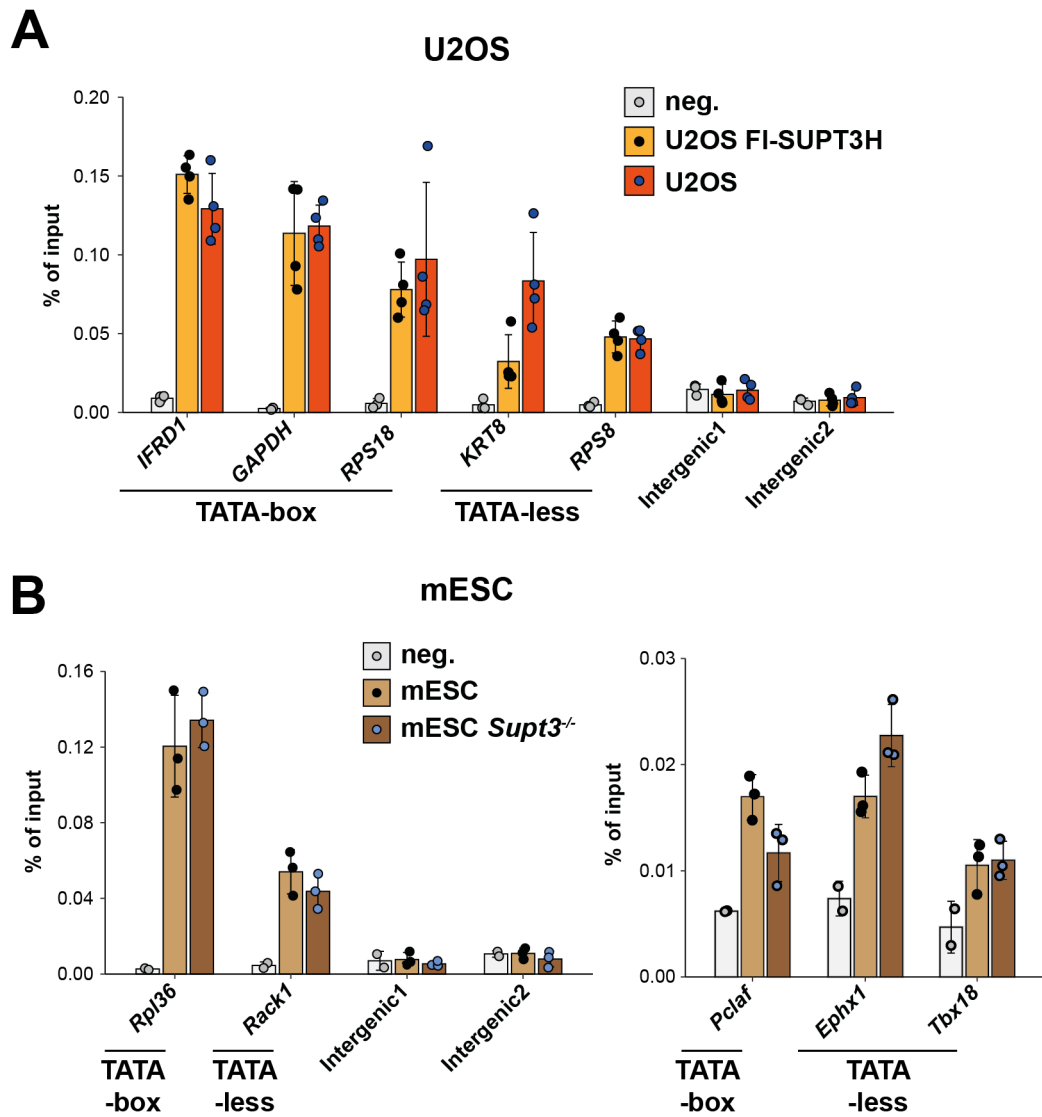
indicated on the graph. **(B)** Growth curve analysis of viable cells comparing WT and *Supt3*<sup>-/-</sup> mouse ESCs grown in FCS+LIF+2i medium for five days. Error bars show mean  $\pm$  standard deviation (SD) of two biological replicates using two independent clones. **(C)** Clonal assays in FCS+LIF medium comparing WT and *Supt3*<sup>-/-</sup> cells. Colonies were stained with crystal violet. Left, representative images; Right, quantification of colony numbers relative to WT cells. The statistical test performed is two-sided Wilcoxon rank-sum test. Error bars show mean  $\pm$  standard deviation (SD) of four replicates from two biological replicates, with two independent clones each. p value is indicated on the graph. **(D)** Quantification of the number of colonies staining positive for alkaline phosphatase (AP<sup>+</sup>) in FCS+LIF medium normalized relative to WT cells. The statistical test performed is two-sided Wilcoxon rank-sum test. Error bars show mean  $\pm$  standard deviation (SD) of four replicates from two biological replicates, with two independent clones each. p value is indicated on the graph. **(E)** Relative mRNA levels of pluripotency transcription factors (as indicated) comparing WT and *Supt3*<sup>-/-</sup> mouse ESCs grown in FCS+LIF medium. Expression was normalized to RNA polymerase III transcribed genes (*Rn7sk* and *Rpph1*) as well as to WT cells. Error bars show mean  $\pm$  standard deviation (SD) of at least 7 biological replicates (mean of three technical RT-qPCR replicates) using two independent clones each. Statistical test performed is two-sided Wilcoxon rank-sum test. Only statistically significant p values ( $p < 0.05$ ) are indicated.



**Figure 5**

**Figure 5: Loss of *Supt3* in mouse ESC has no major effect on Pol II transcription regulation in FCS+LIF+2i medium.** **(A)** Proportion of reads per genomic element for 4sU-seq experiments in two independent wildtype (WT) and *Supt3*<sup>-/-</sup> mouse ESC clones. Exon = reads aligning to exons; Intron = reads aligning to introns, Others = reads aligning to intergenic regions, exon-intergenic junctions and exon-intron junctions. **(B)** UCSC genome browser views of 4sU-seq experiments on two differentially expressed genes (*Atxn1* and *Reep1*) between WT and *Supt3*<sup>-/-</sup> mouse ESCs. Forward and reverse strands are shown. Arrows indicate direction of transcription. **(C)** Volcano plot representation of 4sU-seq experiments with significantly

down- or upregulated genes shown in light blue or yellow respectively with numbers indicated on the top, in two independent wildtype (WT) and *Supt3*<sup>-/-</sup> mouse ESC clones. A threshold of 50 reads was used to define expressed genes. The position of the two differentially expressed genes shown in (B) are indicated by white circles with black border. **(D)** Violin plot representation comparing the distribution of the log<sub>2</sub> fold changes between TATA-less and TATA-box containing genes. The number of genes per category are indicated at the bottom. The statistical test performed is two-sided Welch's t-test.



**Figure 6**

**Figure 6: Loss of SUPT3H in human U2OS cells or mouse ESCs does not significantly affect TBP binding at promoters.** Quantification of TBP ChIP-qPCR experiments in U2OS and U2OS FI-SUPT3H cells **(A)** or in WT and *Supt3*<sup>-/-</sup> mouse ESCs **(B)** at promoters of selected genes (as indicated) and at two independent intergenic regions. **(A and B)** TATA-less and TATA-box-containing promoters are indicated. ChIP-qPCR experiments without antibody were used as negative control (neg.). Error bars show mean  $\pm$  standard deviation (SD) of at four and three biological replicates using two independent clones for human and mouse data, respectively, in each sample the mean shows at least four technical RT-qPCR replicates. Statistical

test performed is two-sided Wilcoxon rank-sum test. Statistical significance ( $p < 0.05$ ) was not reached.

# Etude du rôle de TFIID et de ses complexes partiels dans l'initiation de la transcription

## Résumé

L'ARN polymérase II lit les gènes pour synthétiser des molécules appelées ARN. Ce processus est appelé transcription. Ces ARN peuvent jouer une fonction directe ou sont utilisés pour produire des protéines. TFIID est un autre complexe protéique qui recrute l'ARN polymérase II aux promoteurs pour initier la transcription. La régulation de la transcription est l'ensemble des processus conduisant à l'expression d'un ensemble spécifique de gènes dans la cellule. Le corps humain compte environ 200 types de cellules qui contiennent le même ensemble de gènes. Elles ne diffèrent que par le pool de gènes qu'elles expriment et donc par le pool de protéines. Cela rend la cellule spécialisée pour une fonction spécifique dans le corps. Dans le cas de maladies, la régulation de la transcription est souvent modifiée et les cellules ne peuvent donc pas assurer leur fonction, ce qui entraîne l'apparition de la maladie.

Le TFIID joue un rôle majeur dans la régulation de la transcription. En effet, selon le type de cellule, sa composition est modifiée entraînant le recrutement de l'ARN polymérase II vers des gènes spécifiques. De manière intéressante, dans certains contextes cellulaires spécifiques, même si la perte d'une sous-unité de TFIID altère l'assemblage de la TFIID, cela a un impact mineur sur la transcription des gènes. Il a été proposé qu'il existe au sein des cellules d'autres complexes dont la composition est proche de celle de TFIID et qui pourraient compenser la perte de TFIID. Cependant la composition et la fonction de ces complexes n'a pas été élucidées. Dans le cancer, il semble qu'une modification anormale de la composition de TFIID pourrait être impliquée dans la formation d'un nouveau programme d'expression génétique impliqué dans le développement des cancers. La mutation de certaines sous-unités de TFIID entraîne des maladies neurodégénératives. Mais là encore, aucune étude n'a expliqué la manière dont la variation de la composition de TFIID modifie la régulation de la transcription. Ma thèse porte sur cette question : comment la variabilité de la composition des TFIID contrôle l'expression des gènes.

**Mots clés :** Transcription - TFIID - TAFs - mESC - développement de la souris - SAGA

## Résumé en anglais

RNA polymerase II reads genes to synthesize molecules called RNA. This process is called transcription. These RNAs can have a direct function or are used to produce proteins. TFIID is another protein complex that recruits RNA polymerase II to promoters to initiate transcription. Transcription regulation is the set of processes leading to the expression of a specific set of genes in the cell. The human body has about 200 cell types that contain the same set of genes. They differ only in the pool of genes they express and thus in the pool of proteins. This makes the cell specialized for a specific function in the body. In the case of diseases, the regulation of transcription is often altered and the cells are therefore unable to perform their function, which leads to the development of the disease. TFIID plays a major role in the regulation of transcription. Depending on the cell type, its composition is altered, resulting in the recruitment of RNA polymerase II to specific genes. Interestingly, in some specific cellular contexts, although the loss of a TFIID subunit alters TFIID assembly, this has a minor impact on gene transcription. It has been proposed that there are other complexes within cells with a composition close to that of TFIID that could compensate for the loss of TFIID. However, the composition and function of these complexes has not been elucidated. In cancer, it appears that an abnormal change in the composition of TFIID may be involved in the formation of a new gene expression program involved in the development of cancers. Mutation of certain TFIID subunits leads to neurodegenerative diseases. But again, no study has explained how variation in TFIID composition alters transcriptional regulation. My thesis addresses this question: how variability in TFIID composition controls gene expression.

**Keywords:** Transcription - TFIID - TAFs - mESC - mouse development - SAGA