Functional study of the coactivator SAGA : role in RNA Polymerase II transcription

Tiago Fidalgo Baptista

UNIVERSITÉ DE STRASBOURG

École doctorale des Sciences de la Vie et de la Santé

IGBMC – CNRS UMR 7104 – Inserm U 964

THÈSE présentée par :

Tiago FIDALGO BAPTISTA

soutenue le : **25 septembre 2017**

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

Spécialité : Aspects moléculaires et cellulaires de la biologie

Functional study of the coactivator SAGA: role in RNA Polymerase II transcription

THÈSE dirigée par :

M. DEVYS Didier	MCF-MCUPH, IGBMC et Université de Strasbourg, France

RAPPORTEURS :

M. SHORE David	Professeur, Université de Genève, Suisse
M. MUELLER Ferenc	Professeur, Université de Birmingham, Royaume-Uni

---

AUTRES MEMBRES DU JURY :

Mme SOUTOURINA Julie	DR, I2BC/CEA/CNRS, Gif-sur-Yvette, France
M. SEXTON Thomas	CR, IGBMC, Illkirch, France

Acknowledgements

First of all, I would like to thank Dr Julie Soutourina, Professor David Shore, Professor Ferenc Mueller and Dr Thomas Sexton for accepting to be members of my thesis committee. Thank you all for taking the time to read and evaluate my thesis.

Afterwards, I want to show my gratitude to my supervisor Didier Devys. I truly believe that I could hardly have asked for a better PhD supervisor. These past four years have been exciting and exhilarating, yet exhausting at times. Needless to say that you have been someone that made this entire journey so much smoother. Thank you for being supportive, trustful and encouraging! I also want to thank László Tora to have welcomed me to his group and allowed me to develop my PhD thesis in such extraordinary environment. All the comments and discussions we had for sure helped me to build my character and my work.

I would also like to thank both past and current members of the Tora laboratory. Federica, that started this journey with me, thank you for the laughs and fun moments, Farrah for being one of the most incredible and kind-hearted creature on this planet and Ivanka for the moments we have spent bonding over yeast, wisdom, and friendship. I also want to thank Changwei, Sascha, and Paul, for the incredible moments we have spent, namely those Halloween nights (that were definitely less scarier than doing a PhD), all the members of the DUB club (Veronique, Matthieu and Kenny, besides Didier and Farrah) for the companionship and great discussion we have had over the years, Pooja for being such a good spirit, Eli for the always interesting conversations and Stéphane for being so useful for all mouse-related issues. I also want to thank some past members from the lab: Chen-Yi for introducing me to the *S. cerevisiae* world and for being such a great colleague and friend. Finally, I want to thank Nikolas, Marjorie, and Sarina for the great moments we have shared over the time I had the pleasure to share the same lab with you.

I also have to thank everyone that, one way or another, had an important technical and/or scientific input on my work. I want to thank everyone from the IGBMC Microarray and Sequencing platform, namely Christelle Thibault-Carpentier and Violaine Alunni, for the arrays hybridization. I want to express my gratitude to our collaborators over in Seattle, USA, Professor Steven Hahn, Sebastian Grünberg and Linda Warfield, for being a dream to work with and for making my (our) work that more robust. I also want to thank Professor Marc Timmers, for all the discussions and contributions to my project. I want to thank Vincent Géli and Pierre Luciano (CRCM, Marseille, France) for advices/discussion and for receiving me

in Marseille when I needed some help. Moreover, I want to thank P. Anthony Weill for the antibodies he kindly shared with me. Finally, I want to thank my thesis advisory committee, Dr Izabela Sumara and Professor Steven Johnsen, for their helpful comments during the early steps of my project.

I would also like to thank the Marie Curie actions – European Commission (EU-FP7 PEOPLE-2013 program, PITN-GA-2013-606806, NR-NET) and ARC for funding me during my PhD.

Over the first three years of my PhD I had the pleasure to be part of a Marie-Curie ITN action. During this period, I had the gratification to interact, discuss and have fun with incredible people. I would like to give all of them, with no exception, my expression of gratitude. Specifically, I want to thank Teresa (one of the funniest people in the world) for being such a good fellowship colleague and friend (thank you for the moments we have bonded over our miseries and joys!).

Last but not least, I want to thank everyone outside of my scientific life, but that still had such a great importance in my life over the years. I want to thank every friend from IGBMC. All the great friends I have met during my studies and the friends I keep in my life until today. I want to especially thank Pedro for standing with me over these four years, the support, guidance and all the joyful and memorable moments we shared together. A very special thanks to all my family, that always been there for me no matter what, specifically my mother and brothers. Finally, I want to thank my grandparents, António e Conceição, for being the best examples in life that one could have asked and for giving me some of the best memories of my life.

Abstract

The regulation of eukaryotic transcription involves the tightly regulated assembly of the pre-initiation complex for efficient RNA Polymerase II (RNA Pol II) positioning and activity. A key component of the general transcription machinery is TBP that can be either loaded on promoters as part of TFIID into TFIID-dominated genes or through the interaction with SAGA at SAGA-dominated genes. The coactivator SAGA, that acetylates histone H3 and deubiquitinates histone H2B, was widely accepted as a regulator of the expression of only a small subset of genes in yeast (around 10% of the genome in *S. cerevisiae*). In contrast, the remaining 90% of the genome would rely on TFIID. Nevertheless, discrepancies between binding of the factor and the transcriptional effects upon loss of the complex raised the question whether SAGA would have a broader role on transcription than previously anticipated.

To reevaluate the role of SAGA in RNA Pol II transcription, we used an array of new techniques to address the range of its enzymatic activities (using ChIP-seq to map histone marks regulated by SAGA), direct recruitment to chromatin (ChEC-seq) and impact in transcription *per se* (using analysis of newly-synthesized RNA). Also, we have simultaneously investigated whether TFIID would also play a general role on RNA Pol II transcript, using strains that allow for rapid and efficient depletion of TFIID subunits.

Several new results using independent and different experiments have concordantly challenged the role of these coactivators on transcription: (i) SAGA-dependent changes in H3K9 acetylation revealed that SAGA, through its Gcn5 subunit acetylates most active promoters; (ii) direct localization studies by ChEC-seq indicate recruitment of SAGA to UASs of both SAGA- and TFIID-dominated genes; (iii) SAGA is required for TBP and RNA Pol II recruitment at active promoters of both SAGA- and TFIID-dominated genes; (iv) disruption of the SAGA complex led to a global decrease in mRNA synthesis; (v) Gcn5 acetyltransferase synergizes with Spt3 (TBP-binding) to promote overall RNA Pol II transcription; (vi) disruption of TFIID complex leads to a global decrease in mRNA synthesis.

Hence, our observations do not support the earlier distinction of two different pathways for PIC assembly with TBP delivery at promoters depending on either SAGA or TFIID, leading us to reconsider SAGA and TFIID as general cofactors for RNA Pol II, being both recruited and required for the transcription of a vast majority of genes.

Résumé

Introduction

Chez les eucaryotes, la transcription par l'ARN polymérase II (ARN Pol II) est un mécanisme très précisément régulé. L'initiation de la transcription est définie par le recrutement et la formation sur un promoteur du complexe de pré-initiation (PIC). Le PIC est composé de l'ARN Pol II et des facteurs généraux de la transcription, comme le facteur TFIID pouvant se fixer sur la boite TATA (TATA box) par l'intermédiaire de la sous-unité TBP (TATA box binding protein). Dans le noyau des cellules eucaryotes, l'ADN est enroulé autour des protéines histones pour former les nucléosomes, limitant ainsi l'accessibilité des séquences d'ADN et nécessitant d'autres mécanismes de régulation. La chromatine adopte effectivement différentes conformations correspondant à différents degrés de compaction dont la régulation est cruciale pour tous les mécanismes faisant intervenir l'ADN. Un des nombreux complexes affectant la structure de la chromatine dans les cellules eucaryotes est le complexe SAGA (Spt-Ada-Gcn5 acetyltransferase).

Le complexe SAGA a deux activités enzymatiques indépendantes, une activité d'acétylation des histones portée par la sous-unité Gcn5 et une activité de désubiquitination d'histone portée par Ubp8. SAGA peut également interagir avec la protéine TBP par l'intermédiaire des sous-unités Spt3 et Spt8 alors que d'autres sous-unités telles que Spt7 ou Spt20 forment un module architectural du complexe. Les premières études génomiques réalisées dans la levure *S. cerevisiae* ont mené à la distinction de gènes dits 'dominés par SAGA' et de gènes 'dominés par TFIID', en fonction de leur dépendance principale pour les complexes SAGA ou TFIID. Les gènes 'dominés par SAGA' ne correspondent qu'à environ 10% des gènes de levure. Mais des études plus récentes ont montré que seule une fraction mineure des gènes sur lesquels SAGA est recruté est dérégulée lorsque ce complexe est inactivé. De la même façon, une minorité des gènes régulés par SAGA est lié par des sous-unités du complexe indiquant une déconnexion entre recrutement de SAGA et effet transcriptionnel. Différentes études ont également montré que les activités enzymatiques de SAGA agissent d'une façon globale sur l'ensemble du génome. En effet, la délétion des sous-

unités catalytiques de SAGA entrainent des modifications importantes des niveaux globaux des marques d'histones qu'elles régulent.

A mon arrivée dans le groupe de Laszlo Tora, mes collègues avaient déjà montré que, chez la levure comme chez les mammifères, le complexe SAGA est responsable de l'acétylation de la lysine 9 de l'histone H3 (H3K9ac) sur les promoteurs et de la désubiquitination de l'histone H2B mono-ubiquitiné (H2Bub) dans la région transcrite de tous les gènes actifs. D'autre part une inactivation de SAGA chez *S. cerevisiae* provoque une diminution majeure du recrutement de l'ARN Pol II sur ces gènes. Ces nouvelles connaissances suggèraient que ce complexe co-activateur serait nécessaire pour toute transcription par l'ARN Pol II.

Afin de confirmer cette hypothèse, les buts de ma thèse de Doctorat étaient les suivants :

(a) de quantifier, avec la meilleure exactitude possible, le rôle joué par SAGA (et TFIID) dans la transcription par l'ARN Pol II chez *S. cerevisiae ;*

(b) de résoudre, par l'analyse comparative d'ARN totaux et d'ARN nouvellement synthétisés, les contradictions révélées par les études antérieures d'ARNm.

(c) de décrire le recrutement de SAGA sur l'ensemble du génome, en utilisant des méthodes alternatives à l'immunoprécipitation de la chromatine et séquençage haut-débit (ChIP-seq).

(d) de déterminer les contributions respectives de SAGA et TFIID à la transcription globale par l'ARN Pol II et de comprendre comment ces deux complexes coopèrent.

Résultats

Pour obtenir une carte à haute résolution des sites de liaison de SAGA sur la chromatine chez S. cerevisiae, nous avons utilisé une méthode orthogonale au ChIP-seq appelée clivage endogène de la chromatine couplé au séquençage haut-débit (ChEC-seq). Cette technique utilise une fusion de la MNase avec une protéine d'intérêt, ici différentes sous-unités du complexe SAGA, permettant d'induire des coupures de l'ADN aux endroits où cette protéine est recrutée. Nous avons observé une fréquence élevée de coupures induites par la fusion SAGA-MNase sur les régions régulatrices amont (UASs) de nombreux gènes, aussi bien pour des gènes précédemment caractérisés comme dominés par SAGA que pour des gènes TFIID-dépendants (**Figure 1**). Ces résultats montrent donc que SAGA, comme le

complexe Mediator, lie les régions régulatrices (UASs) de très nombreux gènes, indépendamment de leur classification et que cette liaison est très largement indépendante de l'activité transcriptionnelle de ces gènes.

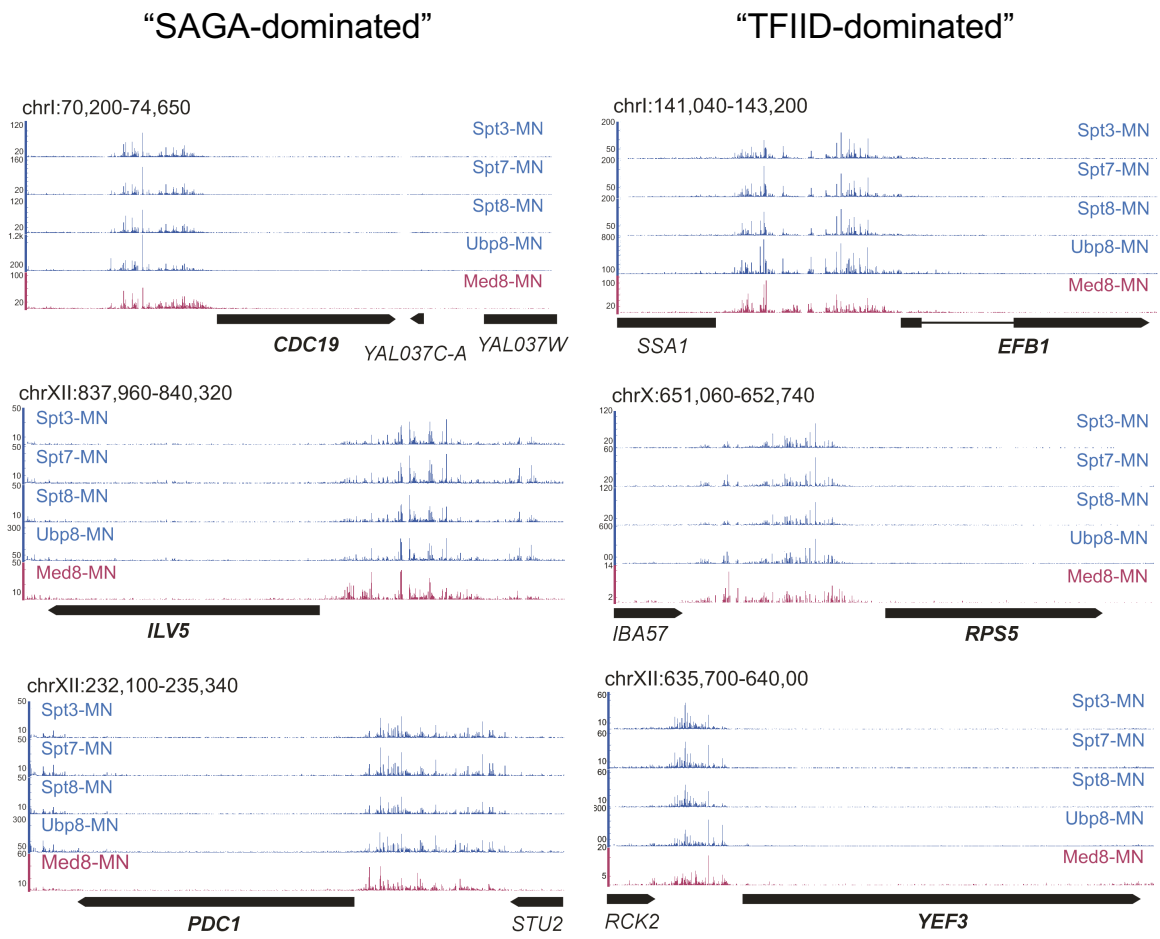


Figure 1- Profils de liaison de sous-unités de SAGA analysé par ChEC-seq. Tracés des signaux de clivages induits par les fusions MNase-Spt3, -Spt7, -Spt8 et Ubp8 sur (A) trois exemples représentatifs de gènes dominés par SAGA (*CDC19*, *ILV5* et *PDC1*) et (B) trois gènes dominés par TFID (*EFB1*, *RPS5* et *YEF3*). Les sites de clivage par la fusion Med8-MNase sont montrés comme référence pour la liaison du complexe coactivateur Médiator sur les UASs (Grunberg et al., 2016).

Plusieurs études récentes ont révélé que des effets globaux sur la transcription par l'ARN Pol II peuvent être compensés par des modifications concomitantes de la dégradation des ARNm, avec une normalisation, par effet tampon, des taux d'ARNm. Par ces mécanismes compensatoires, une diminution globale de la transcription par l'ARN Pol II peut être indétectable par des études transcriptomiques analysant des ARN totaux. Pour confirmer que l'inactivation de SAGA provoque une diminution globale de la transcription par l'ARN Pol II, j'ai adapté au laboratoire un protocole appelé cDTA (pour *comparative dynamic*

transcriptome analysis). Cette méthode permet de mesurer, pour tous les gènes de *S. cerevisiae*, les taux de synthèse et de dégradation des ARNm et de comparer ces taux entre différentes souches de levure. Pour identifier les sous-unités de SAGA ayant un impact sur la transcription, j'ai choisi des souches dans lesquelles les gènes de deux sous-unités structurales sont inactivés par délétion (*spt7*Δ, *spt20*Δ). Après un marquage bref de 6min avec un analogue d'uracile, les ARNm nouvellement-synthétisés sont purifiés et quantifiés soit par RT-qPCR (analyse de gènes sélectionnés) soit par analyse transcriptomique cDTA (pour l'ensemble du génome). Dans ces deux souches, l'analyse des ARN nouvellement synthétisés a révélé une diminution très importante (d'environ 4 fois) de la synthèse de tous les ARNm alors que la dégradation des ARNm était également fortement ralentie (par un facteur équivalent) (**Figure 2**).

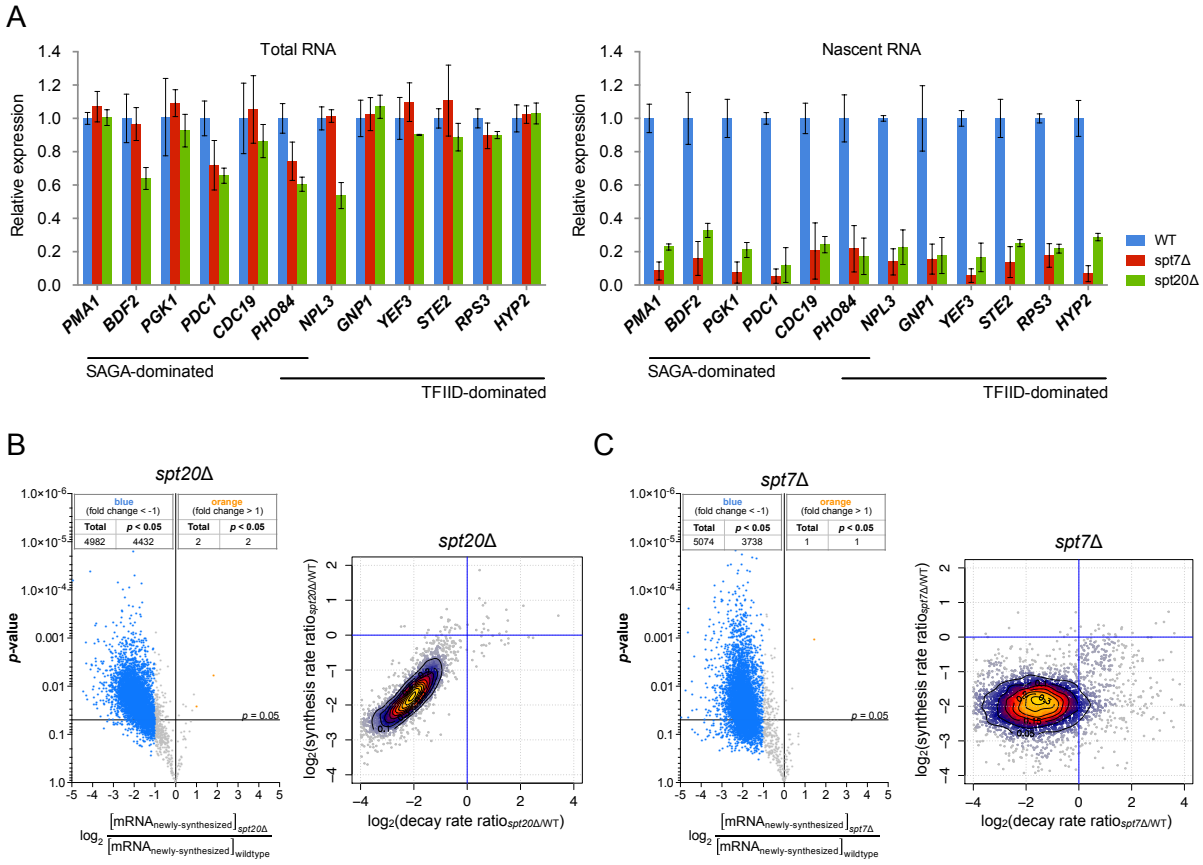


Figure 2 - SAGA est nécessaire pour l'activité de l'ARN polymérase II sur l'ensemble des gènes. **(A)** Les ARNm totaux et nouvellement synthétisés ont été extraits de souches de levure sauvages (WT), *spt7*Δ, and *spt20*Δ et quantifiés par RT-qPCR. (B et C) (Panels de gauche) Volcano plot montrant l'amplitude des modifications des taux d'ARNm nouvellement synthétisés et leur significativité statistique (p-value). Les amplitudes de modification (FC) sont calculées comme le Log2 du rapport entre le taux d'ARNm de chaque gène après normalisation par le signal *S. pombe* dans la souche *spt20*Δ (B) ou la souche *spt7*Δ (C) et le taux du même ARNm dans la souche sauvage *S. cerevisiae*. (Panels de droite) Profils cDTA pour les souches *spt20*Δ (**B**) et *spt7*Δ (**C**). Pour tous les gènes analysés, les modifications de taux de synthèse sont indiquées en fonction des modifications des taux de dégradation. Les taux de modifications sont calculés comme le Log2 des rapports entre mutant et sauvage.

Cette observation permet d'expliquer pourquoi les analyses précédentes, réalisées à partir d'échantillons d'ARN totaux, ne montraient que des modifications transcriptionnelles mineures. Les effets transcriptionnels observés étaient comparables pour des gènes 'dominés par SAGA' et pour ceux 'dominés par TFIID' (**Figure 3**).

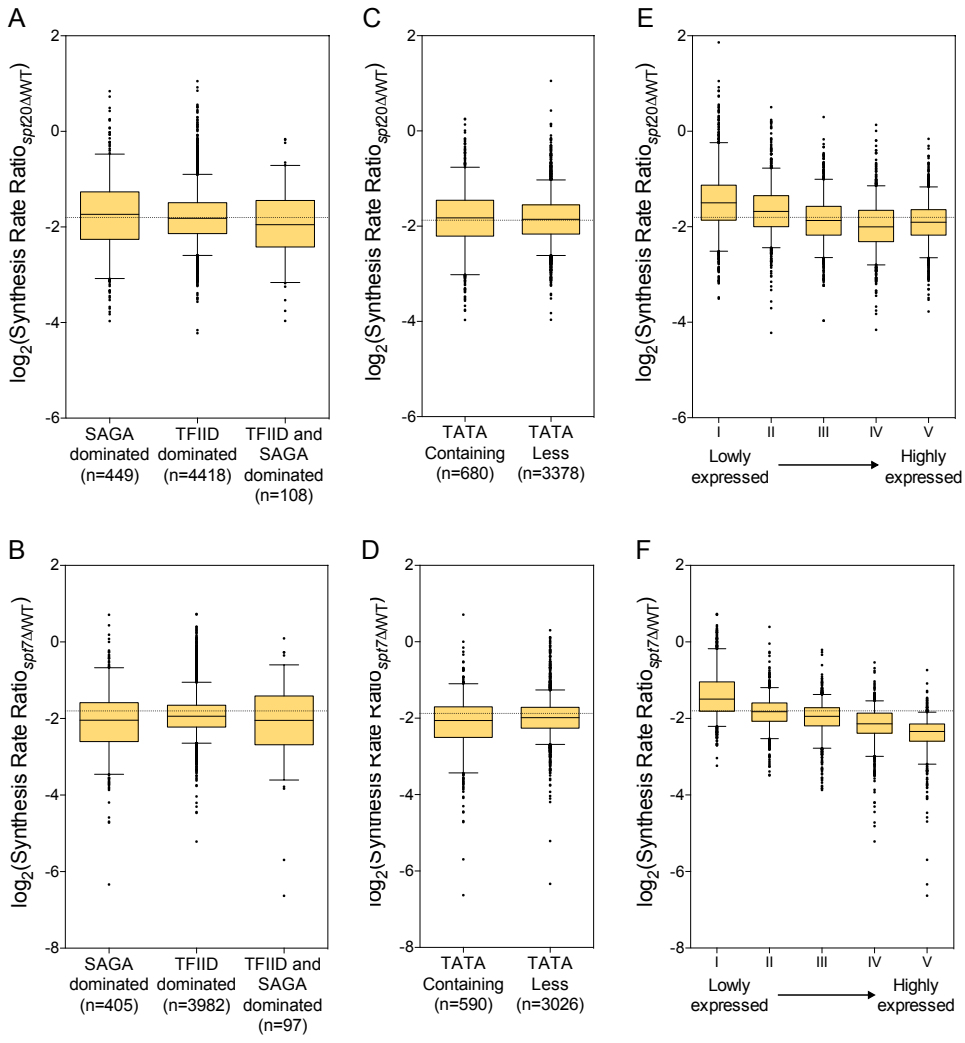


Figure 3 – Changements des taux de synthèse des ARNm dans les souches *spt20*Δ et *spt7*Δ pour différentes classes de gènes. Boites à moustaches montrant la distribution des modifications des taux de synthèse d'ARNm entre les souches *spt20*Δ **(A,C,E)** ou *spt7*Δ **(B,D,F)** et les souches sauvages. Des changements similaires sont observés entre des gènes dominés par SAGA ou par TFIID (A,B) ou entre des promoteur ayant ou non une boite TATA canonique (C,D). Les gènes sont répartis en quintiles en fonction de leur niveau d'expression dans une souche sauvage et les modifications des taux de synthèse des ARNm sont représentées pour chaque quintile (E,F).

Afin de confirmer, par des expériences indépendantes, la mise en place par les cellules d'un mécanisme compensatoire d'augmentation de la demi-vie des ARNm, j'ai mesuré les demi-vies des ARNm en quantifiant la décroissance de ces transcrits après inhibition de la transcription par la thiolutin. Dans les deux souches étudiées (*spt7*Δ, *spt20*Δ), la demi-vie des ARNm testés a été trouvée allongée, confirmant ainsi les modifications de la stabilité de ces ARNm mesurées par cDTA (**Figure 4A et B**).

Les mutants *spt7*Δ et *spt20*Δ présentant un phénotype de croissance fortement ralentie, j'ai voulu confirmer que la réduction de la synthèse des ARNm est bien un effet direct de la perte de SAGA et non un effet secondaire dû à un allongement du cycle cellulaire, comme précédemment observé pour d'autres complexes. Nous avons généré une souche de levure dans laquelle on peut induire une perte nucléaire de la sous-unité Spt7 par rétention dans le cytoplasme en utilisant la technologie 'anchor-away'. L'addition de rapamycine au milieu de culture induira un ancrage de Spt7 à une protéine ribosomale et donc une déplétion nucléaire rapide. L'utilisation de ce système inductible a confirmé les résultats obtenus dans une souche présentant une délétion stable du gène *SPT7* indiquant que cet effet transcriptionnel global résulte directement de la perte d'activité de SAGA. De plus, nous avons observé que les ARNm totaux diminuent en début d'induction, correspondant une phase de diminution de synthèse peu ou mal compensée par une diminution de leur dégradation puis reviennent à des taux normaux lorsque ces mécanismes compensatoires sont efficacement mis en place (**Figure 4C**).

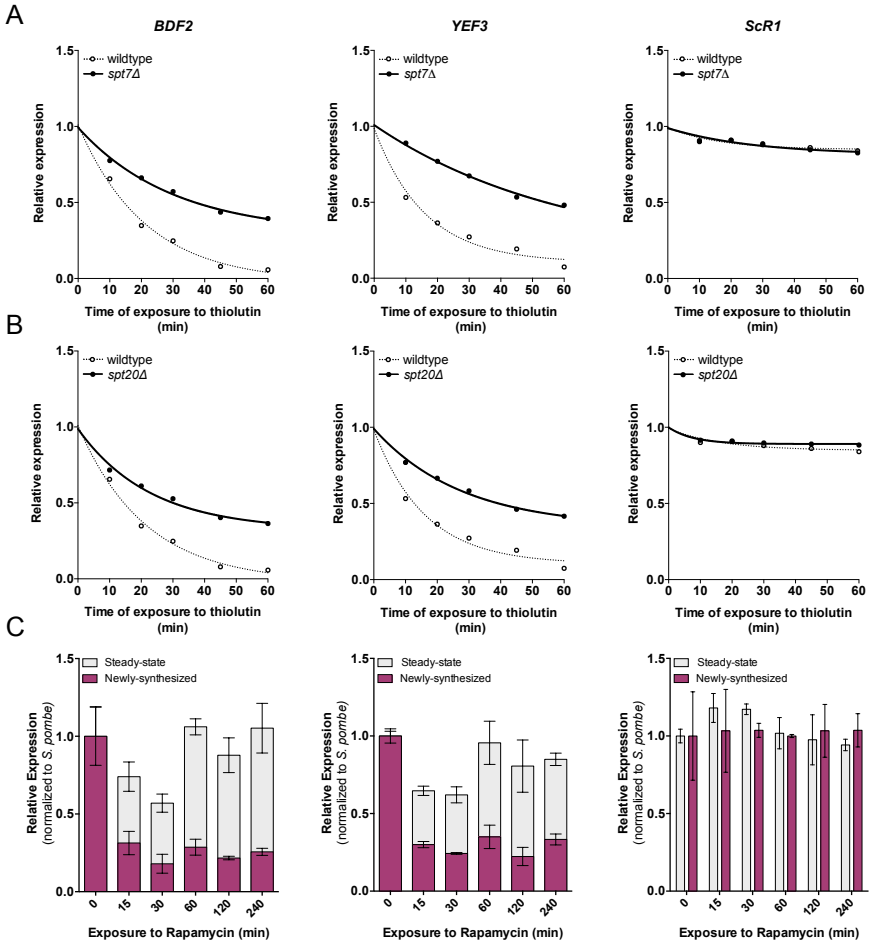


Figure 4 – (A,B) L'inactivation de SAGA entraine une augmentation de la demi-vie des ARNm. Quantification de la demi-vie des ARN par inhibition de la transcription par la thiolutin dans des souches *spt7*Δ **(A)** ou *spt20*Δ **(B)**. **(C)** Analyse en fonction du temps des modifications des taux d'ARNm totaux ou nouvellement synthétisés pour un gène dominé par SAGA (*BDF2*), un gène dominé par TFIID (*YEF3*) et un gène contrôle transcrit par l'ARN pol III (*ScR1*) lors de l'induction d'une déplétion nucléaire de Spt7.

Chez la levure, le complexe SAGA est constitué de 19 sous-unités différentes, organisées en cinq modules fonctionnels et structuraux. Notre but suivant était de déterminer les contributions respectives des différentes activités de SAGA aux effets transcriptionnels observés. Afin de tester ces activités individuellement et de façon combinatoire, j'ai répété les quantifications d'ARNm nouvellement synthétisés purifiés à partir des souches suivantes: *ubp8*Δ, *gcn5*Δ, *spt3*Δ, *spt8*Δ, *ubp8*Δ*gcn5*Δ and *spt3*Δ*gcn5*Δ. Ces analyses suggèrent que les

sous-unités Spt8 et Ubp8 individuellement n'auraient que peu d'effet sur la transcription. Par contre la perte des sous-unités Gcn5 ou Spt3 provoque une diminution globale de la transcription par l'ARN Pol II (d'environ 1,5 fois pour Gcn5 et 2 fois pour Spt3). D'autre part, les différentes sous-unités du complexe semblent agir de façon synergique. En effet la perte simultanée de Gcn5 et Ubp8 a un effet plus sévère que l'addition des effets observés dans les deux souches. Les effets transcriptionnels les plus sévères ont été observés dans la souche *spt3*Δ*gcn5*Δ, avec une réduction médiane d'environ 10 fois des taux d'ARNm, suggérant une action synergique des activités d'acétylation d'histone et d'interaction avec TBP. Cet effet est plus prononcé que celui observé dans les souches avec une perte de sous-unités structurales de SAGA. L'ensemble de ces résultats suggère donc que l'activité transcriptionnelle de SAGA soit due à l'addition de différentes activités agissant de façon synergique au sein d'un seul complexe (**Figure 5**).

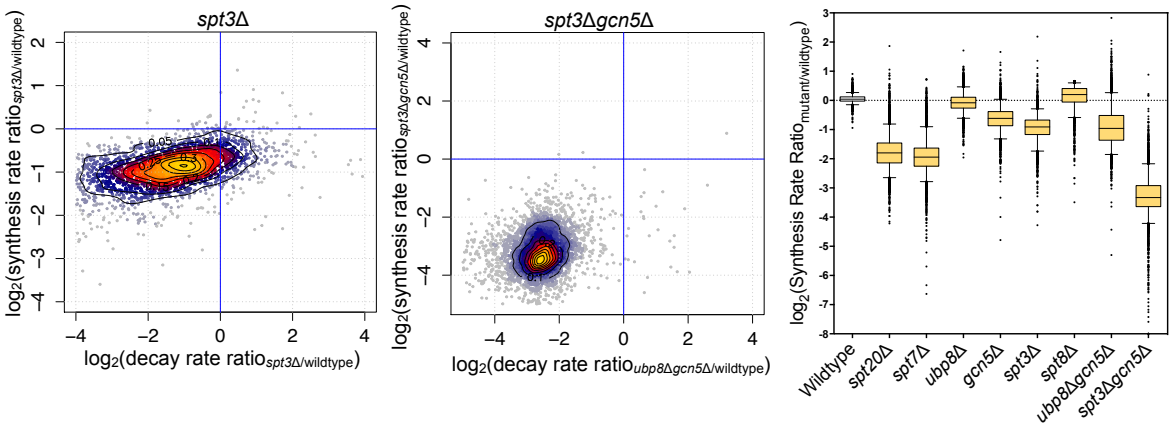


Figure 5 – Analyse par cDTA de souches délétées pour différentes sous-unités de SAGA. ***(Panels de Gauche et Centre)*** Pour chaque ARNm, les taux de synthèse et de dégradation sont mesurés dans les souches *spt3*Δ et *spt3*Δ*gcn5*Δ. Les modifications (calculées comme le Log2 du rapport entre souche mutante et sauvage) des taux de synthèse sont représentées en fonction de celles des taux de dégradation. ***(Droite)*** Boites à moustache montrant l'étendue des modifications des taux de synthèse des ARNm pour les différentes souches analysées.

L'hypothèse que les complexes SAGA et TFIID puissent participer individuellement au recrutement de la protéine TBP sur les promoteurs, permettait de fournir une explication mécanistique à la distinction entre les gènes 'dominés par SAGA' et ceux 'dominés par TFIID'. La diminution globale de la transcription par l'ARN Pol II dans la souche *spt3*Δ suggère au contraire que SAGA participerait de la même manière au recrutement de TBP sur les promoteurs ces deux classes de gènes. J'ai réalisé des expériences d'immunoprécipitation

de la chromatine (ChIP) montrant que le recrutement de TBP sur les promoteurs de gènes dominés par SAGA ou par TFIID est très fortement réduit en l'absence de la sous-unité Spt3. Ces résultats révèlent donc un rôle majeur de SAGA pour l'assemblage de la machinerie transcriptionnelle de base au niveau de tous les promoteurs des gènes transcrits par l'ARN Pol II (**Figure 6**).

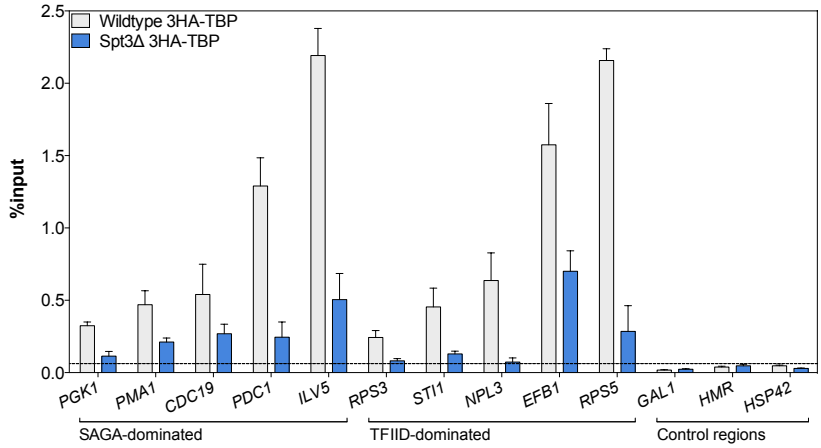


Figure 6 – Analyse par ChIP du recrutement de TBP sur les promoteurs de gènes dominés par SAGA ou par TFIID dans une souche de levure délétée pour *SPT3*. L'enrichissement de TBP sur les promoteurs de 5 gènes dominés par SAGA, 6 gènes dominés par TFIID et 3 régions controles a été quantifié par ChIP-qPCR anti-HA sur des souches *spt3*Δ *3HA-TBP* et sur la souche parentale *3HA-TBP*. Les valeurs (moyenne ± SD de trois expériences de ChIP indépendantes) sont exprimées en pourcentage du signal d'ADN de l'input.

A partir de ces observations, nous avons également voulu réévaluer le rôle joué par différentes sous-unités de TFIID dans la transcription. Nous avons généré des souches permettant d'induire une déplétion nucléaire de Taf4 (une sous-unité spécifique de TFIID) ou de Taf5 (une sous-unité partagée par les complexes SAGA et TFIID). La perte de ces sous-unités provoque, comme nous l'avons observé après déplétion nucléaire de Spt7, une diminution importante de la synthèse des ARNm. En parallèle, notre collaborateur Steve Hahn, a montré que le recrutement de l'ARN Pol II est très fortement diminué sur pratiquement tous les gènes après déplétion de différentes sous-unités de TFIID (**Figure 7**).

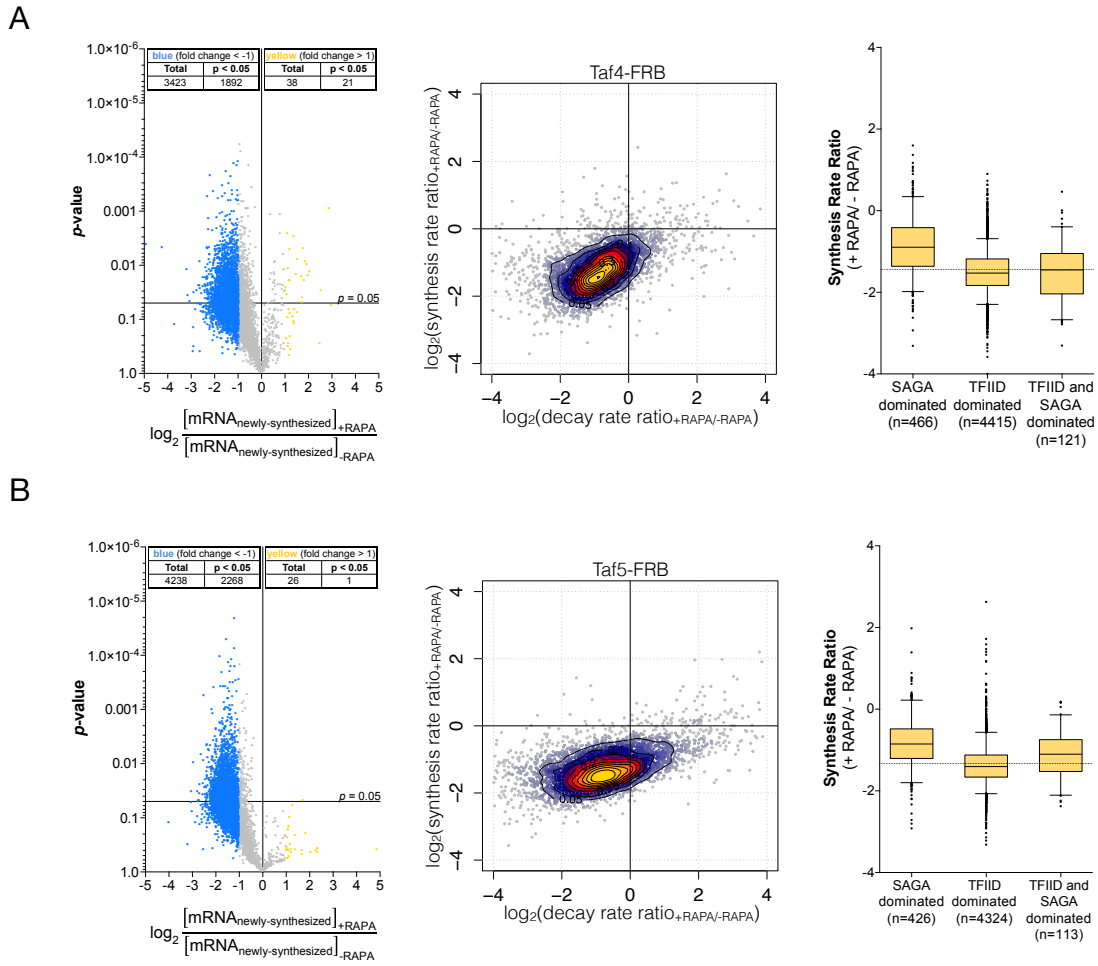


Figure 7 – Analyse cDTA après déplétion conditionnelle de (A) Taf4 et (B) Taf5 (souches anchor-away). ***(Panels de Gauche)*** Volcano plot montrant les modifications les taux d'ARNm nouvellement synthétisés après déplétion nucléaires de **(A)** Taf4 ou **(B)** Taf5 en fonction de leur significativité statistique (p-value). (Panels du centre) Modifications des taux de synthèse d'ARNm pour 5434 gènes après traitement par rapamycin dans une souche **(A)** *TAF4-FRB* ou **(B)** *TAF5-FRB* sont représentées en fonction des modifications des taux de dégradation de ces mêmes ARNm. (Panels de Droite) Boites à moustache montrant pour chaque classe de gènes, la distribution des modifications des taux de synthèse des ARNm après déplétion nucléaire de **(A)** Taf4 ou **(B)** Taf5. La ligne pointillée horizontale indique la valeur médiane du rapport des taux de synthèse.

Conclusions

L'ensemble de mes travaux, reposant sur une combinaison d'approches complémentaires, a permis de montrer que les deux complexes SAGA et TFIID sont recrutés sur les séquences régulatrices d'une très grande majorité de gènes chez la levure *S. cerevisiae,* et jouent un rôle déterminant pour la transcription par l'ARN Pol II. Cette nouvelle vision du rôle des complexes SAGA et TFIID, agissant comme des cofacteurs

généraux requis pour l'ensemble de la transcription par l'ARN Pol II, mène à reconsidérer le modèle précédemment proposé stipulant que chaque complexe contrôlerait majoritairement l'expression d'un groupe de gènes distinct.

Table of contents

List of figures

List of tables

INTRODUCTION

Introduction

The capacity for a given organism to appropriately answer to environmental cues or to correctly commit and maintain a cell identity relies on transcriptional regulation. Central to the process of transcription are the complexes of proteins collectively known as RNA Polymerases (RNA Pol). RNA Pols have been found in all species, but the number and composition of these proteins vary across evolution. For example, in bacteria a single type of RNA Pol is found, while through all eukaryotic life three distinct types exist.

In eukaryotic cells, three distinct nuclear RNA Pol I, II and III carry out transcription of nuclear genes. In fact, each of this individual RNA Pols is responsible for the transcription of specific sets of genes. These genes belong to the class I, II or III depending on the polymerase responsible for their transcription. RNA Pol I is dedicated to the synthesis of ribosomal RNA (rRNA), and this accounts for up to 60% of transcriptional activity in a eukaryotic cell. Pol III synthesizes the transfer RNAs (tRNAs), the 5S rRNA and a variety of other small, untranslated RNAs. In fact, together, the activities of RNA Pol I and III account for 80% of total RNA synthesis in dividing cells. Finally, RNA Pol II is a 12-subunit enzyme that transcribes protein-coding genes to produce messenger RNAs (mRNAs) and many other classes of RNAs, such as small nuclear RNAs (snRNAs).

Importantly, several differences are found between prokaryotic and eukaryotic transcription. For example, prokaryotic transcription and translation are coupled, while in eukaryotes these two processes occur independently and in two distinct compartments (nucleus *versus* cytoplasm). Nevertheless, one of the most evident differences lies on how RNA Pols are recruited. While bacterial RNA Pol does not depend on additional components to be recruited to the transcribed *loci*, eukaryotic RNA Pol II, for instance, relies on a group of protein complexes – collectively known as general transcription factors (GTFs) – in order to be recruited to chromatin. Briefly, GTFs and RNA Pol II recruitment, and assembly of the preinitiation complex (PIC) occur through recognition of specific DNA elements just upstream of the transcription start site (TSS) – the promoter region. Once the PIC is correctly formed, transcription initiation occurs, the RNA Pol II is released and elongates the mRNA.

Moreover, the eukaryotic genome is tightly packed in the nucleus under the form of chromatin. Hence, chromatin, and not DNA itself, is the target of DNA-related processes. The basic structural unit of chromatin is the nucleosome core, a particle consisting of 147 bp

of DNA wrapped around a histone octamer. Each histone octamer is composed of two copies of H2A and H2B (that form two H2A/H2B dimers) and two copies of H3 and H4 (that forms one H3/H4 tetramer). Nevertheless, such extensive packaging of the DNA, that initially acts as an obstacle for the DNA related processes, is indeed compatible with activities that require "reading" of the DNA. In fact, activators – also known as transcription factors – are able to recognize and bind specific sequences within the DNA, ultimately triggering the recruitment of specific coactivators. Coactivators are multisubunit complexes capable of facilitating transcription by RNA Pol II via chromatin remodeling (nucleosome eviction or sliding) or through post-translation modification of histones, such as acetylation or methylation. Altogether, coactivators act in a gene-specific manner to activate transcription. On the opposite side of the spectrum, corepressors are recruited by gene-specific repressor transcription factors that change the chromatin landscape to inhibit transcription. Altogether, the combinatorial action of activators and coactivators, promote GTFs binding at promoters, PIC assembly and transcription initiation.

Importantly, mechanisms of transcription initiation are highly conserved from yeast to mammals. Nevertheless, the introduction will mainly focus on *S. cerevisiae*, with specific mentions to the metazoan context when necessary. Initially, in this chapter we are going to explore how RNA Pol II transcription initiation is tightly regulated and how specific factors – such as GTFs and coactivators like Mediator – are essential for global RNA Pol II transcription *in vivo*. Second, as the work I have been developing focuses on the role of SAGA as a general co-factor for RNA Pol II transcription in budding yeast, more detailed information regarding this complex and its previously reported gene-specific function will be explored. Finally, I will explore how transcription is intricately associated with other biological processes – such as mRNA degradation – and which new approaches we have utilized to readdress the genome-wide role of the coactivators SAGA and TFIID.

1. Transcription initiation by RNA Polymerase II

1.1. Core promoter architecture

The core promoter is the region of the DNA where the recruitment of general transcription factors (GTFs) and preinitiation complex (PIC) assembly occurs. The output of these sequential events will terminate with transcription initiation by RNA Pol II. Initially characterized in the regulatory region of mammalian genes, the core promoter was defined as the "minimal DNA elements required for basal transcription" (Smale and Kadonaga, 2003).

1.1.1. Core promoter elements

RNA Pol II core promoters are recognized as modular in structure and many of them contain specific consensus sequence elements, which are positionally constrained in relation to the transcription start site (TSS). In metazoans, functional sequence elements in core promoters include the TATA element, initiator (Inr), downstream promoter element (DPE), motif 10 element (MTE) and TFIIB recognition element (BRE) (Juven-Gershon and Kadonaga, 2010; Hahn and Young, 2011; Muller and Tora, 2014) (**Figure 8**).

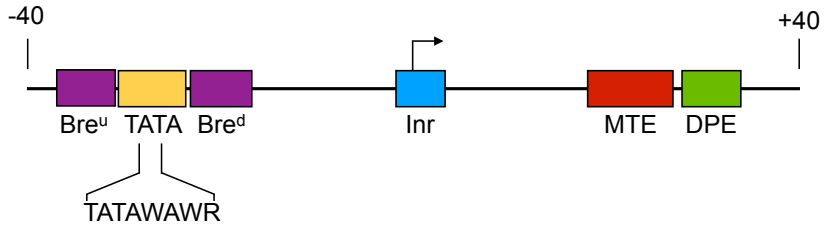


Figure 8 - A scheme of some core promoter elements, showing the locations of the BRE^u and the BRE^d relative to the TATA box, the Inr, MTE and DPE. Adapted from (Juven-Gershon and Kadonaga, 2010).

Nevertheless, the identified promoter elements are scarcer in *S. cerevisiae*. Of the abovementioned motifs, the only one that is clearly conserved from *S. cerevisiae* to metazoan is the TATA consensus sequence (TATAWAWR, with W = A or T and R = A or G) (Basehoar *et al.*, 2004; Juven-Gershon and Kadonaga, 2010; Sugihara *et al.*, 2011). The Inr

encompasses the transcription start site and is probably the most commonly occurring core promoter motif. Interestingly, Inr-like sequences, but not Inr consensus sequences, have been described in *S. cerevisiae* (Yang *et al.*, 2007). Moreover, the BRE, MTE and DPE motifs are conserved in several eukaryotic organisms, but not in budding yeast (Lagrange *et al.*, 1998; Rhee and Pugh, 2011). Nevertheless, it has recently been shown that TFIIB contacts the DNA through 4 different regions [B1 (upstream of the TATA) and B2-B4 (downstream of TATA, extending towards the TSS)], though with no sequence-conservation properties (Rhee and Pugh, 2011).

As previously mentioned, the TATA box is the only conserved element in *S. cerevisiae*. This consensus sequence is known to exist on the promoter of approximately 19% of the yeast genes, while the remnant were classified as TATA-less (Basehoar *et al.*, 2004). The TATA sequence is specifically recognized by the TATA-binding protein (TBP), serving as the docking site for the latter. Once bound to the target DNA, TBP promotes bending of the template DNA, facilitating the assembly of the basal transcription machinery. In fact, TBP is so important for transcription initiation that it is a core component of the transcription machinery of the three RNA Pols. While the TATA-box is present in a minority of promoters, TBP is still required for the expression of TATA-less genes and still delivered at their promoters. This presented itself as a conundrum, due to the failure to identify TBP-recognition elements within the TATA-less genes. However, the group of B. Franklin Pugh identified that TATA-less promoters instead contained a sequence having two or less mismatches to the TATA box consensus. These elements are designated as 'TATA-like', since they do not form a consensus, and together with the canonical TATA consensus sequence, are referred as 'TATA elements' (Rhee and Pugh, 2012). Interestingly, the characterization of a new promoter DNA element, the GA element, which occurs at one-third of TATA-less genes and is reported to be recognized by TBP, partially addresses how TBP can be delivered at TATA-less promoters (Seizl *et al.*, 2011). Nevertheless, it has been recently reported that sequence-specific TBP binding is not required at yeast TATA-less promoters and that promoter specificity must be generated by alternative mechanisms other than TBP recognition (Kamenova *et al.*, 2014). Just like TBP is not exclusive for RNA Pol II transcription, TATA elements are also found in the promoter of class I and class III genes in several organisms (Trivedi *et al.*, 1999; Hamada *et al.*, 2001; Yukawa *et al.*, 2011).

Interestingly, one important difference between yeast and metazoan promoters is the site of transcription initiation in comparison to the TATA box location. In both yeast and

metazoan, the PIC is assembled around the TBP-TATA complex, positioning the active site of the RNA Pol II approximately 30 bp downstream of the TATA, the exact location of the metazoan transcription initiation site. Nevertheless, in *S. cerevisiae* RNA Pol II initiates transcription at preferred sequences, within around 50-120 bp downstream of the TATA (Hahn and Young, 2011). Additionally, in budding yeast, the location of the TATA-box or the TATA-like sequence in relation to the TSS differs: the TSS at TATA-less promoters reside ≈10–20 bp closer to the TATA element than at TATA-box-containing promoters (Rhee and Pugh, 2011).

1.1.2. Promoter nucleosome positioning and transcription regulation

Genome-wide studies using MNase digestion have provided tremendous insights regarding nucleosome positioning. Initially, promoter regions were characterized as being overall depleted of nucleosomes relative to transcribed regions, forming nucleosome-free regions (NFR). NFRs are localized just upstream of the TSS and this pattern of nucleosome positioning was found to be conserved from yeast to fly and human (Yuan *et al.*, 2005; Ozsolak *et al.*, 2007; Mavrich *et al.*, 2008a; Mavrich *et al.*, 2008b; Schones *et al.*, 2008). The first nucleosome downstream of the TSS – +1 nucleosome – is strongly localized and is the nucleosome presenting the tightest position. Remarkably, while the abovementioned pattern was known to be somewhat conserved over eukaryotic life, the position of the +1 nucleosome relative to TSS seems to vary in different organisms, thus reflecting differences in regulatory mechanisms or in the core transcription machinery. Specifically, the center of the +1 nucleosome in *S. cerevisiae* is located approximately 50–60 bp downstream of the TSS, with transcription typically starting around 10 bases into the first nucleosome (Yuan *et al.*, 2005; Mavrich *et al.*, 2008a). Conversely, in *Drosophila*, the center of the +1 nucleosome is found 135 bp downstream of the TSS, with the upstream border of the nucleosome located ≈60 bp downstream of the TSS (Mavrich *et al.*, 2008b). Also in human, the 5′-end border of the +1 nucleosome is located at +40 bp from the TSS in actively transcribed genes (Schones *et al.*, 2008). While this overall view represented the “average gene”, later studies differentiated distinct organizations depending on nucleosome positioning, ultimately offering distinct regulatory pathways for gene transcription.

According to the work of Tirosh and Barkai, promoters in budding yeast can be distinguished according to the plasticity of their expression. In this context, transcriptional plasticity refers to the fluctuation of expression a given gene can experience: genes with low

plasticity are those whose transcription is stable through time, while genes with high transcriptional plasticity modulate their expression in response to cues, for example, exhibiting a big range of expression over time. Interestingly, they have reported that expression plasticity was highly correlated with the nucleosome positioning around the promoters. Hence, they suggest two general mechanisms of regulation of gene transcription by nucleosomes (**Figure 9, *Left***). In general, low-plasticity genes are associated with either a static or well-positioned architecture of the nucleosomes and a substantial and well-evident nucleosome-free region upstream of the TSS. Also, this pattern – depleted proximal-nucleosome pattern also termed as "open promoter pattern" – requires the deposition of H2A.Z/Htz1. Additionally, the regulatory elements of these genes are usually located within the nucleosome-free region and readily accessible to transcription factors, thus concluding that these genes do not necessarily depend on chromatin remodeling and chromatin regulators. This nucleosome position is characteristic of housekeeping genes, whose transcription is constant, hence with low plasticity (Tirosh and Barkai, 2008).

On the contrary, high plasticity genes are associated with high and dynamic nucleosome occupancy across the promoter, namely upstream of the TSS, in comparison to the "open promoters" with their distinctive nucleosome-depleted region. Those promoters assume the so-called occupied proximal-nucleosome pattern or "covered promoters". In fact, such high nucleosome occupancy might compete with the binding of transcription factors to the promoter, indicating that these genes are more prone to the action of chromatin remodelers or chromatin modifiers to be transcribed in comparison to "open promoters", as previously reported. Overall, this classification represents the stress-induced genes, which present high transcriptional plasticity, are characterized by high nucleosome occupancy during uninduced states and display high nucleosome remodeling and histone eviction upon gene activation (Tirosh and Barkai, 2008; Cairns, 2009; Rando and Winston, 2012). Also, occupied proximal-nucleosome promoters have the tendency to contain a TATA consensus sequence, as later sustained (Tirosh and Barkai, 2008; Cairns, 2009; Rhee and Pugh, 2012).

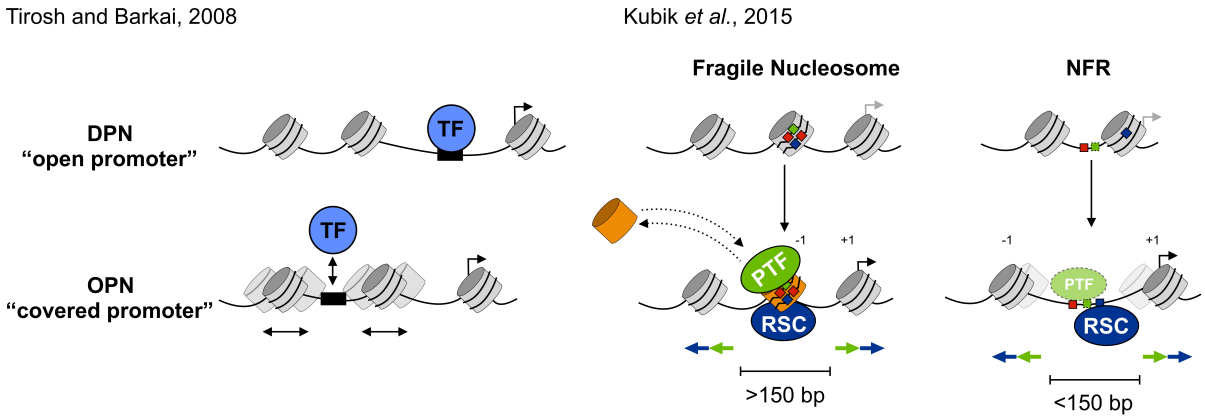


Figure 9 – Models representing distinct types of promoter architectures. (*Left*) Model described by Tirosh and Barkai. Low-plasticity (DPN) promoters, or "open promoters", tend to have well-positioned nucleosomes and a strong NFR right upstream of the TSS. High-plasticity (OPN) promoters, or "covered promoters" have fuzzy nucleosomes whose positions are dynamic, thus competing with transcription factors (TF) for binding. Binding sites are distal to the TSS, and binding of regulators to these sites may influence nucleosome positions proximal to the TSS, thus affecting transcription. Adapted from (Tirosh and Barkai, 2008) **(*Right*)** Model described by Kubik *et al*. Promoters possessing a -1 fragile nucleosome contain a higher number of RSC-associated motifs and pioneer transcription factor (PTF) binding sites. Recruitment of these factors generates an accessible region bigger than 150 bp, which is occupied by a fragile nucleosome. This architecture favors high transcription with low noise. On the other hand, at promoters with a constitutive NFR, a lower density of the motifs and weaker or absent PTF binding results in an accessible DNA region of less than 150 bp. Expression driven from these promoters is generally weaker than from 1 FN promoters, but also with low transcriptional noise. Adapted from (Kubik *et al*., 2017).

Nevertheless, a recent publication has challenged this vision of a NFR flanked by two well-positioned -1 and +1 nucleosomes (Kubik *et al.*, 2015), namely through the notion that NFRs were not completely devoided of nucleosome but instead possess unusually unstable nucleosomes (Jin *et al.*, 2009; Henikoff *et al.*, 2011; Kent *et al.*, 2011). Using MNase titration, the authors were able to obtain a genome-wide view of nucleosome position and stability, allowing the disclosure of a new model for promoter architecture clearly distinctive from the previous one. With this approach, it was specifically shown that 2000 out of the approximately 5000 RNA Pol II promoters in yeast specifically display MNase-sensitive particles upstream of the TSS, in the region previously characterized as a NFR. Those genes having an unstable -1 nucleosomes, called fragile nucleosome (FN), strongly correlated with high levels of gene expression and are particularly enriched in housekeeping genes. More interestingly, the authors also reported that FNs occur in every NFR spanning more than 150 bp, indicating that a fragile nucleosome always forms in the presence of a DNA stretch long enough to do so (Kubik *et al.*, 2015). In fact, the presence of fragile nucleosome particles in

NFRs have been reported in the past (Xi *et al.*, 2011), while disputed by others (Fan *et al.*, 2010; Chereji *et al.*, 2017). Additionally, the authors explored the means of regulation of NFRs or -1 FN promoters and have developed the model represented in **Figure 9, *Right***. Their findings suggest that the RSC chromatin remodeler acts in a genome-wide fashion, by pushing nucleosomes away from promoter regions containing binding sites for RSC or transcription factors. Thus, via this mechanism, RSC establishes the formation of either a true NFR (when motifs are clustered in a constrained region) or a site where a fragile nucleosome will form (when binding motifs are numerous and dispersed) (Kubik *et al.*, 2015; Kubik *et al.*, 2017). The remainder 3000 yeast promoters out of the 5000 analyzed, displayed a -1 stable nucleosome, indicating that a NFR exists in these genes, but never exceeding a length of 150 bp. Hence, NFRs are smaller than previously anticipated (Kubik *et al.*, 2015; Kubik *et al.*, 2017).

1.2. RNA Polymerase II, general transcription factors and preinitiation complex assembly

In eukaryotes RNA Pol II cannot be recruited and drive transcription by itself. In fact, as discussed above, the architecture of the core promoter and other *cis*- and *trans*-regulatory elements serves as a platform for RNA Pol II recruitment, positioning, and transcription initiation. Among the factors collectively known as general transcription factors (GTFs), we can include the transcription factor (TF) IIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH.

1.2.1. PIC assembly: sequential GTF binding and RNA Pol II holoenzyme pathways

GTFs were initially characterized as factors intrinsically and crucially necessary for RNA Pol II transcription from a DNA template *in vitro*. Upon their discovery, massive efforts have been made in order to understand how these elements interact with RNA Pol II and the DNA template to form a stable PIC. From the assortment of results collected throughout the years, two different models have been hypothesized: the sequential assembly pathway and the RNA Pol II holoenzyme pathway.

The sequential pathway relies on the *in vitro* evidence that the PIC seems to be stably assembled in a stepwise fashion even before the addition of dNTPs. Briefly, PIC formation can be divided into different steps (Buratowski *et al.*, 1989; Van Dyke *et al.*, 1989; Cheung

and Cramer, 2012; Grunberg and Hahn, 2013; He *et al.*, 2013; Sainsbury *et al.*, 2015) (**Figure 10**):

(i) Recruitment of TBP/TFIID through recognition of a TATA-box;
(ii) Stabilization of TFIID binding through TFIIA and TFIIB;
(iii) DNA-GTFs interactions anchors RNA Pol II to double-stranded promoter DNA together with TFIIF;
(iv) Formation of a stable complex containing TFIID-TFIIA-TFIIB-RNA Pol II-TFIIF (core PIC);
(v) Recruitment of TFIIE and entry of TFIIH (closed complex conformation);
(vi) DNA melting and formation of the transcription bubble (open complex conformation);
(vii) The open complex initiates RNA synthesis (6-7 nucleotides long), forming the initially transcribing complex;
(viii) As the RNA extends farther, contacts with to the promoter and GTFs are broken, resulting in promoter escape and formation of the elongation complex.

The second model – RNA Pol II holoenzyme pathway – was initially described due to the fact the RNA Pol II could be purified as a preassembled holoenzyme complex containing some GTFs, chromatin remodelers/modifiers enzymes, among others (Ossipow *et al.*, 1995). While the purified preparation would be different from laboratory to laboratory, some have shown that RNA Pol II is associated with a subset of GTFs (TFIIB, TFIIE, TFIIF and TFIIH), chromatin-associated proteins (SWI/SNF remodeler and Gcn5) and suppressors of RNA Pol B mutations (SRBs – subunits of the Mediator complex), but devoid of TFIIA and TFIID. This indicated that TFIID (stabilized by TFIIA) could facilitate the entry of RNA Pol II holoenzyme, due to its promoter-binding properties (Koleske and Young, 1994; Wu and Chiang, 1998; Wu *et al.*, 1999; Thomas and Chiang, 2006).

While these two distinct models are valid in an *in vitro* system, there is no evidence that either one occurs exclusively *in vivo*. Eventually, the two pathways might coexist and be used according to specific environmental cues, for instance. On the other hand, we might be in the presence of an extremely dynamic system, where these two models represent its two extremes.

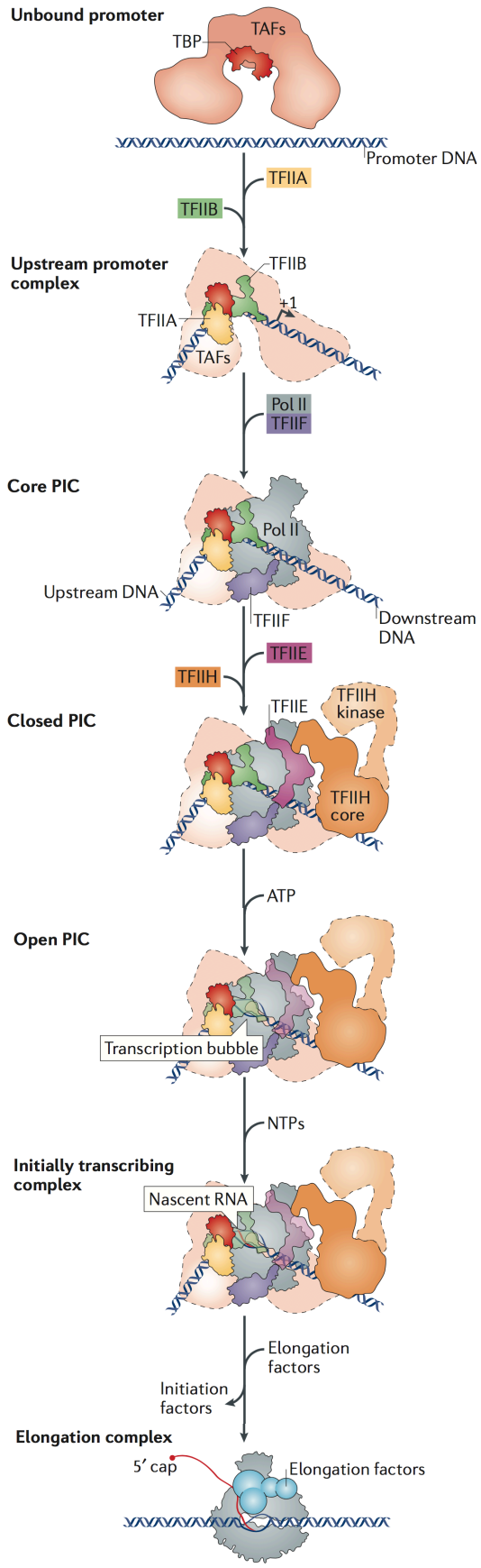


Figure 10 – (Legend on the next page) ➔

Figure 10 – Schematic representation of RNA Pol II transcription initiation according to the canonical model of sequential PIC assembly from GTFs and RNA Pol II on promoter DNA. TFIID/TBP binds to the promoter, inducing the bending of the DNA. The TBP–DNA complex is then stabilized by TFIIB and TFIIA, which flank TBP on both sides. The resulting upstream promoter complex is joined by the RNA Pol II–TFIIF complex, thus forming the core PIC. Subsequent binding of TFIIE and TFIIH terminates PIC assembly. In the presence of ATP, the DNA is opened (forming the "transcription bubble") and RNA synthesis starts. Finally, dissociation of initiation factors enables the formation of the Pol II elongation complex, which is associated with transcription elongation factors. From (Sainsbury *et al.*, 2015).

1.2.2. RNA Polymerase II

RNA Pol II is the multisubunit complex responsible for the transcription of class II genes, including all protein-coding genes. Since it is known to not be recruited to the DNA template by itself, it interacts with an assortment of factors, including the GTFs, to initiate transcription. Over the years, and with advances in technology, the understanding of the structure of RNA Pol II not only allowed an understanding of how those interactions work synergistically to achieve processive transcription but also enlighted the scientific community regarding the function of this complex.

RNA Pol II is composed of 12 highly conserved subunits (Rpb1 to Rpb12). Interestingly, the conservation in sequences and structure is so high that some mammalian subunits can partially or completely substitute their yeast counter parts. While some of the subunits are exclusive to RNA Pol II, 5 of them (Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12) are also present in RNA Pol I and RNA Pol III. However, Rpb4, Rpb7, Rpb9 and the unstructured C-terminal domain of Rpb1 are exclusive to RNA Pol II and no homologous subunits are found in other RNA Pols (Thomas and Chiang, 2006).

Initially, the highest resolution obtained of RNA Pol II was of the 10-subunit structure, which excluded Rpb4 and Rpb7 (Cramer *et al.*, 2001). RNA Pol II is composed of four distinct mobile modules denominated "core", "clamp", "shelf", and "jaw lobe" (**Figure 11A**). The core module comprises Rbp3, 10, 11 and 12 and regions of Rpb1 and Rpb2, forming the active center (Cramer *et al.*, 2001; Gnatt *et al.*, 2001). At the center of the polymerase is a deep cleft where incoming DNA template enters from one side, with the active center buried at the base of the cleft. This cleft is formed by all four mobile modules of RNA Pol II and has been observed in both closed and open conformations in the 10-subunit enzyme. Also, this 10-subunit RNA Pol II is elongation-competent, but it is not able to initiate transcription without Rpb4/Rpb7. Later, work that aimed to resolve the 12-subunit

structure of RNA Pol II revealed that Rpb4/7 binds to a pocket formed by Rpb1, Rpb2, and Rpb6 at the base of the clamp (Armache *et al.*, 2003; Bushnell and Kornberg, 2003) (**Figure 11B**). Interestingly, Rpb7 was shown to act as a block to lock the clamp in the closed conformation, suggesting that double-stranded DNA never enters the active site cleft. Rather, it has been proposed that during the open complex formation, the single-stranded DNA template strand is inserted deep into the cleft to reach the active site. Moreover, Rpb4/7 also provides a binding surface for other factors and for RNA exiting the elongating polymerase (Armache *et al.*, 2003; Bushnell and Kornberg, 2003).

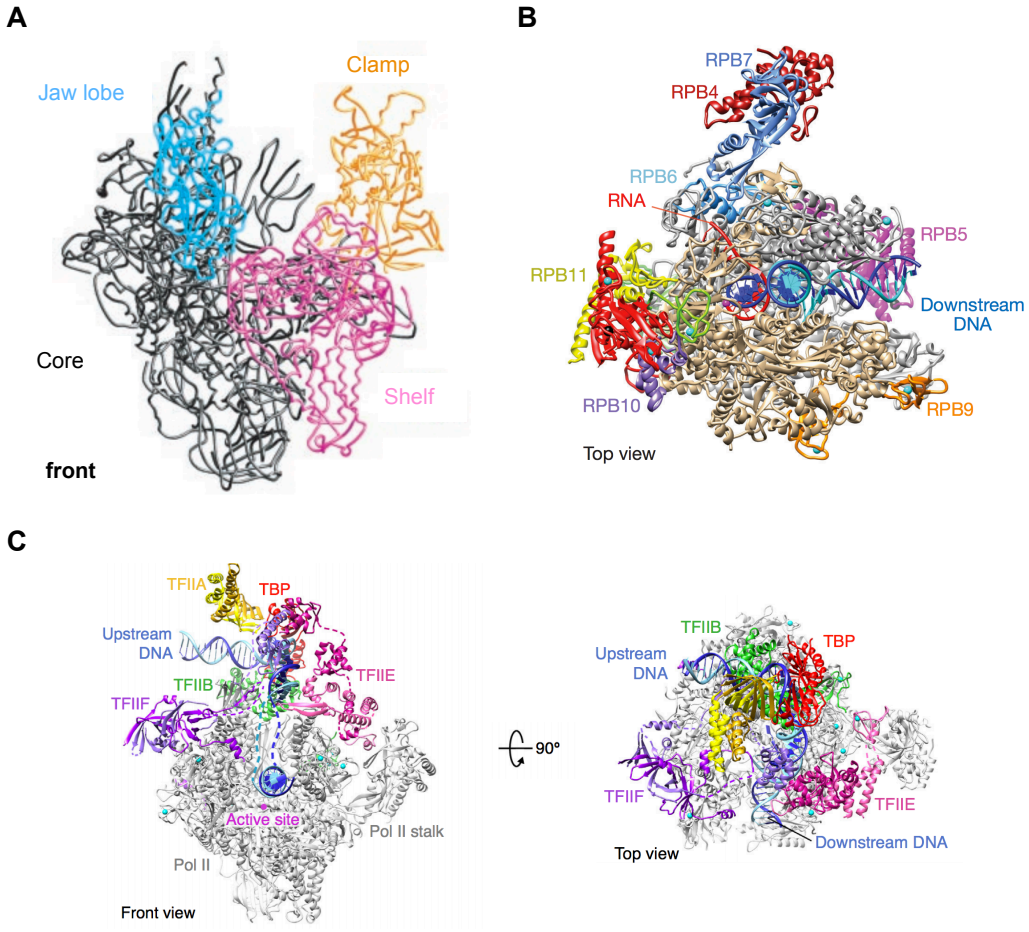


Figure 11 – Structural representation of RNA Pol II. (A) Front view of RNA Pol II depicting the four mobile modules of RNA Pol II. Backbone traces of the core (grey), jaw-lobe (blue), clamp (orange), and shelf (pink) modules. From (Cramer *et al.*, 2001). **(B)** Ribbon model of *S. cerevisiae* RNA Pol II (12 subunits complex). From (Bernecky *et al.*, 2016). **(C)** Conserved structure of the core RNA polymerase II open promoter initiation complex. Representation of RNA Pol II (silver) and GTFs. The views correspond to 'front' and 'top' views of yeast RNA Pol II and are related by a 90° rotation around a horizontal axis. From (Hantsche and Cramer, 2017).

Importantly, over the years, structural information regarding RNA Pol II, alone or in association with GTFs or coactivators, has been incremental and important to understand the function on the polymerase. Specifically, the collective efforts have disclosed insights regarding the transition from open to close conformation of the PIC, surface of interactions between RNA Pol II, the DNA template, GTFs and Mediator, and mechanisms of transition from transcription initiation to elongation (Kostrewa *et al.*, 2009; He *et al.*, 2013; Fishburn *et al.*, 2015; Murakami *et al.*, 2015; Plaschka *et al.*, 2015; He *et al.*, 2016; Louder *et al.*, 2016; Plaschka *et al.*, 2016; Robinson *et al.*, 2016; Hantsche and Cramer, 2017) (**Figure 11C**).

As previously mentioned the CTD of Rpb1 is a highly unstructured region containing tandem repeats of the heptapeptide Tyrosine-Serine-Proline-Threonine-Serine-Proline-Serine ($Y^1S^2P^3T^4S^5P^6S^7$). The number of repeats varies between the organisms. In vertebrates, the CTD contains 52 tandem consensus heptapeptide repeats, *S. cerevisiae* possess 26 repeating units, while other eukaryotes have an intermediate number of repeats. Also, in vertebrate CTDs, 21 out of 52 tandem repeats match the consensus perfectly, while the remaining 31 heptads have one or more substitutions. In budding yeast, 19 out of the 26 repeats match perfectly to the consensus sequence, with the remainder also presenting substitutions (Hsin and Manley, 2012). Also, a 10-residue sequence is found at the very C terminus of the CTD, helping its stabilization (Chapman *et al.*, 2004). Importantly, the CTD is a dynamic structure that has fundamental roles during transcription initiation and elongation, serving as a scaffold for several RNA Pol II interactors. Essentially, there are two distinct forms of the CTD: an unphosphorylated version that corresponds to the RNA Pol II at the early stages of the PIC assembly (designated RNA Pol IIA); a hyperphosphorylated CTD that occurs as transcription initiation and elongation takes place (RNA Pol IIO), with the Serine 2 (S2) and Serine 5 (S5) being specifically modified. Specifically, a generalized model for CTD phosphorylation during transcription establishes that at the beginning of the genes, S5 is phosphorylated by the TFIIH-associated kinase Kin28 (CDK7) (peaking around the TSS), and during elongation S2 is increasingly phosphorylated by Ctk1 (CDK9) or pTEFb, with loss of S5 (Hsin and Manley, 2012).

Both yeast (Kin28) and mammalian CDK7 were initially characterized to be TFIIH-associated kinases, whose activity was necessary for transcription *in vitro* (Feaver et al., 1991; Lu et al., 1992; Akoulitchev et al., 1995). Being part of TFIIH, a GTF initially recruited to form the PIC, CDK7 activity is present during transcription initiation, through phosphorylation of S5 (Hengartner *et al.*, 1998; Rodriguez *et al.*, 2000). Also, CDK7 seems

to be able to phosphorylate serine 7 of the heptapeptide repeat, both *in vivo* and *in vitro* (Akhtar et al., 2009; Glover-Cutter et al., 2009). Altogether, RNA Pol II is phosphorylated both on S5 and S7 during transcription initiation. Interestingly, while CDK8 – a Mediator subunit – was shown to be able to phosphorylate S2 and S5 *in vitro*, TFIIH capacity to enhance/stimulate transcription was shown to be repressed by this kinase, indicating that CDK8 acts as a repressor of transcription initiation. However, *in vivo* data does not support these observations, since deletion of yeast CDK8 (Srb10) does not affect phosphorylation levels on the CTD (Liao *et al.*, 1995; Hengartner *et al.*, 1998; Sun *et al.*, 1998; Rodriguez *et al.*, 2000).

pTEFb is the kinase responsible for CTD phosphorylation during elongation. pTEFb is composed of CDK9 (kinase) and cyclin T. pTEFb was initially characterized as being able to overcome pausing and stimulate transcription (Marshall and Price, 1995; Marshall *et al.*, 1996). In fact, inhibition of the kinase activity of CDK9 by 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) leads to elongation inhibition both *in vitro* and *in vivo* (Peng et al., 1998a; Peng et al., 1998b). pTEFb is also involved in the regulation of two other elongation factor, DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF). Briefly, after transcription initiation, DSIF and NELF associate with RNA Pol II and pause elongation downstream of the TSS. pTEFb then phosphorylates DSIF, NELF and the CTD of Rpb1, with processive elongation occurring upon NELF dissociation from the polymerase (Renner *et al.*, 2001; Wu *et al.*, 2003; Cheng and Price, 2007). Interestingly, in budding yeast there are two CDK9-like kinases (Bur1 and Ctk1), both being able to phosphorylated S2 (Hsin and Manley, 2012). Nevertheless, the mechanism of phosphorylation by Ctk1 and Bur1 is distinct: Bur1, recruited by the CTD phosphorylated on S5, phosphorylates S2 at promoters; subsequent Ctk1 recruitment, stimulated by Bur1, phosphorylates S2 further downstream in the coding region of the transcribed genes (Qiu *et al*., 2009; Zhang *et al*., 2012).

Notably, the pattern of phosphorylation during transcription is highly dynamic and plastic. In fact, the equilibrium between phosphorylation and dephosphorylation is of extreme importance, with previously hyperphosphorylated RNA Pol II reengaging on transcription needing to be "erased". In fact, during the transcription cycle itself, dephosphorylation is already observed. Two phosphatases – Fcp1 and Ssu72 – are primarily responsible for the dynamic dephosphorylation of CTD residues, conserved in yeast and mammals (Hsin and Manley, 2012). Fcp1 (TFIIF-associating CTD phosphatase 1) is known to interact with

TFIIB, Rpb4, and TFIIF, indicating that is responsible for dephosphorylation of S2 during the initial steps of transcription, reverting the IIO conformation for the IIA conformation (Chambers *et al.*, 1995; Cho *et al.*, 2001; Kimura *et al.*, 2002). Also, Fcp1 is known to dephosphorylate S2 at the end of the transcription cycle (Hsin and Manley, 2012). On the other hand, Ssu72, a component of the mRNA 3'-end processing complex cleavage and polyadenylation factor (CPC), has been reported to interact with basal transcription factors and to be responsible for the dephosphorylation of S5 and S7 (Ganem *et al.*, 2003; Krishnamurthy *et al.*, 2004; Zhang *et al.*, 2012). Additionally, the CTD of Rpb1 can be the subject of additional modifications, such as phosphorylation of Tyrosine1 or Tyrosine4, proline isomerization as well as glycosylation and ubiquitination (Kelly *et al.*, 1993; Bregman *et al.*, 1996; Comer and Hart, 2001; Bucheli and Sweder, 2004; Gillette *et al.*, 2004; Heidemann *et al.*, 2013). However, as revealed by recent mass spectrometry analyses, these additional modifications appear to be far less abundant than S2 and S5 phosphorylation (Schuller *et al.*, 2016; Suh *et al.*, 2016).

1.2.3. General transcription factors

1.2.3.1. TFIID

TFIID is a general transcription factor that recognizes and binds the core promoter and nucleates the PIC assembly through either a sequential step or the two-component RNA Pol II holoenzyme pathway. It is composed of TATA-binding protein (TBP) and 13 TBP-associated factors (TAFs).

As clearly stated above, one of the key components of TFIID is TBP (**Figure 12**). In fact, its importance for general transcription is highlighted by the fact that TBP is also part of the core machinery responsible for RNA Pol I and RNA Pol III transcription. TBP has a highly conserved bipartite saddle-like structure, where the concave part of TBP binds to the minor groove of the TATA-element, leading to a 90° bending of the DNA, thus resulting in an asymmetric platform that facilitates PIC assembly.

Interestingly, while affinity to DNA is relatively weak, TBP can still exhibit non-specific binding to DNA, namely in A/T rich regions, which results in the formation of non-productive PICs. Hence, TBP binding has to be tightly regulated. Indeed, there are several mechanisms ensuring negative regulation of TBP binding, specifically through the action of other proteins that either inhibit TBP binding or block the binding of other GTFs, such as

TFIIA or TFIIB. One of the mechanisms that inhibits TBP action is through the binding of TBP by yeast Mot1 (BTAF1 in mammals), initially identified as part of the B-TFIID complex, composed of Mot1 and TBP (Timmers *et al.*, 1992; Pereira *et al.*, 2001; Pereira *et al.*, 2003). Not only does Mot1 (BTAF1) bind the same TBP surface as the DNA template, but it also uses its ATPase activity to remove TBP from promoters (Auble *et al.*, 1997; Chicca *et al.*, 1998; Pereira *et al.*, 2001). Briefly, the mechanism is hypothesized to go as described by Moyle-Heyrman and colleagues: there is the initial formation of the TBP-DNA complex, containing the bent DNA. Once Mot1 binds to the TBP-DNA complex, there is unbending of the DNA, resulting in the formation of a stable ternary complex. Upon ATP binding and hydrolysis by the ATPase domain of Mot1, together with conformational changes, TBP is ultimately displaced and released from its binding site (Moyle-Heyrman *et al.*, 2012).

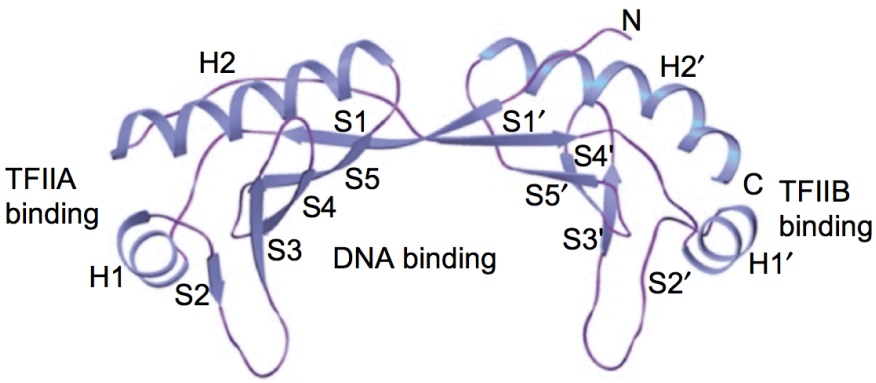


Figure 12 - Ribbon representation of the 3D structure of the TATA-binding protein (TBP) core domain determined by X-ray crystallography. The regions of TBP contacted by TFIIA, TFIIB and DNA are indicated. From (Davidson, 2003).

Another factor involved in negative regulation of TBP binding is negative cofactor 2 (NC2) complex, composed of NC2α and NC2β and conserved in mammals (Goppelt and Meisterernst, 1996). The structural analysis of the NC2-TBP-DNA ternary complex has shown that the N-terminal region of both NC2 subunits bind the DNA on the underside of the previously formed TBP-DNA complex, while the C-terminus domain of NC2β contacts TBP on its convex portion (Kamada *et al.*, 2001). This NC2 complex interaction with TBP/DNA resembles a molecular clamp, gripping both the upper and lower surface of the structure, and

inhibiting binding of TFIIA and TFIIB (Goppelt *et al.*, 1996; Kamada *et al.*, 2001). Besides these two mentioned mechanisms, others, such the formation of DNA-binding-deficient TBP-homodimers or interactions with TAF1, are known to negatively regulate TBP binding, offering compelling evidence of the tightly controlled process of transcription initiation (Thomas and Chiang, 2006). On the opposite end of the spectrum, some factors are known to have an important positive role in TBP stabilization at promoters, namely TFIIA and TFIIB:

- Association of TFIIA to TBP-DNA binary complex competes with the binding of negative regulators, such as BTAF1/Mot1 (Auble and Hahn, 1993; Moyle-Heyrman *et al.*, 2012);
- TFIIA is capable to dissociate TBP-homodimers, leading to increased availability of monomeric TBP (Coleman *et al.*, 1999);
- TFIIB, together with TFIIA, stabilizes the formation of TBP-DNA complex (Imbalzano *et al.*, 1994).

Importantly, both TBP and TAFs are capable of recognizing a broad range of elements within the core promoter: TBP subunit of human TFIID contacts the TATA elements and TAFs from human TFIID can interact with Inr and DPE, for instance. Since initially TFIID was recognized to be more required for TATA-less genes, it was acknowledged that the plethora on interactions between TAFs and DNA/chromatin was essential for PIC assembly at that class of promoters, due to the absence of a TATA-consensus sequence. Nevertheless, it was recently discovered that TATA-less promoters generally contain a deviant TATA-box with up to two mismatches, still potentially allowing recognition by TBP.

Despite being classically defined as a GTF, the fact that TAFs able to directly interact with activators, suggests that TFIID also functions as coactivator. Strikingly, it has been reported that distinct domains on one given activator are capable to interact with distinct TAFs, thus facilitating (i) activator binding to the target sequence and (ii) stabilization of TFIID at promoters (Chiang and Roeder, 1995; Burley and Roeder, 1996; Verrijzer and Tjian, 1996). This offers another mechanism of PIC assembly at TATA-less promoters, through activator-mediated TFIID recruitment. Additionally, TAFs are known to interact with other GTFs and RNA Pol II (Goodrich *et al.*, 1993; Yokomori *et al.*, 1993; Ruppert and Tjian, 1995; Wu and Chiang, 2001). The large assortment of interactions between TFIID/TAFs with other elements of the PIC might globally facilitate nucleation of the PIC,

but also act in a cell-type specific manner, since TFIID complexes containing different TAFs have been reported to exist depending on the cellular context (Muller *et al.*, 2010). Additionally, TFIID recruitment might be facilitated by histone modifications, since some TAFs are known to interact with modified residues in histone tails: the C-terminal plant homeodomain (PHD) zinc finger of TAF3 binds to H3K4me3, which is enhanced when in combination with histone acetylation (Vermeulen *et al.*, 2007); the double bromodomain of human TAF1 is able to interact with acetylated lysine on histone H4 (Jacobson *et al.*, 2000).

1.2.3.2. TFIIA

Yeast TFIIA contains 2 subunits (large and small subunit TOA1 and TOA2 in *S cerevisiae*) (Sainsbury et al., 2015). While it was initially thought that TFIIA was essential for RNA Pol II transcription (Reinberg *et al.*, 1987; Reinberg and Roeder, 1987), today we know that is not required for basal transcription, but instead stimulates both basal and activated transcription (Kang *et al.*, 1995). In fact, the importance of TFIIA in PIC formation is two-fold: on one hand TFIIA binds TBP-DNA and interacts with DNA upstream to the TATA consensus sequence and the underside of the TBP saddle, resulting in the stabilization of the protein-DNA complex (Imbalzano *et al.*, 1994; Bleichenbacher *et al.*, 2003); on the other hand, TFIIA is able to block the influence of negative regulators of transcription, such as Mot1 and NC2, both negative modulators of TBP binding and activity (Auble and Hahn, 1993; Geiger *et al.*, 1996; Tan *et al.*, 1996; Kokubo *et al.*, 1998).

1.2.3.3. TFIIB

Of all the GTFs, TFIIB is the only one that is composed by a single polypeptide only (Sainsbury *et al.*, 2015). The requirements of TFIIB for proper PIC assembly relies on its capacity to both facilitate TBP binding to DNA (through its C-terminal domain – B-core) and RNA Pol II recruitment (through its N-terminal domain – B-ribbon) (Nikolov *et al.*, 1995; Kostrewa *et al.*, 2009). Additionally, the B-reader and B-linker domains of TFIID contribute to opening of DNA at promoters and TSS selection, respectively, suggesting additional roles for TFIIB outside of RNA Pol II recruitment only (Pardee *et al.*, 1998; Cho and Buratowski, 1999; Ranish *et al.*, 1999; Kostrewa *et al.*, 2009).

Indeed, structural studies helped to clarify how TFIIB is relevant in processes of transcription initiation, namely the transition from initiation to elongation (**Figure 13**). Initially, promoter DNA is recruited to RNA Pol II by the B-ribbon domain, which binds the dock, and is positioned above the cleft by the B-core that binds the wall (closed complex formation). Afterwards, B-linker assists on the DNA melting around 20 bp downstream of the TATA consensus sequence. This allows the emerging template strand to slide in the cleft and be positioned in the template tunnel. The B-reader of TFIIB then contributes to the stabilization of the bubble located near the active center, allowing the downstream DNA duplex to be loaded into the downstream cleft, ultimately leading to the formation of the open complex. Subsequently, with the help of the B-reader domain, the DNA template strand is scanned for an Inr motif. Then the first phosphodiester bond is formed, leading to the RNA chain initiation. Finally, in the last two phases, we first observe the formation of short DNA-RNA hybrids that are unstably bound and are frequently ejected – abortive transcription – and ultimately promoter escape, with growth of the RNA beyond seven nucleotides triggering TFIIB B release and formation of the elongation complex (Kostrewa *et al.*, 2009).

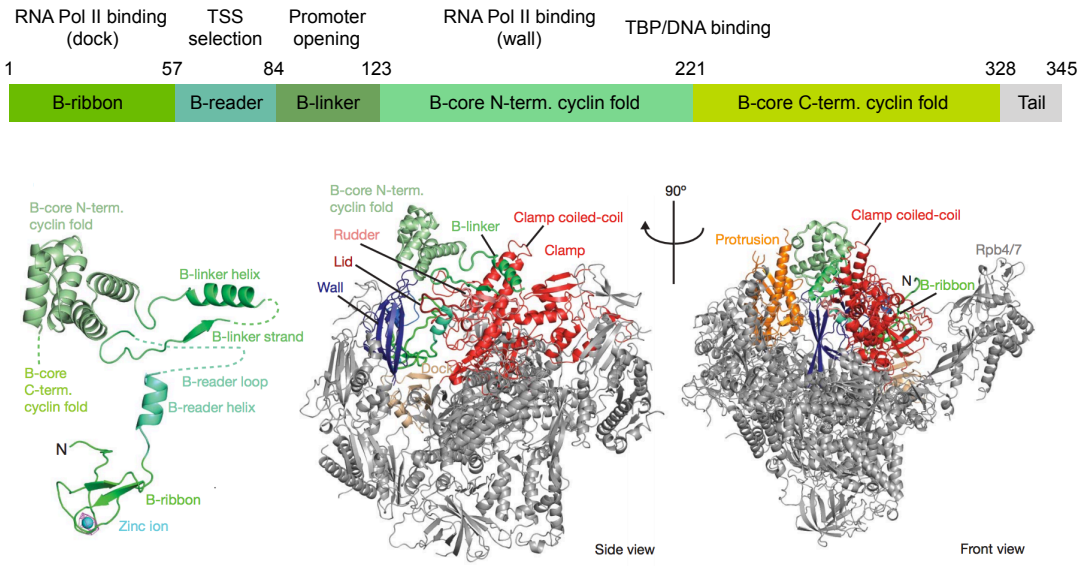


Figure 13 - Structure of RNA Pol II–TFIIB complex. ***(Top)*** Organization of the distinct functional domains of TFIIB. Known interactions and transcription initiation steps associated with each domain are indicated accordingly. ***(Bottom)*** Ribbon model of TFIIB as observed in its complex with RNA Pol II. Ribbon model with RNA Pol II in silver and TFIIB with the same colors as shown above. Side and front views are used. Pol II domains that interact with TFIIB are highlighted. Adapted from (Kostrewa *et al.*, 2009).

1.2.3.4. TFIIF

Yeast TFIIF is formed by the heterodimerization of its subunits Tfg1 and Tfg2 (TFIIα and TFIIβ in mammals, respectively) and Tfg3. However, the latter is specific to yeast and was shown to be non-essential for transcription and less tightly associated with the complex (Henry *et al.*, 1992; Henry *et al.*, 1994). Several functions in transcription initiation have been attributed to TFIIF, including: (i) prevention of non-specific interactions of RNA Pol II with DNA (Conaway *et al.*, 1991); (ii) stabilization of the PIC, namely through stabilization of TFIIB within the PIC (Cabart *et al.*, 2011; Fishburn and Hahn, 2012); (iii) TSS selection (Ghazy *et al.*, 2004); (iv) stimulation of early RNA synthesis (Yan *et al.*, 1999; Khaperskyy *et al.*, 2008); (v) positive regulation of elongation through suppression of transient RNA Pol II pausing and (vi) stabilization of the transcription bubble (Pan and Greenblatt, 1994). Interestingly, *in vitro* studies have shown that transcription can be initiated to a certain extent in the absence of TFIIE and TFIIH, but not TFIIF, indicating the importance of TFIIF in the recruitment of the remnant GTFs (Pan and Greenblatt, 1994).

1.2.3.5. TFIIE

TFIIE is composed of two subunits: yeast Tfa1 and Tfa2 or TFIIEα/β in humans. The primary function of TFIIE (together with TFIIH) is to promote the open conformation establishment, since preopening of DNA abrogates the requirement of these two subunits (Pan and Greenblatt, 1994; Fishburn and Hahn, 2012). Additionally, it has been proposed that TFIIE plays a role in the stabilization of the non-template single-stranded DNA (Okamoto *et al.*, 1998; Grunberg and Hahn, 2013). Also, TFIIH recruitment requires the prior binding of TFIIE suggesting that these two complexes directly and physically interact (Grunberg and Hahn, 2013). Interestingly, it has been shown that the flexible region of TFIIE (and TFIIF) are positioned on either side of the RNA Pol II cleft, not only interacting with each other, but also with RNA Pol II and the TBP–TFIIB–DNA sub-complex, promoting and stabilizing structural transitions during open-complex formation (Hahn and Buratowski, 2016).

1.2.3.6. TFIIH

TFIIH is a multisubunit complex comprising 10 subunits playing a role in transcription, but also involved in nucleotide excision repair for DNA damage. TFIIH accommodates three different enzymatic activities. hXPB/ySsl2 and hXPD/hRad3 are

ATPases. Ssl2 and Rad3 also possess helicase activity. Importantly, XPB is required for promoter opening *in vitro* and *in vivo*, while the helicase function of XPD is required for DNA opening during DNA repair. During PIC assembly, TFIIH is recruited by TFIIE and binds to RNA Pol II. Then, once at the site of PIC formation, XPB unwinds 11 bp of the promoter during the formation of the open complex (Fishburn *et al.*, 2015). This process and configuration are then stabilized by the cleft and jaw domains of RNA Pol II (Louder *et al.*, 2016). Interestingly, XPB subunit of TFIIH does not bind directly to the DNA that will be unwound. Instead, XPB tracks the DNA in a 5'-3' direction and translocates the downstream DNA into the cleft region of RNA Pol II, leading to unwinding of the DNA and recognition of the TSS (Fishburn *et al.*, 2015). Also, according to the PIC structure model, XPB binds downstream of the Inr and contacts both DNA strands in the promoter. Then, the function of XPB in transcription relies on its translocase activity, and this mechanism is a rate-limiting step in the initiation of transcription (Louder *et al.*, 2016; Zurita and Cruz-Becerra, 2016; Alekseev *et al.*, 2017). Interestingly, it has been recently shown that in human, loss of XPB does not impact transcription, but inhibition of its ATPase activity does. This indicates that, while cells can accommodate to its loss, they do not cope with inhibition of its activity (Alekseev *et al.*, 2017). Besides the functions mentioned above, TFIIH is also capable of acetylating S5 from the CTD of Rpb1, as discussed in a previous section.

1.3. Upstream activating sequences and activators

Transcription regulation primarily begins with the recognition of specific DNA elements by specific transcription factors (TFs). This is achieved either by transcription activators that specifically bind to upstream activating sequences (UAS) or, conversely, transcriptional repressors (binding to upstream repression sequences – URS). Today, it is widely recognized that the existence of a UAS – and their respective TFs – are broadly required for RNA Pol II transcription. This means that, by definition, the basal state of the yeast promoter is not active and transcription must be initiated by the binding of TFs (Struhl, 1999). UAS/URS function is generally orientation independent. However, its action is dependent on their location on the 5' side of the promoter. Interestingly, the same is not true in multicellular eukaryotes, where enhancers – higher eukaryotes equivalent to the UAS – can act either 3' or 5' of the promoter. In yeast, UASs can usually be found in the nucleosome-depleted region or on the exposed surface of nucleosomes. Also, different types

of UASs can co-exist, thus allowing combinatorial regulation of transcription (Hahn and Young, 2011).

TFs can be divided into three distinct groups according to the type of DNA-binding domain (DBD) they comprise: zinc (Zn^{2+}) stabilized, helix-turn-helix and leucine zipper type. Interestingly, TFs have a modular structure that comprises multiple and independent functioning domains. The effector domains of a TF may include an activation/repression domain, a nuclear localization sequence (NLS) and a regulatory domain (Hahn and Young, 2011). One of the earliest and most studied model systems for transcription activation refers to the galactose-mediated induction of gene expression by Gal4.

The GAL genes are necessary for yeast to grow when the carbon source is galactose. Among them we can count structural (*GAL1, GAL2, GAL7* and *GAL10*) and regulatory (*GAL3, GAL4* and *GAL80*) genes. The expression of these genes is required for proficient intracellular transport and metabolism of galactose. Induction of GAL genes by galactose relies on the transcriptional activator Gal4, which is known to bind to UAS existent on promoters of different *GAL* genes. In the absence of galactose, Gal4 is inactive due to the binding of Gal80 to its activation domain, masking it. When bound to Gal80, Gal4 is unable to interact with SAGA or NuA4 and incapable of allowing binding of TBP or TFIIB *in vitro* (Wu et al., 1996; Carrozza et al., 2002). In order to be active, binding of Gal4 by Gal80 has to be relieved by Gal3, through a mechanism that is not completely understood. Two distinct models propose the mechanism of action of Gal4 activation in the presence of galactose: one stipulates that Gal3 binds Gal80 without dissociating it from Gal4, though the presence of this tripartite complex at promoters has yet to be proven *in vivo* (Platt and Reece, 1998); the other claims that binding of Gal3 to Gal80 and its arrest in the cytoplasm dissociates the latter from Gal4, thus unmasking the activation domain (Jiang *et al.*, 2009). Once free, Gal4 can recruit coactivators (SAGA, NuA4, Mediator) and general transcription factors (GTFs) to the target genes.

1.4. Transcriptional coactivators and their implication on transcription

Activated transcription is probably one of the most important paths to regulate transcription initiation. As the name indicates, co-activators are the principal targets of activators. Usually, coactivators are large multisubunit complexes that enhance transcription through direct contact with GTFs or through chromatin-modifying activities. In fact, in several instances, coactivators comprise more than one enzymatic activities and can target

chromatin through remodeling or post-translational modifications of histones. Altogether, coactivators can then facilitate exposure of the core promoter and PIC formation. In very general terms, coactivators can be distinguished according to their role: adaptors (like the Mediator complex), chromatin remodelers and chromatin modifying complexes. One very important fact is that, excluding Mediator, most of the coactivators have been suggested to participate in gene expression control of specific sets of genes. This is mainly due to the fact that specific activators will recruit specific coactivators to chromatin. Nevertheless, as we can observe for the case of the SAGA complex, several other chromatin-interacting domains are present on its subunits, indicating that (i) coactivators can be recruited through more than one mechanism and (ii) the impact on transcription might be broader than anticipated.

Another very interesting fact is that coactivators have been conserved throughout eukaryotic evolution, thus indicating their importance in basic transcription mechanisms. Among the most discussed (and conserved) coactivators we can count Mediator, SAGA (discussed in the next section), NuA4, and TFIID, the latter not necessarily following the classic paradigm of a coactivator.

1.4.1. The Mediator complex

yMediator is a coactivator composed by 25 distinct subunits organized into four different modules: “head”, “middle”, “tail” and “kinase”. The “head” and the “middle” modules are essential for yeast viability, while subunits from the “kinase” and “tail” modules are not. Mediator acts as an intermediate between transcriptional regulators and GTFs, bridging transcription activation domains and RNA Pol II. Importantly, while it was suggested that Mediator and RNA Pol II directly interact, Soutourina *et al*. have elegantly shown that not only Rpb3 subunit of RNA Pol II directly interacts with the head module subunit of Mediator Med17, but that this interaction is essential for *in vivo* recruitment and activity of RNA Pol II (Soutourina *et al.*, 2011). Overall, currently, it is well established that Mediator is required for global RNA Pol II transcription and that they interact with each other through different interaction surfaces (Plaschka *et al.*, 2015). In fact, several other studies have also reported cooperative functions between Mediator and TFIID (Johnson *et al.*, 2002; Marr *et al.*, 2006; Grunberg *et al.*, 2016), TFIIA, TFIIB, TFIIE, TFIIF and TFIIH, allowing the recruitment and/or activity of those GTFs and PIC assembly (Johnson and Carey, 2003; Baek *et al.*, 2006; Esnault *et al.*, 2008; Jishage *et al.*, 2012; Eyboulet *et al.*, 2015; Eychenne *et al.*, 2016). Although previous reports have studied the genome-wide localization of

Mediator in the budding yeast genome (Andrau *et al.*, 2006; Ansari *et al.*, 2009; Ansari *et al.*, 2012; Jeronimo and Robert, 2014; Wong *et al.*, 2014; Eyboulet *et al.*, 2015; Paul *et al.*, 2015), recently it has been shown that Mediator broadly and specifically associates with UASs genome-wide, in line with its fundamental role in transcription (Grunberg *et al.*, 2016).

Once all the contacts are made between RNA Pol II, GTFs, and Mediator, leading to the formation of an open complex, interactions between RNA Pol II and GTFs/Mediator need to be broken, to promote polymerase release. Particularly, TFIIH (via Kin28) phosphorylates the CTD of Rpb1 of RNA Pol II, allowing RNA Pol II escape from the promoter. This event is dependent on Mediator, highlighting its importance for transcription initiation by RNA Pol II (Kim *et al.*, 1994; Nair *et al.*, 2005; Boeing *et al.*, 2010). Not only that, but phosphorylation of RNA Pol II by TFIIH facilitates, but is not required, the disruption of Mediator-RNA Pol II interactions, thus promoting transcription initiation (Jeronimo and Robert, 2014; Wong *et al.*, 2014).

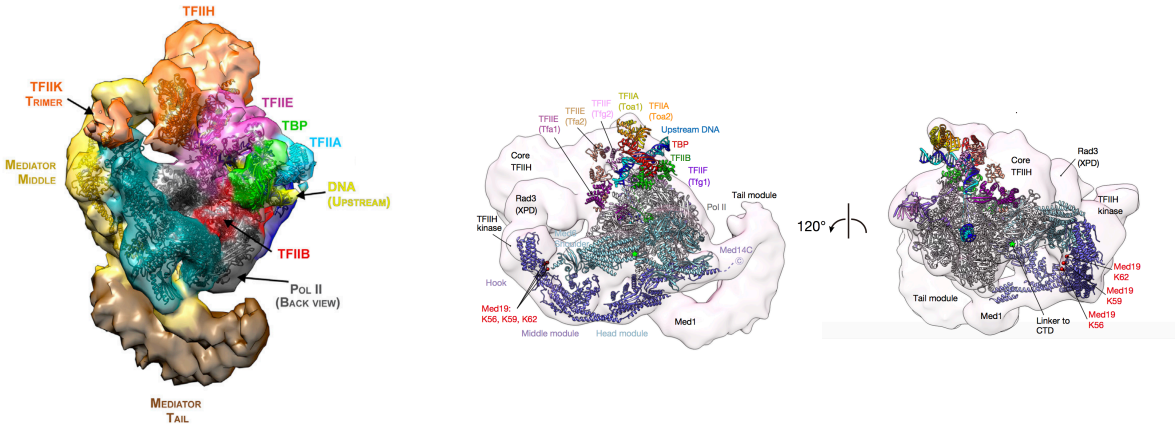


Figure 14 – Structure and model of the PIC-Mediator complex. (*Left*) A surface representation of the cryo-EM map is shown, where it is possible to identify the interaction between Mediator and components of the PIC. From (Robinson *et al.*, 2016). **(*Right*)** The atomic model of the cPIC–cMed complex obtained by the group of Patrick Cramer (ribbon representation) was placed into the low-resolution cryo-EM map of the PIC–Mediator complex represented on the left. From (Nozawa *et al.*, 2017).

Past results using electron microscopy, have disclosed that the four different modules are highly flexible, which prevented the acquisition of crystals for X-ray diffraction and high-resolution structure of the whole complex. Nevertheless, in recent years it was possible to obtain better resolution structures of the head module of fission yeast Mediator (Lariviere *et al.*, 2012). Also, using cryo-electron microscopy (cryo-EM), structures of the core mediator (composed of all the 15 essential subunits of the complex) bound to the core initiation

complex (DNA, TBP, TFIIB, TFIIF and RNA Pol II) have been obtained for *S. cerevisiae* and *S. pombe* (Plaschka *et al.*, 2015; Robinson *et al.*, 2016; Nozawa *et al.*, 2017) (**Figure 14**). Altogether, these studies allowed a better comprehension of Mediator function in PIC assembly and RNA Pol II transcription.

1.4.2. Chromatin remodeler complexes

Chromatin-remodeling factors are multi-protein complexes that use the energy of ATP hydrolysis to promote nucleosome lateral sliding or eviction. In eukaryotes, there are four distinct families of chromatin remodelers: Swi/Snf, Iswi, Chd and Ino80 (Clapier and Cairns, 2009). While all families rely on ATP hydrolysis to promote disruption of DNA-nucleosome interactions, the different classes of remodelers have different impacts on nucleosome positioning or stability. The budding yeast Swi/Snf complex was the first chromatin remodeler to be identified. Interestingly, both Swi/Snf complex and the closely related RSC complex are multisubunit complexes, with ATPase activity and with common subunits. However, RSC complex is one order of magnitude more abundant in the cells than the Swi/Snf complex. Additionally, the RSC complex is essential for viability, while the other is not (Rando and Winston, 2012). Nevertheless, both complexes work in order to promote transcription. Swi/Snf was shown to regulate the expression of approximately 5% of the yeast genome transcribed by RNA Pol II (Holstege *et al.*, 1998). Also, it has been shown that this remodeler is important for the regulation of the expression of ribosomal proteins and heat-shock genes (Shivaswamy and Iyer, 2008). Importantly, it has been shown that Swi/Snf acts in combination with Gcn5 histone acetyltransferase (HAT), subunit of the SAGA coactivators (Pollard and Peterson, 1997; Roberts and Winston, 1997). On the other hand, the RSC is involved in transcriptional regulation of RNA Pol II and RNA Pol III genes (Rando and Winston, 2012). Genome-wide studies have shown that RSC largely binds to RNA Pol II promoters through recognition of specific sequences (Badis *et al.*, 2008; Lorch *et al.*, 2014). There, this chromatin remodeler is responsible for the maintenance of nucleosome positioning. Indeed, RSC has been reported as being pivotal for the formation and maintenance of NFRs or assembly of fragile nucleosomes (Kubik *et al.*, 2015).

1.4.3. Histone modifications and transcription

As previously mentioned, not only chromatin remodeling, but also post-translational modification of histones play a major role in transcription initiation. Among the most studied modifications we can count acetylation, methylation, phosphorylation, and ubiquitination (**Figure 15**). Importantly, these modifications mainly occur in the histone N-terminal tails that protrude from the nucleosome. However, several modifications have been reported in the globular domain of histone. For simplification, here we are going to describe modifications and chromatin modifiers in the context of yeast. Additionally, modifications occur in canonical histones, but also in histone variants. Importantly, histone variants are non-canonical histones expressed throughout the cell cycle and possess specific roles in DNA-related processes (Kouzarides, 2007).

Histone acetylation is one of the best-characterized cases of PTMs and it is generally associated with transcription activation. A direct role for histone acetylation in promoting transcription was proposed due to the neutralization of the positive charge of lysine residues thereby loosening histone-DNA interactions. Not only that, but histone acetylation can also serve as a platform for recruitment of other protein complexes that are capable of interacting with histone modifications (Lee and Workman, 2007). These proteins are generically called "readers" and are known to facilitate transcription. In fact, the recognition of acetylation and its association with transcription is known for a long time. In *S. cerevisiae*, the first HAT to be identified was Hat1, though it did not have any observable role in transcription (Travis *et al.*, 1984; Rando and Winston, 2012). Nevertheless, it was later shown that Hat1 was part of a complex with Hta2, being able to acetylate nascent histone H4 in the cytoplasm (Parthun *et al.*, 1996; Parthun, 2007; McCullough and Marmorstein, 2016). In the meantime, several other HATs were identified, together with histone deacetylases (HDACs), responsible for the erasure of the mark. Remarkably, over the years, specific events have allowed the popularization of the study of HATs and HDACs. Namely the identification of the evolutionary conserved Gcn5, known to be associated with transcription activation, as a HAT, later shown to be part of two distinct complexes: SAGA and ADA complexes (discussed in the next section of the *Introduction*) (Berger *et al.*, 1990; Brownell *et al.*, 1996; Grant *et al.*, 1997). In light of current knowledge, HAT complexes can be grouped according to their catalytic domain: Gcn5 N-acetyltransferases (GNATs) that include Gcn5, PCAF, Elp3, Hat1 Hpa2 and Nut1 (further discussed in the next section); the MYST family of HATs, that include Morf, Ybf2 (Sas3), Sas2 and Tip60/Esa1; the "orphan" class of HATs,

that lack a true consensus HAT domain, such as p300/CBP and Taf1, for example (Lee and Workman, 2007). Altogether, these complexes, responsible for the deposition of the histone marks, are collectively known as "writers". Hence, "writer" complexes deposit the modification, while "readers" recognize the modifications and trigger a response. In such a dynamic processes, "erasers" are also involved and these complexes are capable of abolishing the mark, such as HDAC complexes.

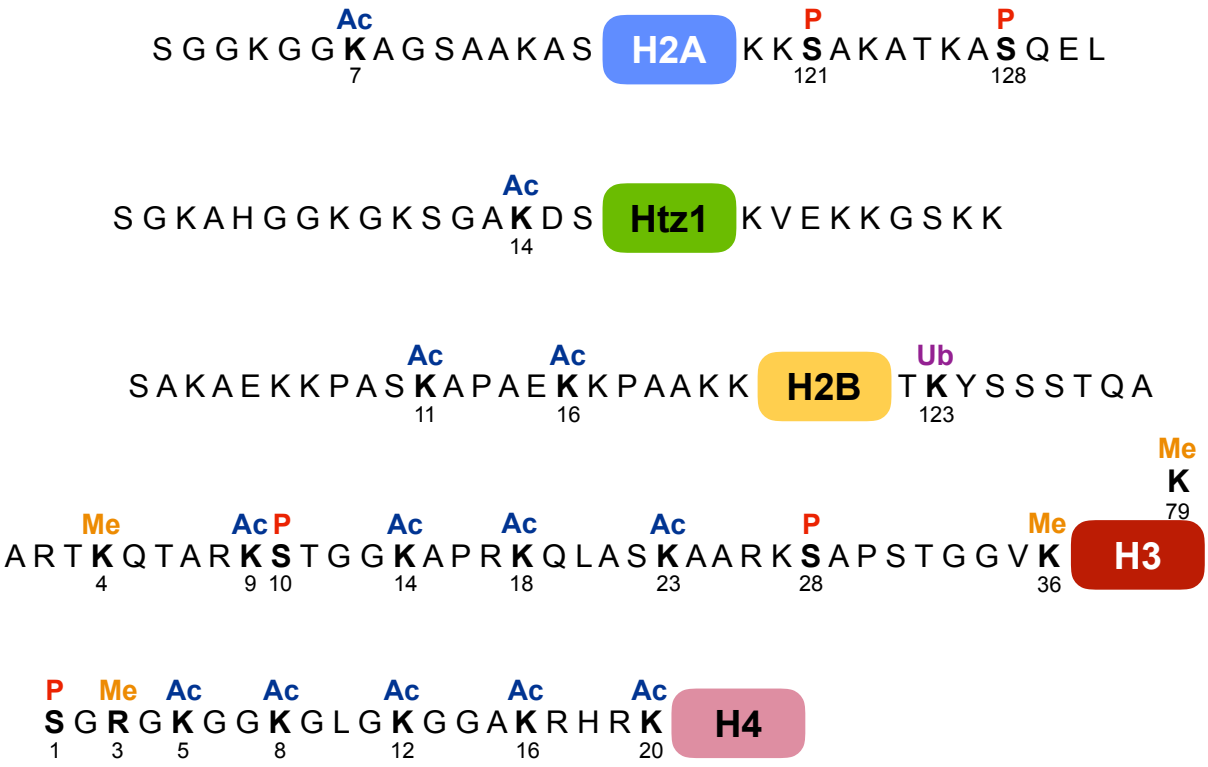


Figure 15 – Summary of histone modifications in *S. cerevisiae*. The core histone fold domains are indicated schematically as hexagons, and the N- and C-terminal tail sequences are shown. The numbers shown under modified residues indicate amino acid positions. Known sites of modification are indicated and color-coded as follows: acetylation (Ac – dark blue); methylation (Me – Orange); phosphorylation (P – red); ubiquitination (Ub – purple).

In the process of identifying HAT complexes, the three MYST-class HATs Sas2 (SAS complex), Sas3 (NuA3 complex) and Esa1 (NuA4 complex) have been identified and characterized in yeast. Specifically concerning NuA4/Esa1, capable of acetylating histones H4, H2A, and Htz1/H2A.Z, it shares some of its subunits with other transcriptional complexes (for example Arp4, that is present in NuA4, Swi/Snf, RSC and SWR1) (Rando and Winston, 2012). Interestingly, SAGA and NuA4 also share a subunit: yTra1/hTRAPP. As it will be possible to understand later in this manuscript, Tra1 is the subunit of

SAGA/NuA4 responsible for the interaction with acidic activators, thus being involved in the recruitment of the complexes to chromatin. The role of NuA4 in transcription initiation has been studied both at specific genes, but also in a genome-wide manner. Overall, it has been shown that NuA4/Esa1 is important for the expression of ribosomal-proteins encoding genes and heat-shock/starvation genes (Reid *et al.*, 2000; Nourani *et al.*, 2004; Robert *et al.*, 2004). Also, it was found to be recruited to all active promoters, correlating to gene expression levels, thus indicating a broader role for this coactivator complex in RNA Pol II transcription (Robert *et al.*, 2004). Also, acetylation by NuA4 has been implicated in transcriptional elongation because (i) it associates to the gene-bodies/coding regions (Ginsburg *et al.*, 2009), (ii) it stimulates H4 acetylation (lysine 5, 12, 16 and 20) and is necessary for normal RNA Pol II elongation (Ginsburg *et al.*, 2009) and (iii) it is involved in recruitment of the chromatin remodeler RSC, which drives nucleosome eviction (Kasten *et al.*, 2004; Carey *et al.*, 2006).

While acetylation is associated with derepression/activation of transcription, the role of histone methylation is not so obvious. One of the most studied examples of histone methylation in yeast is H3K4me3, which is found in the 5' nucleosome of actively transcribed genes. In contrast, mono- and di-methylation of H3K4 extend to the coding region of genes. Set1, member of the COMPASS complex, catalyzes H3K4me and its deposition is incredibly and tightly regulated. First, it requires histone H2B ubiquitination on lysine 123 (H2Bub) for COMPASS to be recruited (Dover *et al.*, 2002; Vitaliano-Prunier *et al.*, 2008). Second, in order for ubiquitination to be catalyzed by Rad6/Bre1 the PAF complex is necessary, thus indicating that H3K4me3 depends both on PAF/Paf1 and H2Bub (Krogan *et al.*, 2003). Third, the kinase Bur1, subunit of pTEFb needed for phosphorylation of DSIF and of the RNA Pol II CTD on S2, is also required for H3K4me3 to take place (Laribee *et al.*, 2005). Considering the degree of regulation of H3K4me and its genome-wide presence at actively transcribed genes, one would expect an important role on transcription *per se*. Nevertheless, in a *set1*Δ or *spp1*Δ, where H3K4me is abrogated, virtually no effects on steady-state or nascent transcription were observed, with only a handful of genes being either up- or down-regulated (Venkatasubrahmanyam *et al.*, 2007; Lenstra *et al.*, 2011; Margaritis *et al.*, 2012; Howe *et al.*, 2017). In fact, H3K4me3 is the classical paradigm of a PTM that triggers the question whether this mark is a cause or an effect of transcription, due to its seemingly absent role on transcription. Among the arguments sustaining the idea that H3K4me3 might be a consequence of transcription we can account: (i) H3K4me3 deposition

is dependent on and downstream of Paf1 elongation factor and H2Bub, both associated with elongation and enriched in gene bodies (Dover *et al.*, 2002; Krogan *et al.*, 2003; Vitaliano-Prunier *et al.*, 2008) and (ii) time-course analyses disclosed that H3K4me3 peaks after the peak of transcription, observed during metabolic cycle and meiosis in yeast (Ng *et al.*, 2003; Kuang *et al.*, 2014). In fact, several reports revealed that H3K4me might have a role that is not necessarily associated with transcription activation, but linked to splicing, transcription memory/responsiveness or transcriptional consistency (Ng *et al.*, 2003; Borde *et al.*, 2009; Howe *et al.*, 2017).

Nevertheless, other histone marks have been associated with transcription. H3K36me3, catalyzed by the methyltransferase Set2, is associated with transcription activation and occurs in the transcribed region of expressed genes, via recruitment by elongating RNA Pol II. Interestingly, this mark is not necessarily required for elongation, but instead is necessary for the activation of the HDAC Rpd3S along the actively transcribe DNA. In fact, while Set2 activity is not necessary for viability, deletion of Set2 promotes increased acetylation on the gene bodies, leading to the appearance of cryptic promoters along the transcribed region (Carrozza *et al.*, 2005; Joshi and Struhl, 2005). On the other hand, H3K79me3, catalyzed by Dot1, is also associated with the coding region of genes, but not correlated with transcription activation. Instead, H3K79me3 is involved in telomeric silencing by Sir proteins: H3K79me3 seems to be a barrier for Sir proteins binding, thus deletion of Dot1 and decreased H3K79me3 leads to unspecific binding of Sir proteins genome-wide, titrating the Sir complex away for normally silent regions (van Leeuwen *et al.*, 2002). Moreover, some histone modifications are related to transcriptional silencing, though these marks are absent from *S. cerevisiae*, but present in *S. pombe* and multicellular eukaryotes. For instance, H3K9 methylation, catalyzed by SETDB1 and the related enzymes SUV39H1 and SUV39H2 (di- and trimethylation) and GLP and G9a (H3K9me1 and H3K9me2), is found in constitutive heterochromatin, specifically H3K9me2 and H3K9me3. Subsequently, H3K9 di- and trimethylation are bound by the chromodomain of Heterochromatin Protein 1 (HP1), which can self-oligomerize and recruit repressive histone modifiers, contributing to heterochromatin compaction and spreading (Becker *et al.*, 2016). H3K27me3 is another mark associated with silencing and is catalyzed by EZH2 (Kuzmichev *et al.*, 2002). EZH2 is part of the Polycomb Repressive Complex 2 (PRC2), which is responsible for the repression of many genes involved in development and cell differentiation, being associated with repression of developmental genes (Boyer *et al.*, 2006;

Bracken *et al.*, 2006). H3K27me3 is also an important mark of the inactive X chromosome (Rougeulle *et al.*, 2004).

2. The SAGA coactivator

The plastic nature of the chromatin landscape is a key regulator of transcription activation or repression. While debatable, histone modifications have been recognized to play a key role in several DNA-dependent biological processes, namely transcription. Major actors of such modifications are coactivator or corepressor complexes, usually found in the form of multi-subunit complexes capable of exerting one or more activities. SAGA is a paradigm of a coactivator where different modules with distinct activities are joined together in one multi-subunit complex. Initially identified in *S. cerevisiae* (Grant *et al.*, 1997), the SAGA coactivator is a highly conserved macromolecular complex of around 1.8 MDa composed of 19 subunits in this organism (**Table 1**).

Table 1 – Composition of the SAGA complex in *S. cerevisiae*.

Module	Subunit
DUB Module (mDUB)	Ubp8 Sgf73 Sgf11 **Sus1**
HAT Module (mHAT)	**Gcn5** **Ada2** **Ada3** **Sgf29**
TBP-interacting subunits	Spt3 Spt8
Interaction with activators	**Tra1**
Structural core	Ada1 Spt7 Spt20 **Taf5** **Taf6** **Taf9** **Taf10** **Taf12**

Representation of the different subunits is organized according to its functional module. In red we have represented the catalytic subunits of each enzymatic module. In **bold** are annotated subunits that are shared between SAGA and other complexes.

2.1. Putting the pieces together: from subunits identification to the birth of the SAGA complex

Initially, when the SAGA complex was still yet to be described, two distinct families of genes/proteins involved in transcription were identified by two independent genetic screens and described: SPT genes and ADA genes. Later on, members of these two groups were found to be part of one multisubunit complex and some subunits were independently identified in both screens.

2.1.1. The SPT family of proteins

Genetic screens in budding yeast are often powerful tools to identify groups of genes that lead to a given phenotype when mutated. One example is provided by the study of transcription-defective mutants caused by the insertion of the transposon Ty or its long-terminal-repeat δ. Insertion of Ty or δ sequence in the 5' region of a gene often abolishes or reduces transcription of the adjacent gene.

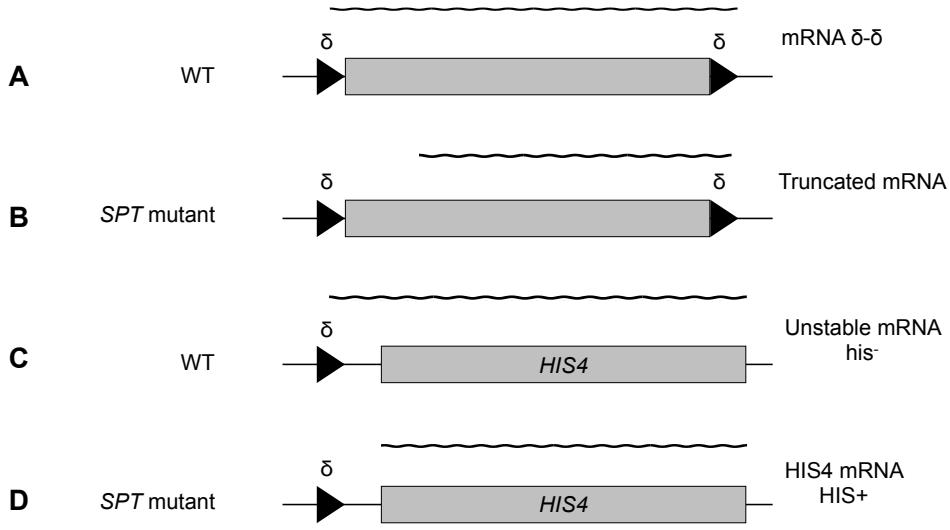


Figure 16 – Representation of the genetic screen for the identification of SPT proteins. (A) To mobilize the Ty retrotransposon, a wild-type strain is capable of transcribing the Ty element immediately after its first δ sequence. **(B)** A *SPT3* mutant is unable to initiate transcription after the δ element and loses its capacity to mobilize the Ty retrotransposon. **(C)** In a wild-type strain, the insertion of a δ sequence of the Ty retrotransposon leads to aberrant transcription. Transcription starts from the δ element upstream of the promoter of *HIS4* and not from the TSS of *HIS4*, leading to auxotrophy for histidine. **(D)** In a situation where *SPT3* is mutated, transcription from the δ element is lost and *HIS4* is expressed normally. Adapted from (Winston *et al.*, 1984b).

The SPT (Suppressor of Ty) family was initially identified by Winston and colleagues, where certain insertional mutations, located in the 5' or upstream region of HIS4 or LYS2, abolished expression of these genes, resulting in a His^- or Lys^- auxotrophy. Then, selection for His^+ or Lys^+ revertants identifies *SPT* mutants (Spt^-) in these strains. Altogether, in the *SPT* mutants analyzed, transcription initiating from Ty or δ was reduced, and conversely, transcription of the adjacent gene was restored to normal levels (Winston *et al.*, 1984a; Winston *et al.*, 1984b; Fassler and Winston, 1988) (**Figure 16**). Hence, a common role for the *SPT* family in promoter function was deduced. Initially studied in the *SPT3*, similar results were also observed for other *SPT* members: *SPT7*, *SPT8* and *SPT15* (with TBP identified as its product) (Winston *et al.*, 1987; Eisenmann *et al.*, 1989; Hahn *et al.*, 1989). Importantly, not only these mutants shared common Ty phenotypes, but they could also be distinguished from other *SPT* members due to additional phenotypes, such as in mating and sporulation. At this point, the first mention to a putative complex integrating *SPT3*, *SPT7*, and *SPT8* products was made: "the three gene products may interact to form a complex that is directly required for activation of transcription from the δ sequence" (Winston *et al.*, 1987).

Interestingly, three mutations of the gene *SPT15* were also identified on the SPT genetic screen. These strains presented a severe phenotype, indicating the pivotal role the product of this gene has in cellular biology. Importantly, this gene was then identified as encoding *TBP*, which is essential for transcription (Eisenmann *et al.*, 1989; Hahn *et al.*, 1989). Finally, a final SPT member, *SPT20*, was also found to have virtually the same transcriptional phenotype as the previously mentioned *SPT*s (Roberts and Winston, 1996). Altogether, it was suggested that Spt20, along with Spt3, Spt7 and Spt8, are required for normal TBP function at certain promoters. It is important to mention that several SPT genes were identified in the process: *SPT3*, *SPT8*, *SPT7* and *SPT20*, belonging to the "TBP sub-group"; *SPT4*, *SPT5*, *SPT6*, *SPT11* and *SPT12*, the "histone group", coding for histones (*SPT11* – H2A and *SPT12* – H2B), complexes involved in transcription elongation (Spt4/Spt5, the yeast counterpart of the human DSIF complex), recombination and chromosome segregation (*SPT4* and *SPT6*) (Yamaguchi *et al.*, 2001).

Importantly, SPT members of the SAGA complex are known to be important for the interaction of this complex with TBP. Specifically, Spt3 and Spt8 are capable to bind TBP in a SAGA-dependent context, both *in vitro* and *in vivo*. Also, Spt8, but not Spt3, is capable to bind TBP in a SAGA-independent manner and competes with TFIIA for TBP binding (Eisenmann et al., 1992; Warfield et al., 2004; Sermwittayawong and Tan, 2006; Laprade et al., 2007; Mohibullah and Hahn, 2008). Nevertheless, Spt3 interaction with TBP is

considered to be the most genetically relevant interaction for gene activation (Eisenmann *et al.*, 1992; Warfield *et al.*, 2004; Sermwittayawong and Tan, 2006; Laprade *et al.*, 2007). In fact, crosslinking experiments demonstrated that Spt3 requires Spt20 and Spt7 to assist on TBP interaction (Han *et al.*, 2014). These results disclose a complex interaction network between the Spt proteins of SAGA and TBP. Additionally, both Spt7 and Spt20 were described to be essential for the whole complex stability and activity, thus being commonly denominated (together with Ada1, see bellow) structural subunits of the SAGA coactivator (Grant *et al.*, 1997; Sterner *et al.*, 1999; Wu and Winston, 2002; Nagy *et al.*, 2009).

2.1.2. The ADA family of proteins

Interestingly, also the ADA (standing for Adaptor) proteins from the SAGA complex were discovered via genetic screenings using the chimeric Gal4-VP16 activator. Indeed, using this system it was shown for the first time the presence of adaptor proteins, proteins that, while not being strictly necessary for basal transcription, are required for enhanced/activated transcription.

Gal4-VP16 is a chimeric protein containing the DNA binding domain of Gal4 fused with the activation domain of the herpes simplex virus protein VP16. Using this chimeric protein, the authors reported that, *in vitro*, transcription from promoters bearing a Gal4-binding sequence was extraordinarily activated. In contrast, this chimeric activator inhibited transcription from promoters lacking a Gal4-binding sequence (containing an UAS dA:dT). With this, the authors suggested that transcription inhibition by Gal4-VP16 binding to unspecific sequences was due to the sequestering of elements of the transcription machinery (Berger *et al.*, 1990). Additionally, activated, but not basal, transcription was blocked when this chimeric activator binds to oligonucleotides containing Gal4-bind sites. Having that, it is possible to conclude that, when the chimeric activator does not bind to the DNA template, but instead it binds to oligonucleotides containing Gal4 UAS, it sequesters a factor that bridges the activator with the transcriptional machinery, demonstrating the existence of adaptor proteins (Berger *et al.*, 1990) (**Figure 17**).

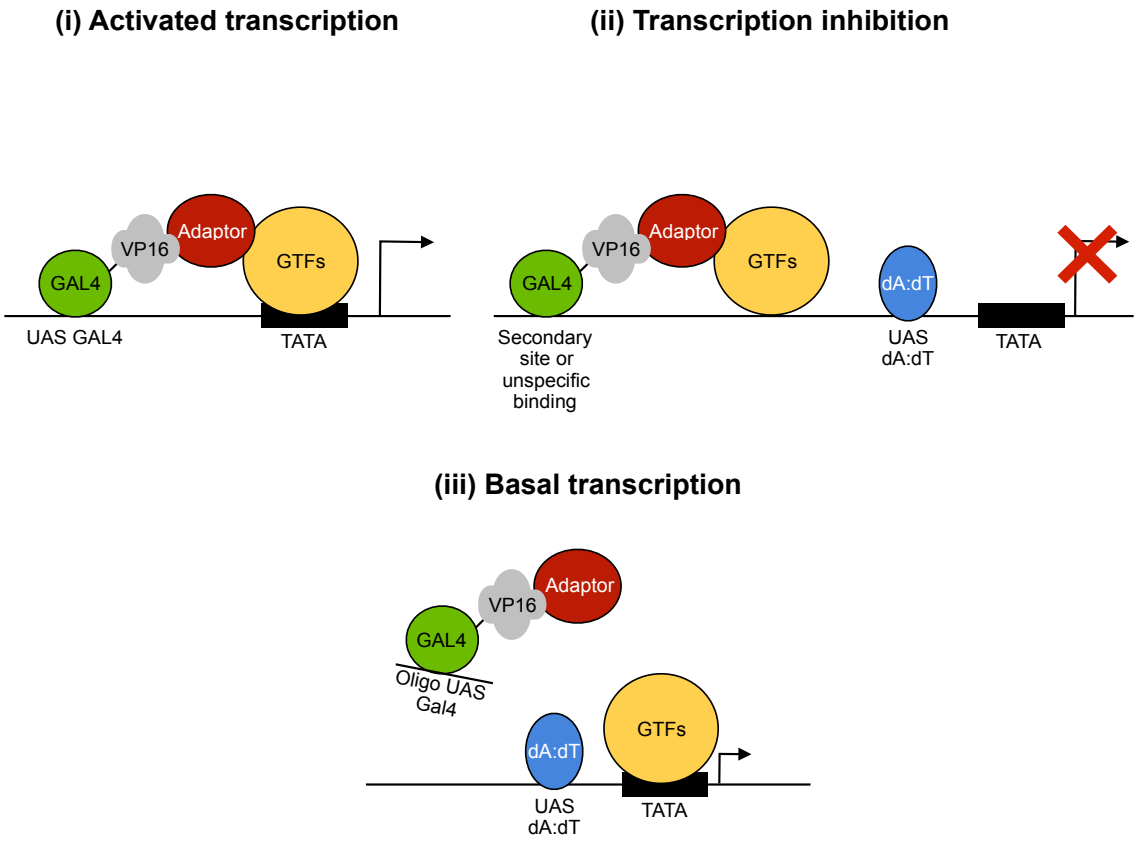


Figure 17 – *In vitro* evidence of adaptor proteins. (*Top left*) When in the presence of a yeast extract, the chimeric activator Gal4-VP16 is able to activate transcription from a DNA template containing a UASs recognized by Gal4. **(*Top right*)** In the absence of an UAS recognized by Gal4, the chimeric activator binds unspecifically to the DNA template and sequesters the basal transcription machinery, thus inhibiting transcription. **(*Bottom*)** A strong interaction between the acidic domain of VP16 and the adaptor titrates the adaptor away from the promoter and prevents activation by the dA:dT activator at the promoter. GAL4-VP16 binds to the adaptor while unbound to template DNA because the GAL4 DNA binding domain has been occupied by an oligonucleotide encoding the GAL UAS. Adapted from (Berger *et al.*, 1990).

The expression of the Gal4-VP16 activator is toxic, due to sequestration of the transcriptional machinery, thus preventing proper transcription. Hence, through genetic screening, one could mutate certain proteins, responsible for the bridging of the chimeric activator and transcriptional machinery, and observe the reversion of the toxic phenotype. The first three ADA members to be identified were Ada1, Ada2, and Ada3 (Berger *et al.*, 1992; Pina *et al.*, 1993; Horiuchi *et al.*, 1997). While Ada1 presented special features, Ada2 and Ada3 shared a common phenotype. First, transactivation by Gal4-VP16 is lost in Ada2 and Ada3 mutants. Second, both mutant strains share a slow-growth phenotype and heat-sensitivity. Third, a strain containing mutation for both Ada2 and Ada3 did not present an

incremental phenotype, suggesting that both proteins were involved in the same pathway (Berger *et al.*, 1992; Pina *et al.*, 1993). Additionally, some results suggested that there was a functional connection between Ada3 and TBP (Pina *et al.*, 1993). Moreover, another gene, Ada4, was identified using the same genetic screening, later being identified as being Gcn5, the catalytic subunit of the mHAT (Marcus *et al.*, 1994). Just like observed for Ada1, Ada5, also identified in the screen, when mutated promoted a much more severe phenotype, such as slower growth and decreased activation by Gal4-VP16 (Marcus *et al.*, 1994). Also, transcription in deletion mutants was affected to an extended number of UASs than those observed for Ada2 and Ada3 deletion strains. Finally, Ada5 was identified as being Spt20 (Marcus *et al.*, 1996; Roberts and Winston, 1996). As for Ada1, the phenotypes resembled more those of some SPT family members, such as Spt7 and Spt20 (Berger *et al.*, 1992; Marcus *et al.*, 1996; Roberts and Winston, 1996; Horiuchi *et al.*, 1997). This means that the two different families of genes, SPT and ADA, initially thought to be independent of each other, presented overlapping phenotypes. In fact, Ada1, Spt7, and Spt20 were hypothesized to exert a function that bridges SPT and ADA activities. In fact, those proteins are recognized as structural subunits of the SAGA complex, essential for its stability as a whole (Marcus *et al.*, 1996; Roberts and Winston, 1996; Horiuchi *et al.*, 1997).

2.1.3. Identification of the SAGA complex in yeast

Workman and colleagues have first shown that recombinant Gcn5, by itself, was not capable of acetylating histone H3 when incorporated within a nucleosome, even though it was able to do so on free histones. This seminal work also showed that Gcn5 acts as the catalytic subunit in two high molecular-weight HAT complexes: one with approximately 1.8 MDa and the other with around 0.8 MDa. Importantly, they demonstrated that both of these complexes contain the adaptor proteins Ada2 and Ada3, validating these complexes as native adaptor complexes. Also, the larger of the two complexes also contains Spt3, Spt7, and Spt20. From there on, the complex is known as Spt-Ada-Gcn5 acetyltransferase, or simply SAGA (Grant *et al.*, 1997).

Later on, mass spectrometry and immunoblotting analyses disclosed that there was a subset of TAFs – namely TAF5, TAF6, TAF9, TAF10 and TAF12 – that were shared between SAGA and TFIID. Importantly, integration of certain TAFs within SAGA highlighted their importance in transcriptional regulation through a mechanism seemingly different of that of TFIID (Grant *et al.*, 1998a). Importantly, as it is observed for TFIID,

TAFs play a major role for SAGA structure and stability and are known to be required for the complex to exert its activities.

Additionally, a high molecular weight protein (approximately 400 kDa) was identified as a SAGA subunit. While it was initially called p400, it was later identified as Tra1 (TRAAP in human), the subunit of SAGA know to mediate the interaction of the complex with activators (Grant *et al.*, 1998b; Saleh *et al.*, 1998; Brown *et al.*, 2001). Tra1 is not exclusive to the SAGA complex, as it is equally found in NuA4 coactivator complex. Work from the laboratory of Steven Hahn has disclosed regions of Tra1 essential for viability, integration of the protein within the complexes and activities of the respective coactivators (Knutson and Hahn, 2011).

The last subunits of SAGA to be recognized were those forming part of the deubiquitinating module (mDUB). First Ubp8 was shown to be stably associated with the SAGA complex and capable of removing ubiquitin from histone H2B when integrated within the coactivator (Gavin *et al.*, 2002; Sanders *et al.*, 2002; Henry *et al.*, 2003). Other components of the mDUB where then identified – Sgf11 and Sus1 – and shown to be necessary for SAGA-dependent H2B deubiquitination (Ingvarsdottir *et al.*, 2005; Lee *et al.*, 2005; Kohler *et al.*, 2006). Finally, the fourth and last member of the mDUB was identified – Sgf73 – and disclosed to be necessary to anchor this module within SAGA, through its N-terminal domain (Kohler *et al.*, 2008; Lee *et al.*, 2009).

2.2. The ADA complex in S. cerevisiae

In *S. cerevisiae* two Gcn5-containing HAT complexes can be found: a high molecular weight one – the extensively discussed SAGA complex – and a lower molecular weight one – the ADA complex. In both cases, Gcn5 can be found in association with Ada2 and Ada3. At that time, being characterized to a lesser extent than SAGA, it was suggested that other subunits should make-up the complex, since it was identified as having an approximate molecular weight of 0.8 MDa (Grant *et al.*, 1997). Also, it could be easily hypothesized that the identified ADA complex was just a stable intermediate complex for SAGA formation. However, a specific subunit of the ADA complex was then identified, allowing to discard the latter hypothesis. That subunit, Ahc1 (for ADA HAT complex component 1), is a specific ADA subunit, thus not found in SAGA, and is essential for the stability of the complex (Eberharter *et al.*, 1999). More recently, a study by the laboratory of Michael P. Washburn has fully characterized the ADA complex subunits and architectural organization. Ultimately,

ADA complex was found to be composed of Gcn5, Ada2, Ada3 and Sgf29 – the same mHAT found in SAGA -, plus Ahc1 and Ahc2 (Lee *et al.*, 2011).

In fact, the knowledge regarding the ADA complex is extremely limited. However, it is known that both purified SAGA and ADA complexes are capable of acetylating *in vitro* H3K14 and H3K18, but only SAGA acetylates H3K9 (Grant *et al.*, 1999). Also, the deletion of the bromodomain of Gcn5 alters the acetyltransferase activity of SAGA, but not that of ADA, indicating that in the latter functions in a specific and distinct manner than SAGA (Eberharter *et al.*, 1999).

2.3. The SAGA-like (SLIK)/ SALSA complex in budding yeast

As the name indicates, the SAGA-like complex (SLIK/SALSA) is a complex that, in most part, is identical to SAGA. It was initially found in an attempt to purify HAT complex in yeast, where two complexes that highly resembled each other were purified in different fractions. One was identified as SAGA and the other SLIK. Whilst these two complexes would share most of their subunits, and are capable of acetylating the same histone residues and to regulate transcription, some differences between them could be determinant for their specific roles (Pray-Grant *et al.*, 2002).

The first and major difference between SAGA and SLIK is the lack of Spt8 in the latter. Since complexes purified from strains lacking Spt8 are unable to interact with TBP *in vitro*, this might indicate that SLIK has much less affinity for TBP (Sterner *et al.*, 1999; Sterner *et al.*, 2002). A second relevant distinction is that a truncated version of Spt7, lacking its C-terminal domain, is found within SLIK. This truncated form of Spt7, resulting from the proteolytic cleavage of full-size Spt7 by Pep4, lacks the domain that allows the interaction between Spt7-Spt8, thus justifying the non-incorporation of Spt8 in SLIK (Sterner *et al.*, 1999; Wu and Winston, 2002; Spedale *et al.*, 2010). Interestingly, this truncation and absence of Spt8 makes SLIK more similar to the metazoan SAGA than to the yeast SAGA itself. Nevertheless, the fact that Pep4 is a cytoplasmic protease might suggest that SLIK could be an artifact of SAGA purification, thereby the *in vivo* existence of SLIK is questionable. Also, a specific subunit is found in SLIK but not SAGA, Rtg2, which is important for the stability of the complex. Rtg2 has been reported as having an important role in the regulation of genes whose expression is altered in cells with mitochondrial dysfunction (Pray-Grant *et al.*, 2002). Importantly, deletion of Spt7, shared between SAGA and SLIK, has a more severe phenotype than deletion of either *Spt8* or *Rtg2*. While *spt8*Δ and *rtg2*Δ strains share a milder and

somewhat similar phenotype, specific phenotypes related to either deletion of *Spt8* or *Rtg2* indicated that, although SAGA and SLIK might have partially overlapping activities, these two complexes are not necessarily functionally redundant (Pray-Grant *et al.*, 2002).

2.4. SAGA becomes more complex(es) in the metazoan context

Due to the increasing functional complexity during evolution, diversification of protein complexes is a well-known and established process, namely through duplication and/or mutation of ancestor genes. There are several examples of transcription-related complexes whose complexity has increased throughout organisms' evolution: the single Set1/COMPASS complex in yeast has diverged into 6 distinct Set1/MLL complexes in mammals (Miller *et al.*, 2001; Roguev *et al.*, 2001; Smith *et al.*, 2011); Sin3, which exists in the form of one only complex in yeast, has three distinct but related complexes (Sin3A and Sin3B) in metazoans (Hayakawa and Nakayama, 2011). Indeed, SAGA is another paradigm of a complex that has diverged through evolution in several distinct protein complexes in metazoans. Among them we can count the SAGA variants and the more distantly related ATAC, collectively known as SAGA-like complexes.

2.4.1. Evolution of the SAGA complex

While SAGA composition is remarkably conserved throughout the eukaryotic kingdom, some exceptions in terms of subunits content and domains within SAGA subunits are visible (summarized in **Figure 18**).

Regarding differences in subunits content, two very evident changes in SAGA in metazoans is the absence of an ortholog of Spt8 and a truncated homolog of yeast Spt7, lacking the C-terminus. In fact, as previously mentioned, in yeast there is a variant of SAGA – SLIK/SALSA – where a C-terminally truncated form of Spt7 is found. This truncated form of Spt7 lacks its Spt8-interacting domain, leading to the exclusion of Spt8 from this complex (Belotserkovskaya *et al.*, 2000; Pray-Grant *et al.*, 2002; Sterner *et al.*, 2002; Wu and Winston, 2002). Having that, the metazoan SAGA more strictly resembles the SLIK/SALSA complex from yeast. One the implications of the lack of Spt8 might be the interaction between the metazoan SAGA and TBP, which might be weakened in metazoan in comparison to yeast.

Concerning changes in protein domains in SAGA subunits, several points can be made (**Figure 18**). These changes, despite being in a relatively small number, might have an

impact on the structure and/or activities of the complex. Since the metazoan Spt7 lacks the C-terminal domain existent in yeast, this ultimately led to the loss of the histone fold domain and bromodomain in the *D. melanogaster* ortholog. Since the histone-fold domain is known to mediate the interaction between Spt7 with Taf10, one can expect that this loss might lead to structural changes of the complex. Additionally, the loss of the Spt7 bromodomain in flies might reduce the interaction of SAGA with acetylated histones (Hassan *et al.*, 2002). In humans, the situation is slightly different, since a bromo-associated domain, resembling a histone-fold, is found, instead of the classical histone-fold domain or bromodomain.

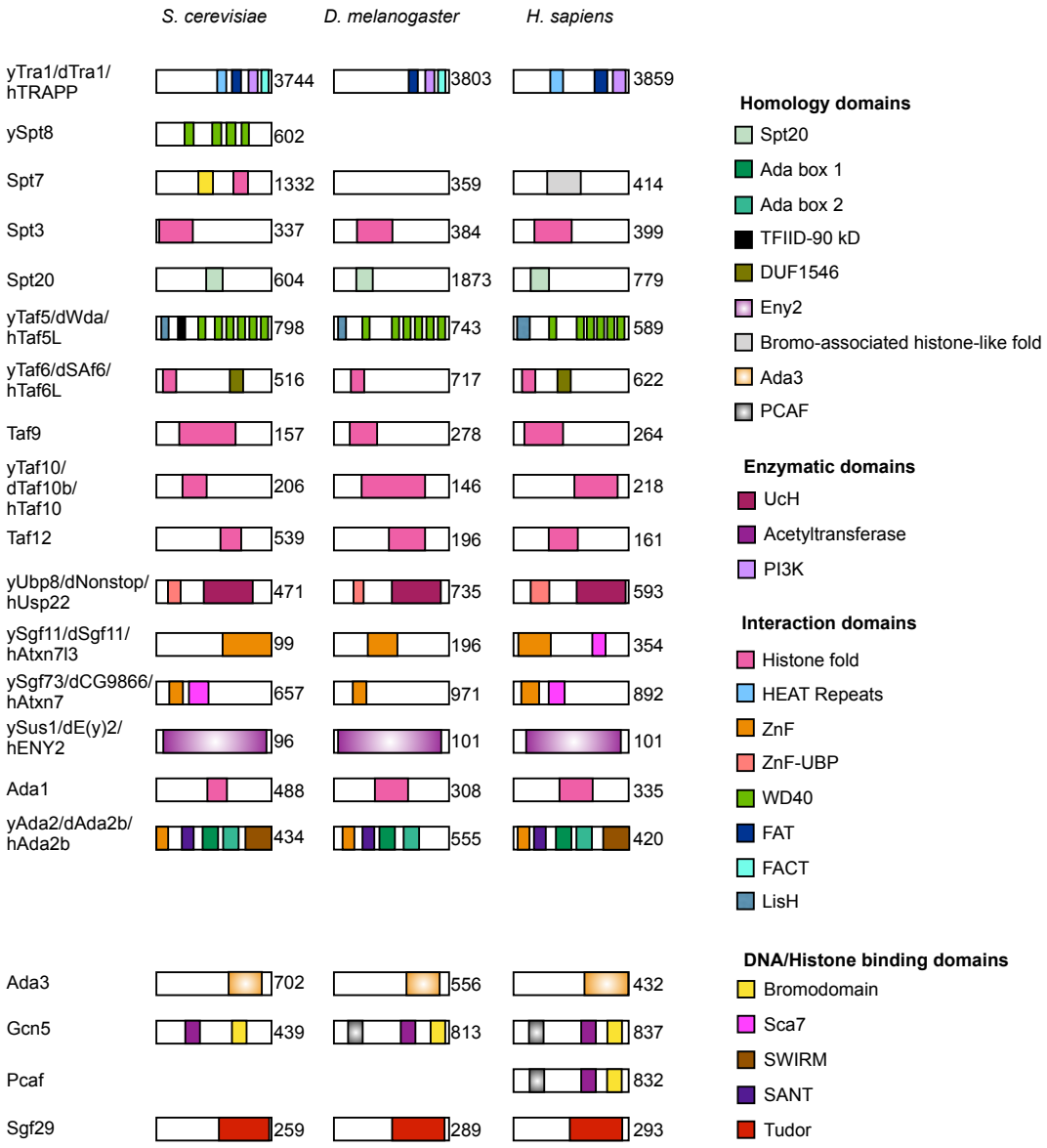


Figure 18 – Composition and domain organization of SAGA in *S. cerevisiae*, *D. melanogaster* and *H. sapiens*. Adapted from (Spedale *et al.*, 2012b).

A LisH domain – implicated in dimerization – found in yeast Taf5 is lost in the different orthologs, WD5 (*D. melanogaster*) and TAF5L (*H. sapiens*) (Ogryzko *et al.*, 1998; Guelman *et al.*, 2006). Similarly, the DUF1546 domain found in yeast Taf6 and human TAF6L is not found in its counterpart in flies (Weake *et al.*, 2009). Also, the human ATXN7L3 contains an atypical zinc-finger (SCA7) domain, which is not present in yeast (ySgf11) or fly (dSGF11) (Kohler *et al.*, 2008; Bonnet *et al.*, 2010). Moreover, yeast Sgf73 and the human ortholog ATXN7 both contain a SCA7 domain that, once again, is not found in flies (Kohler *et al.*, 2008; Bonnet *et al.*, 2010). Finally, Ada2b in flies lacks a SWIRM domain that, this time around is found in both yeast and human (Muratoglu *et al.*, 2003).

As mentioned above, another mechanism that allows divergence of complexes throughout evolution is gene duplication. Specifically, gene duplication has occurred for *TAF5*, *TAF6*, *TAF9*, *TAF10*, *GCN5*, and *ADA2*. In yeast, there is a single copy of each of the Taf genes, leading to the incorporation of each and every Taf in both SAGA and TFIID. In mammals, gene duplication occurred and, while *TAF5* and *TAF6* products are integrated in TFIID, TAF5L and TAF6L are specific to SAGA (Ogryzko *et al.*, 1998). In flies, *TAF5L* and *TAF6L* genes exist, but their products are not components of the SAGA complex. Instead, Wda and Saf6 are part of the SAGA coactivator, rather than Taf5/Taf5L or Taf6/Taf6L, respectively (Guelman *et al.*, 2006; Weake *et al.*, 2009). The *TAF9* gene has duplicated in humans – *TAF9* and *TAF9B* – and both are integral part of TFIID and SAGA (Frontini *et al.*, 2005). Regarding the TAF10 gene, it has only duplicated in *D. melanogaster*: *TAF10* (TFIID) and *TAF10B* (SAGA). In yeast, only one *GCN5* and one *ADA2* gene exist. However, in vertebrates *GCN5* has duplicated, leading to the existence of *PCAF* gene. Both GCN5 and PCAF can be assembled within SAGA in a mutually exclusive fashion (Krebs *et al.*, 2010). The *ADA2* gene has also duplicated through eukaryotic evolution, resulting in two distinct genes: *ADA2A* and *ADA2B*. While ADA2b is part of the mHAT of the SAGA complex, ADA2a is specific for ATAC, a complex that is specific to multicellular eukaryotes (Barlev *et al.*, 2003; Wang *et al.*, 2008; Orpinell *et al.*, 2010; Suganuma *et al.*, 2010).

2.4.2. ATAC, a metazoan GCN5-containing HAT complex

ATAC (Ada-Two-A-Containing) has emerged early in metazoan evolution and, as previously mentioned, is exclusive to multicellular eukaryotes. While one single variant of ATAC is found in flies, two are present in vertebrate through duplication of GCN5 (GCN5- and PCAF-containing ATAC).

The detailed analysis of the SAGA complex in fly and human showed that ADA2b is part of this complex, while ADA2a would never be incorporated in it. Nevertheless, it was recognized that ADA2a interacted with both GCN5 and PCAF (Muratoglu *et al.*, 2003). In fact, the identification of proteins associated with Ada2a in *D. melanogaster* allowed disclosing the existence of a new HAT complex distinct from SAGA, containing Ada2a, Gcn5 and Ada3. Using mass spectrometry tools in purified ATAC complexes, namely MudPIT (multidimensional protein identification technology), unveiled new subunits of this complex: HCF and Atac1 (Guelman *et al.*, 2006). In order to fully characterize the complex, the laboratory of Jerry Workman has used a more systematic MudPIT analysis, and, among other subunits, it was possible to identify Atac2, an ATAC subunit with an acetyltransferase domain. This has showed that ATAC possesses two distinct HAT proteins: Gcn5 and Atac2 (Suganuma *et al.*, 2008). Moreover, it was later shown that ATAC was not exclusive to flies, but it also existed in mouse and human (Wang *et al.*, 2008; Guelman *et al.*, 2009; Nagy *et al.*, 2010).

In vitro analyses have shown that Atac2 is capable of acetylating H3 and H4 when present as free histones, while Gcn5 only acetylates H3. Also, the enzymatic activity of the former is 40-fold reduced in comparison to the latter (Guelman *et al*., 2009). Likewise, purified ATAC was reported to acetylate free H3 and free H4, but also H3/H4 within nucleosomes *in vitro* (Suganuma *et al.*, 2008; Guelman *et al.*, 2009; Nagy *et al.*, 2010). In fact, in contrast to SAGA, ATAC in flies shows a strong specificity for H4 on nucleosomal preparations (Suganuma *et al.*, 2008). Regarding its enzymatic activities towards histone acetylation *in vivo*, while there is a scarcity of studies, one could say that there is an overall lack of clarity concerning acetylation by ATAC. Through knockdown of Atac2, ATAC complexes have been reported to be necessary for overall levels of H3K9ac, H4K5ac, H4K12ac, and H4K16ac in mammals (Guelman *et al*., 2009). Using a knockdown of ATAC-specific subunit Ada2a, a distinct study reported that, instead, H3K9ac and H3K14ac were reduced in mammalian cells (Nagy *et al*., 2010). The only common ground between these two reports targeting different subunits of ATAC is represented by H3K9ac. On the other hand, in flies, mutation of Atac2 results in specific decreased of H4K16ac levels (Suganuma *et al*., 2008), while a separate study has shown that only levels of H4K5ac and H4K12ac were reduced in Ada2a mutant (Ciurciu *et al*., 2008). Altogether, these results indicate partially similar roles for Atac2 in the regulation of global acetylation levels in flies and mammals.

2.5. Modularity and structure of the SAGA coactivator

The first effort to comprehend the structure of SAGA came from Patrick Schultz laboratory (Brand *et al.*, 1999a; Wu *et al.*, 2004). Specifically, Schultz and colleagues resorted to electron microscopy (EM) to obtain a low-resolution 3D model of the *S. cerevisiae* coactivator complex (**Figure 19**). At first analysis, the SAGA complex appeared as an elongated particle that was absent in an *ada1*Δ mutant, confirming the specificity for the complex. Precisely, the complex seemed to be formed by four different domains (I to IV) rearranged linearly, with a highly flexible fifth domain (V) at the tip of the complex. Two large clefts were observed, one formed between domains I and II + III and another one formed between domains II+III and IV. In the same work, the authors also performed immunolabeling of 9 SAGA subunits, in order to specifically localize those subunits within the structure obtained (see **Figure 19**, ***left panel***). Taf5, Taf6, and Taf10 revealed a spatial distribution similar to that of TFIID, where they would form a scaffold for SAGA assembly.

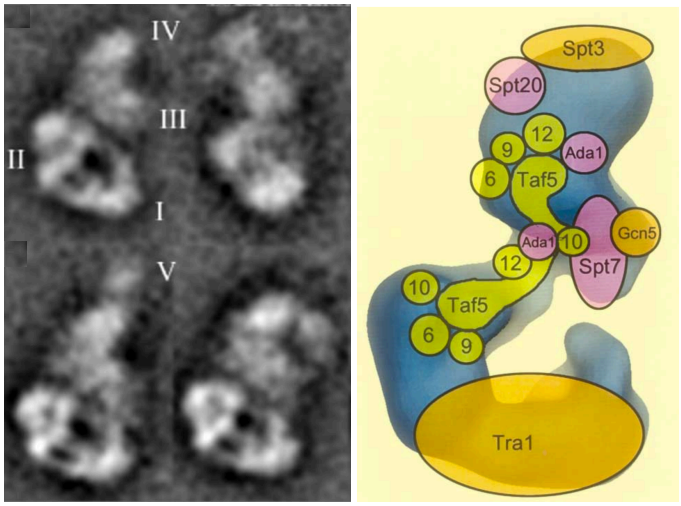


Figure 19 – Molecular Architecture of the S. cerevisiae SAGA Complex. (*Left*) Representative SAGA views obtained by electron microscopy. In the view shown in the top left panel four domains are revealed and identified as I–IV. On the bottom left a mobile fifth domain (V) is visible. **(*Right*)** Schematic representation of the localization of the SAGA subunits on the 3D model of the complex. From (Wu *et al.*, 2004).

Together with Spt7, Spt20, and Ada1, those subunits were mapped to the central part of the complex (domains II, III and IV), overall supporting their role in the architectural stability of the complex. Also, the localization of both Spt7 and Ada1 together with their partners Taf10 and Taf12, respectively, supported their interaction through their histone fold domain (HFD). Moreover, SAGA subunits associated with transcription regulation, namely Spt3, Gcn5, and Tra1, were mapped within completely distinct domains, indicating the existence of a modular organization that reflects their distinct regulatory roles (**Figure 19, *right panel***). It is also important to mention that, at that time, the obtained structure of SAGA highly resembled the mammalian TFTC, later on identified as the human version of the SAGA complex (Martinez *et al.*, 1998; Brand *et al.*, 1999a; Brand *et al.*, 1999b; Wu *et al.*, 2004).

A more recent study coming from Michael P. Washburn's laboratory used different biochemical approaches, quantitative proteomics, and computational analyses to better appreciate the architectural organization of Gcn5-containing complexes (SAGA and ADA) in *S. cerevisiae*. Specifically, they used several TAP-tagged baits and deletions of SAGA subunits to clearly define the existing modules and their interconnectivities (Lee *et al.*, 2011) (**Figure 20**). These results strongly supported the modularity of the SAGA complex and, among other things, it was shown that:

(i) Ada2 is localized within the center of the mHAT;

(ii) SPT module is in close proximity of the mDUB, predicting that Sgf73 is the subunit of the mDUB anchoring it to the complex, in line with previous studies (Kohler *et al.*, 2010);

(iii) Sgf29 is a core subunit of the mHAT of SAGA;

(iv) The SPT module is located centrally within the complex and is necessary for multiple (if not all) activities of SAGA, as previously indicated (Grant *et al.*, 1997; Horiuchi *et al.*, 1997; Roberts and Winston, 1997; Sterner *et al.*, 1999);

(v) Although deletion of structural subunits like Spt20 or Ada1 might lead to the disruption of SAGA as a whole, assembly of "free" mDUB or mHAT might prevail, leading to residual deubiquitination or acetylation activities.

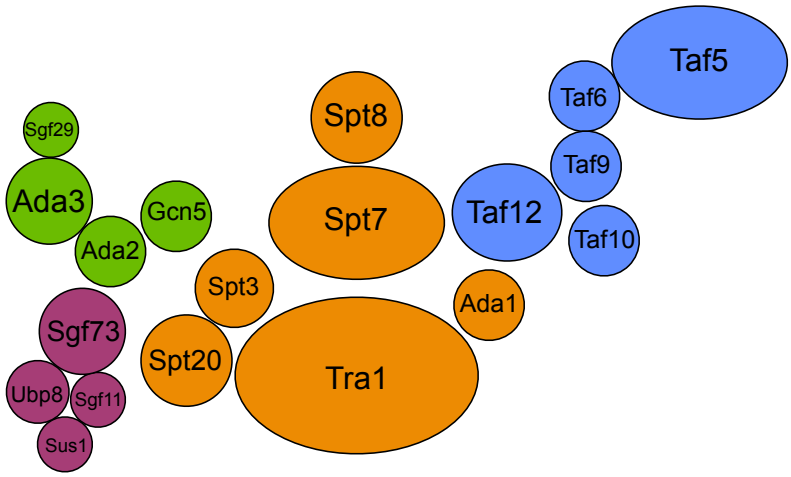


Figure 20 – Deletion interaction network and the macromolecular assembly of the SAGA complex. Based upon all deletion purifications, all proteins of the SAGA complex were organized into modules and consequently a macromolecular model was assembled. Adapted from (Lee *et al*., 2011).

A different approach was carried out in Steven Hahn's group. Here, Han *et al.* decided to use a combination of crosslinking followed by mass spectrometry, together with classical genetic and biochemical analyses. Altogether, the obtained results revealing physical proximity between crosslinked aminoacids allowed the disclosure of additional information regarding the molecular interaction between SAGA subunits and its modules (Han *et al.*, 2014) (**Figure 21**). According to their results and model, and in light of previous reports (Bieniossek *et al.*, 2013), the SAGA core contains two pairs of each histone fold containing pairs Taf6-Taf9 and Ada1-Taf12, together with two copies of Taf5. In fact, these results regarding the core of the SAGA complex highly resemble that of TFIID structure, where two copies of Taf6-Taf4 and Taf4-Taf12 are found together with two copies of Taf5 (Bieniossek *et al.*, 2013). To this complex, the addition of one copy of Spt7-Taf10, just like observed for Taf8-Taf10 in TFIID, forms a core structure that serves as a platform for the other modules or peripheral subunits of the complex (Han *et al.*, 2014). They could observe that, besides crosslinking with the N-terminal domain of Taf5, the C-terminal domain of Spt7 also crosslinks with Spt8. Additionally, Sgf73 extensively crosslinks with Ada1 and Taf12 (through their histone fold domains) and N-terminus of Taf5, anchoring the mDUB to the SAGA core, corroborating previous reports (Kohler *et al.*, 2008; Lee *et al.*, 2009). Also, Sgf73 crosslinked to the mHAT through Ada2 and Ada3. Although this could suggest that both modules could be functionally associated, analyses of DUB mutants disclosed little to

no impact of HAT stabilization within SAGA, supporting previous evidence that these modules were functionally independent. On the other hand, Ada3, subunit that anchors the mHAT to the SAGA complex, was found to crosslink Ada1 and Taf12 (through their histone-fold domains), Taf5 (through its WD40 repeats) and Taf6 (near the HEAT repeat domain). Moreover, Spt20, Spt3, and Tra1 highly crosslinked with the histone-fold domains of Ada1, Taf9, Taf10, and Taf12, as well as the N-terminal portion of Taf5.

Additionally, the abovementioned work also offered some insights regarding TBP-SAGA interactions. Spt8 is known to genetically and physically interact with TBP and competes with TFIIA for binding to TBP (Warfield *et al.*, 2004). In fact, in this study they have shown that Spt8 interacts with the positively charged groove of TBP, just like TFIIA. Although Spt3-TBP binding is genetically the most relevant for gene activation, Spt3 by itself is unable to bind TBP, unlike Spt8, that is capable to bind TBP in a SAGA-independent fashion (Eisenmann *et al.*, 1992; Warfield *et al.*, 2004; Sermwittayawong and Tan, 2006; Laprade *et al.*, 2007). Since Spt20 crosslinked with either side of the Spt3 histone-fold domain involved in TBP binding, it was reasoned that Spt20 might promote Spt3-TBP binding through conformational changes of Spt3 (Han *et al.*, 2014).

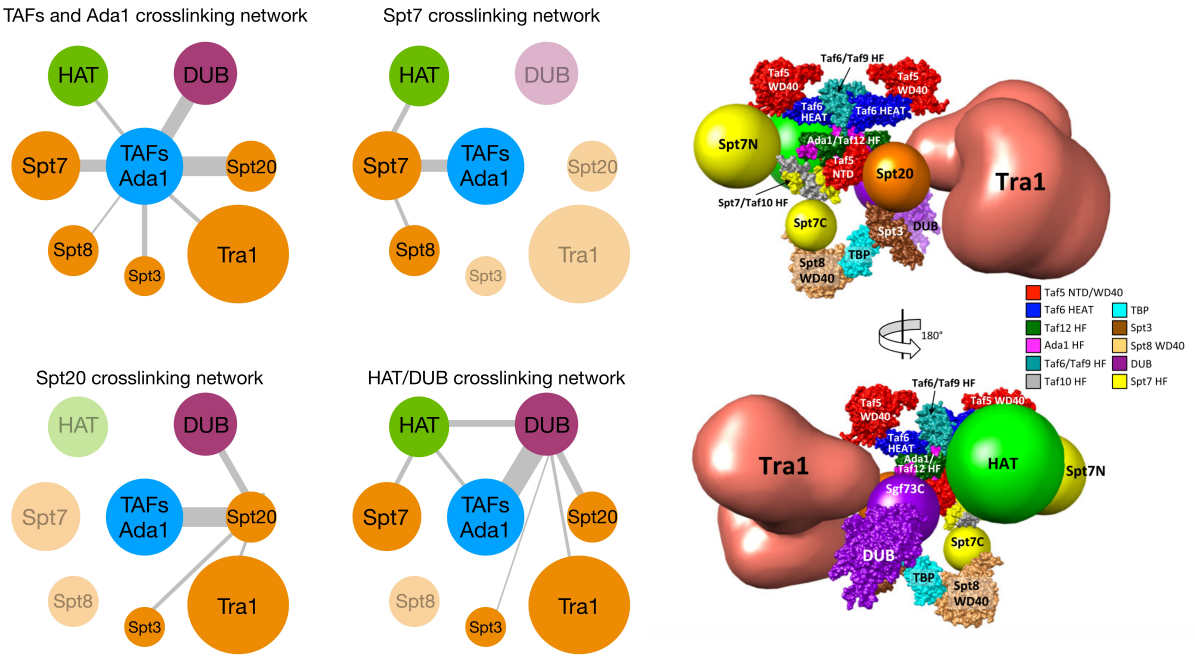


Figure 21 – SAGA complex architecture obtained through crosslinking of the complex followed by mass-spectrometry. (*Left*) Crosslinking network for TAFs/Ada1, Spt7, Spt20, HAT, and DUB modules. Grey lines connect crosslinked SAGA components, with the width of the lines proportional to the number of intermolecular crosslinks. **(*Right*)** Model for the molecular architecture of the SAGA complex according to the crosslink results. Adapted from (Han *et al.*, 2014).

More recently, a study by Setiaputra *et al.* developed a comprehensive localization strategy to understand the arrangement of subunits within SAGA, combining their new electron-microscopy labeling with the results obtained by Hahn's group (Setiaputra *et al.*, 2015). In this work it was possible to delineate three different structural conformations for SAGA – arched, curved and donut. Importantly, they observed that disruption of different modules would differently impact the conformation of the complex: disruption of the mDUB prevents the formation of donut conformation, whereas loss of the mHAT caused the shoulder region to be more mobile (**Figure 22, *left panel***). Interestingly, their findings that mDUB absence affects the flexibility and the presence of the tail region, where the mHAT is found, might indicate an indirect synergy between the two distinct catalytic activities, as proposed by others for both yeast and human (Henry *et al.*, 2003; McMahon *et al.*, 2005; Palhan *et al.*, 2005; Atanassov *et al.*, 2009; Burke *et al.*, 2013; Han *et al.*, 2014; Setiaputra *et al.*, 2015).

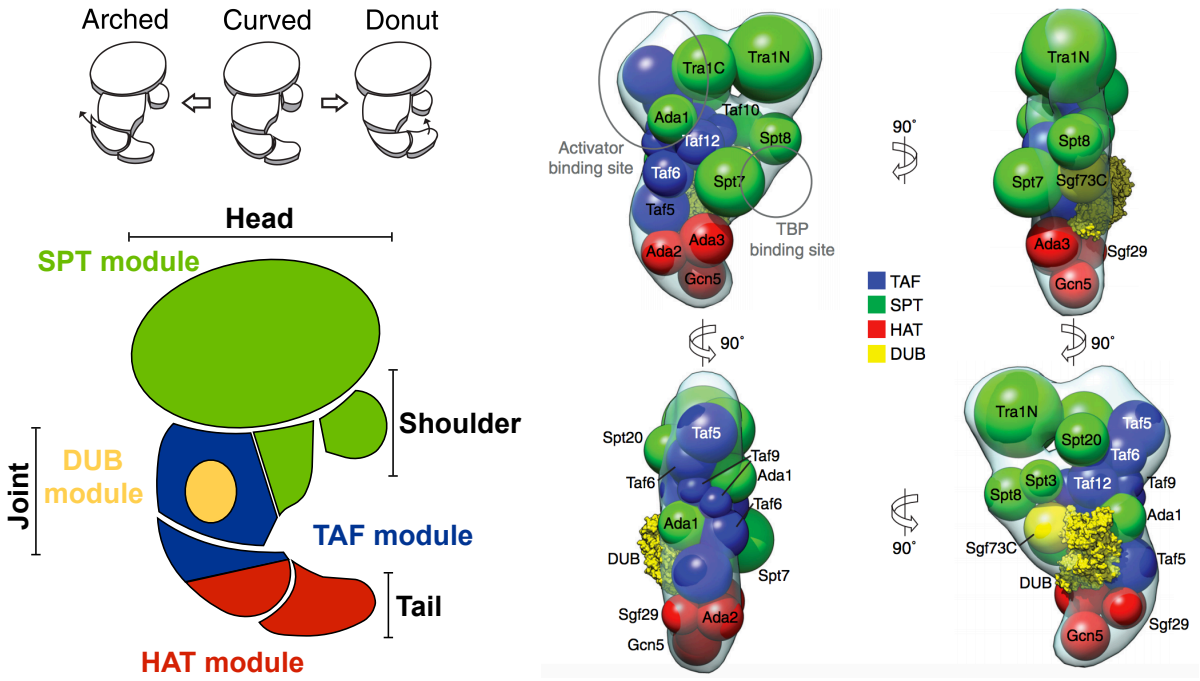


Figure 22 - Subunit arrangement of the modular yeast SAGA complex. (*Left*) Schematic representation of the SAGA complex highlighting the different conformations of the complex (arched, curved and donut conformations) and localization of the different domains and modules. **(*Right*)** Spatial arrangement of SAGA subunits. Results from EM-based subunit localization experiments and crosslinking mass spectrometry were combined with the curved SAGA three-dimensional reconstruction to propose an overall arrangement of all 19 complex subunits. Adapted from (Setiaputra *et al.*, 2015).

More importantly, in this study the authors were able to generate a detailed model reflecting the structural organization of the whole SAGA. Their results highly resembled those of Han *et al*., with the central module being composed of Spt-Taf proteins together with the mDUB, flanked at opposite ends by the Tra1 and mHAT layers (Setiaputra *et al.*, 2015) (**Figure 22, *right panel***). Another study by Patrick Schultz mapped other subunits of the SAGA complex, and while some results were overlapping with Setiaputra *et al*., others seemed to be in disagreement, namely the location of Spt20, Spt3 and the mDUB (Durand *et al.*, 2014).

Considering the multiple approaches used to allow a better understanding of the structure of the SAGA complex, it is important to somehow integrate all the information collected over the time. Importantly, these distinct studies have some common conclusions, such as the fact that SAGA is organized within modules related to specific activities. Namely, the subunits belonging to the mDUB (Ubp8, Sgf11, Sgf73, and Sus1) and the mHAT (Gcn5, Sgf29, Ada2, and Ada3) are always found together, forming two discreet modules within the complex. Also, the structural studies agreed on the fact the TAFs, Spt7, Spt20 and Ada1 have an essential role for complex integrity, as previously shown through purely biochemical and genetics approaches (Horiuchi *et al.*, 1997; Grant *et al.*, 1998a; Sterner *et al.*, 1999; Kirschner *et al.*, 2002; Wu and Winston, 2002). Nevertheless, there are some distinct conclusions reached by the different studies. On one hand, some studies situate TAFs in a central position, as a platform for the remnant subunits to be assembled, just as described in the initial study by electron microscopy (Wu *et al.*, 2004; Han *et al.*, 2014; Setiaputra *et al.*, 2015). Not only that, but also some studies agreed on the fact that Spt3 was found to be in close proximity of Spt20, while Spt7 was located close to the mHAT/Gcn5 (Han *et al.*, 2014; Setiaputra *et al.*, 2015). However, the same was not concluded by the work of Lee and colleagues, where (i) Spt proteins and TAFs clustered in two different groups and (ii) TAFs were positioned in a more peripheral part of the complex (Lee *et al.*, 2011). In spite of the fact that most of the mapping obtained by crosslink/mass spectrometry is in good agreement with the mapping obtained by electron microscopy, a big discrepancy between these studies is the location of Tra1: while Han et al., and later Setiaputra *et al*., showed that Tra1 is in close proximity of Spt20 (Han *et al.*, 2014; Setiaputra *et al.*, 2015), Wu and colleagues positioned Tra1 and Spt20 at opposite ends of SAGA (Wu *et al.*, 2004) (see **Figure 19**, **Figure 21** and **Figure 22**). Also, the model proposed by Steven Hahn's publication distinguishes itself from that of Michael P. Washburn, since the latter proposed that Tra1, together with Spt subunits and Ada1, were present in a single group, central to the complex (Lee *et al.*, 2011; Han *et al.*, 2014) (see **Figure 20** and **Figure 21**). Notably, while all the

information gathered over the years is precious and fundamental for the understanding of both structure and function of SAGA, there is still a long way to go in order to have a complete and high-resolution characterization of the SAGA complex. However, the rise of increasingly more refined methods in the field of structural biology opens an exciting door to resolve some of the unanswered questions posed in recent years.

2.6. Regulation of SAGA recruitment to chromatin

One of the first steps of gene transcription regulation by SAGA is its recruitment to chromatin. For instance, in *S. cerevisiae*, SAGA is recruited to the *GAL1* promoter prior to recruitment of GTFs and RNA Pol II and PIC assembly (Bhaumik and Green, 2001; Larschan and Winston, 2001; Bryant and Ptashne, 2003). Indeed this observation is found for several genes, both in yeast and metazoan (Nagy *et al.*, 2009; Sellam *et al.*, 2009; Helmlinger *et al.*, 2011; Lang *et al.*, 2011; Venters *et al.*, 2011; Weake *et al.*, 2011). Since SAGA is a multi-subunit complex with several subunits or domains that were reported to interact with chromatin, several mechanisms were proposed to regulate the limiting step of SAGA recruitment.

The first and most obvious model for SAGA recruitment to its regulatory regions is through its Tra1 subunit (**Figure 23A**). In yeast it has been extensively shown that Tra1 directly interacts with activators, recruiting SAGA to promoters in a gene-specific manner (Brown *et al.*, 2001; Helmlinger *et al.*, 2011). Not only that, but in multicellular eukaryotes the mammalian ortholog of Tra1, TRAPP, has been reported to directly interact with transcription factors such as c-Myc and E2F (McMahon *et al.*, 1998). Whether Tra1/TRAPP subunit of the SAGA complex is enough to target the complex to genes is highly debatable. For instance, in *S. pombe* this has been addressed more carefully. It is important to mention that there is a great advantage of studying this specific question in *S. pombe*: while in *S. cerevisiae* Tra1 is component of both SAGA and NuA4 and is essential for survival, in fission yeast gene duplication led to the existence of two distinct genes coding for either Tra1 (non-essential) and Tra2 (essential), the first integrated specifically within SAGA and the latter specifically into NuA4. What Helmlinger and colleagues have shown was that, although SAGA is recruited to some genes through Tra1, recruitment of SAGA to other genes was independent of its activator-interacting subunit (Helmlinger *et al.*, 2011). Having that, by itself, Tra1 does not account for all events of SAGA recruitment to chromatin. Interestingly, crosslinking studies performed in yeast have reported that Taf12, subunit of

both SAGA and TFIID, interacts with acidic activators, thus potentially being involved in the recruitment of both complexes (Fishburn *et al.*, 2005; Reeves and Hahn, 2005). Additionally, interaction between SAGA and components of the basal transcription machinery might also collaborate in this process. Specifically, the interaction between TBP and Spt3/Spt8 SAGA subunits might mediate the recruitment of the coactivator to the promoters of its target genes (**Figure 23B**) (Larschan and Winston, 2001; Laprade *et al.*, 2007; Mohibullah and Hahn, 2008).

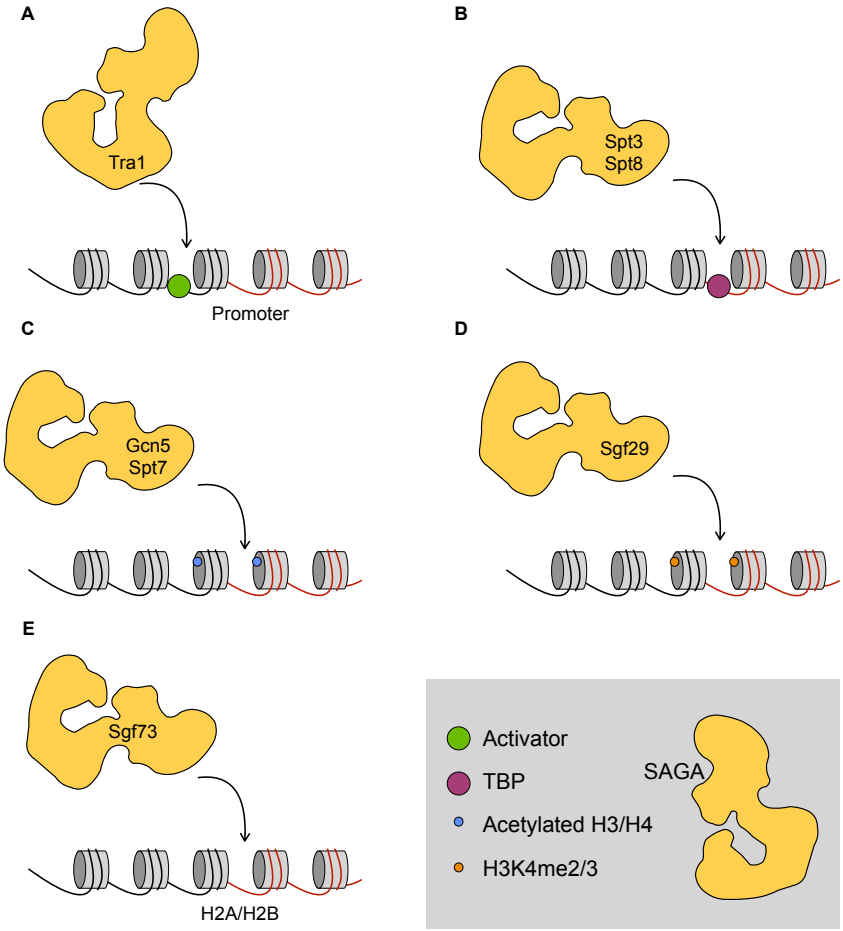


Figure 23 – Multiple mechanisms of SAGA coactivator recruitment to chromatin. Multiple SAGA subunits interact with either activators or histone marks. **(A)** Tra1 interacts with activators while **(B)** Spt3 and Spt8 interact with TBP. The **(C)** bromodomains of Gcn5 and Spt7 interact with acetylated nucleosomes, **(D)** the double Tudor domain of Sgf29 binds H3k4me2/4 and **(E)** the SCA7 domain of Sgf73 binds H2A/H2B dimers. Adapted from (Weake and Workman, 2012).

Importantly, SAGA can also be recruited to chromatin via mechanisms that do not encompass interaction with either activators and/or members of the transcription machinery. In fact, SAGA can potentially be recruited or retained at promoters through interactions with modified histones enriched at SAGA binding sites. Particular examples are the

bromodomains of both Spt7 and Gcn5 that are able to interact with acetylated histones and were shown to participate in the retention of SAGA at promoters when transcriptional activators are absent (Hassan *et al.*, 2002) (**Figure 23C**). Another subunit of the SAGA complex that has the capacity to read histone modifications and be involved in SAGA recruitment/retention is Sgf29 (**Figure 23D**). Sgf29 contains a double Tudor domain, granting a capacity to bind H3K4me2/3. Importantly, in human cells, the localization of Sgf29 at promoters correlates with H3K4me3 existence, as revealed by ChIP-seq analyses (Vermeulen *et al.*, 2010). Also, in budding yeast loss of Sgf29 leads to a genome-wide mislocalization of SAGA (Vermeulen *et al.*, 2010; Bian *et al.*, 2011). Both of these results corroborate the implication of Sgf29 in SAGA recruitment. Finally, Sgf73 (or its human ortholog ATXN7) contains a Zn-binding fold within its SCA7 domain, rendering it the capacity to bind H2A-H2B dimers, but not H3-H4 tetramers (Bonnet *et al.*, 2010) (**Figure 23E**). Also, ATXN7L3 contains a similar domain but without the same histone/nucleosome binding properties (Bonnet *et al.*, 2010).

Besides the abovementioned domains, other subunits of the SAGA complex have domains known to be capable of mediating an interaction with chromatin. For instance, Ada2 contains a SANT domain, predicted to bind to histone tails, and a SWIRM domain predicted to regulate transcription through protein-protein interactions. Moreover, Spt8 contains a WD40 domain, involved in interactions with methylated lysines (Lee and Workman, 2007). Impressively, according to the model constructed by Setiaputra *et al*., subunits of SAGA with chromatin-binding properties are clustered in proximity to each other, along one side of the complex. Their proximity and location within the complex suggests a major interaction interface with the chromatin template, specifically in a region with high degree of flexibility (Setiaputra *et al*., 2015) (**Figure 24**).

Altogether, it is quite recognizable the different mechanisms that might dictate SAGA recruitment at genes. More importantly, the action of these distinct paths is not necessarily mutually exclusive. Instead, they might work synergistically and in accordance to the epigenetic landscape in order to ultimately result in SAGA binding and/or retention at its target *loci*.

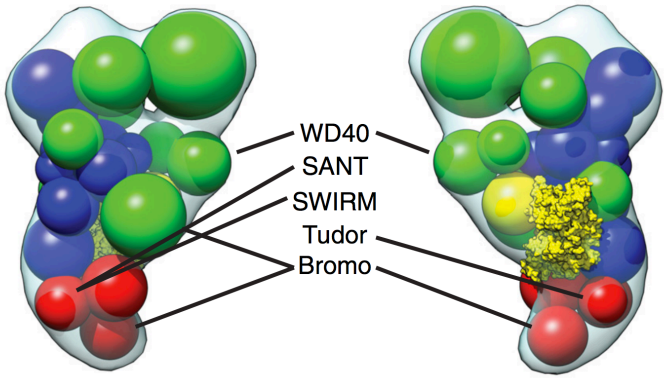


Figure 24 – Chromatin-binding domains of SAGA cluster around one edge of the complex. SAGA subunits with domains known to bind histones are shown in the spatial arrangement model, as indicated in Figure 22. From (Setiaputra *et al*., 2015).

2.7. SAGA and transcriptional regulation

Considering the implication that SAGA has been having on transcription, three distinct seminal studies have analyzed RNA Pol II transcription upon deletion or conditional loss of SAGA subunits. Importantly, two of these studies specifically addressed the interplay/redundancy between SAGA and TFIID due to their shared subunits and activities (discussed in detail below) (Lee *et al.*, 2000; Huisinga and Pugh, 2004). The third study, the most recent among them, analyzed the impact of all non-essential subunits of SAGA (and other transcription complexes) on the transcriptome of *S. cerevisiae* (Lenstra *et al.*, 2011).

Specifically, Lenstra and co-workers generated microarray expression profiles for viable deletion mutants of chromatin regulators in budding yeast, including nucleosome remodelers, histone modifiers and coregulators of transcription by RNA Pol II, including SAGA. Interestingly, globally the SAGA mutants' expression profiles do not result in one single gene expression signature. Instead, the clustering of the observed profiles resembled the modular and functional organization of the complex reported by others (Wu *et al.*, 2004; Lee *et al.*, 2011; Han *et al.*, 2014; Setiaputra *et al.*, 2015), meaning that subunits belonging to the same module or having a similar activity within the complex would share the same expression profile. Additionally, while the number of genes affected by SAGA depletion was very small, there was a range within the different subunits, with some being more impactful than others (Lenstra *et al.*, 2011).

Deletion of the structural subunits of the complex (Spt7, Spt20, and Ada1) resulted in a similar expression profile, being clustered in the same group. Also, the loss of these

subunits was the one causing the strongest transcriptional phenotype. Nevertheless, and as mentioned in the previous paragraph, transcriptional dysregulation only happened for a small fraction of genes (from 300-400). Additionally, *spt3*Δ and *spt8*Δ – forming one cluster – and *gcn5*Δ, *ada2*Δ, and *ada3*Δ – forming another cluster – were shown to present a distinct and limited gene expression changes signature, indicating that different activities of the complex affect different set of genes. Not only that, but the clustering is in good agreement with the modular nature of this coactivator. In contrast, some subunits seemed not to have any relevant transcriptional role, as observed on *sgf29*Δ (from the mHAT) and strains lacking mDUB subunits (like *ubp8*Δ and *sgf11*Δ). Actually, from the mDUB only Sgf73 seemed to have a modest, but observable, impact on transcription (Lenstra *et al.*, 2011). These effects might reflect the loss of the whole DUB, since Sgf73 anchors this module to the complex (Lee *et al.*, 2009). On the other hand, all mDUB subunits are strictly required for its assembly and activity (Rodriguez-Navarro, 2009), so these Sgf73 specific effects could be due not to the enzymatic activity itself, but could be explained by an altered conformation of the complex in the absence of the C-terminal part of Sgf73 known to interact with the core SAGA complex.

Another interesting observation is that when gene expression changes were identified, some genes were found down-regulated although a significant number of genes were up-regulated. Since SAGA is a coactivator, the expected result would be that genes would be predominantly down-regulated, with exceptions that could be due to experimental variability, for example. Interestingly, this could potentially reflect (i) the existence of secondary effects that would drive up-regulation of some genes in response to down-regulation of others, (ii) a compensatory mechanism that will be further discussed in *section 3* of the Introduction - *A facelift on transcription: how to address old questions using new approaches*, or, more remotely, (iii) that SAGA has some repressor activity. It is important to mention that the same was not so apparent in other genome-wide transcriptomic studies on SAGA (Huisinga and Pugh, 2004).

2.8. SAGA and TFIID: a shared but distinct effort to drive transcription

As mentioned before, SAGA and TFIID share some common subunits, namely Taf5, Taf6, Taf9, Taf10, and Taf12. In fact, those subunits seem to be involved in the formation of a structural core of the SAGA complex that highly resembles that of TFIID. Additionally, both share the capacity to deliver TBP at promoters. Considering that both complexes share common features, it has been questioned whether their activities overlap genome-wide or, instead, if they have distinct functional roles in RNA Pol II transcription, each controlling the expression of a given set of genes.

2.8.1. Functional redundancy between SAGA and TFIID

The first study to systematically address this question came from the laboratory of Richard A. Young. Specifically, using high-density oligonucleotide arrays, the authors investigated the implication of both SAGA and TFIID on transcription. For that, they performed their analysis in thermo-sensitive mutants for all shared TAFs (Lee *et al.*, 2000). While Tafs were previously described as having a general role in class II transcription, from what they could observe none of the common Tafs by itself was required to the same extent as RNA Pol II itself. Instead, each Taf seemed to affect a specific group of genes, having an overall impact on transcription ranging from 8 to 59% of the genome. However, combining genes that are affected by one or more of these analyzed Tafs, they could determine that SAGA and TFIID are together responsible for the expression of 70% of the budding yeast genome (**Figure 25A** and **B**), consistent with previous reports (Holstege *et al.*, 1998). Nevertheless, this information by itself could not distinguish the activities of SAGA and TFIID. To further understand the specific role of each complex, the same type of analysis was performed in thermo-sensitive mutants for TFIID-specific subunits (Taf1 and Taf2) and thermo-sensitive mutants for SAGA-specific subunits (Spt3, Spt20, and Gcn5). From that they observed that (i) Spt3, Gcn5 and Spt20 together regulate around 12% of the entire budding yeast genome, (ii) Spt3 and Gcn5 regulate the expression of a small and distinct set of genes, (iii) Spt20 had a broader role in transcription, but still very limited, (iv) genes affected by loss of Spt3 were included within the group of genes regulated by Spt20, (v) while the genes regulated by either Gcn5 or Spt20 partially overlapped, indicating specific roles for Gcn5 outside of the SAGA complex (Lee *et al.*, 2000) (**Figure 25C** and **D**). On the other hand, conditional loss of Taf1 or Taf2 was responsible for the regulation of 27% of the genes, with the authors reasoning that it is either a subunit-specific effect or that SAGA

might be able to somewhat compensate for the loss of the TFIID-specific subunit (**Figure 25D)**. When deleting Gcn5 in the Taf1 thermo-sensitive background, changes in expression were much bigger than when these subunits were separately absent, suggesting that SAGA and TFIID are redundant (Lee *et al.*, 2000). The authors advocated that this more impactful effect was due to the loss of the acetyltransferase activities of both Gcn5 and Taf1. Nevertheless, it is important to mention that Taf1 acetyltransferase activity was only shown *in vitro* and direct evidence about this matter are weakly supported *in vivo*, as no point mutant specifically affecting the putative HAT domain was characterized.

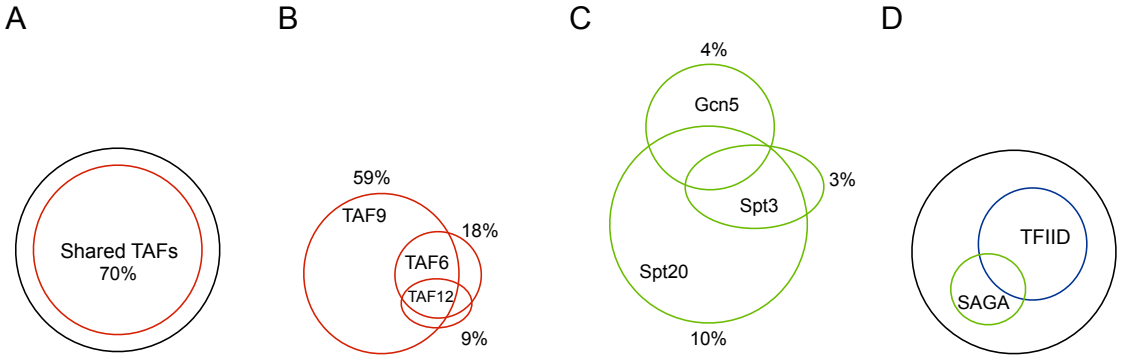


Figure 25 – Transcriptional effects promoted by depletion of TFIID and/or SAGA subunits. (A) The sum of individual subunit contributions reveals a global role for shared TAFIIs. **(B)** Genes affected by shared Tafs overlap significantly, although each of the Taf affects genome-wide transcription to different extents. **(C)** Expression of distinct sets of genes depends on individual subunits of SAGA. **(D)** Venn diagrams showing genes dependent on TFIID-specific and SAGA-specific subunits. The black circle represents the whole transcriptome. Adapted from (Lee *et al.*, 2000).

In 2004, Huisinga and Pugh decided to further explore this question. For that, they used different strains in which the isolated effect of subunit depletion of either complex was analyzed (Taf1, Gcn5, and Spt3) or in combination with each other (Taf1 + Gcn5, Taf1 + Spt3). Regarding the strain deleted for Gcn5, they observed a global and uniform decrease in transcription, with the average fold-change centering around 2, with 60% of the genes presenting a 1.7-fold reduction, which the authors classified as modest. Interestingly, the authors exclude that the effects are attributed to the acetyltransferase activity, since a catalytic dead mutant presented no significant changes on RNA Pol II transcription. In the situation where Spt3 was absent, no global changes were observed in comparison to the wild-type. Nevertheless, it was possible to delineate a discrete set of genes (representing around 11% of the genome) whose transcription was dependent on Spt3 (Huisinga and Pugh, 2004). On the other hand, in the absence of functional Taf1, transcription was globally affected by

an average factor of 2.4, with 84% of the analyzed genes having a decrease in transcription higher than 1.7-fold. More importantly, when depletion of Taf1 was combined with deletion of Spt3, transcription was virtually abrogated, with a decrease of around 4-fold and with a profile similar to that of a mutant lacking a functional Rpb1 (RNA Pol II). From this, the authors concluded that both Spt3 and TAF1 contributed to the expression of almost all measured transcripts. Furthermore, they also observed that Gcn5 and Taf1 are partially redundant, as previously reported (Lee *et al.*, 2000; Huisinga and Pugh, 2004). Altogether, the results from this work allowed the categorization of the genes according to their requirement of either TFIID or SAGA to deliver TBP at promoters. Hence, on one hand, we have the TFIID-dominated genes, representing around 90% of the yeast genome, whose transcription is predominantly dependent on TFIID/Taf1. By opposition, the SAGA-dominated genes represent the remaining 10% of *S. cerevisiae* genome and, as indicated by the nomenclature, their expression is mainly dependent of SAGA/Spt3 (**Figure 26**). Nevertheless, the authors do not exclude that one factor might be involved in the transcription of genes belonging to the contrasting group (Huisinga and Pugh, 2004).

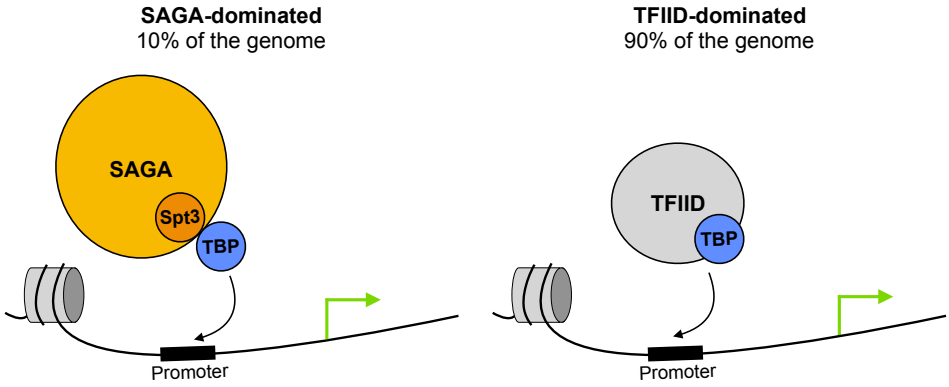


Figure 26 – TFIID and SAGA regulate the expression of different sets of genes. SAGA-dominated genes, predominantly regulated by SAGA, represent 10% of the budding yeast genome, while TFIID-dominated genes comprise the remainder 90% of the genome.

While some of the conclusions of these two studies are grossly similar, some obvious differences are observed, specifically regarding the number of genes that were detected to be dependent on TFIID/SAGA. While one could argue that the differences observed in the Spt3 mutants (3% *versus* 11%) are either due to the approach used (high-density oligonucleotide arrays *versus* microarrays) or the different genetic background of the strain, the transcriptional changes in the Taf1 mutants are tremendously contrasting (27% *versus* 84%).

This is not fully discussed in the report by Huisinga *et al*., only stating: "when SAGA is present, the bulk of the genome does not display an absolute dependence on Taf1 (TFIID), which is in agreement with previous reports".

2.8.2. TATA-containing versus TATA-less/-like show different dependency on SAGA or TFIID

As previously referred, the TATA box plays an important role in the assembly and nucleation of the PIC at promoters. Due to the chromatin status itself, in several cases this DNA element might be occluded. According to the study by Huisinga and Pugh, the coactivators SAGA and TFIID might be responsible to conduct delivery of TBP at promoters. However, the TATA element is not equally present at RNA Pol II promoters, allowing the distinction between TATA-containing and TATA-less promoters. Still, the genome is globally dependent on the partially redundant activities of both SAGA and TFIID complexes for efficient transcription (Pugh and Tjian, 1991; Lee *et al.*, 2000; Huisinga and Pugh, 2004) and TBP was found to be recruited at both TATA-containing and TATA-less promoters (Kuras and Struhl, 1999; Li *et al.*, 1999). After several attempts (Singer *et al.*, 1990; Cliften *et al.*, 2003; Kellis *et al.*, 2003), in 2004, the laboratory of B. Franklin Pugh published a seminal paper regarding the identification of the TATA element and regulation of genes containing a canonical TATA-box (Basehoar *et al.*, 2004). There, they have reported that the consensus sequence of the TATA motif is TATA(A/T)A(A/T)(A/G) and that it is present in approximately 19% of promoters of the budding yeast genome, rendering the remainder genes TATA-less. Regarding the regulation of genes containing, or not, a TATA box, they aimed at understanding how the TBP regulators – SAGA/Spt3 and TFIID – differently control the expression of those genes. From their analyses they could conclude that: (i) Taf1-independent genes are mainly TATA-containing; (ii) Taf1-dependent genes were in the vast majority TATA-less; (iii) SAGA/Spt3-dependent genes had a high proportion of genes containing a TATA. From this, it was possible to conclude that, considering both SAGA and TFIID as the main deliverers of TBP at promoters, TFIID predominantly delivers TBP at TATA-less/TFIID-dominated genes and SAGA does it at TATA-containing/SAGA-dominated genes (Basehoar *et al.*, 2004; Huisinga and Pugh, 2004) (**Figure 27**).

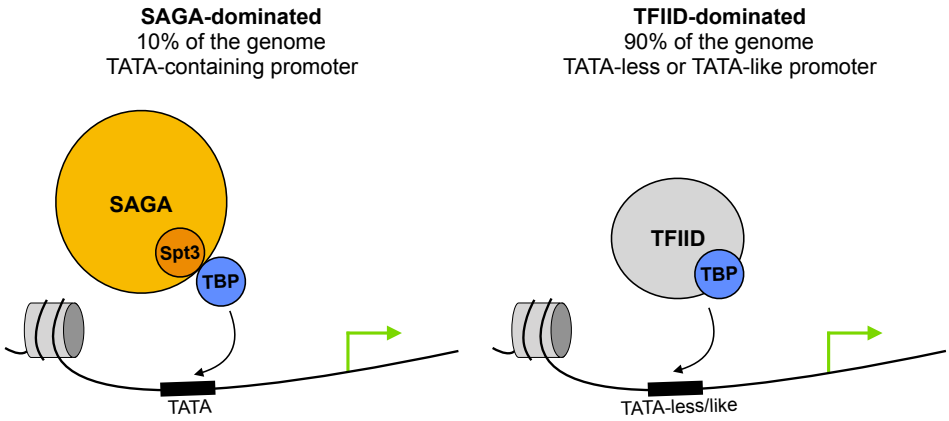


Figure 27 – TATA-containing and TATA-less promoters show different dependency on either SAGA or TFIID. SAGA-dominated genes tend to have a TATA consensus sequence on their promoter, while TFIID-dominated genes usually lack one.

2.8.3. Stress-induced genes *versus* housekeeping genes

Since SAGA is nonessential (all yeast strains deleted for SAGA-specific subunits are viable) and regulates only a small fraction of the yeast genome, this raised a question regarding the physiological role of the coactivator and what type of genes it regulates. Through comparison of publicly available data regarding transcriptional changes upon general environmental stresses with the categorization of those genes according to their dominance of either SAGA or TFIID, the authors reported that: (i) genes that are up-regulated during stress (heat, oxidation, acidity, DNA damage nutrient starvation) were strongly biased toward being SAGA-dominated; (ii) genes that are down-regulated/repressed during stress tend to be housekeeping genes and TFIID-dominated (Huisinga and Pugh, 2004) (**Figure 28**).

On the other hand, a different study from the same group has performed the same type of analysis, but this time using the TATA-containing/TATA-less as the reference. Importantly, it was observed that genes that were up-regulated during stress were biased towards the existence of a TATA box on their promoter (50% of the induced genes contained a TATA sequence). In contrast, the vast majority of the stress-inhibited genes (94% of the analyzed genes) lacked a TATA consensus sequence (Basehoar *et al.*, 2004). Interestingly, most of the stress-inhibited genes have the tendency to be general housekeeping genes (Gasch *et al.*, 2000; Causton *et al.*, 2001), thus indicating that housekeeping genes are prone to be deprived of a TATA box. Considering all the collected information, it is possible to conclude that:

- SAGA-dominated genes (10% of *S. cerevisiae* genome) comprise mainly stress-induced, TATA-containing genes;
- TFIID-dominated genes (90% of *S. cerevisiae* genome) comprise mainly housekeeping, TATA-less genes.

However, it was still puzzling how TFIID would deliver TBP to TATA-less promoters, considering the lack of a binding sequence. However, the more recent work from Rhee and Pugh identified that TATA-less promoters instead contained a sequence having two or fewer mismatches to the TATA box consensus. These elements are designated as 'TATA-like', since they do not form a consensus, and together with the canonical TATA consensus sequence, are referred as 'TATA elements'. As expected, these TATA-like sequences are found on the promoters of TFIID-dominated genes (Rhee and Pugh, 2012). Nevertheless, it has been recently shown that sequence-specific TBP binding is not required at yeast TATA-less promoters and that promoter specificity must be generated by alternative mechanisms other than TBP recognition (Kamenova *et al.*, 2014).

2.8.4. Acetylation of histones allows the differentiation of SAGA- and TFIID-dominated genes

SAGA, via the bromodomains of Gcn5 and Spt7, can interact with acetylated histones. Furthermore, TFIID interacts with the bromodomain-contacting protein 1 (Bdf1) protein, thus also being indirectly capable of interacting with acetylated lysines, specifically modified histones. Interestingly, since it was reported that Bdf1 interacts specifically with acetylated histones (Jacobson *et al.*, 2000; Ladurner *et al.*, 2003; Matangkasombut and Buratowski, 2003), it was hypothesized that genes containing high level of acetylated H3/H4 would specifically recruit TFIID and would thus be TFIID-dominated (Huisinga and Pugh, 2004). Interestingly, the authors observed that hypoacetylated regions tended to be SAGA-dominated, while genes with high levels of acetylated H4 tended to be TFIID-dominated, indicating that the acetylation state of histone H4 differentially contributes to the regulation of SAGA- and TFIID- dominated genes (**Figure 28**). A similar correlation was not observed for intergenic H3 acetylation. Interestingly, in *hda1*Δ and *rpd3*Δ strains (Hda1 and Rpd3 are histone deacetylases, so the corresponding deletion strains present hyperacetylation of both H3 and H4), genes that were the most highly acetylated tended to be SAGA-dominated. This indicates that Hda1 and Rpd3 play a role in maintaining the levels of acetylation low in histones surrounding SAGA-dominated genes (Huisinga and Pugh, 2004). Results regarding

the relationship between histone marks and regulation of TATA-containing and TATA-less were consistent with the previous results: acetylation of histone tails inhibited the expression of mainly TATA-containing genes (Basehoar *et al.*, 2004).

2.8.5. Highly regulated genes *versus* lowly regulated genes

Stress-induced genes are more dominated by SAGA activities and housekeeping gene expression relies mainly on TFIID. This supposes that SAGA-dominated genes might be more regulatable than the TFIID-dominated ones, due to the plastic nature of stress-responsive genes. Indeed it has been revealed that SAGA-dominated genes are the targets of either negative or positive regulation by several chromatin-associated complexes (Basehoar *et al.*, 2004; Huisinga and Pugh, 2004) (**Figure 28**):

- Ssn6-Tup1, a repressor complex that interacts with non-acetylated histones H3 and H4 and broadly associated with repression of transcription (Keleher *et al.*, 1992; Malave and Dent, 2006), tends to repress genes that are SAGA-dominated with high frequency of a TATA consensus sequence;
- Srb10, subunit of the Mediator complex associated with transcription inhibition or nuclear exclusion/turnover of stress-response activators (Ste12 and Gcn4, among others) (Hengartner *et al.*, 1998; Chi *et al.*, 2001; Prelich, 2002; Raithatha *et al.*, 2012; Gonzalez *et al.*, 2014), mainly has an inhibitory role on genes that are regulated by SAGA and contain a TATA-box;
- Msn2 and Msn4 stress-induced activators have the tendency to positively regulate SAGA-dominated genes.

Also, removal of TBP from promoters and inhibition of the formation of a processive PIC can be mediated by two distinct complexes: Mot1 and the heterodimeric complex NC2 (NC2α and NC2β). While Mot1 uses its ATPase activity to displace and redistribute TBP along the promoters, NC2 interacts with TBP and obstructs the binding of the GTFs TFIIA and TFIID, thus inhibiting the formation of the PIC. Interestingly, regarding transcriptional effects, genes that are more up-regulated upon mutation or conditional nuclear depletion of either Mot1 or NC2 subunits are those that are SAGA-dominated and TATA-containing, indicating that the negative regulators of TBP activity act in a more SAGA-dependent context (Basehoar *et al.*, 2004; Huisinga and Pugh, 2004; Spedale *et al.*, 2012a). In contrast, genes that are down-regulated upon abrogation of Mot1 and NC2 activities are generally TFIID-dominated (Spedale *et al.*, 2012a). Additionally, genome-wide binding studies have also

shown that there is a modest tendency for genes that possess a TATA box to be more bound by SAGA and Mot1, but not by NC2α and NC2β. Nevertheless, when considering the strength of the binding by those factors, NC2β, but not NC2α, tends to bind strongly to TATA-containing genes. Instead, the latter has a stronger binding biased towards Taf1-enriched, TATA-less/like genes (van Werven *et al.*, 2008). Despite some discrepancies, the results suggest that negative TBP regulators collaborate with SAGA to regulate gene expression, as suggested by other studies besides those mentioned (Collart, 1996; Madison and Winston, 1997; van Oevelen *et al.*, 2005).

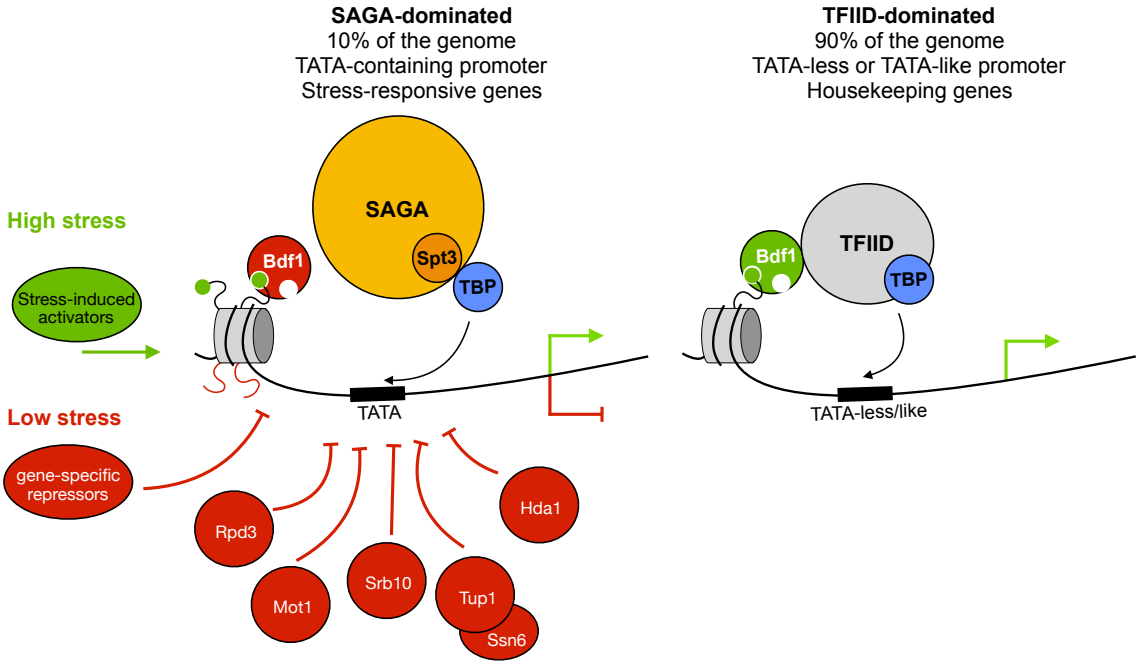


Figure 28 – SAGA- and TFIID-dominated genes are differently regulated. SAGA-dominated genes, mainly comprising stress-induced genes, are tightly regulated, either positively or negatively. Environmental stress leads to up-regulation of a large subset of SAGA-dominated genes (in green). Factors illustrated in red are among some of the factors that play a negative role on the regulation/expression of SAGA-dominated genes. On the other hand, TFIID-dominated genes are less regulated and mainly represent housekeeping genes. Its regulation is positively regulated by H4 acetylation represented as green ball perturbing from the nucleosomes). Adapted from (Huisinga and Pugh, 2004).

2.8.6. Nucleosome positioning and transcriptional plasticity

As previously mentioned, promoter nucleosome architecture was initially defined within two different states (Tirosh and Barkai, 2008), although this vision has been recently challenged and rewritten (Kubik *et al.*, 2015). Nevertheless, housekeeping genes, that tend to

be TFIID-dominated and TATA-less, are typically characterized by a well-defined NFR, found upstream of the coding region and flanked by two well-positioned nucleosomes, enriched for binding sequences. Also, these promoters have less transcriptional noise and are coined as low-plasticity promoters (Basehoar *et al.*, 2004; Huisinga and Pugh, 2004; Tirosh and Barkai, 2008; Rhee and Pugh, 2012). On the other hand, promoters containing a TATA-box, thus SAGA-dominated, exhibit a more variable promoter architecture. In fact, these promoters lack a defined NFR and have fuzzy nucleosomes upstream of the TSS. In this scenario, TFs binding sites are likely occluded by nucleosomes, thus reflecting a competition between nucleosome and TF binding and explaining a high degree of transcriptional noise for the corresponding genes (Tirosh and Barkai, 2008; Rhee and Pugh, 2012). Nevertheless, binding sites become free through the activity of chromatin remodelers and/or chromatin modifiers, which are also more required for the expression of the genes adjacent to these promoters, thus conferring to this genes a high degree of transcriptional plasticity (Boeger *et al.*, 2008; Tirosh and Barkai, 2008).

2.9. Discrepancies between SAGA recruitment, activity and impact on transcription

For the past years, the role of SAGA has been consistently studied. From its enzymatic activities, recruitment to chromatin and impact on transcription, today we have in our hand a tremendous amount of knowledge that perfects our understanding of the complex specifically, and on transcription in a broader sense. Nevertheless, the combination of all collected results is far from being in agreement. Simplistically, one would expect that genes that are transcriptionally dependent on SAGA would also be bound by the complex. Also, considering its enzymatic activities, genes regulated and bound by SAGA are expected to be punctuated by this coactivator complex. However, as it will be possible to see here, not everything is as easy to interpret.

Regarding the impact of the complex on transcription by RNA Pol II, the different studies seem to agree on several points (Lee *et al.*, 2000; Huisinga and Pugh, 2004; Lenstra *et al.*, 2011):

- The Spt3 activity of the SAGA complex (TBP recruitment) seems to have a minor impact on steady-state transcription;
- Genes are globally more regulated by TFIID than by SAGA;

- The Gcn5 activity (HAT) had slightly more conflicting results, with some reporting a minimal impact and others reporting a broader role. Still, studies agree that the impact, even if general, is modest.

Although the results do not point towards exactly the same direction, there is a general agreement concerning the existence of SAGA- and TFIID-dominated genes. So far, this classification has been accepted by the scientific community as the two pathways for TBP deliver and PIC assembly.

On the other hand, when looking at the genome-wide distribution of the complex, the results are, once again, dispersed. For instance, Robert *et al*. have shown that SAGA/Gcn5, together with NuA4, was broadly and globally bound to actively transcribed genes, thus suggesting a genome-wide recruitment of the coactivator (Robert *et al.*, 2004). Also a work from the group of H. Th. Marc Timmers has reported a binding of the SAGA complex, through chromatin immunoprecipitation of Spt20, which would largely surpass the 10% of genes whose transcription is affected by the loss of the complex (van Werven *et al.*, 2008). Altogether, and judging by recruitment alone, these results indicate a less restricted role of the SAGA complex in RNA Pol II transcription. More recently, the laboratory of B. Franklin Pugh, that initially characterized the SAGA- and TFIID-dominated genes, has provided a comprehensive genome-wide binding map of chromatin regulatory proteins, including SAGA (Venters *et al.*, 2011). Here the authors distinguish between SAGA and TFIID regulated pathways of transcription, which, according to their results, seem to have distinctive features. Among other conclusions, such as that SAGA pathway is more specialized and is generally bound by a bigger number of chromatin modifiers than TFIID-dominated genes, the authors also reported that genes tend to use both TFIID and SAGA in different extents, since components of both complexes were located to thousands of genes (Venters *et al.*, 2011).

Interestingly, a report from 2012 has analyzed both the recruitment of the complex studied by Venters *et al.* and the genome-wide transcriptional effects upon SAGA loss reported by Lenstra *et al.* (Lenstra et al., 2011). Here, the authors reported an enormous disconnection between the globally similar binding patterns of chromatin factors and the transcriptional effects provoked during their loss. This was true for the SAGA complex, but also for a plethora of other complexes. Importantly, the binding pattern of the different cofactors analyzed was far less specific than the transcriptional effects observed. Specifically for SAGA, the authors indicate that there is a significant negative correlation between binding and effects of three SAGA subunits. However, this correlation had an $R < -0.15$,

which one could argue to be biologically irrelevant. Also, only a fraction of bound genes is dysregulated (either up- or down-regulated) upon deletion of SAGA subunits, and, in contrast, only a fraction of dysregulated genes is bound by the corresponding subunits (Lenstra and Holstege, 2012) (**Figure 29**). In fact, the same behavior is observed for histone marks: although some marks, like H3K4me3, are always associated with active transcription, its loss through Set1 deletion has virtually no impact on gene expression (Howe *et al.*, 2017). Also, it is important to mention that in metazoan, the binding sites for SAGA complex, in comparison to the whole genome, are very limited, as described by ChIP in mammalian cells (Vermeulen *et al.*, 2010; Krebs *et al.*, 2011). Specifically, in Krebs *et al.* the authors immunoprecipitated subunits of both ATAC (ZZZ3) and SAGA (Spt20). While some patterns of differential binding of the two complexes were identified, the total number of high-confidence binding sites did not exceed 400. These results indicated that these complexes would only be required for a very limited number of *loci* (both promoter and enhancer regions) (Krebs *et al.*, 2011).

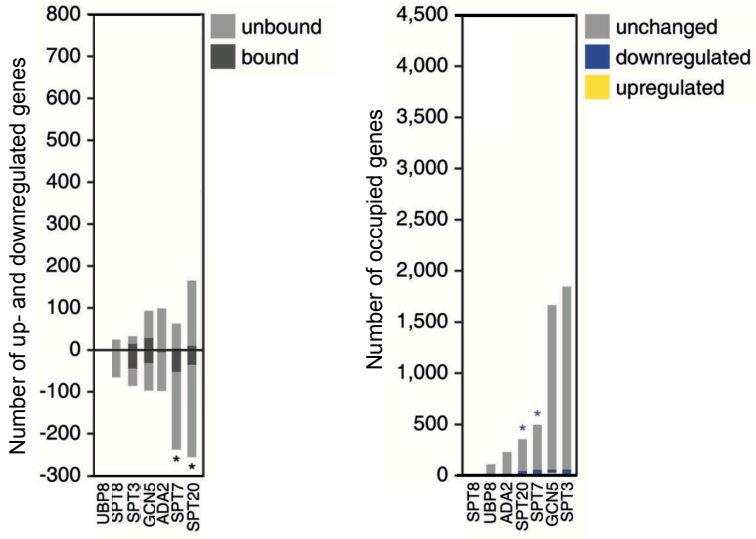


Figure 29 – Overlap between occupancy and expression of different SAGA complex subunits. **(*Left*)** Number of significant up- and down-regulated genes for deletion mutants of indicated SAGA subunits shown on the *x*-axis. Genes that are also bound by the same factor are colored dark gray **(*Right*)** Number of genes occupied by the factors shown on the *x*-axis. Genes that also show expression effects in the corresponding deletion mutants are colored yellow (up-regulation) or blue (down-regulation). Adapted from (Lenstra and Holstege, 2012).

Despite the fact that recruitment does not necessarily translate into a measurable effect, what is the complex doing there? Also, is the complex just there by chance, which

seems unlikely, or does it have a proper role? How can we explain the seemingly repressive effect of the SAGA coactivator on a subset of genes? As previously mentioned, one would expect a coherence between binding, activity and transcriptional effects. While recruitment of SAGA indicates a broader role for the complex, specific transcriptional effects indicate that the coactivator is only necessary for the expression of a subset of genes. What about its enzymatic activities?

Table 2 – Overall results obtained in distinct studies regarding the SAGA transcriptional regulation, binding and extent of enzymatic activity.

	Transcriptional effects	Binding/ Recruitment	Enzymatic activity
Lee *et al.*, 2000	**Limited** (12% of the yeast genome)		
Huisinga and Pugh, 2004	**Limited** (10% of the yeast genome)		
Lenstra *et al.*, 2011	**Limited** (≈ 60 – 270 down-regulated genes; ≈ 20 – 180 up-regulated genes)		
Robert *et al.,* 2004		**Global** (recruitment of Gcn5 correlated with transcription)	
Venters *et al.*, 2011		**Relatively broad** (≈100 – 1800 bound genes, depending on the subunit analyzed)	
Lenstra and Holstege, 2012	**Limited** (down-regulated genes ≈ 60 – 270 up-regulated genes ≈ 20 – 180)	**Relatively broad** (≈100 – 1800 bound genes, depending on the subunit analyzed)	
Vermeulen *et al.*, 2010		**Limited**	
Krebs *et al.*, 2011		**Limited** (approximately 400 binding sites)	
Henry *et al.*, 2003			**Global**
Bian *et al.*, 2010			**Global**
Jin *et al.*, 2011			**Global**

Several studies have indicated that the enzymatic activities of SAGA are prolific genome-wide, both in yeast and mammals. In a study performed in *S. cerevisiae*, Henry *et al.* reported that Ubp8 globally deubiquitinates H2B in a SAGA-dependent manner. In fact, loss of Ubp8 resulted in an overall accumulation of H2Bub by 10-fold (Henry *et al.*, 2003). Similar observations were disclosed in human cells, where loss of mDUB activity led to an enrichment of monoubiquitination of H2B (Lang *et al.*, 2011). Also, the fact that H2Bub is enriched in the transcribed region of actively transcribed regions in the mammalian system (Minsky *et al.*, 2008; Vethantham *et al.*, 2012), serving as a platform for the SAGA complex, indicates that the coactivator might be active genome-wide. Interestingly, also the HAT activity of SAGA/Gcn5 has been implicated to occur in a broad way. For instance, in a study conducted in budding yeast, Bian and colleagues have reported that deletion of Gcn5, Ada3 or Sgf29, subunits of the mHAT, led to a global decrease of histone H3 acetylation *in vivo*, mainly H3K9ac and H3K18ac (Bian *et al.*, 2011). Moreover, something similar is observed in mammalian cells: deletion of GCN5 and PCAF, mutually exclusive catalytic subunits of the SAGA/ATAC complexes, terminates in complete loss of H3 acetylation, specifically H3K9ac (Jin *et al.*, 2011). Also, the fact that H3K9ac is consistently found on the promoters of actively transcribed genes, and since Gcn5 seems to be the main HAT responsible for its acetylation, one can conclude that SAGA/Gcn5 is active genome-wide.

Altogether these results collectively show that there is a lack of correlation between binding/recruitment, enzymatic activities and transcriptional effects (**Table 2**). Some of the possible explanations for the discrepancies observed are:

- Limited recognition of SAGA-regulated genes can be due to analysis of steady-state RNA: as we can see in the next chapter of the *Introduction*, analysis of total levels of RNA by themselves might occlude the real effects observed at the level of nascent transcription;
- Broader binding of the factor to chromatin in comparison to transcriptional effects upon loss of SAGA subunits might indicate that (i) while the complex is recruited to transcribed *loci*, it does not possess a regulatory role or (ii) the activity of the complex might be compensated by another factor;
- One of the drawbacks of ChIP is its inefficiency to capture very transient interactions between factor and chromatin. Thus the few genomic *loci* bound by SAGA, as

observed in mammalian cells, does not necessarily reflect the real requirement of SAGA.

Having that, despite the existence of a big collection of information regarding the functional role of the SAGA complex, that information is dispersed and lacks coherence. With that in mind, it is imperative to develop new strategies that would finally address the requirements of SAGA for RNA Pol II transcription.

3. A facelift on transcription: how to address old questions using new approaches

Transcriptomic studies approach transcription as a steady-state process, where total levels of mRNA might reflect the impact of a given factor. We know today that transcription is not an isolated biological process. In fact, transcription *per se* is highly and intricately associated with processes that occur simultaneously (mRNA processing) or after it (mRNA export, translation, and degradation). Nowadays, transcription should actually be regarded has a complex process part of a major regulatory axis, that includes mRNA synthesis, processing, export, translation, and degradation. In recent years, several studies have dug deep into this question and revolutionized the way transcription is analyzed. Prior to disclosing some of these new approaches, it is important to see that there is way more on transcription than initiation and elongation itself and that degradation has an important role on the transcriptional read-out. For the sake of keeping the focus, here we are going to overview the interplay between mRNA transcription and degradation. However, it is important to mention that mRNA export, for example, has been implicated in the regulation of cytoplasmic levels of RNA, trough transcription-coupled mechanisms (Molina-Navarro *et al.*, 2011; Babour *et al.*, 2016; Raices and D'Angelo, 2017).

3.1. mRNA synthesis and decay

Over recent years, there has been an accumulation of evidence showing that a crosstalk between synthesis and decay exists in a genome-wide fashion. This process has been evident in both budding yeast and mammalian cells (Helenius *et al.*, 2011; Sun *et al.*, 2012; Rodriguez-Molina *et al.*, 2016). For example, in mouse fibroblasts, deletion of Mat1, a subunit of TFIIH, did not result in dramatic changes in transcription when analyzing steady-state (or total) levels of RNA. In fact, only a small group of genes was affected, either up- or down-regulated, highly resembling the classical results of transcriptomic analyses. However, since deletion of Mat1 inhibits phosphorylation of serine 5 of the CTD of RPB1 (RNA Pol II), the expected results would be a more striking and generalized decrease of transcription. Importantly, using a method that allows the analysis of specifically nascent transcripts, the authors disclosed that, indeed, transcription was overall decreased (Helenius *et al.*, 2011). Moreover, the same was true in *S. cerevisiae*. The laboratory of Aseem Z. Ansari reported

that, upon inhibition of the kinase activity of TFIIH, effects on transcription were only clearly observable when quantifying newly-transcribed RNA, since a compensatory mechanism masked the real impact on transcription (Rodriguez-Molina *et al.*, 2016).

In fact, the work of Rodriguez-Molina *et al*. used an experimental method delineated in the group of Patrick Cramer. Interestingly, a work published by the laboratory of the latter, was the starting point of my work. There, the authors reported that a point mutation on RNA Pol II tremendously reduces its elongation rate without dramatically affecting the levels of steady-state RNA. Using the pipeline developed by them – comparative dynamic transcriptome analysis (cDTA), discussed below – they have shown that (i) the rate of mRNA synthesis in that mutant was globally affected and that (ii) a compensatory mechanism allows buffering the global cellular levels of mRNA through reduction of their degradation. Not only that, but the same was true in a situation where subunits of the Ccr4-Not complex (involved in mRNA degradation) were deleted, with decreased decay being compensated by a decrease in RNA Pol II transcription (Sun *et al.*, 2012). Importantly, the existence of this interplay between the mRNA synthesis and decay masked the existence of transcriptional effects, which could be disclosed only through the use of appropriate techniques.

Consequently, it is important to (i) understand the mechanisms of mRNA degradation and how they act, (ii) how transcription and decay coordinate their activities and (iii) what are the benefits of such coupling.

3.2. Mechanisms of mRNA decay

Correctly processed mRNAs, possessing a 7-methylguanine cap at their 5’ end, are recognized by the eukaryotic initiation factor 4E (eIF4E), recruiting the mRNA to eIF4F and forming a complex with the small 40S ribosome, initiator Met-tRNA and initiation factors. This culminates in translation initiation of those mRNAs. When not associated with polysomes, mRNAs may associate with P-bodies and stress granules, cytoplasmic sites of storage and degradation. Usually, the first step of degradation of mRNA is shortening of their poly(A) tail. This process, denominated deadenylation, is primarily catalyzed by either Ccr4-Not complex or Pan2-Pan3 complex. Once deadenylation is terminated, the mRNA can be efficiently processed through its 3’end by the exosome, resulting in an oligonucleotide whose cap is removed by a salvage pathway promoted by Dcs1. Additionally, the deadenylation process also promotes decapping by the Dcp complex and auxiliary factors, particularly Dhh1 (DEAD box helicase). In this case, that process culminates in the 5’-3’ degradation of the

transcript by Xrn1, a 5'-3' exoribonuclease that contains an RNA binding domain and a RNase catalytic domain. Apart from these highly conserved mechanisms, there are alternative mechanisms that do not rely on shortening of the Poly(A) tail of the mRNA, collectively known as deadenylation-independent processes (**Figure 30**).

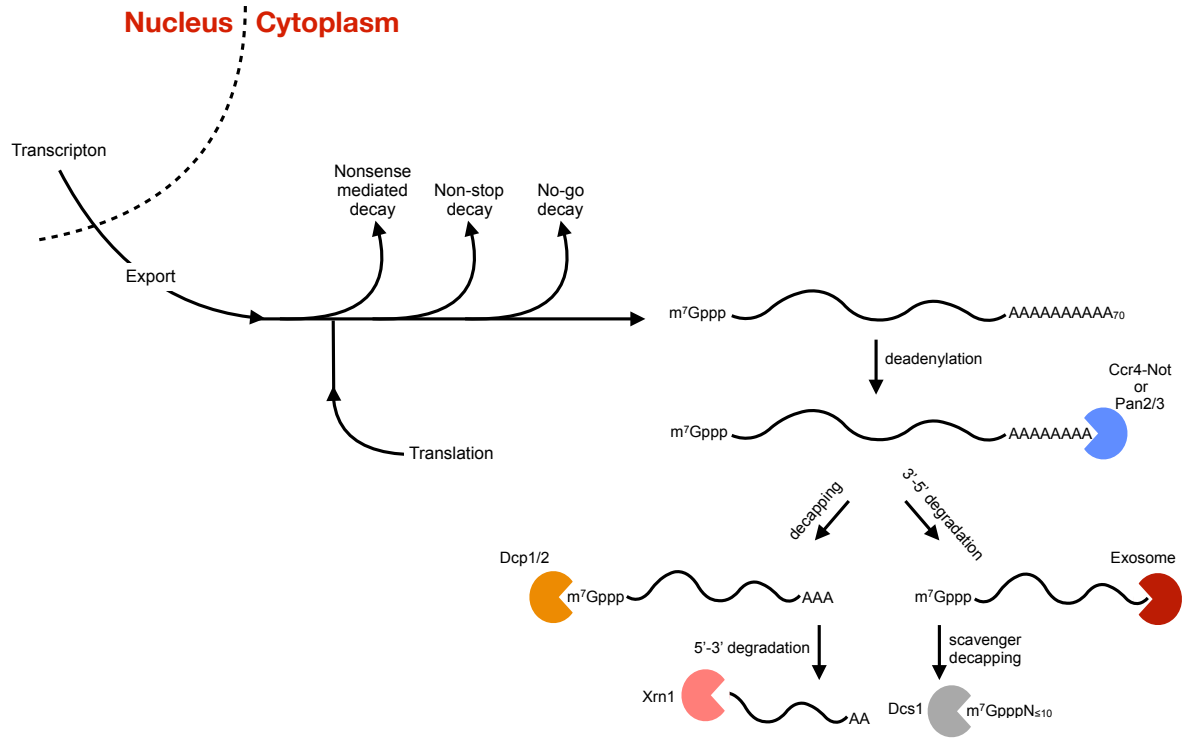


Figure 30 – Cytoplasmic mRNA decay pathways. Surveillance pathways can identify and initiate degradation of defective RNAs and RNPs, including non-stop decay (NSD), no-go decay (NGD) and nonsense-mediated decay (NMD). General cytoplasmic mRNA decay includes sequential deadenylation, decapping, and degradation by exonucleases from both ends. Adapted from (Braun and Young, 2014).

3.2.1. mRNA deadenylation

Two distinct complexes are responsible for the shortening of the poly(A) tail of mRNAs: Ccr4/Pop2/Not and Pan2/Pan3 complex, with the first one being the most predominant deadenylase (Daugeron *et al.*, 2001; Tucker *et al.*, 2001). Ccr4/Pop2/Not complex possesses two distinct active 3'-5' exonucleases – Ccr4 and Pop2/Caf1. Furthermore, it includes Not1, Not2, Not3, Not4, Not5, Caf40 and Caf130 proteins (Denis and Chen, 2003). While both Ccr4 and Pop2/Caf1 exhibit deadenylation activity *in vitro*, the activity of the whole complex is dependent on Ccr4 (Goldstrohm *et al.*, 2007). In fact, Pop2 is required for deadenylation by the Ccr4-Not complex not necessarily due to its enzymatic

activity, but mainly because it interacts and enhances the role of Ccr4 (Tucker *et al.*, 2002; Viswanathan *et al.*, 2004). For the remnant subunits of the complex, they either act as adaptor proteins for the activity of the complex or they might be implicated in functions of the complex outside of deadenylation, such as transcription initiation and elongation and co-translational assembly of macromolecular complexes (Deluen *et al.*, 2002; Tucker *et al.*, 2002; Swanson *et al.*, 2003; Kruk *et al.*, 2011; Kassem *et al.*, 2017). A second complex involved in mRNA deadenylation in the PAN complex, composed of Pan2 (catalytic subunit) and Pan3 (Boeck *et al.*, 1996). Interestingly, the PAN complex requires the binding of Pab1 to exert its activity, interacting through Pan3 subunit, while this process impedes the activity of Ccr4 (Boeck *et al.*, 1996; Tucker *et al.*, 2002). Additionally, PAN and Ccr4-Not complexes seem to act in a temporal manner: Pan2/Pan3 initially shortens the poly(A) tail from approximately 90 residues to 65 and Ccr4-Not appears to carry out the continued deadenylation of the mRNAs (Brown and Sachs, 1998; Tucker *et al.*, 2002). When the poly(A) tail of one mRNA reaches a size of around 10 residues, the transcript can be targeted for decapping and binding of Pat1/Lsm1-7 at the 3' end (Tharun and Parker, 2001; Chowdhury *et al.*, 2007). Once deadenylated, the transcript can follow two different pathways: 3'-5' degradation by the cytoplasmic exosome, or decapping and 5'-3' degradation by Xrn1.

3.2.2. mRNA decapping

mRNA decapping is promoted by the scavenger decapping factors DcpS (Dcp1/Dcp2). The Dcp1/Dcp2 complex is composed of two subunits: Dcp2, the catalytic subunit, and Dcp1, enhancer of Dcp2 activity (Parker, 2012). Hence, these two proteins form a complex where Dcp1 acts as a co-factor for the m^7G-cap-hydrolysis reaction promoted by Dcp2 (Coller and Parker, 2004). Dcp2 contains a NUDIX/MutT domain that can bind the 5'-cap and the RNA body. Indeed, the catalytic activity of Dcp2 is enhanced once bound to the RNA through the NUDIX domain (Dunckley and Parker, 1999; Deshmukh *et al.*, 2008). Importantly, actively translated mRNAs are prevented from decapping, since cap-binding protein eIF4E, the poly(A)-tail and poly(A)-binding proteins constrain decapping activity (Tucker and Parker, 2000; Ramirez *et al.*, 2002). Also, decapping is actively regulated. For instance, the loss of the poly(A) binding protein Pab1 cause decapping prior to deadenylation, suggesting a role of Pab1 in decapping regulation. Also, decapping can be promoted through the activities of complexes that lead to the inhibition of translation. One of example is Dhh1,

a DEAD-box helicase member of the DEAD family of ATPases. In fact, Dhh1 interacts with subunits of the translation initiation complex and repress translation, ultimately leading to the dissociation of the mRNA from polysomes and accumulation in P-bodies, where they can be stored or degraded. Experimental evidence show that Dhh1 inhibits the formation of 48S initiation complex *in vitro* and that it enhances decapping in deadenylated mRNA (Coller and Parker, 2005; Swisher and Parker, 2010; Carroll *et al.*, 2011). Once decapped, the mRNAs can be targeted for degradation by Xrn1.

3.2.3. 5'-3' mRNA degradation by Xrn1

Transcripts are degraded by Xrn1 in a 5'-3' direction. In fact, it has been shown that Xrn1 nuclease prefers mRNA substrates with a free 5'-monophosphate end, a product formed after decapping (Stevens, 2001). Xrn1 is present in the cytoplasm, specifically at P-bodies. Due to is localization, it is known that Xrn1 interacts with Dcp1/Dcp2, Lsm proteins, Pat1 and Dhh1 (Parker and Sheth, 2007). Interestingly, this exoribonuclease has been implicated in nuclear activities, namely regulation of antisense non-coding RNAs and Xrn1-dependent transcription coupling with degradation (discussed below) (van Dijk *et al.*, 2011; Haimovich *et al.*, 2013b; Sun *et al.*, 2013).

3.2.4. 3'-5' mRNA degradation by the Exosome

The second pathway of RNA degradation following deadenylation is the 3' to 5' degradation by the exosome. The exosome is a multisubunit complex composed of 10 proteins, including six members of the RNase PH protein family and three small RNA-binding proteins. Together, these 9 subunits form the core of the complex, a ring-like structure analogous to the bacterial PNPase (Allmang *et al.*, 1999; Liu *et al.*, 2006). In this structure we can also find a tenth protein, the catalytic subunit Dis3. Dis3 binds to the exosome core in the cytoplasm and degrades deadenylated mRNAs in 3'-5' direction (Lebreton *et al.*, 2008; Schaeffer *et al.*, 2009). For the cytoplasmic degradation of mRNAs, the exosome also relies on cofactors, such as Ski proteins (Anderson and Parker, 1998). Interestingly, the exosome can also be present in the nucleus. There it associates with Rrp6 (another 3'-5' exonuclease), Rrp47 and Mpp6 and is involved in several processes, such as nuclear RNA processing and degradation (Mitchell and Tollervey, 2003; Milligan *et al.*,

2008; Synowsky *et al.*, 2009). Once degraded by the Exosome, the remaining cap is degraded by the decapping enzyme Dcs1 (Braun and Young, 2014).

3.3. From the cradle to the grave: mechanisms of coupling of mRNA synthesis and decay

The first question that might pop into one's mind is: how can two processes completely separated in space be mechanistically linked? In recent years, several studies performed in budding yeast have suggested several mechanisms through which nuclear transcription is coupled to cytoplasmic stability or degradation. These processes can be mediated through both *cis*-acting sequences and *trans*-acting elements.

3.3.1. Coupling through RNA Pol II core proteins

Two different subunits of the core RNA Pol II – Rpb4 and Rpb7 – have been implicated in mRNA imprinting and coupling of mRNA transcription and degradation. The first indications that these two subunits were involved in functionally linking transcription and decay were two-fold: Rpb4 and Rpb7 were reported to form a heterodimer independent of RNA Pol II and the inclusion of these two subunits into RNA Pol II appeared sub-stoichiometric (Kolodziej *et al.*, 1990). Additionally, the heterodimer formed by Rpb4/Rpb7 was found to easily dissociate from RNA Pol II (Edwards *et al.*, 1991; Orlicky *et al.*, 2001). Also, it was observed that this heterodimer was involved in the recruitment of factors involved in 3'-end processing and proper usage of polyadenylation sites (Runner *et al.*, 2008), mRNA export (Farago *et al.*, 2003) and, more importantly for this section, mRNA translation/decay. Furthermore, Rpb4/7 heterodimer has been reported to shuttle between the nucleus and the cytoplasm in a transcription-dependent fashion (Selitrennik *et al.*, 2006) and its cytoplasmic post-transcriptional activities are dependent on prior association with RNA Pol II (Goler-Baron *et al.*, 2008). In fact, two distinct studies have delivered evidence that, while elongation occurs, the extended RNA exiting RNA Pol II encounters Rpb7 and forms contacts with it, thus imprinting the mRNA (Ujvari and Luse, 2006; Chen *et al.*, 2009). Interestingly, while shuttling of Rpb4 and Rpb7 depend on each other (Selitrennik *et al.*, 2006), Rpb4 has been reported as having a role in the regulation of degradation of a specific set of transcripts, Rbp7 was shown to have a much broader role on this process (Lotan *et al.*, 2005; Lotan *et al.*, 2007).

While the mechanistic insight through which Rpb4/Rpb7 regulate mRNA decay has yet to be unveiled, some speculation can be made based on several reports (Lotan *et al.*, 2005; Lotan *et al.*, 2007). First, both subunits affect the deadenylation process of transcripts whose decay is regulated by the heterodimer. Second, both Rpb4 and Rpb7 were found to interact directly with mRNA decapping complex Pat1-Lsm1-7. Also, Rpb4/Rbp7 heterodimer was found to co-localize with Lsm1-7 and Dcp1, in P-bodies. Finally, mutations on either Rpb4 or Rbp7 led to an alteration of the number of P-bodies within the cell. While this might be an attractive, yet provocative, explanation, some loose ends exist. For instance, Rpb4/Rpb7 interacts with Pat1, a platform for several decay factors and involved in the recruitment of Lasm1-7, Dcp1/2, and Xrn1 to the transcript (Tharun and Parker, 2001; Nissan *et al.*, 2010). However, none of these explains the requirement for deadenylation, since the heterodimer does not interact with Ccr4-Not complex or Pan2-3, nor Pat1 is linked with shortening of poly(A) tail (Hatfield *et al.*, 1996). Interestingly, a recent report by the group of Patrick Cramer has shown that, instead, Rpb4 is a *bona fide* RNA Pol II subunit with functions in transcription, arguing against the previous results demonstrating direct function of Rpb4 in mRNA degradation. Additionally, the cytoplasmic localization of Rpb4 was not observed (Schulz *et al.*, 2014).

3.3.2. Coupling transcription and decay via gene promoters and UAS

The first indication that promoters could efficiently impact on mRNA stability was actually performed in a mammalian system (Enssle *et al.*, 1993). Here, they have shown that a gene containing a premature translation-termination codon would result in lower levels of this transcript in comparison to the wild-type one. One could hypothesize that the mutated mRNA was being vastly degraded only due to changes in splicing. Also, they noted that there was no changes on the TSS utilized and the 5'-UTR. However, they observed that swapping of the endogenous promoter of the β-globin for a viral one would ultimately result in the rescue of the mRNA levels, indicating a diminished degradation of the transcript when under the control of the viral promoter (Enssle *et al.*, 1993).

In yeast, Bregman and colleagues have found that the presence of Rap1 binding sites in gene regulatory regions affects the stability of the corresponding transcripts (Bregman *et al.*, 2011). These Rap1 binding sites are short sequences found in the UASs in several genes allowing the recruitment of the Rap1 activator. Specifically, they have shown that either mutation of endogenous Rap1 sites or swapping of the endogenous promoter for that of the

ACT1 gene, that does not contain a Rap1 binding site, would lead to stabilization of the RNA. However, if they used a version of the *ACT1* promoter containing a Rap1 binding site, the destabilization would be restored. Hence, the binding of the transcription activator Rap1 is not only essential to regulate the stability *RPL30* mRNA, where it was initially investigated, but also Rap1 target genes (Bregman *et al.*, 2011). Nevertheless, the mechanism through which Rap1 'marks' *RPL30* mRNA is still unknown.

Later on, another study further supported the idea that regulatory sequences have an impact in mRNA stabilization, showing that exchanging of the upstream *cis*-regulatory sequence of two orthologous genes in two related yeast species would lead to changes in RNA stabilization (Dori-Bachash *et al.*, 2012). Also, the laboratory of Robert H. Singer has shown that two mitotically regulated transcripts, products of *SWI5* and *CLB2*, also displayed promoter-dependent mRNA stabilization (Trcek *et al.*, 2011). There it was reported that Dbf2, a mitotic kinase, which interacts with the Ccr4-Not complex, is the coordinator of this process. However, how Dbf2 is recruited to promoters and exerts its activity remains elusive (Trcek *et al.*, 2011).

Another paradigm of promoter-related mRNA stabilization is what is observed in the *GAL* system in *S. cerevisiae*. The addition of galactose to a culture growing in a non-fermentable carbon source induces a transcriptional response that permits galactose metabolism. On the other hand, readdition of glucose causes a shutdown of this transcriptional response with simultaneous degradation of the mRNA produced from the induced genes (Lombardo *et al.*, 1992; Cereghino *et al.*, 1995; Cereghino and Scheffler, 1996; Prieto *et al.*, 2000). Nevertheless, if the endogenous promoter of *GAL7* is swapped with the promoter of the constitutively expressed gene *ADH1*, this glucose-induced decay does not occur (Munchel *et al.*, 2011). While one could predict that this change of stability would be due to the absence of carbon source transition, another possible explanation could be that enhanced stability of *GAL7* transcripts under the influence of the *ADH1* promoter is due to the promoter, since it contains a Rap1p binding site. As previously mentioned, Rap1p could then stabilize those mRNAs (Bregman *et al.*, 2011). Altogether, the *cis*-regulatory elements and the associated transcription factor would be responsible for the stabilization and not the nutritional conditions *per se*.

3.3.3. Coupling promoted by mRNA decay factors

Besides the previously mentioned mechanisms, three components of the mRNA decay machinery have been implicated in the interplay between transcription and degradation: Ccr4-Not complex, Dcp2, and Xrn1. While for the first two the mechanisms behind it are still unknown, for the latter one two distinct models were proposed.

As previously mentioned, Xrn1 is the major cytoplasmic 5'-3' exoribonuclease. Recently, two different studies have provided explanations for Xrn1 implication in the functional coupling of transcription and degradation. The first study by Haimovich *et al.* attributed a direct stimulatory role on transcription by Xrn1, immediately reflected on transcription (Haimovich *et al.*, 2013b) (**Figure 31, *left***). Here, they have shown that strains lacking either full-length Xrn1 or harboring mutated versions of the gene had impaired RNA Pol II activity. Unexpectedly, they have also found that Xrn1, together with other degradation factors, was present upstream of the TSS of the genes most affected by the deletion of *XRN1*. Also, cytoplasmic shuttling of those degradation factors (Lsm1 and Dcp2) were dependent on Xrn1. Altogether, they proposed that "synthetic and decay processes represent two arms of a larger machinery, the synthegradosome" (Haimovich *et al.*, 2013b).

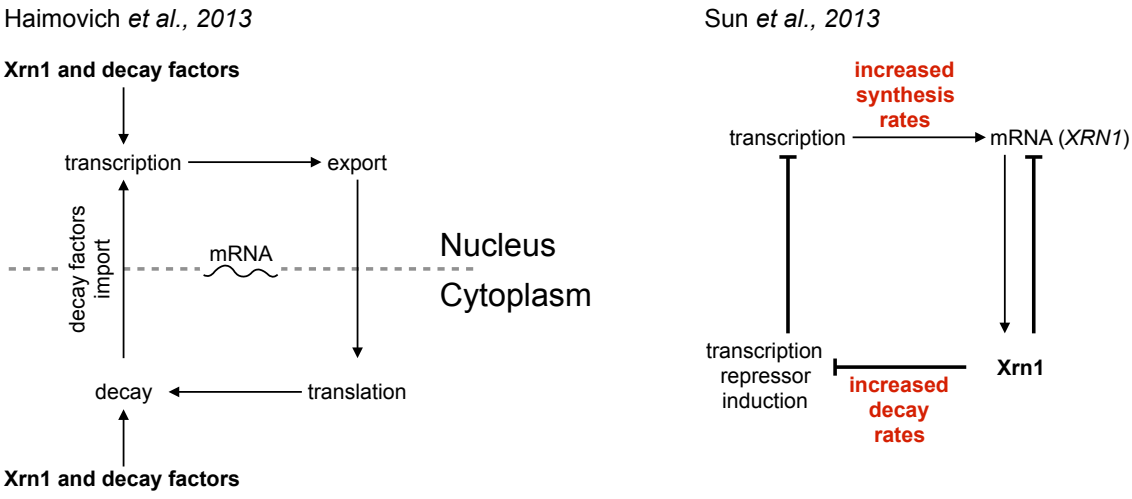


Figure 31 – Circular regulation of transcription and mRNA decay by Xrn1. (*Left*) The model of Haimovich *et al*. Here, the authors link transcription and decay together by the mRNA decay factors that shuttle between the cytoplasm and nucleus to regulate both processes. **(*Right*)** The model of Sun *et al.* The mRNA levels are controlled by feedback regulation of *XRN1* mRNA levels. Xrn1 protein level is maintained by translation and degradation of *XRN1* mRNA. The global synthesis rates are controlled by Xrn1-dependent transcription repressor induction.

Conversely, Sun and colleagues described, instead, a negative effect of Xrn1 on transcription rates (Sun *et al.*, 2013). In a situation where *XRN1* was deleted, it was possible to observe that global decrease of decay rates was accompanied by an increased synthesis rates, leading to an increase of around 3.2-fold of steady-state levels of RNA. Strikingly, among all the 46 mutants (targeting factors known to be involved in the control of mRNA stability) analyzed, this imbalance was only observed in that lacking Xrn1. Overall, the authors propose the following model: when Xrn1 mRNA levels augment, levels of Xrn1 protein also increase, leading to a subsequent decrease of its mRNA levels; however, in a situation where Xrn1 mRNA levels decrease, Xrn1 protein levels equally decrease, leading to mRNA stabilization and thus an increase of the mRNA level. Considering that Xrn1 acts globally, this simple feedback loop can control all mRNA levels. Additionally, they attribute the negative transcriptional role of Xrn1 to the indirect action of the transcriptional repressor Nrg1, whose mRNA levels are *XRN1*-dependent (Sun *et al.*, 2013), although one could imagine that several transcription factors could be affected (**Figure 31, *right***).

Altogether, while both studies disagree on the mechanism behind coupled transcription and degradation by Xrn1, they both agree that mRNA levels can be buffered from changes on either transcription or degradation by compensatory mechanisms.

3.3.4. Benefits of functionally coupling mRNA transcription and decay

The functional coupling of transcription and degradation plays an essential role in the coordination of gene expression patterns, allowing a fast fluctuation in the levels of hundreds to thousands of genes in a very short period of time. This allows an efficient response to several physiological and environmental cues: cell cycle, cell proliferation, cellular response to stress, evolution and much more (Haimovich *et al.*, 2013a). One example of this circumstance is the regulation of both specific and global steady-state levels of RNA during osmotic stress in budding yeast: mild osmotic stress leads to rapid up-regulation of transcriptional activity and a simultaneous decrease in degradation of transcripts associated with osmotic response, thus resulting in their stabilization (Romero-Santacreu *et al.*, 2009). Something similar is observed during heat-shock or DNA damage, where induced genes are stabilized with concomitant degradation of repressed transcripts (Shalem *et al.*, 2008; Castells-Roca *et al.*, 2011). Hence, one of the great advantages of coupling mRNA synthesis and decay is indeed to trigger a fast response to a new scenario (stress, for instance), allowing cells to cope with this new situation.

Additionally, besides promoting a mechanism of adaptation for a stressful environment, coupling of synthesis and decay might provide a mechanism to preserve an adequate dosage of the global transcriptome under a specific physiological context. This idea is supported by several studies mentioned above, where global changes on transcription and degradation due to genetic mutation/deletion promotes overall levels of mRNA to be maintained (Helenius *et al.*, 2011; Sun *et al.*, 2012; Sun *et al.*, 2013; Rodriguez-Molina *et al.*, 2016). Importantly, this reveals a mutual feedback mechanism where the interplay between two antagonistic processes – synthesis and decay – is necessary to maintain global gene dosage in different strains of budding yeast.

Interestingly, this interplay between mRNA synthesis and decay is also responsible to modulate steady-state levels of RNA during cell-growth, to compensate for changes in cell size/volume (Garcia-Martinez *et al.*, 2016). In fact, this process is not only important for maintenance of RNA concentration in one strain during its life-cycle. For example, let's observe what happens in *S. cerevisiae* and the fission yeast *S. pombe*:

- The median mRNA synthesis rates are virtually the same between the two species;
- The average half-lives of mRNAs in *S. pombe* are roughly three times longer than those of *S. cerevisiae* (59 minutes *versus* 12 minutes);
- The overall RNA levels are three times higher in *S. pombe* than in *S. cerevisiae*;
- *S. pombe* cells are approximately three times bigger than *S. cerevisiae* (median volume of 115 mm^3 *versus* 42 mm^3);
- mRNA concentration per cell between the two different species is very similar.

Altogether, these evidence indicate that *S. pombe* cells generally contain more stable mRNAs than *S. cerevisiae* cells to reach similar mRNA concentrations at similar mRNA synthesis rates, despite their larger volume. Hence, in this context, we have highlighted the fact that mRNA synthesis and decay cooperate to maintain universal and equivalent concentrations of RNAs between different species (Sun *et al.*, 2012).

Another interesting advantage of coupling mRNA synthesis and decay is that it might facilitate evolution (Haimovich *et al.*, 2013a; Das *et al.*, 2017). Specifically, the existing coupling of mRNAs synthesis and decay reduces the number of mutations necessary for an organism to respond to new environments, facilitating evolvability. In other words, if the processes were not mechanistically linked, each individual process – synthesis or decay – would require specific and separate sets of mutations, which, by itself, would decrease the efficiency of the evolutive process. On the other hand, coordination of these two processes

supports a more plastic transcriptome and proteome, promoting a better selective advantage and allowing the organism to answer positively to evolutionary cues (Dahan *et al.*, 2011; Haimovich *et al.*, 2013a; Haimovich *et al.*, 2013b; Das *et al.*, 2017).

3.4. Readdressing the role of SAGA in transcription

As seen in the previous chapter, the information regarding the functional role of the SAGA complex is far from being in agreement. Importantly, most of the analyzed data relied on either ChIP of SAGA subunits or transcriptomic analyses of steady-state levels of RNA upon deletion of SAGA subunits. Considering that crosslinked ChIP of the coactivator might underestimate the SAGA binding events and that steady-state RNA does not really reflect the transcription state of the genes, we proposed to approach the functional study of the yeast SAGA complex using different methods from those used in the past:

- ChIP on histone marks to understand the dynamics of those marks (H3K9ac and H2Bub) upon loss of the SAGA complex;
- Chromatin endogenous cleavage followed by high-throughput sequencing (ChEC-seq) to analyze the direct recruitment of complex to chromatin;
- Analysis of nascent transcription upon deletion of several SAGA subunits to disclose the gene expression profile promoted by the loss of activity of the complex.

Here we can find a brief overview of comparative dynamic transcriptome analysis (cDTA) to analyze nascent transcription and ChEC-seq to map SAGA binding site avoiding crosslinking and antibody usage. Notably, the cumulative results obtained through these different approaches were not exclusively collected by me. Instead, these results were also gathered by past members from our laboratory, namely Jacques Bonnet and Chen-Yi Wang, and by our collaborator Sebastian Grünberg, from Steven Hahn's group. Nevertheless, I had the opportunity to make sense of these collected results and could interpret these as a whole providing a more coherent appreciation of SAGA function in RNA Pol II transcription. When necessary, a mention to the person that performed the experimental analysis will be done (see the *Results* section of the manuscript).

3.4.1. Comparative dynamic transcriptome analysis (cDTA)

As stated above, steady-state levels of mRNA are the reflex of the equilibrium between synthesis of a given RNA and its degradation. Since global changes on transcription

can lead to a compensatory mechanism that ultimately terminates in the buffering of total levels of transcripts, one has to use tools that specifically allow the analysis of the synthesis of mRNA uncoupled from degradation (**Figure 32**). Like that, it is possible to correctly ascertain the role of a given transcription-associated factor. One way to achieve that purpose is through the use of comparative dynamic transcriptome analysis (cDTA).

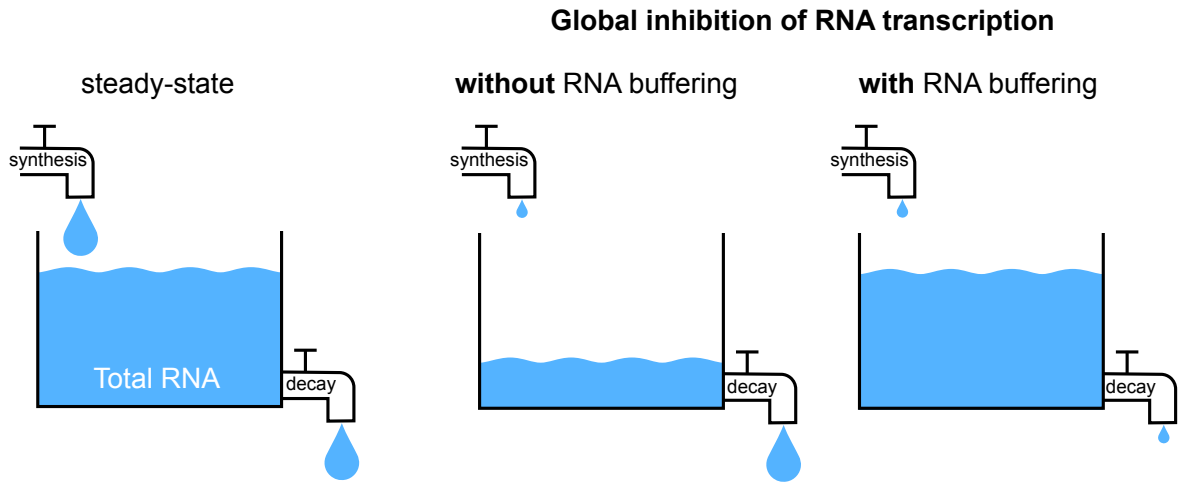


Figure 32 – Mechanisms of mRNA buffering. (*Left*) Steady-state levels of mRNA are dictated by a tight equilibrium between mRNA synthesis and decay. In a situation where transcription is globally affected (by deletion of the SAGA complex, for example), one of two scenarios can be verified: on one hand, **(*Center*)** decreased synthesis rate is simultaneous to unchanged decays rates, thus resulting in global decrease of total mRNA (no buffering); on the other hand, **(*Right*)** decrease in synthesis rate is simultaneously compensated by similar or partial decrease in decay rates, thus resulting in unchanged mRNA levels (with buffering).

cDTA is based on the specific labeling of nascent/newly-transcribed RNAs, followed by its purification and analysis (**Figure 33**). Briefly, the nucleoside analog 4-thiouridine (4sU) is taken up by eukaryotic cells, phosphorylated by uridine kinases and specifically incorporated into RNA during RNA Pol II transcription (Melvin *et al.*, 1978). Afterwards, the thiol-labeled newly-synthesized transcripts can be biotinylated and specifically purified using magnetic beads coated with streptavidin (Cleary *et al.*, 2005; Dolken *et al.*, 2008; Miller *et al.*, 2011; Sun *et al.*, 2012). Importantly, this variation of the method using 4sU can only be readily used metazoan and plant cells, but not in budding or fission yeast. This is due to the fact that 4sU can only be uptaken by cells expressing a nucleoside transporter, like the human equilibrative nucleoside transporter (hENT1). In fact, if hENT1 is expressed in *S. pombe* or *S. cerevisiae*, 4sU can be used to label nascent RNA, allowing the efficient separation of the newly-transcribed RNA from the pool of total RNAs (Miller *et al.*, 2011). Nevertheless, an

easier approach can be used to perform cDTA analyses in yeast: instead of using 4sU, one can use the modified base 4-thiouracil (4tU). In fact, both *S. pombe* and *S. cerevisiae* cells take up 4tU, without needing the expression of a nucleoside transporter (Miller *et al.*, 2011; Munchel *et al.*, 2011; Sun *et al.*, 2012). In fact, metabolism of 4-tU requires the activity of the enzyme uracil phosphoribosyltransferase (UPRT). In several organisms, UPRT is essential for a pyrimidine salvage pathway, recycling uracil to uridine monophosphate using phosphoribosyl pyrophosphate. Nevertheless, this salvage pathway is not present in mammalian or insect cells, but it is present in yeast (Friedel and Dolken, 2009). Importantly, the metabolic labeling using 4tU does not affect normal cell physiology (Miller *et al.*, 2011).

One of the major disadvantages in conventional transcriptomics studies is the unknown normalization factor between different samples. For instance, in a situation where we want to compare the transcriptome of mutant strains to that of a wild-type sample, a plethora of deviating factors can be introduced: efficiency of cell lysis, differences in the extraction and recovery of RNA, variances in scanner calibration when collecting information from microarrays, among others. Taking everything into consideration, it is mandatory to have a pipeline that allows the estimation of "normalization factors". Indeed, using cDTA this estimation is possible simply by an internal normalization standard. For that, prior to RNA extraction, a defined ratio (1:3) of the distantly related fission yeast *S. pombe* is added to the *S. cerevisiae* samples, either wild-type or mutant cells (Sun *et al.*, 2012). Importantly, it is imperative that the *S. pombe* cells grown and labeled have the same origin, so, usually, a big culture of fission yeast is made and stocked. Subsequently, the mixture of cells is lysed, total RNA extracted, labeled RNA purified and the pool of RNAs (both steady-state/total and newly-synthesized RNAs from *S. pombe* and *S. cerevisiae*) are either quantified using microarray chips containing probes for the whole transcriptome of both *S. pombe* and *S. cerevisiae* (Affymetrix® GeneChip® Yeast Genome 2.0 Array) or sequenced (Sun *et al.*, 2012).

Altogether, cDTA delivers a quantification of absolute rates of mRNA synthesis and decay in budding yeast, using *S. pombe* as the internal standard, or *vice-versa*, providing a tool for direct comparison of RNA synthesis and decay rates among distinct samples.

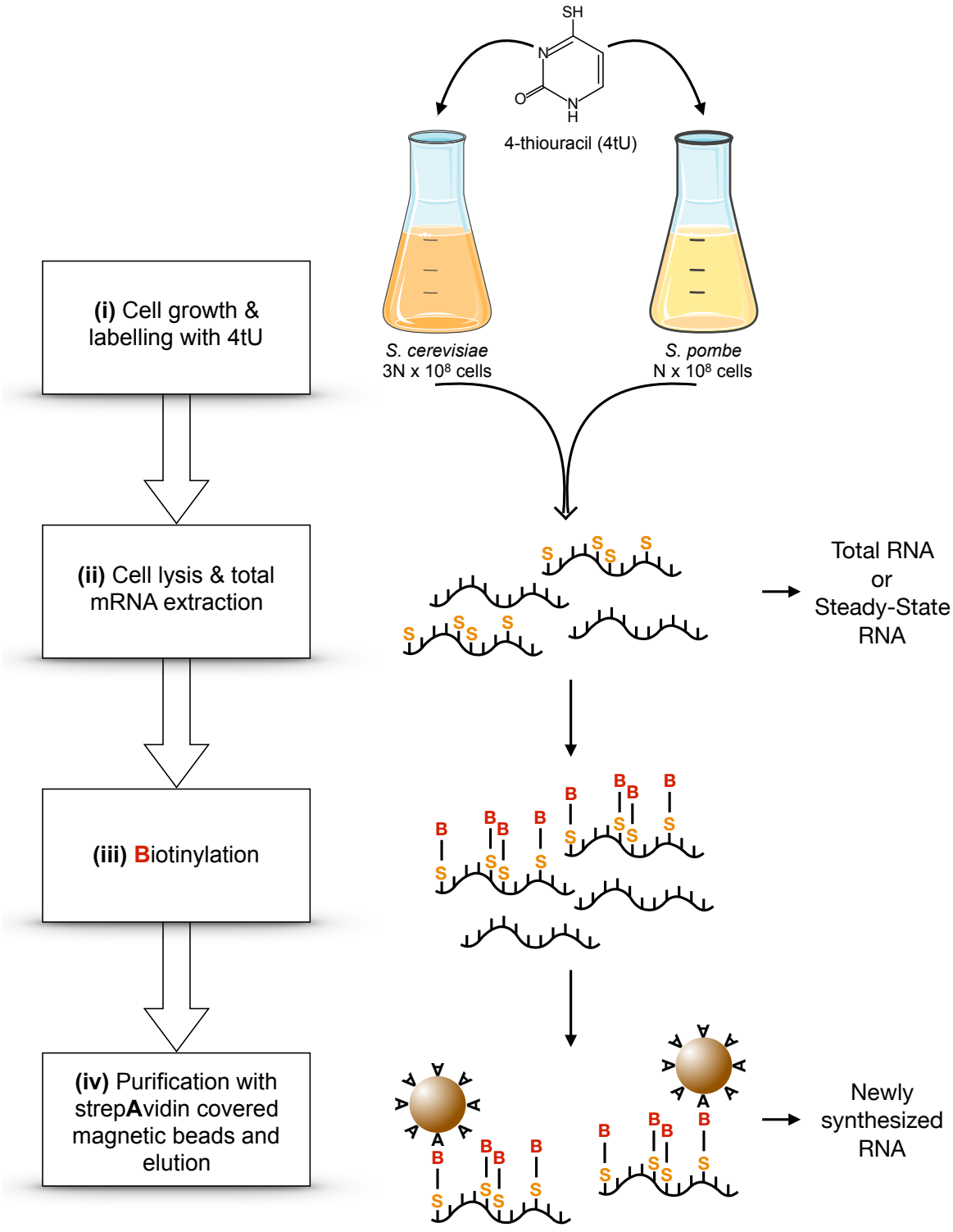


Figure 33 – Design of a comparative dynamic transcriptome analysis (cDTA) experiment. The *S. cerevisiae* and *S. pombe* cells are labeled by addition of 4-tU into the media. The cells are then counted and mixed in a proportion of 3:1 (*S. cerevisiae*:*S. pombe*). Cells are then lysed, RNA is extracted, biotinylated, and labeled RNA separated. Microarrays containing probes against both *S. cerevisiae* and *S. pombe* transcripts are used to quantify both total and labeled RNA.

3.4.2. Chromatin endogenous cleavage followed by high-throughput sequencing (ChEC-seq)

Defining the genome-wide binding profile of a protein to chromatin is essential for the understanding of the role and impact of that factor in DNA-related cellular processes. While the most widely used ChIP-seq approach has been responsible to gain tremendous insights regarding DNA-related processes, it possesses several limiting steps. First, crosslinking by formaldehyde (Zentner and Henikoff, 2014), which preferentially leads to protein-protein crosslinks, might mask epitopes, thus decreasing the efficiency to capture an interaction. Also, transient interacts might be excluded due to crosslinking of chromatin. Not only that, but hyperChIPability of some genomic *loci* might also lead to erroneous conclusions (Park *et al.*, 2013; Teytelman *et al.*, 2013). Second, one of the most rate- and resolution-limiting steps of ChIP is chromatin shearing/sonication (Zentner and Henikoff, 2014). Nevertheless, the usage of nucleases might allow to partially overcome this issue (Rhee and Pugh, 2011; He *et al.*, 2015). Third, one can perform ChIP in the absence of prior crosslinking. While this might be an extraordinary option for the investigation of histone marks, for instance, the low solubility of chromatin-associated factors in the gentle to mild extraction conditions might be problematic (Kasinathan *et al.*, 2014; Skene *et al.*, 2014). Fourth, the dependence on antibodies not only makes ChIP more expensive, but it also makes it difficult to compare different datasets, due to different antibody efficiency. Some techniques to replace ChIP have been developed, like DNA adenine methyltransferase identification (DamID), where the target protein is fused to a bacterial adenine methyltransferase and then the recognition of methylated adenines in GATC sequences is made. Importantly, DamID allows the *in vivo* study of protein-chromatin interaction, without crosslinking of immunoprecipitation (van Steensel and Henikoff, 2000; Vogel *et al.*, 2007; Filion *et al.*, 2010). However, the main problem of DamID is the extremely low resolution (resolution is limited to kilobase-sized regions) and the constitutive and permanent activity of the methyltransferase, which might lead to increased background.

A method that can circumvent all the abovementioned problems is ChEC/ChEC-seq (Schmid *et al.*, 2004; Schmid *et al.*, 2006; Zentner *et al.*, 2015; Grunberg *et al.*, 2016; Grunberg and Zentner, 2017). In ChEC-seq, the protein of interest is fused with the MNase, which is able to degrade unprotected DNA in the presence of calcium (**Figure 34**).

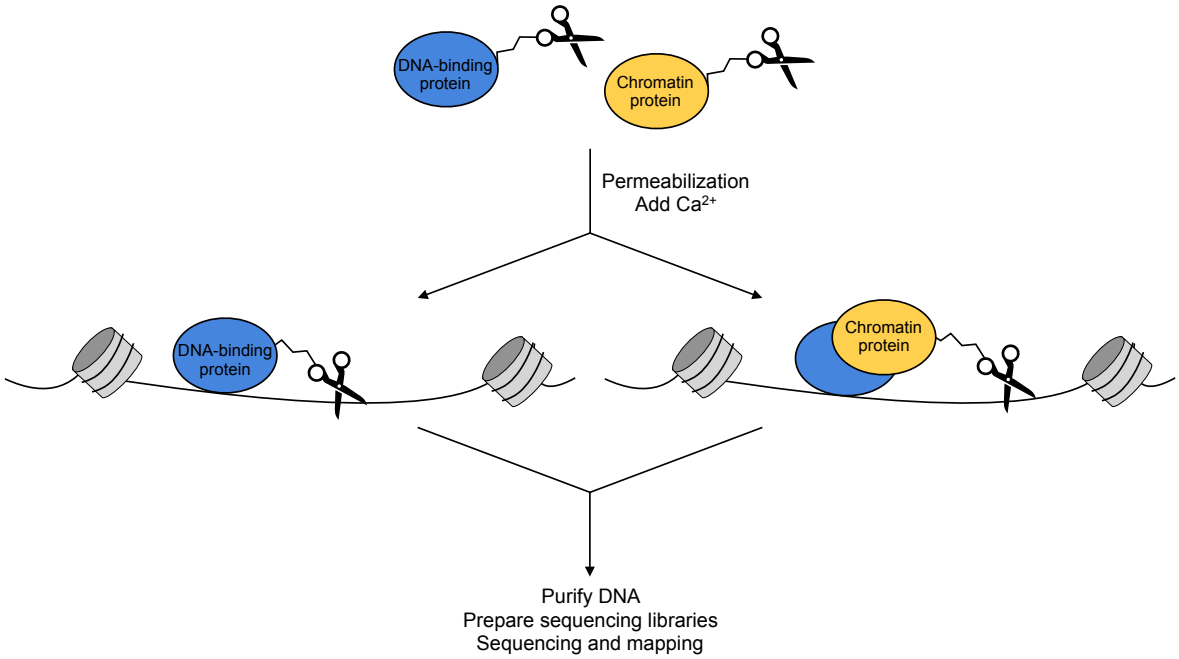


Figure 34 – ChEC-seq workflow. Yeast cells expressing the DNA-binding protein or chromatin protein of interest genetically fused to MNase are permeabilized with digitonin and calcium is added to induce cleavage by the fusion protein. Digested DNA is then purified and prepared for high-throughput sequencing.

While some of the advantages of ChEC are the same as DamID, it also allows the acquisition of higher resolution (close to base-pair resolution). Not only that, but the MNase activity is controllable, since it requires induction through the addition of calcium (Zentner *et al.*, 2015; Grunberg *et al.*, 2016; Grunberg and Zentner, 2017). In fact, MNase is only active in the presence of millimolar concentrations of calcium, which by far exceeds the concentrations found intracellularly (Dunn *et al.*, 1994; Grienberger and Konnerth, 2012).

AIMS

Aims

The SAGA co-activator complex exhibits two distinct and functionally independent catalytic activities as histone acetyltransferase and deubiquitinase, promoted by Gcn5 and Ubp8, respectively. Moreover, SAGA also comprises TBP-binding subunits (Spt3 and Spt8) and architecture module (Spt7, Spt20 among others). Early genome-wide studies in yeast differentiated the so-called SAGA-dominated genes, from the TFIID-dominated genes. Importantly, SAGA-dominated genes represent merely 10% of the yeast genome. Nevertheless, recent studies in *S. cerevisiae* revealed that only a minor fraction of SAGA-bound genes is deregulated upon deletion of SAGA subunits, and, conversely, only a minuscule fraction of misregulated genes is bound by the corresponding subunits, suggesting a complete disconnection between complex binding and transcriptional effect. Additionally, several results indicated that the SAGA coactivator has a broader role on transcription than previously anticipated.

In line with that, the aims of my PhD thesis were:

(a) To comprehend the extent of recruitment of the SAGA complex to chromatin;

(b) To measure accurately the role of SAGA in global RNA Pol II transcription in budding yeast;

(c) To solve the conundrum of the results obtained between steady-state *versus* newly-synthesized mRNA, through the measurement of changes in mRNA synthesis and decay upon loss of SAGA subunits;

(d) To disclose which of the different activities of the SAGA complex are implicated in transcriptional regulation of RNA Pol II-transcribed genes;

(e) To understand the respective contributions of SAGA and TFIID in RNA Pol II transcription and how these two complexes cooperate.

RESULTS

Results

1. The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA Polymerase II transcription. (J Bonnet*, CY Wang*, T Baptista *et al.*; Genes & Development, 2014)

Results regarding the role of SAGA were never conclusive and some discrepancies were detected. Nevertheless, the overall results indicated that, in yeast, only 10% of the genome would be regulated by SAGA. Also in mammalian cells, SAGA seemed to have a very limited role on RNA Pol II transcription. More impressively, histone marks either deposited (H3K9ac) or removed (H2Bub) by SAGA are scattered around the transcribed genome and bulk changes were observed when SAGA-specific subunits were deleted. Hence, since these last results indicated that SAGA potentially has a broader impact on transcription, we decided to tackle this question. But this time, instead of ChIPing directly the complex as it has been done in the past, we opted to ChIP the histone marks associated with the coactivator, either in the presence or absence of the subunits responsible for depositing or removing them. Notably, this approach was performed both in yeast and human cells.

Importantly, this genome-wide analyses allowed to disclose two extremely important conclusions:

- The HAT activity of the SAGA complex is active genome-wide, being capable of acetylating histone H3 on lysine 9 in the promoters of all actively transcribed genes.
- The mDUB of SAGA is responsible for the removal of ubiquitin from histone H2B on the transcribed regions of expressed genes.

The reported SAGA DUB activity on the gene bodies, while acetylation occurred at promoters, made us wonder about the mechanisms of SAGA recruitment and activity on gene-bodies. Our results clearly demonstrated that removal of ubiquitin from H2B is uniquely carried out by the SAGA DUB. In addition, SAGA can act in the transcribed region of active genes independently of elongating Pol II. Importantly, our observations reveal an extremely transient and dynamic association of SAGA with active chromatin, as H2Bub can be erased genome-wide by the SAGA DUB activity in a few minutes. This fast and dynamic interaction

with chromatin might justify why past studies might have underestimated the genome-wide binding of the SAGA complex.

Considering that the SAGA coactivator is enzymatically active throughout the whole transcribed genome, we wondered what was the impact on transcription *per se*. Hence, to partially address **aims (b)** and **(c)**, we followed two different strategies:

(i) Chen-Yi Wang performed RNA Pol II ChIP-seq in wild-type strain and mutant strains lacking structural subunits of the SAGA complex (*spt7*Δ and *spt20*Δ);

(ii) I established the protocol for the analysis of newly-transcribed mRNA and addressed whether nascent transcription in the mutants mentioned above was affected in a given number of genes, independent of their requirement for either SAGA or TFIID.

Remarkably, we have observed that deletion of Spt7 or Spt20 led to a dramatical impairment of RNA Pol II recruitment to expressed genes. Similarly, through analyses of newly-transcribed mRNA, I was able to disclose that nascent transcription was severely decreased for all tested genes, by an average factor of four-fold. Additionally, the observed effects were evident for both SAGA- and TFIID-dominated genes.

Our results reveal a highly dynamic and transient association of SAGA with the entire active chromatin that is required for Pol II recruitment and transcription of both SAGA- and TFIID-dominated genes.

These results were on the published on the 15th of August 2014, in *Genes & Development*.

Author's contributions

Data referring to the genome-wide activity of the SAGA complex were produced by Jacques Bonnet (using human cells) and Chen-Yi Wand (using yeast cells). I was fully responsible for the execution of the experiments presented in the manuscript (Figures 2 and 7 and Supplementary Figure 3). Hence, I established (and optimized) the protocol for the analysis of newly-transcribed mRNA and derived mESCs and performed all the experiments related to those cells. I was also in charge of analyzing and interpreting the data acquired from my experiments. Additionally, other authors and I were involved in the design and conceiving of this work. In the end, and once all the data was collected, the manuscript was written with the input of all the authors. Specifically:

Jacques Bonnet – Designed the study. Analyzed the data. Assisted on the writing of the manuscript. Performed all experiments in mammalian cells, except for experiments in mouse embryonic stem cells;

Chen-Yi Wang - Designed the study. Analyzed the data. Assisted on the writing of the manuscript. Performed all experiments in yeast cells, except for analysis on newly-synthesized RNA;

Tiago Baptista – Participated in the designed the study. Analyzed the data. Assisted on the writing of the manuscript. Establish the mouse embryonic stem cells (mESCs) and performed the experiment with it. Established and performed analysis of newly-transcribed RNA in the yeast strains indicated;

László Tora – Conceived the study. Wrote the manuscript.

Didier Devys – Conceived the study. Analyzed the data. Wrote the manuscript.

All the other authors contributed with reagents, expertise and assistance when necessary.

The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription

Jacques Bonnet, Chen-Yi Wang, Tiago Baptista, et al.

Genes Dev. 2014 28: 1999-2012
Access the most recent version at doi:10.1101/gad.250225.114

The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription

Jacques Bonnet,[1,2,3,4,6] Chen-Yi Wang,[1,5,6] Tiago Baptista,[1,2,3,4] Stéphane D. Vincent,[1,2,3,4] Wei-Chun Hsiao,[5] Matthieu Stierle,[1,2,3,4] Cheng-Fu Kao,[5] László Tora,[1,2,3,4] and Didier Devys[1,2,3,4]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France; [2]UMR7104, Centre National de la Recherche Scientifique, 67404 Illkirch, France; [3]U964, Institut National de la Santé et de la Recherche Médicale, 67404 Illkirch, France; [4]Université de Strasbourg, 67404 Illkirch, Cedex, France; [5]Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taiwan

The SAGA (Spt–Ada–Gcn5 acetyltransferase) coactivator complex contains distinct chromatin-modifying activities and is recruited by DNA-bound activators to regulate the expression of a subset of genes. Surprisingly, recent studies revealed little overlap between genome-wide SAGA-binding profiles and changes in gene expression upon depletion of subunits of the complex. As indicators of SAGA recruitment on chromatin, we monitored in yeast and human cells the genome-wide distribution of histone H3K9 acetylation and H2B ubiquitination, which are respectively deposited or removed by SAGA. Changes in these modifications after inactivation of the corresponding enzyme revealed that SAGA acetylates the promoters and deubiquitinates the transcribed region of all expressed genes. In agreement with this broad distribution, we show that SAGA plays a critical role for RNA polymerase II recruitment at all expressed genes. In addition, through quantification of newly synthesized RNA, we demonstrated that SAGA inactivation induced a strong decrease of mRNA synthesis at all tested genes. Analysis of the SAGA deubiquitination activity further revealed that SAGA acts on the whole transcribed genome in a very fast manner, indicating a highly dynamic association of the complex with chromatin. Thus, our study uncovers a new function for SAGA as a bone fide cofactor for all RNA polymerase II transcription.

[*Keywords*: SAGA; chromatin; transcription; deubiquitinase; acetyltransferase; RNA polymerase II]

Supplemental material is available for this article.

Received January 21, 2014; revised version accepted August 15, 2014.

Through their different activities, coactivator complexes convey the function of transcription activators bound to DNA elements to drive specific gene expression profiles. SAGA (Spt–Ada–Gcn5 acetyltransferase) is an evolutionarily conserved multifunctional coactivator complex organized in functional and structural modules (for review, see Rodríguez-Navarro 2009; Koutelou et al. 2010; Weake and Workman 2012). A number of biochemical and structural studies of SAGA architecture in *Saccharomyces cerevisiae* (y) consistently identified histone acetyltransferase (HAT), deubiquitinase (DUB), and TATA-binding protein (TBP) regulatory and structural modules (Wu et al. 2004; Lee et al. 2011). The HAT and DUB activities within SAGA are carried out by Gcn5 and Ubp8, respectively, which interact with other SAGA subunits to form a HAT module composed of Gcn5, Ada2, Ada3, and Sgf29 and a DUB module composed of Ubp8, Sgf73, Sgf11, and Sus1. These modules are conserved in humans (h) and are composed of homologous proteins (GCN5/PCAF, ADA2b, ADA3, and SGF29 in the HAT module; USP22, ATXN7, ATXN7L3, and ENY2 in the DUB module). The full assembly of the given modules is crucial to increase the substrate recognition and specificity of yGcn5/hGCN5 and enhance the catalytic activity of yUbp8/hUSP22 on nucleosomes (Gamper et al. 2009; Kohler et al. 2010; Samara et al. 2010; Bian et al. 2011; Lang et al. 2011). Interestingly, deletion of subunits of the DUB or the HAT module affected the stability of the corresponding module but not the overall integrity of the SAGA complex (Lee et al. 2011).

[6]These authors contributed equally to this work.
Corresponding authors: devys@igbmc.fr, laszlo@igbmc.fr
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.250225.114.

In *S. cerevisiae*, the Tra1 subunit of SAGA interacts with different activators, leading to its recruitment at a subset of promoters. At genes where it is recruited, the yeast SAGA complex was suggested to stimulate preinitiation complex formation through interactions with TBP and facilitate transcription initiation through Gcn5-mediated acetylation of their promoters (for review, see Weake and Workman 2012). However, it is still unclear how the multiple activities of SAGA are integrated to regulate gene expression. Indeed, analysis of gene expression changes in yeast strains deleted for all nonessential SAGA subunits revealed that the subunits of each module (HAT, DUB, and TBP-binding) tend to cluster together, although some differences were observed between *S. cerevisiae* and *Schizosaccharomyces pombe* (Helmlinger et al. 2011; Lenstra et al. 2011).

Genome-wide location studies of subunits of the SAGA complex in different organisms revealed only a few hundred binding sites for SAGA (Vermeulen et al. 2010; Bian et al. 2011; Krebs et al. 2011; Venters et al. 2011; Weake et al. 2011). In most cases, SAGA subunits were detected at a subset of promoters. Such specific targeting of ySAGA could explain the regulation of "SAGA-dominated genes," which were identified as genes that are down-regulated following the deletion of Spt3, a ySAGA-specific subunit, but unmodified upon inactivation of Taf1, a yTFIID-specific subunit (Huisinga and Pugh 2004). These genes, which preferentially use SAGA rather than TFIID to assemble the transcription machinery, account for ~10% of the yeast genes and correspond mainly to stress-induced genes with a TATA box in their promoters (Basehoar et al. 2004). However, recent studies in yeast revealed that only a fraction of bound genes is misregulated upon deletion of SAGA subunits, and, conversely, only a fraction of misregulated genes is bound by the corresponding subunits (Lenstra et al. 2011; Venters et al. 2011; Lenstra and Holstege 2012). Furthermore, a wider distribution of the complex is suggested by chromatin immunoprecipitation (ChIP) experiments on a few model genes in yeast that also detected SAGA in the coding region and by ChIP-coupled high-throughput sequencing (ChIP-seq) analyses in flies that revealed a colocalization of SAGA with RNA polymerase II (Pol II) in the body of a subset of genes (Govind et al. 2007; Johnsson et al. 2009; Weake et al. 2011). In addition to its gene-specific regulatory role, it has been proposed that SAGA acetylates histones throughout the genome in a global and untargeted manner (Vogelauer et al. 2000). In good agreement, inactivation of the SAGA HAT or DUB activities was reproducibly reported to affect the global levels of histone H3K9 acetylation (H3K9ac) and histone H2B monoubiquitination (H2Bub), two marks associated with active transcription at promoters and in gene transcribed regions, respectively (Henry et al. 2003; Bian et al. 2011; Jin et al. 2011).

To better understand the genome-wide function of SAGA, we investigated the distribution of this coactivator complex by tracking chromatin modifications deposited or removed by its two different enzymatic activities. The genome-wide comparison of H3K9ac or H2Bub densities in wild-type cells and cells with impaired SAGA HAT or DUB activity revealed that SAGA modifies the promoter and the transcribed region of all expressed genes in both *S. cerevisiae* and human cells. Our study reveals a highly dynamic and transient association of SAGA with the entire active chromatin that is required for Pol II recruitment and transcription of almost all expressed genes.

Results

SAGA deubiquitinates H2B on the transcribed region of all expressed genes in human cells

As an indicator of the genome-wide SAGA-binding profile, we aimed to localize histone marks regulated by the two enzymatic activities of the complex. We first analyzed the H2B DUB of the human SAGA complex. Former analyses in mammals revealed that H2Bub was extensively enriched in the transcribed region of expressed genes, in agreement with an association of the H2B ubiquitination enzymes with the elongating Pol II (Minsky et al. 2008; Shema et al. 2008; Vethantham et al. 2012). However, it remains to be determined whether SAGA can deubiquitinate all H2Bub-containing nucleosomes, which represent a large fraction of the genome, or whether its DUB activity would be restricted to the promoters of SAGA-regulated genes. To answer these questions, we first performed H2Bub ChIP-seq experiments on HeLa cells that revealed significant H2Bub signal exclusively on the transcribed region of different genes enriched for marks associated with active transcription (Fig. 1A). Scatter plot analysis of all human genes according to their mRNA levels and H2Bub densities in gene bodies (from transcription start site [TSS] to transcription termination site [TTS]) revealed two different gene populations with respect to H2B ubiquitination. Nonexpressed genes displayed background levels of H2Bub similar to that observed on intergenic regions, whereas significant levels of H2Bub were found at almost all expressed genes that were also characterized by the presence of a Pol II peak at their promoter regions (Fig. 1B; Supplemental Fig. S1B,C).

To determine at which genomic locations SAGA would remove ubiquitin from H2B, we generated and compared H2Bub ChIP-seq profiles in control and *ATXN7L3* knockdown HeLa cells after normalization of the two data sets by using large intergenic regions (see the Supplemental Material). We used our previously generated cell line expressing an *ATXN7L3* shRNA in which the DUB module is dissociated from the SAGA complex, leading to an increase of global H2Bub levels (Supplemental Fig. S1A; Lang et al. 2011). After inactivation of the SAGA DUB, H2Bub densities in the whole transcribed regions of expressed genes were higher than that in control cells, but the overall H2Bub distribution was preserved. In contrast, comparable low background H2Bub levels were observed in nonexpressed genes in both *ATXN7L3* knockdown and control HeLa cells (Fig. 1A). It has been previously suggested that the very

low levels of H2Bub at gene promoters were due to the high DUB activity of SAGA recruited to promoters. However, the absence of any H2Bub signal in wild-type conditions and upon depletion of SAGA DUB activity at the promoter of a representative gene (*USP25*) indicates that H2B was not targeted by the ubiquitinating enzymes at this promoter (Fig. 1C). Similar observations were made on all expressed genes upon depletion of the SAGA DUB activity, suggesting that the H2B ubiquitinating enzymes are not active on these promoters (Fig. 1D). In contrast, we observed an increase of H2Bub levels in the bodies of expressed genes in *ATXN7L3* knockdown as compared with control HeLa cells (Fig. 1E). Metagene analysis revealed a similar H2Bub profile in both cell lines, but the signal was increased by ~1.7-fold throughout gene bodies upon SAGA DUB depletion (Fig. 1F).

As a quantification of SAGA DUB activity, we measured the H2Bub ratio in *ATXN7L3* knockdown cells versus control cells in intergenic regions as well as in the transcribed region of all expressed genes. This analysis clearly differentiated intergenic regions from expressed genes, which had higher H2Bub densities and increased H2Bub ratios (Fig. 1G). Nonexpressed genes share the same profile as intergenic regions, with background levels of H2Bub and mean H2Bub ratios close to 1 (Supplemental Fig. S2A). To determine whether gene expression levels influence the SAGA DUB activity, we compared gene categories characterized by low, medium, and high expression (Supplemental Material). The three groups displayed strikingly similar H2Bub ratios that were

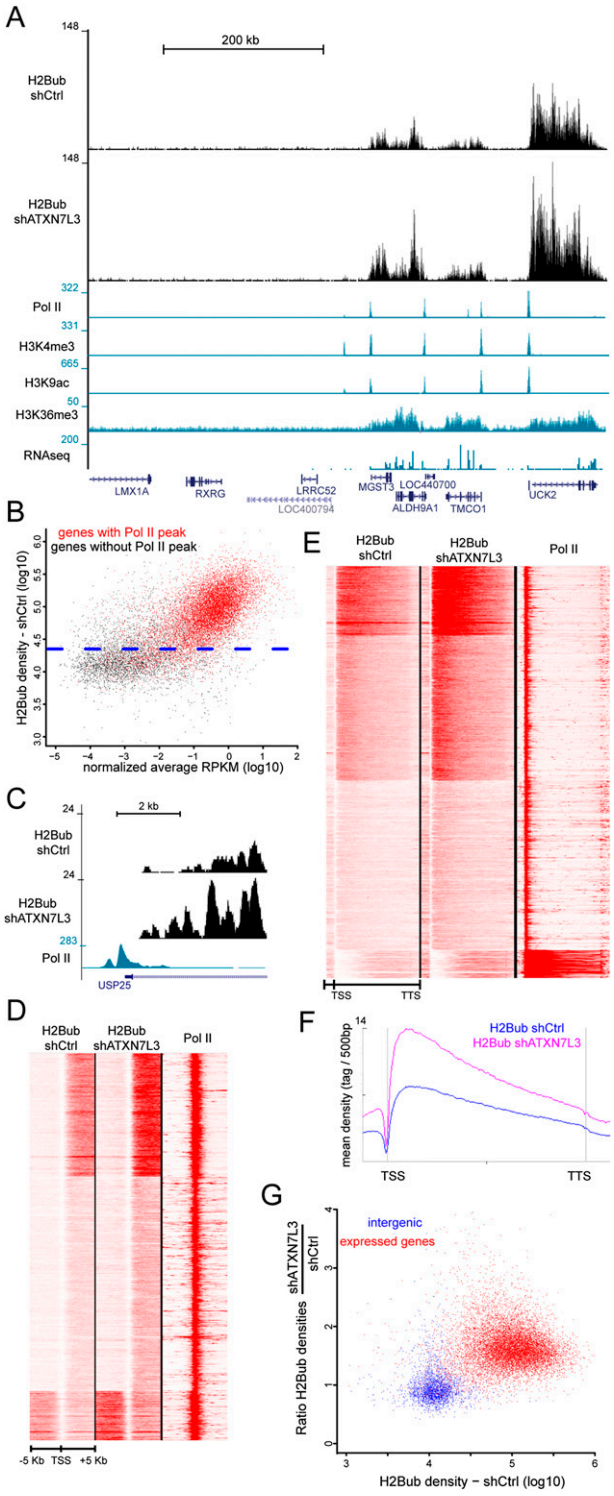


Figure 1. SAGA acts in the transcribed region of all expressed genes in HeLa cells. (*A*) Genome browser tracks depicting H2Bub distribution at a representative region in control and *ATXN7L3* knockdown HeLa cells. H2Bub localized in the transcribed region of expressed genes that are characterized by Pol II, H3K4me3, and H3K9ac peaks at their promoter; H3K36me3 signal on the gene body; and RNA sequencing (RNA-seq) reads on exons. (*B*) Gene expression levels (based on normalized average RPKM [reads per kilobase per million mapped reads]) were plotted versus the H2Bub densities on the corresponding gene bodies. The blue dotted line indicates the background density of H2Bub in control HeLa cells (95% of intergenic regions have lower H2Bub densities). Significant H2Bub and RNA-seq reads are found at most of the 10,934 genes that have a Pol II peak at their promoter (red dots) but not at the majority of the 8363 genes devoid of Pol II (black dots). When considering expressed genes, a very weak correlation (Pearson correlation coefficient, 0.20) could be observed between H2Bub levels and gene expression, indicating that H2B ubiquitination is not solely related to gene expression levels. (*C*) Absence of H2Bub at the promoter of an expressed gene in control and *ATXN7L3* knockdown HeLa cells. (*D*,*E*) Heat maps showing the distribution of H2Bub in control HeLa cells (H2Bub shCtrl) and after the inactivation of SAGA DUB activity (H2Bub shATXN7L3) around the TSS (TSS −5 kb/+5 kb) (*D*) and on bodies of expressed genes (from TSS to TTS) (*E*). Nine-thousand-six-hundred-thirty genes with a Pol II peak at the TSS were considered. (*F*) Average profiles depicting H2Bub distribution on gene bodies. Upon inactivation of SAGA DUB activity, H2Bub is not increased at the promoter of expressed genes (*C*,*D*) but increases on the body of expressed genes (*E*,*F*). (*G*) Scatter plots representing H2Bub densities in control cells relative to SAGA DUB activity (ratio of H2Bub densities in shATXN7L3 and shCtrl cells). Two-thousand-eight-hundred-forty-nine intergenic regions (blue dots) were compared with 10,486 expressed genes (red dots).

clearly different from those of intergenic regions (Supplemental Fig. S2B). Furthermore, 50% of expressed genes had H2Bub ratios ranging from 1.5 to 2.0, indicating a homogenous DUB activity, although these genes had much more heterogenous H2Bub densities, varying by 10-fold (Fig. 1G; Supplemental Fig. S2C). Our results together demonstrate that the SAGA DUB activity is directed toward the transcribed region of almost all expressed genes and is directly proportional to H2Bub levels, suggesting that the modification itself is the main driving force for this activity.

SAGA can be recruited in gene bodies independently of Pol II to induce a very fast reversal of H2B ubiquitination

The above reported distribution of the SAGA DUB raises the question of the mechanisms for SAGA recruitment and activity in gene bodies. An obvious hypothesis posits that following its recruitment at promoters, SAGA would interact with the elongating Pol II and thus travel through the whole transcribed region. To test this hypothesis, we measured H2Bub bulk levels in HeLa cells at different time points following treatment with actinomycin D, an inhibitor of transcription elongation. Transcription inhibition suppresses the deposition of the mark as the H2B ubiquitination machinery is recruited in gene bodies by the elongating Pol II. Thus, changes in the levels of already deposited H2Bub upon actinomycin D treatment can be explained only by the activity of DUBs that can remove ubiquitin from H2B.

Indeed, we observed a very fast loss of H2Bub that was dramatically reduced after 10 min and undetectable after 3 h of treatment in wild-type HeLa cells, indicating an extremely dynamic action of DUBs (Fig. 2A). We could not detect such extremely rapid effects on other chromatin marks analyzed (H2Aub, H3K4me3, or H3K9ac), suggesting a unique and dynamic versatility of histone H2B ubiquitination. Importantly, the loss of H2Bub was significantly delayed in *ATXN7L3* knockdown cells, indicating a role for SAGA DUB in H2Bub removal (Fig. 2A). To demonstrate unequivocally the role of the SAGA DUB in H2Bub turnover, we analyzed H2Bub kinetics in cells in which this activity was completely suppressed. We therefore generated mouse embryonic stem cells (mESCs) homozygous for an *Atxn7l3*-null allele (*Atxn7l3*$^{-/-}$) (Supplemental Fig. S3). While H2Bub was almost undetectable after 10 min of actinomycin D treatment in wild-type mESCs, the H2Bub global levels stayed completely unchanged in *Atxn7l3*$^{-/-}$ mESCs at all of the analyzed time points (Fig. 2B). This result demonstrates unambiguously that the fast disappearance of H2Bub in control cells is exclusively carried out by the SAGA DUB activity, ruling out the role of other DUBs in this process. In addition, SAGA can act in the transcribed region of active genes independently of elongating Pol II. Importantly, our observations reveal an extremely transient and dynamic association of SAGA with active chromatin, as H2Bub can be erased genome-wide by the SAGA DUB activity in a few minutes.

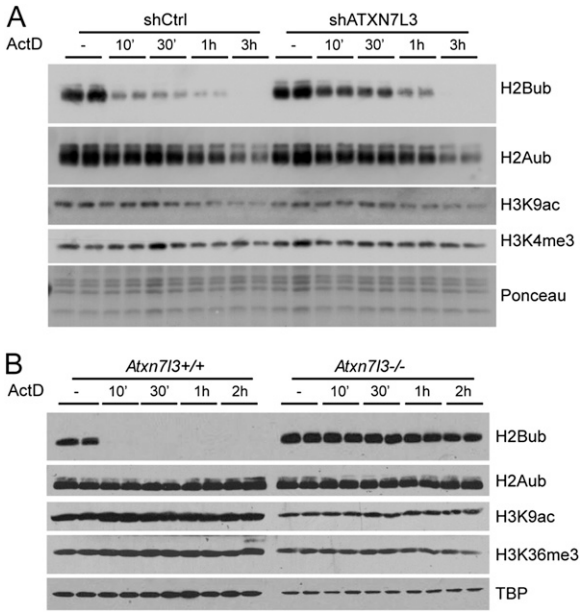


Figure 2. SAGA mediates a very fast reversal of H2B ubiquitination upon transcription inhibition. Control (*left* panels) and *ATXN7L3* knockdown (*right* panels) HeLa cells (*A*) or *Atxn7l3* knockout mESCs (*B*, *right* panels) were treated with actinomycin D for the indicated time points. Acidic extracts were analyzed by Western blotting using the indicated antibodies.

SAGA deubiquitinates the bodies of all expressed genes in yeast

Next, we wanted to confirm in other organisms the wide distribution of the SAGA DUB activity that we observed in human cells. Former ChIP-on-chip experiments in *S. cerevisiae* revealed a colocalization of H2Bub with H3K79me3 in the coding region of only a subset of genes and suggested that Ubp8 would mainly remove ubiquitin from H2Bub in the 5′ region of genes at H3K4me3 peaks (Schulze et al. 2009, 2011). This would indicate different regulations for H2B ubiquitination between yeast and other eukaryotes, as indicated by differential H2Bub distribution (Kharchenko et al. 2011; Roudier et al. 2011; Vethantham et al. 2012). To compare the activity of the evolutionarily well-conserved DUB module of the yeast and the human SAGA complexes, we aimed to compare H2Bub profiles in wild-type and *ubp8*Δ strains. To get an accurate map of H2Bub in *S. cerevisiae*, we took advantage of the highly specific anti-H2Bub monoclonal antibody that was raised against a branched peptide corresponding to the ubiquitinated human histone H2B (Minsky et al. 2008) but did not recognize H2Bub in budding yeast. We thus humanized the yeast H2B through substitutions of R119 and S125 to K and T, respectively (Fig. 3A). This modified H2B was Flag-tagged and expressed in yeast depleted for endogenous H2B. Western blot analysis of yeast extracts with an anti-Flag antibody revealed that this humanized H2B (hH2B) was efficiently ubiquitinated to the same extent as a Flag-tagged yeast H2B (yH2B) (Fig. 3B). The anti-H2Bub antibody revealed

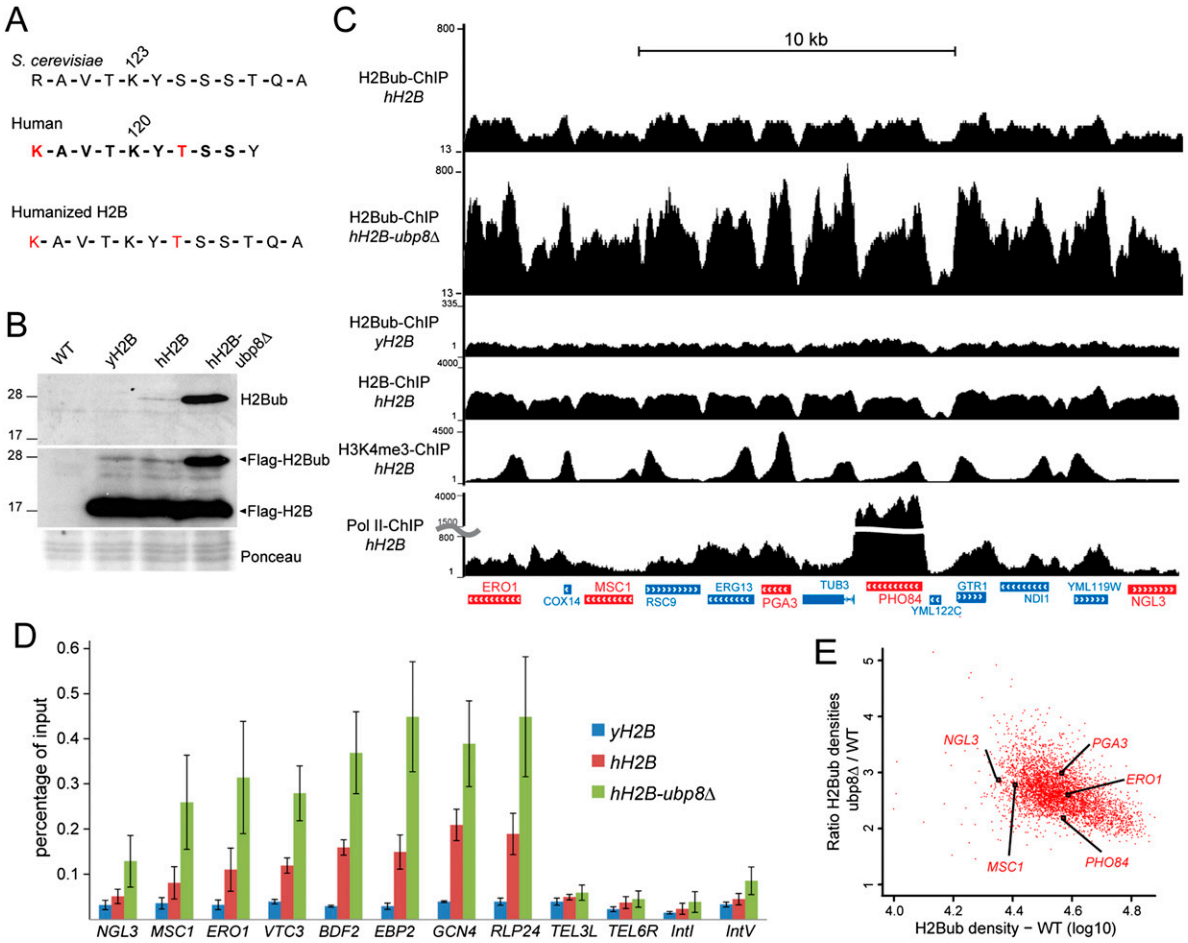


Figure 3. H2Bub ChIP-seq analyses in budding yeast expressing a humanized histone H2B reveal a global distribution of the SAGA DUB activity. (*A*,*B*) A humanized version of the yeast histone H2B is efficiently ubiquitinated and recognized by an anti-human H2Bub antibody. (*A*) *S. cerevisiae* and human H2B protein sequence comparison around their ubiquitination sites. The peptide used to raise the anti-H2Bub antibody (in bold) contains two residues (in red) that are not conserved in the yeast H2B. hH2B was obtained by mutating these two residues into their human counterpart. (*B*) H2B ubiquitination in *htb1*Δ *htb2*Δ yeast strains expressing Flag-tagged *yH2B* or *hH2B* and after further *UBP8* deletion (*hH2B-ubp8*Δ) was analyzed by immunoblotting using an anti-H2Bub (*top* panel) or an anti-Flag (*middle* panel) antibody. (*C*) H2Bub ChIP-seq analyses in *S. cerevisiae* expressing hH2B (H2Bub-ChIP *hH2B*) and after further *UBP8* deletion (H2Bub-ChIP *hH2B-ubp8*Δ) or the unmodified yH2B (H2Bub-ChIP *yH2B*). Genome browser tracks of a representative genomic region showing H2Bub enrichment in gene bodies. Representative genes shown in *E* are indicated in red. H2B, H3K4me3, and Pol II profiles in the *hH2B* background are shown in the *lower* tracks. (*D*) H2Bub ChIP was performed on *yH2B*, *hH2B*, and *hH2B-ubp8*Δ strains. H2Bub levels were quantified by real-time quantitative PCR (qPCR) on bodies of active genes and control intergenic regions (*TEL3L*, *TEL6R*, *IntI*, and *IntV*). The values (mean ± SD of three independent ChIP experiments) are expressed as percentage of input DNA signal. (*E*) Scatter plot showing a homogenous SAGA DUB activity on the 3916 expressed genes. The ratios of H2Bub densities on gene bodies between hH2B-*ubp8*Δ and hH2B strains were plotted against the H2Bub densities.

a single band of 25 kDa in the hH2B, whose intensity was further increased upon *UBP8* deletion. These results together demonstrate that the regulation of H2B ubiquitination in *S. cerevisiae* was not affected by the yeast H2B humanization.

ChIP-seq experiments in hH2B-expressing yeast cells revealed that H2Bub was specifically enriched in gene bodies of almost all active genes and depleted from intergenic regions (Fig. 3C; Supplemental Fig. S4A). The signal specificity was demonstrated by the detection of background signal on the whole genome of cells expressing yH2B. Moreover, significant variation of H2Bub densities from gene to gene contrasted with a homogenous enrichment of H2B on all genes, indicating that H2B local densities cannot be the sole determinant of H2B ubiquitination (Fig. 3C; Supplemental Fig. S4B).

As the vast majority of the yeast genome can be ubiquitinated on H2B (see above) and thus could be a substrate for the SAGA DUB, we next determined the distribution of this activity. To this end, we analyzed H2B ubiquitination in yeast strains expressing hH2B and after further deletion of *UBP8*. H2Bub levels were increased by about threefold to fivefold in an *ubp8*Δ strain when compared with the parental strain, suggesting that SAGA has a global DUB activity on the yeast genome (Fig. 3B; Supplemental Fig. S5A). We next performed

H2Bub ChIP-seq experiments in the same strains. To get an accurate normalization of the two data sets, we quantified H2Bub levels on a number of selected genes and intergenic regions by ChIP combined with quantitative PCR (ChIP-qPCR) (Fig. 3D). A mean 2.5-fold increase of H2Bub levels in gene bodies upon *UBP8* deletion was calculated from three independent ChIP-qPCR experiments and was applied to normalize our two ChIP-seq data sets. On a representative region, the H2Bub levels were found increased in all gene bodies upon *UBP8* deletion, in good agreement with our observation in human cells (Fig. 3C). Importantly, very similar results were observed when the ChIP-seq experiments were performed with the addition of *S. pombe* chromatin as spike-in control for normalization (Supplemental Fig. S5B,C). For all active gene bodies, we calculated the ratio of H2Bub densities between the *ubp8*Δ and parental strains and observed a twofold to threefold increase of the ratio between the two strains (Fig. 3E). This observation and the very high correlation between H2Bub densities between the two strains (Pearson correlation coefficient, 0.84) suggest a global and homogenous activity of Ubp8 in all regions that display significant H2Bub in wild-type cells. Similar to our observations in human cells, the yeast SAGA complex deubiquitinates H2B on the transcribed region of all expressed genes.

GCN5-containing complexes acetylate H3K9 at the promoter of all expressed genes in HeLa cells

The observed localization of the SAGA DUB activity prompted us to determine whether its HAT activity also displays a broad genome-wide distribution. In human cells, GCN5 and its paralog, PCAF, are found in two different HAT complexes, SAGA and ATAC, in which they are regulated by interactions with other subunits of the HAT module, such as ADA3 (Lee and Workman 2007). We could observe a global decrease of H3K9ac levels upon ADA3 depletion in HeLa cells, corresponding to an inactivation of both GCN5-containing complexes (Fig. 4A). In the same conditions, H2Bub levels were unchanged. Conversely, H3K9ac was unchanged, but H2Bub levels were increased upon *ATXN7L3* down-regulation, indicating that the two enzymatic activities of SAGA are independent (Fig. 4A).

To determine the genome-wide distribution of GCN5-containing complexes, we profiled and quantified the distribution of H3K9ac by ChIP-seq in cells expressing either control or *ADA3* shRNA (Supplemental Fig. S6A, B). In control HeLa cells, H3K9ac peaks were detected

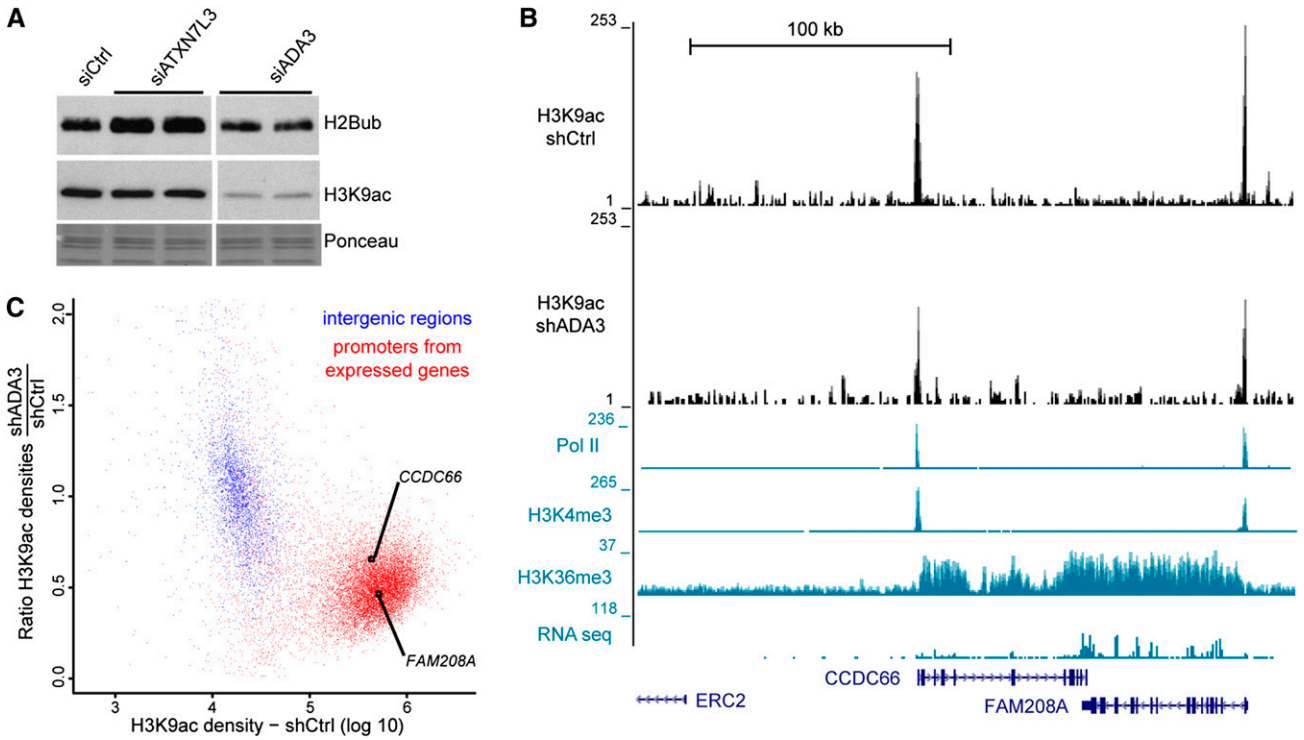


Figure 4. All active Pol II promoters are acetylated by GCN5-containing complexes in human cells. (*A*) Western blot analyses of changes in H2B ubiquitination and H3K9ac upon inactivation of the corresponding enzymes. Total histones were purified by acidic extractions from HeLa cells transfected with the specified siRNA and immunoblotted with the indicated antibodies. (*B*,*C*) Analysis of H3K9ac ChIP-seq profiles in control and *ADA3* knockdown HeLa cells. (*B*) Genome browser tracks at a representative region reveal that H3K9ac peak intensity at the promoter of expressed genes in HeLa cells (H3K9ac-shCtrl) drops upon *ADA3* knockdown (H3K9ac-shADA3). (*C*) GCN5 HAT activity (ratio of H3K9ac densities in shADA3 and shCtrl cells) was plotted relative to H3K9ac densities in 3189 intergenic regions (blue dots) and 11,300 regions surrounding the TSS of expressed genes (TSS −1.5 kb/+3.5 kb; red dots). Promoters of expressed genes shown in *B* are indicated.

exclusively at the promoters of expressed genes together with Pol II and H3K4me3 but were absent from silent genes (Fig. 4B). The control and *ADA3* knockdown data sets were normalized by using intergenic regions where only background levels of H3K9ac could be detected (Supplemental Material). In ADA3-depleted cells, the overall profile of H3K9ac was conserved, but the peak heights were significantly reduced (Fig. 4B). As a measure of GCN5 HAT activity, we calculated the ratio of H3K9ac densities obtained in *ADA3*-expressing versus control shRNA-expressing cells both in intergenic regions and at the promoters of all expressed genes. This analysis clearly separated the two categories of genomic loci: (1) All intergenic regions had low levels of H3K9ac, and the H3K9ac ratios were centered on 1, and (2) the vast majority of promoters of active genes were enriched for H3K9ac and displayed decreased ratios of H3K9ac, indicating that GCN5-containing complexes significantly acetylate H3K9 at these locations (Fig. 4C).

GCN5 acetylates H3K9 at all active promoters in yeast

As *GCN5* deletion in *S. cerevisiae* induced a decrease in global H3K9ac (Fig. 5A), we further investigated Gcn5 distribution in yeast by measuring H3K9ac levels in wild-type and *gcn5*Δ strains by ChIP-seq. In wild-type cells, H3K9ac peaks colocalized with H3K4me3 around the TSS of active genes. In agreement with our results in human cells, H3K9ac peaks were completely lost at active promoters in the *gcn5*Δ strain (Fig. 5B; Supplemental Fig. S6C). We measured the ratio of H3K9ac densities in *gcn5*Δ versus wild-type cells at all active promoters (as defined in Supplemental Fig. S6D). Mid-gene bodies (25%–75% of the coding region) of large genes (>2 kb) with only background levels of H3K9ac were used to normalize the two data sets. All active promoters could be easily differentiated from mid-gene bodies, as they were enriched for H3K9ac in wild-type cells and had a <1 ratio of H3K9ac densities (Fig. 5C; Supplemental Fig. S6E,F). Our results together indicate that, in both yeast and human cells, GCN5-containing complexes acetylate H3K9 specifically at the promoters of active genes and thus strongly suggest a wide distribution of the SAGA HAT activity.

SAGA is required for RNA Pol II recruitment at all transcribed genes and for nascent mRNA synthesis

Based on our results that revealed an unexpected wide action of SAGA on the transcriptionally active genome,

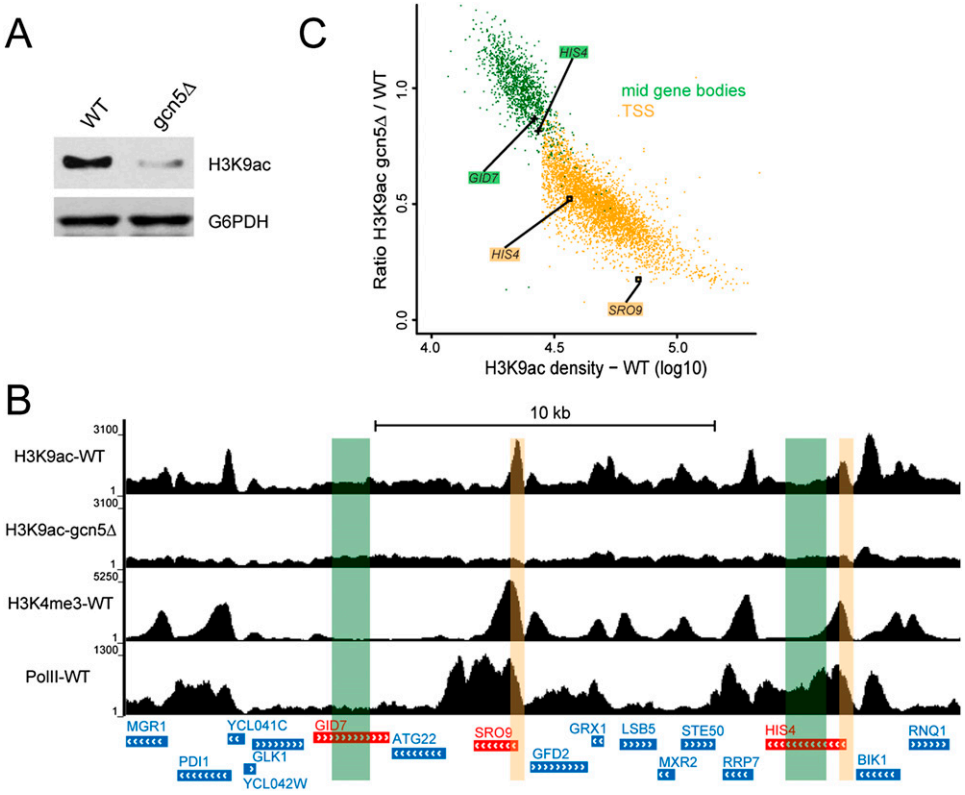


Figure 5. Gcn5 acetylates H3K9 at the promoter of all expressed genes in yeast. (*A*) Whole-cell extracts from wild-type (WT) and *gcn5*Δ yeasts were prepared in an 8 M urea buffer and probed as indicated. (*B*,*C*) Analysis of H3K9ac ChIP-seq profiles in wild-type and *gcn5*Δ yeast cells. (*B*) Genome browser tracks of a representative yeast genomic region. H3K9ac peaks detected in wild-type cells (H3K9ac-WT) are absent when *GCN5* is deleted (H3K9ac-*gcn5*Δ). Expressed genes are characterized by an H3K4me3 peak at their promoter (H3K4me3-WT) and a Pol II signal on the gene body (Pol II-WT). (*C*) Scatter plot representing Gcn5 acetylation activity (ratio of H3K9ac densities in *gcn5*Δ and wild type) versus H3K9ac density in 3916 regions centered on the first codon (±200 base pairs [bp]) of expressed yeast genes as defined by the presence of H3K4me3 and H3K9ac peaks at the TSS (yellow dots) or in 1184 mid-gene bodies of large genes (from 25% to 75% of gene bodies >2 kb; green dots). Promoters and control regions highlighted in *B* are indicated.

we asked whether this complex would have a more global function in transcription regulation than previously appreciated. We hypothesized that SAGA would behave as a general cofactor required for the transcription of all Pol II genes rather than being involved only in the regulation of stress-induced genes. Besides its enzymatic activities, other functions of SAGA could influence Pol II recruitment and transcription. Therefore, we analyzed Pol II recruitment upon deletion of *SPT7* or *SPT20*, which are known to result in the disruption of the SAGA complex in yeast. Using an anti-Pol II antibody that recognizes both modified and unmodified forms of the C-terminal domain of Rpb1, the largest subunit of Pol II, background levels of Pol II were detected in control intergenic or telomeric regions in wild-type *S. cerevisiae* and were unmodified in *spt7*Δ or *spt20*Δ strains. In contrast, in all 14 genes tested that were previously defined as either SAGA- or TFIID-dominated (Huisinga and Pugh 2004), Pol II occupancy was strikingly reduced upon *SPT7* or *SPT20* deletion when compared with that observed in the parental strain (Fig. 6A). Comparable effects were observed for genes that were found to be down-regulated as well as other genes (either SAGA- or TFIID-dominated) whose expression was found to be unchanged upon deletion of these subunits (Lenstra et al. 2011). Two independent chromatin extractions gave similar results, and H2B or H3K4me3 ChIP on the same samples revealed comparable signals in wild-type, *spt7Δ*, and *spt20Δ* strains (Supplemental Fig. S7A). These results suggest that SAGA might be crucial for Pol II recruitment on a large number of active genes.

Next, we analyzed Pol II occupancy on all genes by performing ChIP-seq experiments that revealed a global decrease of the Pol II signal upon *SPT20* deletion (Fig. 6B; Supplemental Fig. S7C). For most genes that displayed a significant Pol II occupancy in wild-type cells, the signal

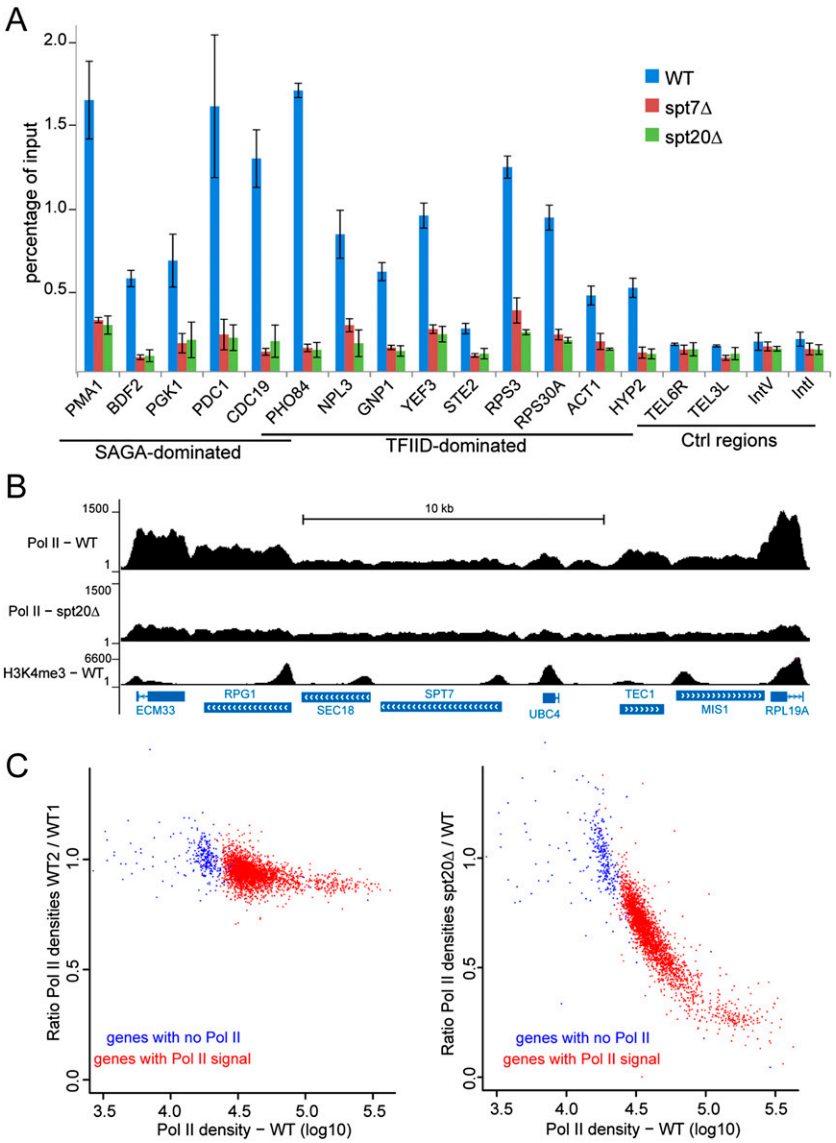


Figure 6. SAGA is required for RNA Pol II recruitment at both SAGA- and TFIID-dominated genes. (*A*) Pol II ChIP-qPCR performed on chromatin extracted from wild-type (WT), *spt7*Δ, and *spt20*Δ yeast cells on SAGA- and TFIID-dominated genes or in control intergenic regions (mean ± SD, $n = 2$). (*B*,*C*) Pol II ChIP-seq analysis in *S. cerevisiae* wild type and *spt20*Δ. (*B*) Genome browser tracks of a representative genomic region show decreased Pol II occupancy in the *spt20*Δ strain. (*C*, *left* panel) Scatter plot showing the reproducibility of Pol II ChIP performed on two independent wild-type chromatin preparations. (*Right* panel) A global loss of Pol II recruitment in *spt20*Δ cells is observed on 2606 active genes with H3K9ac and H3K4me3 peaks at the TSS and significant Pol II signal (red dots). Data were normalized on 291 genes that have background levels of Pol II and lack H3K4me3 and H3K9ac (blue dots).

was strongly reduced in mutant cells. We identified ~300 genes with background levels of Pol II signal and devoid of H3K4me3 and H3K9ac that were further used to normalize the two data sets (Supplemental Material). Next, we compared Pol II occupancy in wild-type and *spt20*Δ cells on ~2600 active genes in which a clear Pol II signal was detected (Supplemental Fig. S7B). On all of these genes, we observed that the Pol II occupancy was strongly reduced upon *SPT20* deletion (Fig. 6C). Very similar conclusions were made when normalization of the two data sets was performed by using Pol II densities measured on Pol III transcribed genes (Supplemental Fig. S7D). Therefore, our results indicate that SAGA was required for Pol II recruitment at all transcribed genes where a significant Pol II occupancy could be detected.

This result was unexpected, as previous measurements of mRNA expression changes in ySAGA mutant strains identified only a small subset of SAGA-dominated genes. However, measuring steady-state mRNA levels may not fully reflect the SAGA-dependent transcriptional effects, as cells can compensate for global transcription changes (Haimovich et al. 2013). To overcome this limitation, we analyzed how SAGA affects gene expression through the quantification of newly synthesized mRNA that measure mRNA synthesis uncoupled from mRNA degradation. Wild-type, *spt20*Δ, and *spt7*Δ *S. cerevisiae* strains cultured in YPD were metabolically labeled with 4-thiouracil for 6 min. Labeled cells were then mixed in a ratio of 3:1 with labeled *S. pombe* cells that provided an internal standard. After total RNA extraction, newly synthesized (labeled) RNAs were purified, and both fractions were analyzed by RT-qPCR. Surprisingly, we observed a threefold to 10-fold reduction in newly synthesized mRNA production from the 12 genes analyzed in *spt20*Δ or *spt7*Δ strains when compared with wild-type cells (Fig. 7A). In both strains, SAGA- or TFIID-dominated genes were down-regulated to the same extent. These changes in mRNA synthesis reflected the decreased Pol II occupancy that was observed on the same genes in *spt20*Δ or *spt7*Δ strains (Fig. 6A). In contrast, mRNA quantification performed on total RNA from the *spt20*Δ strain revealed unchanged levels for eight of these genes and a moderate decrease of less than twofold for four other genes, in good agreement with a previous study (Lenstra et al. 2011). Similarly, in the *spt7*Δ strain, mRNA steady-state levels from only two genes were found to be slightly decreased

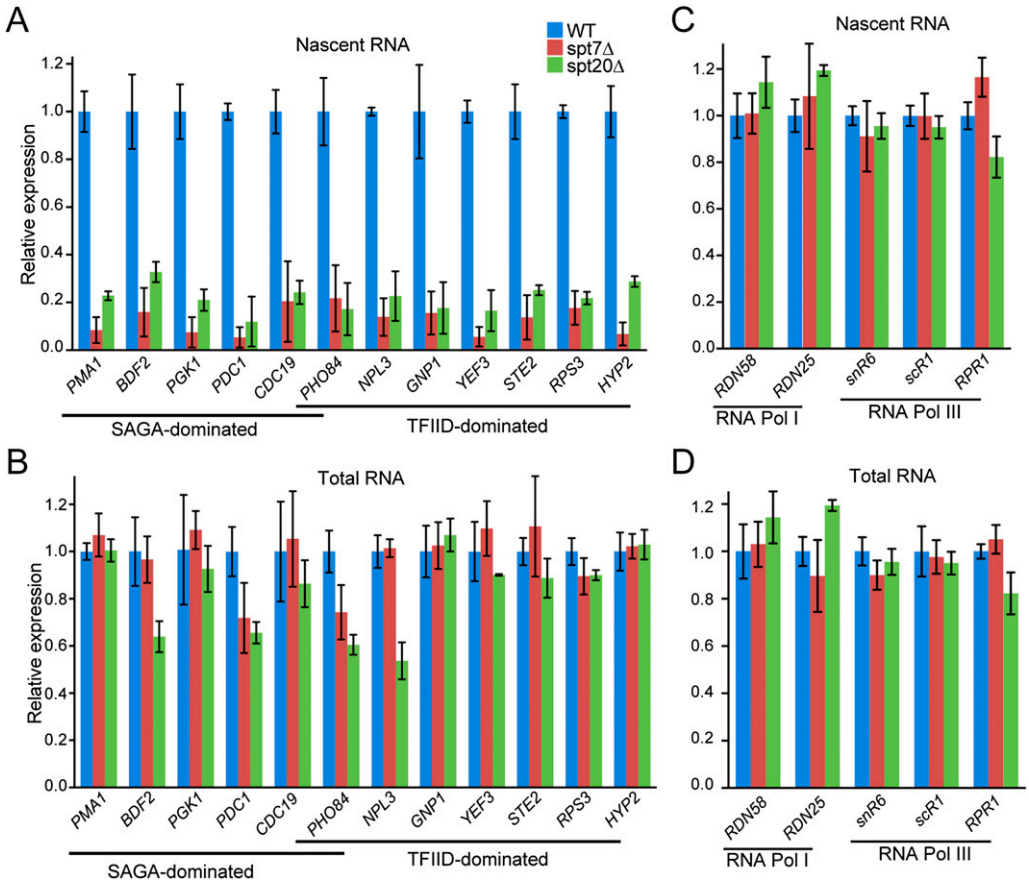


Figure 7. SAGA is required for RNA Pol II transcription of either SAGA- or TFIID-dominated genes. Newly synthesized (*A*,*C*) and total (*B*,*D*) mRNA were extracted from wild-type (WT), *spt7*Δ, and *spt20*Δ yeast cells. mRNA from SAGA- and TFIID-dominated genes (*A*,*B*) or RNA Pol I and RNA Pol III genes (*C*,*D*) were quantified by real-time PCR. Results were normalized to *S. pombe* tubulin expression and are presented as fold variation in comparison with the wild type (mean ± SD, $n = 3$).

(Fig. 7B). The levels of RNA Pol I and RNA Pol III nascent transcripts used as controls were not affected upon *SPT20* or *SPT7* deletion (Fig. 7C,D). Altogether, our results indicate that the SAGA complex plays a role in Pol II recruitment and transcription on a vast majority of Pol II genes independently, whether they were defined as SAGA and/or TFIID dominated.

Discussion

In this study, we provide a series of results that coherently indicate that the SAGA complex behaves as a general cofactor that is required for Pol II transcription of almost all active genes. First, we analyzed the genome-wide location of the HAT and DUB activities of SAGA in both *S. cerevisiae* and human cells. These analyses revealed that SAGA is acting at all active genes, both at their promoters and in their whole transcribed regions. Such a distribution of the two SAGA enzymatic activities was highly unexpected, as it appears much broader than previously anticipated. Second, we showed that Pol II recruitment at all active genes was severely impaired upon deletion of SAGA structural subunits. Third, through the quantification of newly synthesized RNA, we demonstrated that SAGA inactivation induced a strong decrease of mRNA synthesis for all genes that were analyzed. In good agreement with this broad function, our analysis of the SAGA DUB activity upon transcription inhibition indicates that SAGA can act on a large fraction of the genome in a very fast manner. This underscores that SAGA association with chromatin occurs in a very transient and dynamic manner, indicating a new paradigm for the dynamic distribution and broad activity of this coactivator.

SAGA acts on a very large fraction of eukaryotic genomes

Our analyses of changes in H2B ubiquitination upon depletion of the SAGA DUB activity directly reflect SAGA localization at all expressed genes. Indeed, ChIP-seq analyses in the *ubp8*Δ yeast strain or ATXN7L3-depleted HeLa cells revealed H2Bub changes that can only be explained by a loss of the SAGA DUB activity. Actually, Ubp8 has to be incorporated within SAGA to be catalytically active and appears to be a unique constituent of SAGA (Ingvarsdottir et al. 2005; Lee et al. 2011). Similarly, USP22 was shown to be active in vitro only when interacting with all subunits of the DUB module, including ATXN7L3, and was found to be fully active in vivo only as part of the SAGA complex (Lang et al. 2011; Armour et al. 2013). Although it was recently shown that the proteasome regulatory particle can separate a DUB module from the yeast SAGA complex in vitro, such a free DUB module has no apparent effect on the regulation of H2Bub in vivo (Lim et al. 2013). In agreement with a role for SAGA in gene transcribed regions, different studies using ChIP on candidate genes previously suggested the recruitment of different SAGA subunits in gene bodies (Govind et al. 2007; Wyce et al. 2007), and genome-wide ChIP analyses in *S. pombe* and *Drosophila* embryos localized Gcn5 or Ada2b in the transcribed regions of a subset of expressed genes (Johnsson et al. 2009; Weake and Workman 2012). In contrast, our analyses of H2B ubiquitination upon SAGA DUB inactivation unambiguously demonstrate that SAGA is dynamically recruited in the transcribed region of all expressed genes. Although our experiments do not distinguish Gcn5 activity as part of the different Gcn5-containing complex, we demonstrate here that these complexes are recruited at the promoter of all expressed genes in budding yeast and human cells, further suggesting a broad genome-wide recruitment of the SAGA complex.

Former genome-wide location studies of the SAGA complex subunits in different organisms revealed only a few hundred SAGA-binding sites (Vermeulen et al. 2010; Krebs et al. 2011; Venters et al. 2011; Weake et al. 2011). These analyses rely on ChIP experiments of a subset of SAGA-specific subunits that likely capture the most stable associations with chromatin. In contrast, population-based approaches that average the signal across heterogenous cellular states would miss most of the locations where SAGA is only very transiently recruited. However, such short-lived associations can be revealed by tracking the chromatin changes mediated by the enzymatic activities of SAGA, as demonstrated by this study. In agreement with this conclusion, the analysis of the SAGA DUB activity upon transcription inhibition revealed that its association with the chromatin occurs in an extremely dynamic fashion.

SAGA is required for all Pol II transcription, in agreement with its broad genome-wide action at all transcribed regions

The observed distribution of the two enzymatic activities of SAGA contrasts with the current model that posits a recruitment of SAGA at a small subset of genes whose expression is regulated by the different activities of the complex. As indicated above, this is likely explained by the different modes of SAGA interaction with chromatin that are not equally detected, depending on the experimental methods. A broad distribution of SAGA on all active chromatin is in good agreement with our observation of a role for SAGA in Pol II recruitment and function on most active genes. Here, we show that the transcriptional role of SAGA is independent of its mode of recruitment. Indeed, upon SAGA depletion, Pol II recruitment and mRNA synthesis appear to be similarly affected at all genes whether ChIP of SAGA subunits detected a stable association of SAGA at the promoter or not. Our observation describing a general cofactor activity of SAGA at all Pol II genes appears to be in contradiction to the numerous expression studies reporting a low percentage of genes whose expression is deregulated in SAGA mutants (Huisinga and Pugh 2004; Lenstra et al. 2011). However, these gene expression analyses measured steady-state mRNA levels in wild-type and mutant strains. Nevertheless, several studies recently demonstrated that global

effects on mRNA synthesis are accompanied by compensatory mechanisms such as parallel changes in mRNA degradation rates, resulting in mRNA level buffering (Helenius et al. 2011; Schulz et al. 2014). Such a mechanism has been previously demonstrated in a yeast strain expressing a point mutant of Rpb1 subunit of Pol II that causes decreased mRNA synthesis as expected but also decreases mRNA decay rates (Sun et al. 2012).

Two enzymatic activities of the same complex act on different functional genomic elements

Our study surprisingly revealed distinct genomic locations for the two enzymatic activities of the same complex. While H3K9ac by the HAT module of SAGA is restricted to the promoter of active genes, the DUB acts on gene bodies and is excluded from promoters in human cells. It is unclear how SAGA can be recruited at all active genes at both their promoters and transcribed regions. This observation also raises the question of a restriction of the HAT activity to promoter regions when the complex has a much wider distribution. Two alternative mechanisms can be proposed: (1) The specific localization of H3K9ac at the promoter could result from the cumulative actions of SAGA at both promoter and gene bodies and that of histone deacetylases such as Rpd3 acting in the transcribed region of these genes (Vogelauer et al. 2000; Govind et al. 2010; Weake et al. 2011). (2) Alternatively, at active promoters, the combination of transcription factors and histone marks known to interact with different domains of SAGA subunits could create a favorable environment for SAGA recruitment and/or retention and could position the HAT module for optimal substrate recognition (for review, see Weake and Workman 2012).

In contrast to GCN5, which lays a mark at the histone H3 tail, the SAGA DUB removes ubiquitin from H2Bub. Therefore, the DUB activity is restricted to locations where H2Bub is present, and this mark is clearly excluded from the promoter region in human cells. It is remarkable that H2B ubiquitination is very widely distributed in different genomes. Virtually all H2B molecules from any nucleosome, except those at telomeric or silenced regions, are potentially ubiquitinated in *S. cerevisiae*. In human cells, H2Bub is found in the transcribed region of all expressed genes that would correspond to as much as 20% of the whole genome. Our results demonstrate that SAGA can act on any H2Bub molecule genome-wide, raising the question of SAGA recruitment and dynamics on all targeted regions. Two different scenarios can be considered. In a first hypothesis, following its recruitment at active promoters, SAGA would then travel through the gene transcribed region either by binding to ubiquitinated substrates or via an interaction with elongating Pol II, as previously suggested (Wyce et al. 2007; Weake and Workman 2012). An alternative hypothesis could be that SAGA is not actively recruited on chromatin but would be diffusing rather freely in the nucleoplasm and would have a high affinity to any ubiquitinated H2B-containing region. In favor of such a model is our observation that, upon inhibition of transcription elongation, SAGA can completely remove ubiquitin from the whole active chromatin in a few minutes (Fig. 2). Furthermore, we showed that H2B ubiquitination per se is the main driving force for SAGA DUB, as its activity is directly proportional to the local H2Bub levels, which are poorly correlated with gene expression in human cells (Fig. 1; Supplemental Fig. S2).

In summary, our study reveals that SAGA is recruited at all active genes and will acetylate H3K9 at the promoters and deubiquitinate H2Bub in the gene bodies in both *S. cerevisiae* and human cells. This demonstrates an unexpected dynamic association of SAGA with the chromatin that can hardly be detected by conventional ChIP of subunits of the complex, suggesting a multifactorial recruitment of SAGA to chromatin that cannot be explained by previous models. Our work leads to the discovery of a new function for SAGA and indicates a role for SAGA as a general cofactor required for Pol II recruitment and transcription.

Materials and methods

Antibodies

Antibodies used in this study were as follows: H3K9ac (Abcam, ab4441), H2Bub (NRO3; Medimabs, MM-0029), H3K4me3 (Abcam, ab8580), H2B (Active Motif, 39237), H2Aub (Cell Signaling Technologies, XP rabbit mAb #8240), H3K36me3 (Abcam, ab9050), and Flag (Sigma, M2). RNA Pol II (PB-7G5) and TBP (3TF1-3G3) antibodies have been described earlier (Karmodiya et al. 2012).

siRNA transfection

Nontargeting Pool (catalog no. D-001810-10) as well as ON-TARGET plus SMART pool anti-ADA3 (catalog no. L-017508-00) siRNA were purchased from Dharmacon. Additionally, anti-ATXN7L3 (siRNA ID no. 271961) siRNA were purchased from Ambion. All siRNA were transfected in HeLa cells using Lipofectamine 2000 according to the manufacturer's instructions.

shRNA cell line construction

HEK293T cells were cotransfected with pLKO.1-puro (vector containing the shRNA sequence), pLP1, pLP2, and pLP/VSVG plasmids with FuGENE (Promega). Forty-eight hours later, the medium containing lentiviruses was used to infect HeLa cells, and positive cells were selected with 2 µg/mL puromycine. The following MISSION shRNA constructs (Sigma) were used: shRNA anti-ADA3 (TRC no. TRCN0000015734) and shRNA nontarget control (product no. SHC002). shATXN7L3 and the corresponding shCtrl cells were described (Lang et al. 2011).

mESC derivation

The *Atxn7l3* mouse strain used for this research project was created at the Institut Clinique de la Souris from the *Atxn7l3*$^{tm1(KOMP)Wtsi}$ ESC clone EPD0321_4_E11, obtained from the Knockout Mouse Project (KOMP) Repository (http://www.komp.org) and generated by the Wellcome Trust Sanger Institute (WTSI). Targeting vectors used were generated by the Wellcome Trust Sanger Institute and the Children's Hospital Oakland Research Institute as part of the Knockout Mouse Project (3U01HG004080) (Skarnes et al. 2011). At 3.5 d post-conception, blastocysts obtained by crossing *Atxn7l3*$^{+/-}$ heterozygous mice

were isolated and cultured on a monolayer of feeder cells in 2i medium (50% DMEM/F12, 50% neurobasal, 1× N2, 1× B27 with vitamin A, 10 μg/mL insulin, 1 mM glutamine, 25 μg/mL BSA type V, 0.1 mM β-mercaptoethanol, 40 μg/mL gentamycin, 3 μM CHIR 99021, 1 μM PD0325901). One week after blastocyst collection, the inner cell mass outgrowths were dissociated with 0.05% Trypsin-EDTA and plated into new feeder cell-containing plates. After three passages, cells were frozen (10% DMSO, 20% FCS ESCs tested, 70% 2i medium), and DNA was extracted from each clone for genotyping purposes. Genotypes were determined by PCR analysis using standard procedures. Three primers were used for this purpose: 5′-CAAGAAGCAGCATGCTTGGTCA AG-3′ (forward), 5′-CCAACAGCTTCCCCACAACGG-3′ (reverse 1, specific for the mutant allele), and 5′-CAGGAAGAAGTAG CCACACTTAACAGC-3′ (reverse 2, specific for the wild-type allele). Expected sizes of the amplicons were 277 base pairs (bp) for the mutant allele and 215 bp for the wild-type allele. Mouse care and procedures were in accordance with institutional and national guidelines.

Newly synthesized RNA analysis

Wild-type, *spt7*Δ, and *spt20*Δ *S. cerevisiae* cells were grown in YPD medium to an OD600 ≈ 0.8. Newly synthesized RNAs were labeled for 6 min by adding 4-thiouracil (Sigma-Aldrich) until a final concentration of 5 mM. In parallel, wild-type *S. pombe* cells were similarly grown in YES medium and labeled to be used as a spike-in across all samples. Cells were immediately pelleted and flash-frozen in liquid N_2 and stored at −80°C until further use. *S. cerevisiae* and *S. pombe* cells were mixed in a ratio of 3:1. Total RNAs were extracted using RiboPure yeast kit (Ambion, Life Technologies) according to the description provided by the manufacturer. Subsequently, RNA biotinylation was carried out on 200 μg of total RNA using 200 μL of 1 mg/mL EZ-link HPDP-Biotin (Pierce) in 100 μL of biotinylation buffer (100 mM Tris-HCl at pH 7.5, 10 mM EDTA) and 600 μL of DEPC-treated RNase-free water (Sigma-Aldrich) for 3 h at room temperature, protected from light. After chloroform extraction and isopropanol precipitation, purified RNAs were suspended in 100 μL of DEPC-treated RNase-free water (Sigma-Aldrich).

Newly synthesized biotinylated RNAs were bound to 100 μL of μMACS streptavidin microbeads (Miltenyi Biotec) for 90 min at room temperature with gentle shaking. Purification of labeled RNA was then carried out using μMACS streptavidin starting kit (Miltenyi Biotec). Columns were first equilibrated with 1 mL of washing buffer (100 mM Tris-HCl at pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween20). Samples were passed through the columns twice and washed five times with increasing volumes of washing buffer (600, 700, 800, 900, and 1000 μL). Ultimately, labeled RNAs were eluted twice with 200 μL of 100 mM DTT. Following ethanol precipitation, RNAs were resuspended in 10 μL of DEPC-treated RNase-free water (Sigma-Aldrich).

cDNA synthesis was performed on 1 μg of total RNA or 10 μL of labeled RNA using random hexamers and AMV reverse transcriptase (Roche) according to the manufacturer's instructions. Real-time qPCR were performed using SYBR Green I Master (Roche). A list of all of the primers used can be found in Supplemental Table S2. All samples were run in triplicate. After qPCR, all raw values were corrected for the expression of *S. pombe* tubulin. Finally, results were represented graphically as a relative comparison between the wild-type (set to 1), *spt7*Δ, and *spt20*Δ samples.

Yeast strains and protein extraction

Different yeast strains used in this study were derived from the previously described FY406 (Hirschhorn et al. 1995). Deletion mutants were generated as described previously (Janke et al. 2004) or by transformation with PCR products amplified from genomic DNA of the corresponding strain obtained from the *Saccharomyces* Genome Deletion (SGD) strain database. The plasmid expressing the humanized version of yeast HTB1 was generated by site-directed mutagenesis of *pRS413-Flag-HTB1-HTA1*. The genotypes of the strains used in this study are listed in Supplemental Table S3. Yeast whole-cell extracts were prepared as described (Gardner et al. 2005) with minor changes. Cells (3×10^8) from log-phase yeast cultures were harvested by centrifugation and lysed in 400 μL of SUME buffer (8 M urea, 1% SDS, 10 mM MOPS at pH 6.8, 10 mM EDTA, 0.01% bromophenol blue) by mechanical shearing.

ChIP in HeLa cells or *S. cerevisiae*, library preparation, sequencing, and bioinformatics analyses of the data are detailed in the Supplemental Material.

Accession numbers

The sequencing data have been deposited in the Gene Expression Omnibus database under the accession number GSE59370. All of the data sets that were used in this study are listed in the Supplemental Table S1.

Acknowledgments

We thank F. Winston for the gift of material. We thank B. Jost and the Institut de Génétique et de Biologie Moléculaire et Cellulaire microarray and deep sequencing platform for high-throughput sequencing, and Y. Tao and S. Le Gras for their help with bioinformatics analyses. We thank the Institut Clinique de la Souris for the generation of the *Atxn7L3* knockout mouse line. We are grateful to J. Müller, M. Timmers, R. Schneider, and D. Umlauf for critically reading the manuscript. This study was supported by funds from the Agence Nationale de la Recherche (ANR-10-INTB-1201) to D.D. and the NSC (100-2911-I-001-510 and 100-2923-B-001-001-MY3) to C.F.K. This study was also supported by grant ANR-10-LABX-0030-INRT under the frame program labeled ANR-10-IDEX-0002-02.

References

Armour S, Bennett E, Braun C, Zhang X-Y, McMahon S, Gygi S, Harper J, Sinclair D. 2013. A high-confidence interaction map identifies SIRT1 as a mediator of acetylation of USP22 and the SAGA coactivator complex. *Mol Cell Biol* **33:** 1487–1502.

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

Bian C, Xu C, Ruan J, Lee K, Burke T, Tempel W, Barsyte D, Li J, Wu M, Zhou B, et al. 2011. Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J* **30:** 2829–2842.

Gamper A, Kim J, Roeder R. 2009. The STAGA subunit ADA2b is an important regulator of human GCN5 catalysis. *Mol Cell Biol* **29:** 266–280.

Gardner RG, Nelson ZW, Gottschling DE. 2005. Ubp10/Dot4p regulates the persistence of ubiquitinated histone H2B: distinct roles in telomeric silencing and general chromatin. *Mol Cell Biol* **25:** 6123–6139.

Govind C, Zhang F, Qiu H, Hofmeyer K, Hinnebusch A. 2007. Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. *Mol Cell* **25:** 31–42.

Govind C, Qiu H, Ginsburg D, Ruan C, Hofmeyer K, Hu C, Swaminathan V, Workman J, Li B, Hinnebusch A. 2010.

Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. *Mol Cell* **39:** 234–246.

Haimovich G, Choder M, Singer RH, Trcek T. 2013. The fate of the messenger is pre-determined: a new model for regulation of gene expression. *Biochim Biophys Acta* **1829:** 643–653.

Helenius K, Yang Y, Tselykh TV, Pessa HK, Frilander MJ, Mäkelä TP. 2011. Requirement of TFIIH kinase subunit Mat1 for RNA Pol II C-terminal domain Ser5 phosphorylation, transcription and mRNA turnover. *Nucleic Acids Res* **39:** 5025–5035.

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

Henry KW, Wyce A, Lo WS, Duggan LJ, Emre NC, Kao CF, Pillus L, Shilatifard A, Osley MA, Berger SL. 2003. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17:** 2648–2663.

Hirschhorn JN, Bortvin AL, Ricupero-Hovasse SL, Winston F. 1995. A new class of histone H2A mutations in *Saccharomyces cerevisiae* causes specific transcriptional defects in vivo. *Mol Cell Biol* **15:** 1999–2009.

Huisinga KL, Pugh BF. 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.

Ingvarsdottir K, Krogan NJ, Emre NC, Wyce A, Thompson NJ, Emili A, Hughes TR, Greenblatt JF, Berger SL. 2005. H2B ubiquitin protease Ubp8 and Sgf11 constitute a discrete functional module within the *Saccharomyces cerevisiae* SAGA complex. *Mol Cell Biol* **25:** 1162–1172.

Janke C, Magiera MM, Rathfelder N, Taxis C, Reber S, Maekawa H, Moreno-Borchart A, Doenges G, Schwob E, Schiebel E, et al. 2004. A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21:** 947–962.

Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE, Wang C, Brindle PK, Dent SY, 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.

Johnsson A, Durand-Dubief M, Xue-Franzén Y, Rönnerblad M, Ekwall K, Wright A. 2009. HAT–HDAC interplay modulates global histone H3K14 acetylation in gene-coding regions during stress. *EMBO Rep* **10:** 1009–1014.

Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L. 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* **13:** 424.

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471:** 480–485.

Kohler A, Zimmerman E, Schneider M, Hurt E, Zheng N. 2010. Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. *Cell* **141:** 606–617.

Koutelou E, Hirsch C, Dent S. 2010. Multiple faces of the SAGA complex. *Curr Opin Cell Biol* **22:** 374–382.

Krebs A, Karmodiya K, Lindahl-Allen M, Struhl K, 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.

Lang G, Bonnet J, Umlauf D, Karmodiya K, Koffler J, Stierle M, Devys D, Tora L. 2011. The tightly controlled deubiquitination activity of the human SAGA complex differentially modifies distinct gene regulatory elements. *Mol Cell Biol* **31:** 3734–3744.

Lee K, Workman J. 2007. Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol* **8:** 284–295.

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

Lenstra T, Holstege F. 2012. The discrepancy between chromatin factor location and effect. *Nucleus* **3:** 213–219.

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, Margaritis T, van de Pasch LA, van Heesch SA, Brok MO, Groot Koerkamp MJ, et al. 2011. The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell* **42:** 536–549.

Lim S, Kwak J, Kim M, Lee D. 2013. Separation of a functional deubiquitylating module from the SAGA complex by the proteasome regulatory particle. *Nat Commun* **4:** 2641.

Minsky N, Shema E, Field Y, Schuster M, Segal E, Oren M. 2008. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nat Cell Biol* **10:** 483–488.

Rodríguez-Navarro S. 2009. Insights into SAGA function during gene expression. *EMBO Rep* **10:** 843–850.

Roudier F, Ahmed I, Berard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, et al. 2011. Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J* **30:** 1928–1938.

Samara NL, Datta AB, Berndsen CE, Zhang X, Yao T, Cohen RE, Wolberger C. 2010. Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science* **328:** 1025–1029.

Schulz D, Pirkl N, Lehmann E, Cramer P. 2014. Rpb4 functions mainly in mRNA synthesis by RNA polymerase II. *J Biol Chem* **289:** 17446–17452.

Schulze JM, Jackson J, Nakanishi S, Gardner JM, Hentrich T, Haug J, Johnston M, Jaspersen SL, Kobor MS, Shilatifard A. 2009. Linking cell cycle to histone modifications: SBF and H2B monoubiquitination machinery and cell-cycle regulation of H3K79 dimethylation. *Mol Cell* **35:** 626–641.

Schulze JM, Hentrich T, Nakanishi S, Gupta A, Emberly E, Shilatifard A, Kobor MS. 2011. Splitting the task: Ubp8 and Ubp10 deubiquitinate different cellular pools of H2BK123. *Genes Dev* **25:** 2242–2247.

Shema E, Tirosh I, Aylon Y, Huang J, Ye C, Moskovits N, Raver-Shapira N, Minsky N, Pirngruber J, Tarcic G, et al. 2008. The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev* **22:** 2664–2676.

Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, et al. 2011. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474:** 337–342.

Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Larivière L, Maier K, Seizl M, Tresch A, Cramer P. 2012. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* **22:** 1350–1359.

Venters B, Wachi S, Mavrich T, Andersen B, Jena P, Sinnamon A, Jain P, Rolleri N, Jiang C, Hemeryck-Walsh C, et al. 2011. A comprehensive genomic binding map of gene and chromatin regulatory proteins in *Saccharomyces*. *Mol Cell* **41:** 480–492.

Vermeulen M, Eberl H, Matarese F, Marks H, Denissov S, Butter F, Lee K, Olsen J, Hyman A, Stunnenberg H, et al. 2010. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142:** 967–980.

Vethantham V, Yang Y, Bowman C, Asp P, Lee JH, Skalnik DG, Dynlacht BD. 2012. Dynamic loss of H2B ubiquitylation without corresponding changes in H3K4 trimethylation during myogenic differentiation. *Mol Cell Biol* **32:** 1044–1055.

Vogelauer M, Wu J, Suka N, Grunstein M. 2000. Global histone acetylation and deacetylation in yeast. *Nature* **408:** 495–498.

Weake V, Workman J. 2012. SAGA function in tissue-specific gene expression. *Trends Cell Biol* **22:** 177–184.

Weake VM, Dyer JO, Seidel C, Box A, Swanson SK, Peak A, Florens L, Washburn MP, Abmayr SM, Workman JL. 2011. Post-transcription initiation function of the ubiquitous SAGA complex in tissue-specific gene activation. *Genes Dev* **25:** 1499–1509.

Wu P-YJ, Ruhlmann C, Winston F, Schultz P. 2004. Molecular architecture of the *S. cerevisiae* SAGA complex. *Mol Cell* **15:** 199–208.

Wyce A, Xiao T, Whelan KA, Kosman C, Walter W, Eick D, Hughes TR, Krogan NJ, Strahl BD, Berger SL. 2007. H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. *Mol Cell* **27:** 275–288.

Supplemental Material

Supplemental Experimental Procedures

Chromatin immunoprecipitation in HeLa cells

Cells were cross-linked with 1% formaldehyde at room temperature for 10 min and resuspended in swelling buffer (25 mM Tris-HCl pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% NP40, 1 mM DTT and protease inhibitor cocktail (Roche)) for 10 min at 4°C. After douncing, nuclei were resuspended in sonication buffer (50 mM Tris-HCl pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 1% SDS and protease inhibitor cocktail (Roche)) and sonicated using a Covaris machine until an average DNA fragment size of 200 bp was achieved. Supernatants were diluted in ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl and 16.7 mM Tris-HCl pH 8.0) followed by preclearing with non-saturated protein G-Sepharose (for H2Bub antibody) or protein A-Sepharose (for H3K9ac antibody). The pre-cleared samples corresponding to 50 µg of chromatin were incubated overnight at 4°C with either 2 µg of anti-H3K9ac or 4 µg of anti-H2Bub antibody, and then beads saturated with tRNA and fish skin gelatin were added for 4 h to pull down specific protein-DNA complexes. The following washes were carried out at 4°C: twice with low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 8.0), twice with high salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl, 20 mM Tris-HCl pH 8.0), twice with LiCl wash buffer (250 mM LiCl, 1% NP-40, 1% Na-deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0), and finally twice with Tris-EDTA buffer (10 mM Tris-HCl pH 7.5 and 1 mM EDTA). Bound chromatin was eluted with 300 µL of elution buffer (1% SDS and 0.1 M $NaHCO_3$) for 1 h at 65°C with shaking. RNase A treatment and cross linking reversion were performed overnight in 300 mM NaCl at 65°C. Immunopurified material was incubated with 20 µg proteinase K for 1 h at 42°C. DNA was phenol-chloroform extracted and precipitated with ethanol. Prior to DNA sequencing, purified material was used for qPCR validation using a Roche LightCycler 480 with SYBR green master mix (Qiagen or Roche).

Chromatin immunoprecipitation in *Saccharomyces cerevisiae* cells

Cells from log-phase yeast cultures were cross-linked with 1% formaldehyde at room temperature for 15 min and then quenched with 0.125 M glycine for 5 min. Cells were

harvested, washed with PBS and resuspended in FA lysis buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.1% sodium deoxycholate, 1% Triton X-100, 0.4 mM DTT, yeast protease inhibitor cocktail (Sigma)). Cells were lysed by mechanical shearing using acid- washed glass beads for 45 mins at 4°C. The pellets were collected, washed and then sonicated in FA lysis buffer using a Covaris machine to achieve an average DNA fragment size of 200 bp. For each ChIP reaction, 250 µg of chromatin extract were used. After pre-clearing with protein-G sepharose, chromatin extracts were incubated with antibodies overnight and then with protein-G sepharose saturated with tRNA and fish skin gelatin for 4 hrs. Antibodies used for each ChIP reaction: 4 µg α-H2Bub, 0.9 µg α-H3K4me3, 2 µg α-H3K9ac, 4 µl α-H2B, 5 µl α-RNA pol II. The resins were washed with the following buffers once at room temperature: FA lysis buffer, FA lysis buffer containing 0.5 M NaCl, LiCl-containing buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.25 M LiCl, 0.5% NP-40, 0.5% Sodium deoxycholate) and finally Tris-EDTA buffer (10 mM Tris-HCl pH 7.5 and 1 mM EDTA). Bound chromatins were eluted with 250 µl of elution buffer (50 mM Tris-HCl, pH7.5, 10 mM EDTA, 1% SDS) for 30 min at 65°C with shaking and then reverse cross-linked with RNase A treatment at 65°C overnight. Immunopurified material was incubated with 20 µg proteinase K for 1 h at 55°C. DNA was phenol-chloroform extracted and precipitated by ethanol.

For ChIP-qPCR experiments, three independent sets of H2Bub, H3K4me3 and H2B ChIP (WT vs *ubp8*Δ) and two independent sets of Pol II ChIP (WT vs *spt20*Δ or *spt7*Δ) were performed and analyzed by real-time qPCR (Roche). The primers used in the qPCR analyses are listed below.

Library preparation and sequencing

Library preparation for sequencing was performed as described by the manufacturer. Illumina systems were used for DNA sequencing. FASTQ files have been generated using CASAVA v1.8.2. ChIP-seq reads were aligned to the hg19 or sacCer3 genome assembly using the Bowtie aligner 0.12.8. Only sequences that mapped uniquely to the genome with maximum of two mismatches were used for further analyses.

Bioinformatic procedures

Creation of density files for genome browser data visualisation

Raw BED files were used as input to generate WIG files by WigMaker3, an in-house program developed by the microarray and sequencing platform of the IGBMC. Reads

are directionally extended of their theoretical length (200 bp), and 25 bp bins are created. In each bin, the maximal number of overlapping reads is computed. The different tracks were then uploaded and displayed on the UCSC genome browser (Kent et al. 2002).

Data analysis of ChIP-seq results obtained in human cells

Definition of different gene expression categories

ENCODE/CSHL RNA-seq data from HeLa-S3 cells (Long PolyA+ RNA) were downloaded from the UCSC server (UCSC accession: wgEncodeEH000173). The list of 51045 gene coordinates found in the Transcript Gencode V7 view of the RNA-seq data was considered as a reference list of human genes (Ensembl is used as a reference database). The scores (normalized average RPKM (Reads Per Kilobase per Million mapped reads) [((RPKM1 + RPKM2)/2) *1000]/max((RPKM1 + RPKM2)/2) found in this file were used to classify genes according to their expression level. Genes were considered as non-expressed if their score was lower than 0.031. The expressed genes were arbitrarily categorized into three subsets: genes with low (score between 0.031 and 0.1), medium (score between 0.1 and 1) and high expression levels (score higher than 1).

Definition of human intergenic region coordinates

The coordinates of intergenic regions were obtained, using Refseq as a reference database, with the Table Browser of the UCSC server (creation of an intersection between the complement coordinates of Refseq genes with themselves). In our analyses, intergenic regions are used to estimate the average background of each dataset. To consider only regions without any specific signal, we have selected 3369 intergenic regions that are larger than 100 kb and we have deleted 50 kb from each side. These non-overlapping intergenic regions cover in total ~1000 Mb.

Calculation of density values

The number of aligned reads per region of interest was determined with the option "Enrichment Based method" of the program SeqMINER using BED files as input (Ye et al. 2011). Note that the end of each aligned read was extended to 200 bp in the direction of the read by the program. Density values were then defined as follows: density = [(number of aligned reads in a region of interest)/ (length of the region of interest in Mb)] / (total number of aligned reads in the dataset * 10^{-8}).

For the analysis of H3K9ac (shCtrl and shADA3) and RNA Pol II (GSM935395) datasets, the regions of interest correspond to a window surrounding each TSS from -1.5 kb to + 3.5 kb (called TSS window). The H2Bub densities (datasets: shCtrl and shATXN7L3) were calculated on the entire gene body (from the transcription start site to the transcription termination site). For all datasets, intergenic regions were used to estimate the density of the background.

Criteria of filtration

38061 genes (from our reference list of human genes) which are not included in larger genes were taken into account. TSS windows and intergenic regions with a non-null density were considered for each subsequent analysis. For H2Bub datasets (shCtrl and shATXN7L3), we considered only gene bodies and intergenic regions containing more than 20 reads.

Generation of the plots

Scatter plots and boxplots were generated with R version 3.0.2. R scripts are available upon request.

Normalization between datasets

When two datasets were compared, the ChIP experiments were performed together with the same antibody and the DNA samples were also sequenced in parallel. We then considered that the background signal for the two experiments was equal. For the scatter plots created in R, the ratio of the densities at intergenic regions was calculated and the median value of this population was used as a normalization factor. To visualize our data on the genome browser, modified BED files were generated to obtain the same number of reads on intergenic regions in the two datasets that were compared (shCtrl vs shATXN7L3 and shCtrl vs shADA3).

Generation of average profiles and heat maps

Average profiles and k-mean clustering were generated with the program SeqMINER (Ye et al. 2011). The end of each aligned read was extended to 200 bp in the direction of the read. For the analyses around promoters, the tag density was extracted in a 10 kb window centred on each TSS. For average gene profiles, each gene body was divided into 160 equal bins (the absolute size depending on the gene length). Moreover 20 equally sized bins (250 bp) were created upstream and downstream of genes. Densities were collected for each dataset in each bin. For these analyses, promoters or gene bodies of 9630 expressed genes (normalized average RPKM >

0,031) with a significant RNA Pol II density at their TSS (more than 111 reads in the TSS window in the GSM935395 dataset, see also Supplemental Fig. S1) were used.

Data analysis of ChIP-seq results obtained in *S. cerevisiae*

Coordinates of yeast genes

The coordinates of yeast open reading frames from the Saccharomyces Genome Database (SGD) were downloaded on the UCSC Table Browser. 4927 unique open reading frames annotated as "Verified" by SGD, with non-null densities of H3K9ac and H3K4me3 densities at their promoter (First codon +/- 200 bp) and with an entry in the gene expression analysis from Lenstra et al. (Lenstra et al. 2011) (excluding the mitochondrial genes) were used as a reference list for yeast genes.

Calculation of density values

Density values were determined as described for the human data except that yeast values were divided by an additional factor (266; corresponding to the ratio between the size of the human and the yeast haploid genomes), in order to get densities in the same order of magnitude for the human and yeast datasets. For the analysis of H3K9ac (WT and *gcn5*Δ) and H3K4me3 (WT) datasets, densities were calculated on regions surrounding the transcription start site of genes (First codon +/- 200 bp) and on mid-gene bodies of large genes (from 25% to 75% of coding regions larger than 2 kb). The H2Bub, H2B and RNA Pol II densities (in WT, *ubp8*Δ or *spt20*Δ strains) were calculated on the entire coding region (from the first codon to the stop codon).

Criteria of filtration

Out of the 4927 yeast genes that are present in our reference list, 3916 were defined as active genes by the presence of significant H3K9ac and H3K4me3 signals at their promoter (First codon +/- 200 bp) (see Supplemental Fig. S6D). 2606 genes were further selected as genes with highly significant levels of Pol II on their coding regions and were taken into account in the Pol II analysis (see Fig. 6 and Supplemental Fig. S7B).

Generation of the plots

Scatter plots were generated with R version 3.0.2. R scripts are available upon request.

Normalization between datasets

To compare the different H3K9ac datasets, we used 1184 mid-gene bodies of large genes (>2 kb) which cover about 15% of the yeast genome (~1.8 Mb) and which have

background levels of H3K9ac that should be equal in all conditions. As a normalization factor, we used the median value of the ratios of H3K9ac densities between the two datasets of interest in these regions. In parallel, for the visualization of the H3K9ac datasets on a genome browser, normalized BED files were generated by removing randomly reads from the raw BED files of interest to obtain the same number of aligned reads on the mid-gene bodies of large genes for all conditions (WT_1, WT_*2, gcn5*Δ_1 and *gcn5*Δ_2).

291 genes (covering 3.5% of the yeast genome (0.45 Mb)) without any H3K9ac and H3K4me3 peak at their TSS and with background levels of Pol II on their transcribed region in WT conditions (Fig S7B) were used to compare the different Pol II datasets (WT_1, WT_2 and *spt20*Δ). Assuming that the background levels of Pol II should be similar in all conditions, Pol II datasets were normalized as described above for H3K9ac datasets using these 291 genes as a reference.

For the two H2Bub ChIP-seq datasets that are compared in this study (WT vs *ubp8*Δ), it was not possible to characterize regions containing only background levels of H2Bub. We therefore generated a normalized BED file for the H2Bub_*ubp8*Δ dataset containing 2.5 times more aligned reads than the H2Bub_WT BED file (such a difference of 2.5 in H2Bub levels between the two strains was observed in three independent H2Bub ChIP-qPCR experiments).

Supplemental Figures

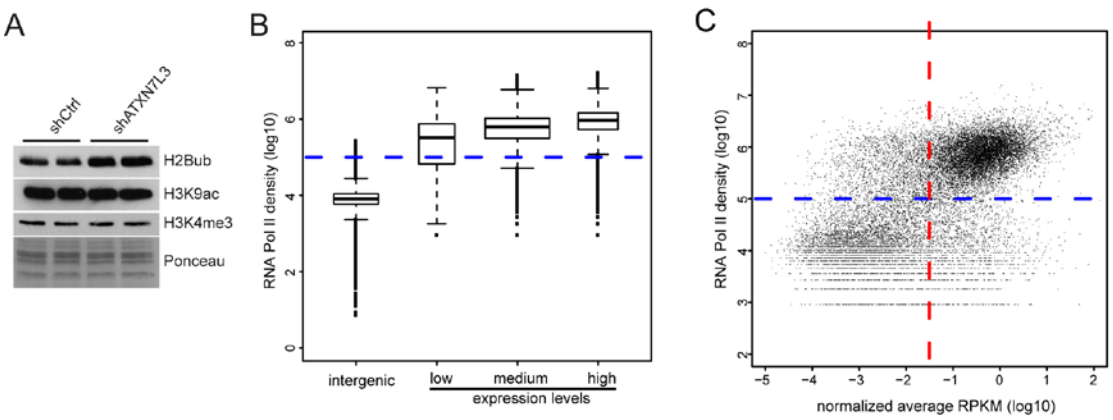


Figure S1

Selection of human genes with a significant RNA Pol II peak at their promoter, related to Figure 1

(A) Western blot analyses of changes in H2B ubiquitination upon inactivation of the SAGA DUB activity in HeLa cells. Total histones were purified by acidic extractions from HeLa cells expressing an *ATXN7L3* shRNA and immunoblotted with the indicated antibodies. (**B**) Boxplot representation of RNA Pol II density (in HeLa cells) on 3189 intergenic regions and on genes with low (2076 genes), medium (6692 genes) or high (2558 genes) expression levels. The blue dotted line defines the threshold for significant RNA Pol II density around the TSS (TSS - 1.5 kb / + 3.5 kb) (99.5% of intergenic regions have lower RNA Pol II density). **(C)** Gene expression levels (based on normalized average RPKM) were plotted relative to the RNA Pol II density around the TSS of 18838 genes with non-null Pol II density and non-null normalized average RPKM. The threshold values for significant RNA Pol II density and for significant gene expression are indicated by a blue and a red dotted line, respectively. The vast majority of expressed genes have a significant Pol II density at their promoter (upper right corner) whereas most non-expressed genes are devoid of Pol II (lower left corner).

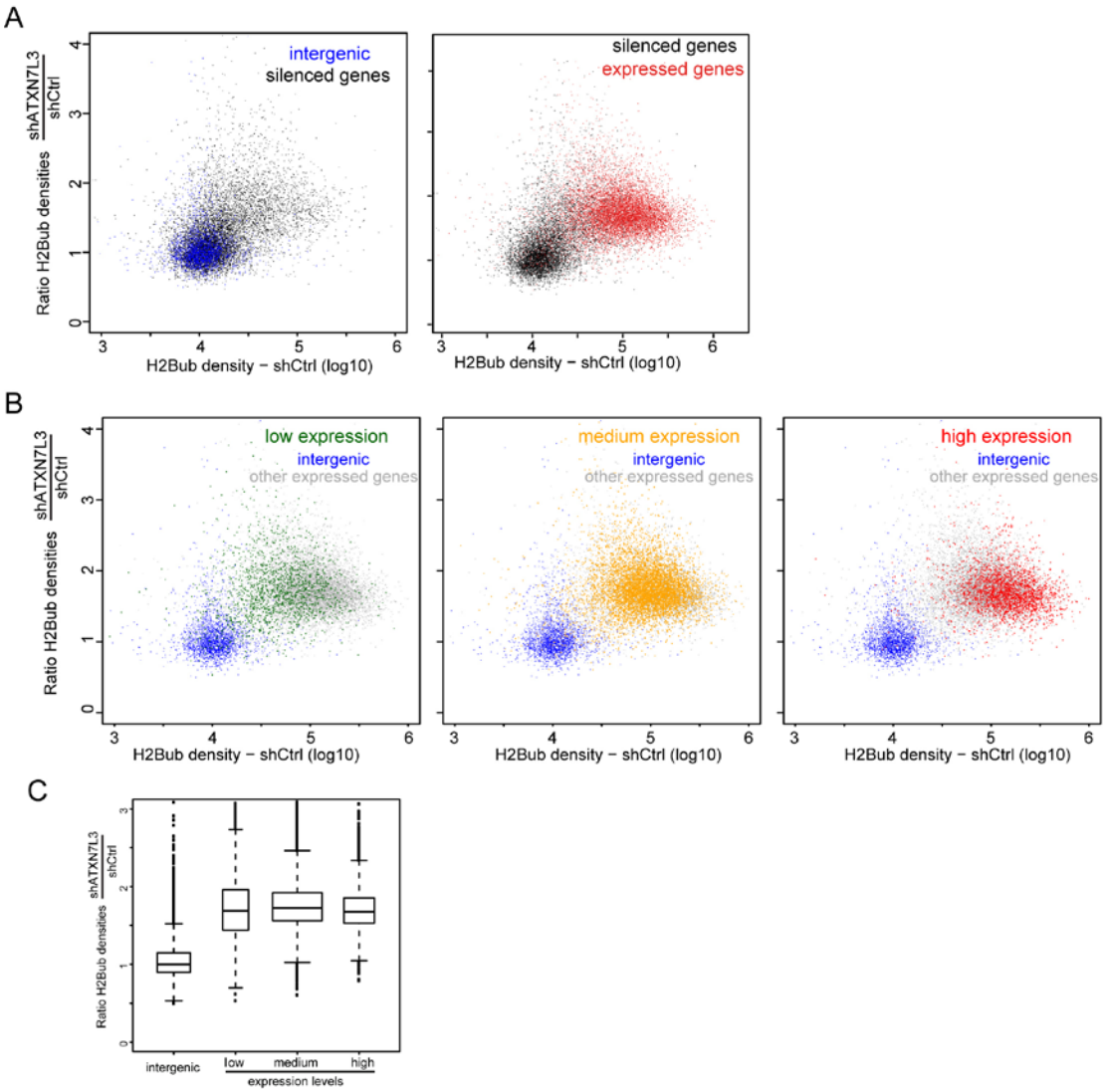


Figure S2

SAGA deubiquitinates all expressed genes in HeLa cells, related to Figure 1

(A,B) Scatter plots representing the H2Bub density in control cells relative to SAGA DUB activity (Ratio of H2Bub densities in shATXN7L3 and shCtrl cells). **(A)** The distribution of the 8811 silenced genes (black dots) largely overlaps that of the 2849 intergenic regions (blue dots, left panel) but strikingly differs from that of 10486 expressed genes (red dots, right panel). Non-expressed genes with higher H2Bub densities and SAGA DUB activity may contain H2Bub signal from nearby expressed genes. **(B)** 2849 intergenic regions (blue) are compared with 10486 expressed genes (grey). Genes with low (1722 genes), medium (6337 genes) or high (2427 genes) expression are highlighted in green, yellow and red (in the left, middle and right panel, respectively). (**C**) Boxplot representation of SAGA DUB activity at intergenic regions and at genes with low (1722 genes), medium (6337 genes) or high (2427

genes) expression. Expressed genes had similar distribution of H2Bub ratios indicating that the SAGA DUB activity is proportional to the local H2Bub levels regardless of the expression levels of the target genes.

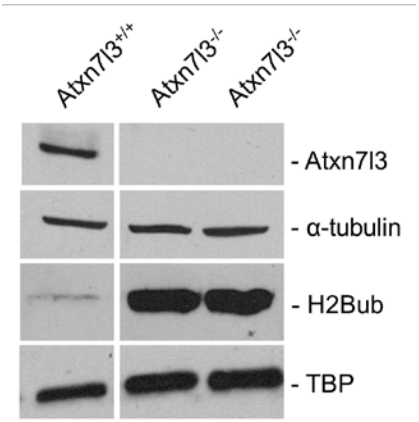


Figure S3

Characterization of *Atxn7l3* knockout mESCs, related to Figure 2.

Whole cell extracts (two upper panels) and acidic extractions (two lower panels) from WT (*Atxn7l3*$^{+/+}$) and *Atxn7l3* inactivated (*Atxn7l3*$^{-/-}$) mESCs were immunoblotted with the indicated antibodies. Atxn7l3 was expressed in WT mESCs but was undetectable in *Atxn7l3*$^{-/-}$ mESCs. H2Bub global levels were strikingly increased upon Atxn7l3 deletion.

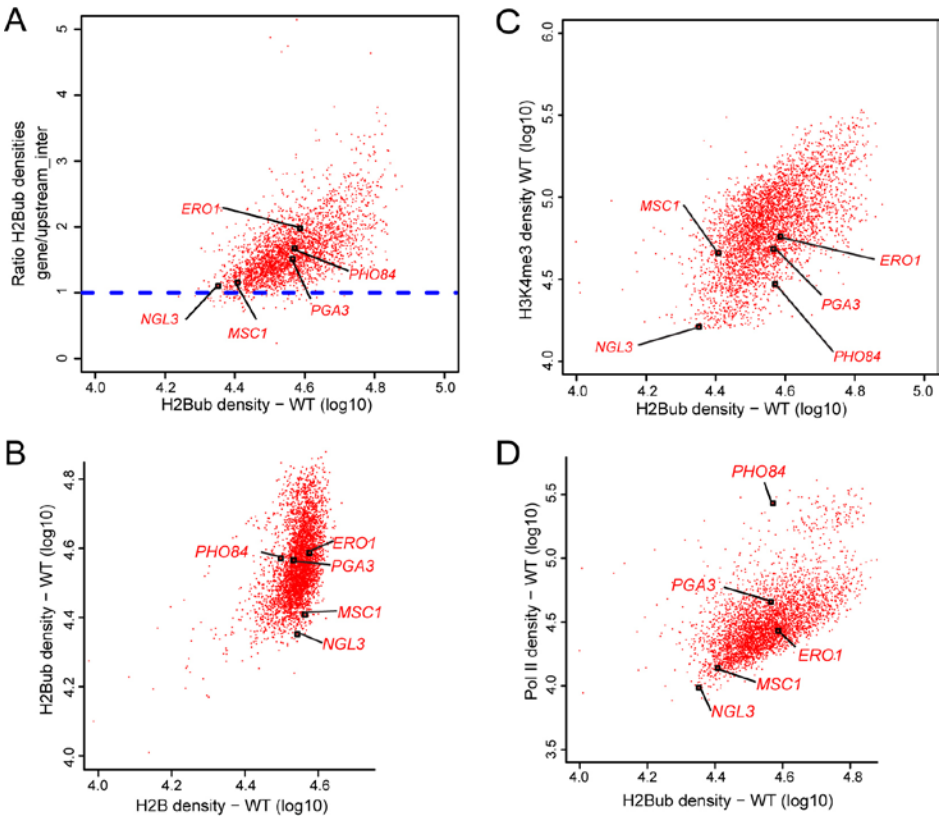


Figure S4

Characterization of H2Bub distribution in budding yeast, related to Figure 3

(A) H2Bub is enriched on the transcribed region of almost every gene in WT cells. H2Bub densities on coding regions of WT cells are represented relative to the fold enrichment of H2Bub on coding regions as compared to the corresponding upstream intergenic region (Ratio of H2Bub densities genes/upstream_inter). For most of the genes, the density of H2Bub is higher on the gene body than on the upstream intergenic region (Ratio H2Bub densities gene/upstream_inter > 1). This suggests that even the genes with low H2Bub densities (eg *NGL3*) harbor a significant level of the mark on their transcribed region. 2937 active genes with a H2B density higher than 4.3 (on the log10 scale) and with an enrichment of H2B on the coding region of at least 1.15 fold were used for this analysis. Representative genes from Fig. 3 are indicated. (**B**) This scatter plot representing H2Bub density versus H2B density reveals a high variation of H2Bub levels between genes whereas H2B occupancy is much more homogenous. On five selected genes, the H2Bub levels were low on two (*MSC1* and *NGL3*) and higher on the three others (*ERO1*, *PGA3* and *PHO84*), whereas these genes had similar H2B occupancy. **(C,D)** Scatter plots depicting a

correlation between H2Bub and H3K4me3 (**C**) or RNA Pol II (**D**) respective enrichment on 3916 genes with a significant peak of H3K9ac and H3K4me3 at their TSS and non-null H2Bub density. The Pearson correlation between H2Bub density and H3K4me3 or Pol II densities were 0.61 and 0.56, respectively, indicating a link between H2B ubiquitination and gene expression. However, similarly to our observations in human cells, the local variations of H2Bub densities cannot be explained exclusively by the transcription rate of a gene.

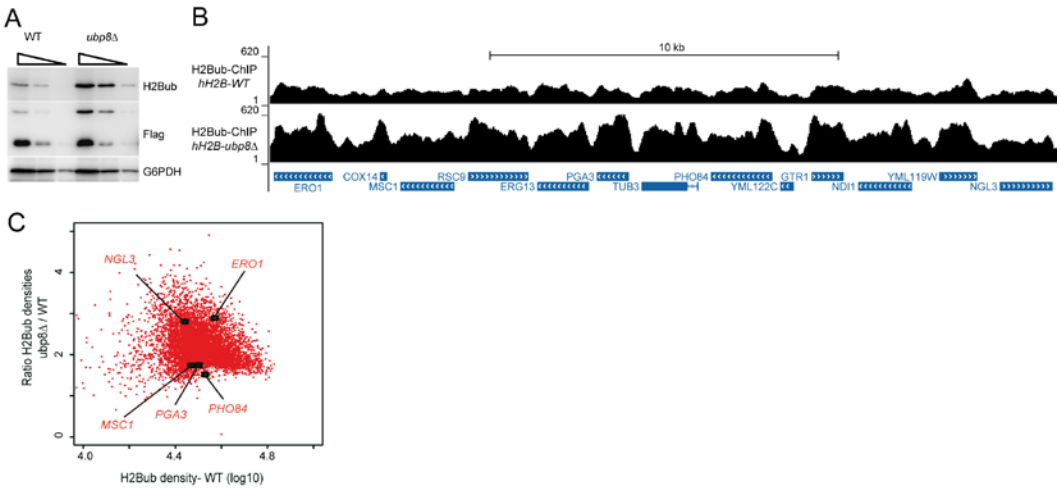


Figure S5

Genome-wide analysis of H2Bub enrichment in *ubp8Δ S. cerevisiae* using *S. pombe* chromatin as spike-in control, related to Figure 3.

(**A**) 3-fold serial dilutions of lysates from stains expressing a humanized H2B and deleted for *UBP8* were immunodetected with the indicated antibodies. **(B)** Genome browser tracks showing increased H2Bub levels in *ubp8Δ* cells as compared to the parental hH2B expressing cells. **(C)** Scatter plot representing the SAGA DUB activity (ratio of H2Bub densities on gene bodies between hH2B-*ubp8*Δ and hH2B-WT strains) relative to H2Bub densities in WT cells. Analysis was performed on 4927 genes. ChIP-seq experiments were performed by using 25 µg of *S. pombe* chromatin mixed with 250 µg of chromatin extracted from hH2B-WT or hH2B-*ubp8*Δ *S. cerevisiae* strains. The two datasets were normalized by the total number of reads aligned to the *S. pombe* genome and revealed a 2.1 average increase of H2Bub levels upon *UBP8* deletion.

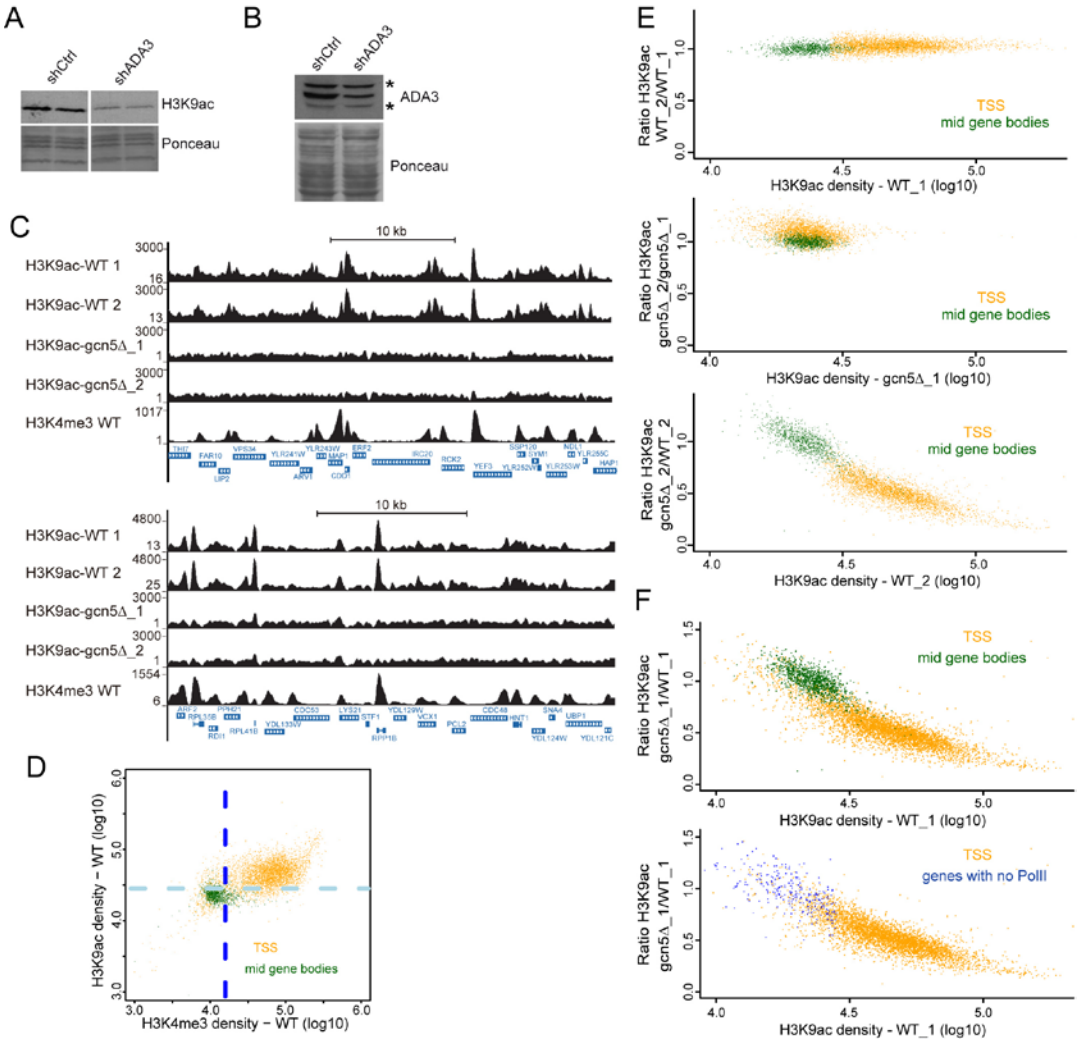


Figure S6

Gcn5-containing complexes acetylate H3K9 at the promoter of all expressed genes in yeast and human cells, related to Figure 4 and 5

(A,B) Western blot analyses of changes in H3K9 acetylation upon inactivation of SAGA HAT activity in HeLa cells expressing an *ADA3* shRNA. Acidic extracts (**A**) and whole cell extracts (**B**) were immunoblotted with the indicated antibodies. Equal loading was verified by Ponceau staining. **(C)** Genome browser tracks of two representative genomic regions reveal a high degree of similarity between the two replicates of H3K9ac ChIP-seq in WT (H3K9ac-WT_1 and WT_2) and in *gcn5*Δ (H3K9ac-*gcn5*Δ_1 and *gcn5*Δ_2) yeast cells. Highly similar H3K9ac peaks are detected at the promoter of expressed genes in the two wild-type replicates but are absent in the two *gcn5*Δ datasets. Expressed genes are characterized by the presence of a peak of H3K4me3 at their promoter (H3K4me3-WT). **(D)** Scatter plot showing H3K4me3 densities relative to H3K9ac densities in WT cells in 1184 mid-gene

bodies of large genes as defined in Fig. 5C (green dots) and in 4927 regions surrounding the TSS (First codon +/- 200 bp) of genes with non-null H3K9ac and H3K4me3 densities (yellow dots). The dark and light blue dotted lines define the thresholds for significant H3K4me3 and H3K9ac densities at promoters. These thresholds separate 3916 genes with high levels of H3K4me3 and of H3K9ac (upper right corner) from 395 genes with low H3K4me3 and H3K9ac densities (lower left corner) similarly to most of mid-gene bodies. **(E)** Global comparison of the two WT and the two *gcn5*Δ replicates. In 3916 regions surrounding the TSS (First codon +/- 200 bp) of expressed genes as defined by the presence of H3K9ac and H3K4me3 peaks (yellow dots) and in 1184 mid-gene bodies of large genes (green dots), the ratios of H3K9ac densities for each couple of replicates (WT_2/WT_1 and *gcn5*Δ_2/*gcn5*Δ_1) are plotted relative to the H3K9ac densities of the first WT and *gcn5*Δ replicates, respectively (upper and middle panel). In the lower panel, H3K9ac densities from the second wild-type replicate (WT_2) are plotted versus SAGA acetylation activity (Ratio H3K9ac densities *gcn5*Δ_2/ WT_2). This graph is very similar to the panels in Fig. 5C in which the first replicates WT_1 and *gcn5*Δ_1 were used. **(F)** Scatter plots representing SAGA acetylation activity (Ratio H3K9ac densities *gcn5*Δ_1/WT_1) versus H3K9ac densities in WT (WT_1) in 4927 regions surrounding the TSS (First codon +/- 200 bp) of genes with non-null H3K9ac and H3K4me3 densities (yellow dots). 1184 mid gene bodies (green dots) are also represented (upper panel) and 291 genes without any H3K9ac, H3K4me3 and Pol II significant enrichment are colored in blue (lower panel). The average position of the mid gene bodies and the 291 genes with no or low transcription looks very close validating the usage of these genes to normalize the different Pol II datasets in Fig. 6 (see also Fig. S7).

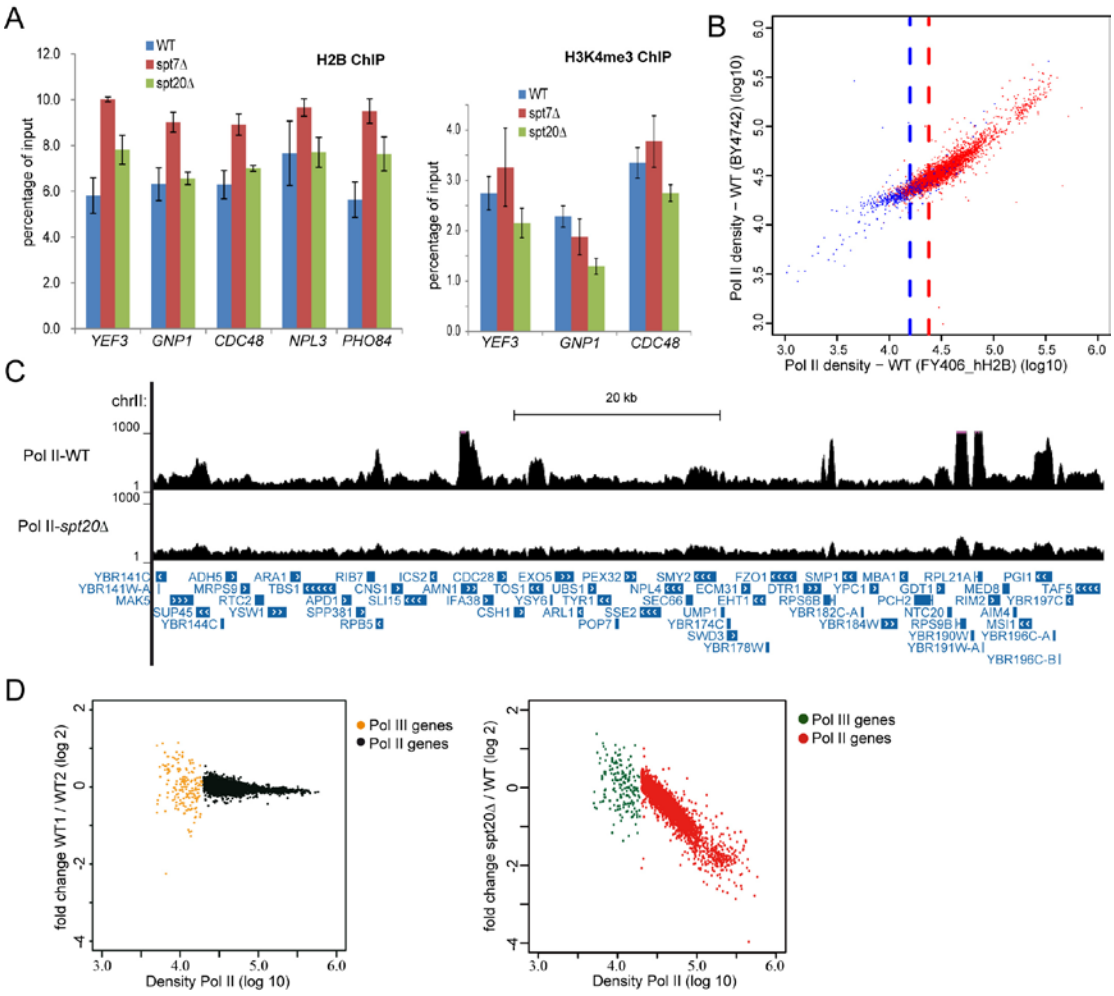


Figure S7

Analysis of the genome-wide Pol II occupancy, related to Figure 6

(A) Comparable enrichment for H3K4me3 and H2B were detected in WT, *spt7Δ* and *spt20Δ* strains as analyzed by ChIP-qPCR experiments performed on the same chromatin preparations as those used in Fig. 6. **(B)** Selection of genes based on their Pol II densities in WT cells. Scatter plot depicting the density of Pol II in two different WT backgrounds (FY406_hH2B and BY4742). The active genes with peaks of H3K9ac and H3K4me3 at their promoter are represented in red and the genes devoid of these two marks are in blue (see Supplemental Fig. S6D). To consider only the genes with background or with highly significant levels of Pol II, two additional thresholds based on Pol II densities in the FY406_hH2B WT strain were defined. The 291 genes with the lowest Pol II densities (blue dots to the left of the blue dotted line) were considered as genes carrying background levels of H3K9ac, H3K4me3 and Pol II. In addition, 2606 genes (red dots to the right of the red dotted line) with a Pol II

density 1.5 fold higher than the threshold defining background levels of Pol II were selected and used for the analysis of the Pol II ChIP-seq datasets. **(C)** Genome browser tracks of a representative yeast genomic region showing that RNA Pol II peaks detected in wild-type cells (Pol II-WT) are absent when *SPT20* is deleted (Pol II-*spt20*Δ). **(D)** Genome-wide analysis of Pol II occupancy in *spt20*Δ and WT strains, using Pol III-transcribed genes as a normalization factor. Scatter plot representing RNA Pol II occupancy (Fold changes of Pol II densities in *spt20*Δ compared with WT in log2 value). Pol II ChIP performed on two independent WT chromatin preparations were highly reproducible (left panel). A global loss of Pol II recruitment in *spt20*Δ cells (right panel) is observed at 4927 verified genes. RNA Pol II densities at Pol III-transcribed genes were used to normalize the two datasets.

Supplemental Tables

Table S1

Description of the high throughput sequencing datasets (GSE59370), related to Fig. 1,3,4,5 and 6

Antibody	Species / Cell type	Platform	Aligned reads	Geo accession ID
S. cerevisiae **(GSE52997)**				
input	WT-yH2B (YCW027)	IGBMC	36702996	GSM1279925
input	WT-hH2B (YCW012)	IGBMC	55002283	GSM1279926
input	*ubp8*Δ-hH2B (YCW015)	IGBMC	40407638	GSM1279927
input	*gcn5*Δ (YCW037)	IGBMC	37156891	GSM1279928
input	BY4742-WT	IGBMC	12651502	GSM1437203
input	BY4741-*spt20Δ* (YCW048)	IGBMC	11495154	GSM1437204
H3K9ac	WT (replicate 1) (YCW027)	IGBMC	43487417	GSM1279929
H3K9ac	WT (replicate 2) (YCW027)	IGBMC	46805328	GSM1279930
H3K9ac	*gcn5*Δ (replicate 1) (YCW037)	IGBMC	51094282	GSM1279931
H3K9ac	*gcn5*Δ (replicate 2) (YCW037)	IGBMC	52904651	GSM1279932
H2Bub	WT-yH2B (YCW027)	IGBMC	3767391	GSM1279933
H2Bub	WT-hH2B (YCW012)	IGBMC	8538243	GSM1279934
H2Bub	*ubp8*Δ-hH2B (YCW015)	IGBMC	23491688	GSM1279935
H2B	WT-hH2B (YCW012)	IGBMC	90242433	GSM1279936
H3K4me3	WT-hH2B (YCW012)	IGBMC	35858331	GSM1279937
RNA Pol II	WT-hH2B (YCW012)	IGBMC	19908756	GSM1279938
RNA Pol II	BY4742-WT (replicate1)	IGBMC	13279022	GSM1437200
RNA Pol II	BY4742-WT (replicate2)	IGBMC	12606125	GSM1437201
RNA Pol II	BY4741-*spt20Δ* (YCW048)	IGBMC	8315249	GSM1437202
Human cells **(GSE52860)**				
H3K9ac	HeLa shCtrl	IGBMC	37489532	GSM1277114
H3K9ac	HeLa shADA3	IGBMC	36739501	GSM1277115
H2Bub	HeLa shCtrl	IGBMC	22742054	GSM1277116
H2Bub	HeLa shATXN7L3	IGBMC	23207032	GSM1277117
RNA Pol II	HeLa-S3	Yale	22079975	GSM935395

H3K4me3	HeLa-S3	Broad	19815009	GSM733682
H3K9ac	HeLa-S3	Broad	21594880	GSM733756
H3K36me3	HeLa-S3	Broad	21594880	GSM733711

Table S2

Primers used for the qPCR experiments, Related to Fig. 3,6,7

Region	Name	Sequences
VTC3	VTC3-5F	CGACTATGAAAGGTTGAAGAAAT
	VTC3-5R	CAACTATCCACGGAACTACG
BDF2	BDF2-mF	CTGAAGAAAATGGAGGTTGAAT
	BDF2-mR	CTTCCTCTTCCTTTCCTTCG
TEL3	TEL3L-F	AATATCTATATCTCATTCGGCGAC
	TEL3L-R	AGTGGTTAGTATATGGTGTAAAAGT
EBP2	EBP2-5F	ATATCAAAGCCAAGCCCT
	EBP2-5R	TTCGGTAGCGTCCTGTT
GCN4	GCN4-F	CACCTGATGCCAAGACC
	GCN4-R	GACGAAAAGAAAGATTCCACTAC
RLP24	RLP24-F	TTGGCTGTGGATTCTACCTTGA
	RLP24-R	GTGGCAACTAACTCTCTATTGT
NGL3	NGL3-F	GCCCAGCACATCAAAAAATAC
	NGL3-R	CTGTCTCTTCTGAGCACG
ERO1	ERO1-F	TTTCTTCAAATACTTTCGGCTGG
	ERO1-R	GACATCAACAGAGCAAGCG
MSC1	MSC1-F	CAGCAGTAGCGTGAGCAA
	MSC1-R	CCAGGTGTCAAAGAGCC
TEL6R	TEL6R-F	CGTGTGTAGTGATCCGAACTCAGT
	TEL6R-R	GACCCAGTCCTCATTTCCATCAATAG
IntI	IntI-F	GCCCTGATCTACGCATAAT
	IntI-R	GTTTTCCTCTCTTTTCCCGA
IntV	IntV-F	GTTACCGACATTGGAGGAT
	IntV-R	TACCGTCAAGATTCAACCATAG
YEF3	YEF3-5F	AGAAGTTATCTGTTGCCACTG
	YEF3-5R	TACCATTCAAGAAAGAAGCGAC
GNP1	GNP1-5F	CGTAATGGGAAACATCGTC
	GNP1-5R	TGGGCGGAATAATGAGGG
CDC19	CDC19-F	CCAAAGACCAACAACCC
	CDC19-R	ATTCGTAAGAACCGTGAGAG
PDC1	PDC1-F	ATTCACCGACACCGAAG

	PDC1-R	TTACGCCGCTGATGGTT
STE2	STE2-F	GACTTACGCTCTCACCG
	STE2-R	AGAAGCCACAAGAAGGAC
PGK1	PGK1-F	AGCGTGTCTTCATCAGAG
	PGK1-R	TGGCAAAGCAGCAACAA
NPL3	NPL3-F	CACCACCGTCAAGAAGGA
	NPL3-R	CAAAGATTTCATTCAACTCGGAT
ACT1	ACT1-F	ATTTCAAGCCCCTATTTATTCCA
	ACT1-R	CCTTTTTTTACTATTTTTCACTCTCCC
RPS30A	RPS30A-F	GCAACAGCGTGGACTTA
	RPS30A-R	AACTTCACTCTGCTATTTTCTCA
HYP2	HYP2-F	TTGAAACTGCTGACGCT
	HYP2-R	TCTTGATGACAACGAAACCG
RPS3	RPS3-F	ATTGTTGAACGGTTTGGC
	RPS3-R	CCCTTAGCACCAGATTCCATA
PMA1	PMA1-5F	CTCATCAGCCAACTCAAGAAA
	PMA1-5R	CGTCATCGTCAGAAGATTCA
PHO84	PHO84-5F	GTGTTGGTTTCTTGACAGATTC
	PHO84-5R	GCATACTACCGTGCCAG
RDN58-2	RDN58-2_F	AACGGATCTCTTGGTTCTCG
	RDN58-2_R	GTGCGTTCAAAGATTCGATG
RDN25-2	RDN25-2_F	TGGCAGTCAAGCGTTCATAG
	RDN25-2_R	CGCTTACCGAATTCTGCTTC
snR6	snR6_F	CGAAGTAACCCTTCGTGGAC
	snR6_R	TCATCCTTATGCAGGGGAAC
scR1	scR1_F	CCTTTGGGCAAGGGATAGTT
	scR1_R	TTTACGACGGAGGAAAGACG
RPR1	RPR1_F	TTGTTCCGTTTGACTTGTCG
	RPR1_R	TGGAACAGCAGCAGTAATCG
***S. pombe* tubulin**	Tubulin_F	CCGCTGGTGGAAAAGTATGTT
	Tubulin_R	GCCAATTCAGCACCTTCAGT

Table S3

Genotypes of the yeast strains

Name	Strain	Genotype	Reference
FY406	FY406	*MAT***a** *(hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1* l*ys2-128δ his3Δ200 trp1Δ63* [*pSAB6-(HTA1-HTB1, URA3)*]	Hirschhorn et.al., 1995

BY4742	BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	Euroscarf
WT-yH2B	YCW027	*MAT***a** *(hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63* [*pRS413-(HTA1-Flag-HTB1,CEN, HIS)*]	this study
WT-hH2B	YCW012	*MAT***a** *(hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63* [*pRS413-(HTA1-Flag-HTB1,CEN, HIS)*]	this study
ubp8Δ-hH2B	YCW015	*MAT***a** *ubp8Δ::KANMX4 (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2 128δ his3Δ200 trp1Δ63* [*pRS413-(HTA1-Flag-HTB1,CEN, HIS)*]	this study
gcn5Δ	YCW037	*MAT***a** *gcn5Δ::HPH (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63* [*pRS413-(HTA1-Flag-HTB1,CEN, HIS)*]	this study
BY4741-*spt7Δ*	YCW047	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 spt7Δ::KANMX4*	SGD deletion library
BY4741-*spt20Δ*	YCW048	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 spt20Δ::KANMX4*	SGD deletion library
FY406-*spt7Δ*	YCW060	*MAT***a** *spt7Δ::KAN MX4(hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63* [*pRS413-(HTA1-Flag-HTB1,CEN, HIS)*]	this study
FY406-*spt20Δ*	YCW046	*MAT***a** *spt20Δ::HPH (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63* [*pRS413-(HTA1-Flag-HTB1,CEN, HIS)*]	this study

* hHTB1 = HTB1-R119K-S125T

Supplemental References

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996-1006.

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, Margaritis T, van de Pasch LA, van Heesch SA, Brok MO, Groot Koerkamp MJ et al. 2011. The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell* **42**: 536-549.

Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, Tora L. 2011. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res* **39**: e35.

2. SAGA is a general cofactor for RNA Polymerase II transcription. (Baptista *et al.*; Molecular Cell, 2017)

The regulation of eukaryotic transcription involves the tightly regulated assembly of the PIC for efficient RNA Pol II positioning and activity. A key component of the general transcription machinery is TBP, which can be either loaded on promoters as part of the multi-subunit TFIID into TFIID-dominated genes or through the interaction with SAGA complex at SAGA-dominated genes. In fact, many studies have suggested that these two specific classes of genes had specific features:

– TATA-containing promoters are often highly regulated by stress, are typically modulated by chromatin remodeling and/or covalent histone modifications and are often classified as Taf1-depleted and SAGA-dominated;

– Promoters lacking a consensus TATA box are typically in the housekeeping class, have the +1 nucleosome adjacent to the site of PIC formation, are enriched in H4 acetylation, and nearly always classified as Taf1- enriched and TFIID-dominated.

However, results from our laboratory, and presented here as part of this thesis, have indicated that SAGA is capable of actively modify the chromatin genome-wide and is required for optimal RNA Pol II recruitment. Additionally, analysis of newly-transcribed mRNA in a reduced number of genes has indicated that SAGA is necessary for the transcription of both SAGA- and TFIID-dominated genes (Bonnet *et al.*, 2014).

Having that, collaborators and I decided to address and clarify past results regarding the differential contribution of SAGA for RNA Pol II transcription:

- To address **aim (a)**, in collaboration with Sebastian Grünberg from Steven Hahn's group, we performed analysis of genome-wide recruitment of SAGA using ChEC-seq;
- To address **aim (b)** and **(c)**, I performed analysis of both steady-state and newly-synthesized mRNA levels upon deletion of SAGA-specific subunits involved in the structural integrity of the complex (*spt7*Δ and *spt20*Δ);
- To address **aim (d)**, I extended the analysis of changes in the budding yeast transcriptome (both steady-state and newly-synthesized mRNAs) to SAGA mutants depleted for several of its activities. For that, I generated mutants either lacking only

one of the activities (*ubp8*Δ, *gcn5*Δ, *spt3*Δ and *spt8*Δ) or deleted for more than one activity simultaneously (*ubp8*Δ*gcn5*Δ and *spt3*Δ*gcn5*Δ);

- Still in line with **aim (d)** and to partially address **aim (e)**, I performed ChIP-qPCR of TBP in a SAGA mutant strain where interaction with TBP has been abrogated (*spt3*Δ).

Among the different obtained results, we observed that SAGA is recruited to the UASs at a majority of yeast genes. In good agreement, SAGA was found to be required for mRNA synthesis of essentially all genes transcribed by RNA Pol II. We additionally observed a compensatory increase of the half-life of a majority of mRNAs upon SAGA depletion explaining the limited changes in steady-state mRNA levels in the different SAGA mutant strains. I further analyzed the relative contributions of the different SAGA activities to Pol II transcription, thereby revealing a synergistic role for Spt3 and the acetyltransferase Gcn5. Altogether, our data lead us to propose that both SAGA is a general cofactor for PIC recruitment that is required for transcription of the vast majority of Pol II genes (Baptista *et al.*, 2017).

These results were submitted and accepted for publication in *Molecular Cell* (online publishing: 14th of September 2017; *in print* date: 5th of October 2017). Below are detailed the contributions of every author.

Author's contributions

SAGA is a general cofactor for RNA Polymerase II transcription. T Baptista, S Grünberg, N Minoungou, MJE Koster, HTM Timmers, S Hahn, D Devys and L Tora.

This work as a collaborative effort between Steven Hahn's and our groups. Specifically, the laboratory of Steven Hahn produced and analyzed the ChEC-seq data (Figures 1 and 2 and Supplementary Figure 1). I was responsible for the execution of all the experiments presented in the manuscript (Figures 3 to 7 and Supplementary Figures 2 to 7), except for the recruitment analysis. Hence, I generated several of the strains used in the present work, I established (and optimized) the protocol for the analysis of newly-transcribed mRNA, and performed all the remaining experiments described in the manuscript (ChIP-PCR, RNA half-life quantifications, immunoprecipitations, among others). I was also in charge of analyzing and interpreting the data acquired in our laboratory. Additionally, I was involved in the design and conceiving of this work, together with Didier Devys and the rest

of the authors. In the end, and once all the data was collected, the manuscript was written with the input of all the authors, with the bulk of the writing being done by Didier Devys and I. Specifically:

Tiago Baptista –Designed the study. Performed all experiments except ChEC-seq analyses. Performed computational analyses for all the experiments, except ChEC-seq. Analyzed and interpreted the data. Wrote the manuscript.

Sebastian Grünberg – Designed the study. Performed ChEC-seq experiments and computational analysis. Analyzed and interpreted the data. Wrote the manuscript.

Nadège Minoungou – Assisted on ChEC-seq experiments;

Maria JE Koster – Provided anchor-away strain.

HT Marc Timmers – Provided anchor-away strain.

Steven Hahn – Conceived the study. Wrote the manuscript.

Didier Devys – Conceived the study. Analyzed and interpreted the data. Wrote the manuscript.

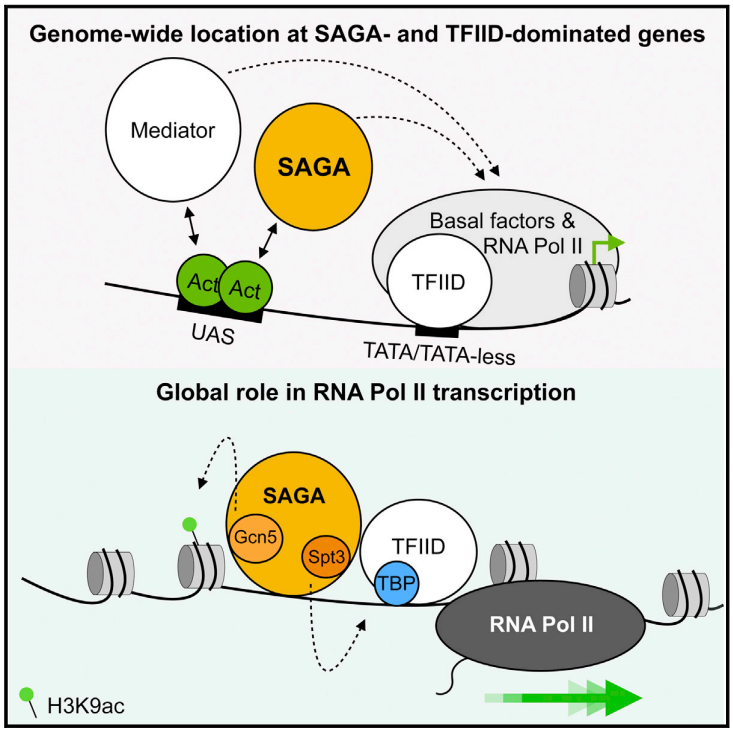
László Tora – Conceived the study. Wrote the manuscript.

Article

Molecular Cell

SAGA Is a General Cofactor for RNA Polymerase II Transcription

Graphical Abstract



Authors

Tiago Baptista, Sebastian Grünberg, Nadège Minoungou, ..., Steve Hahn, Didier Devys, László Tora

Correspondence

devys@igbmc.fr (D.D.),
laszlo@igbmc.fr (L.T.)

In Brief

Baptista et al. show that SAGA, a transcriptional coactivator conserved in all eukaryotes, is involved in overall RNA polymerase II transcription in budding yeast. Using ChEC-seq, SAGA was shown to be recruited to both TATA-containing and TATA-less genes. In agreement, inactivation of SAGA leads to dramatic effects on nascent transcription.

Highlights

- SAGA is recruited to the regulatory regions of both SAGA- and TFIID-dominated genes
- Inactivation of SAGA leads to a global decrease of RNA polymerase II transcription
- SAGA is required for stable TBP recruitment at promoters
- Gcn5 acetyltransferase synergizes with Spt3 (TBP-binding) to promote transcription

Data Resources

GSE96849
GSE97379

Baptista et al., 2017, Molecular Cell *68*, 1–14
October 5, 2017
http://dx.doi.org/10.1016/j.molcel.2017.08.016

CellPress

Please cite this article in press as: Baptista et al., SAGA Is a General Cofactor for RNA Polymerase II Transcription, Molecular Cell (2017), http://dx.doi.org/10.1016/j.molcel.2017.08.016

Molecular Cell
Article

CellPress

SAGA Is a General Cofactor for RNA Polymerase II Transcription

Tiago Baptista,[1,2,3,4] Sebastian Grünberg,[5] Nadège Minoungou,[5,6] Maria J.E. Koster,[7] H.T. Marc Timmers,[7,8] Steve Hahn,[5] Didier Devys,[1,2,3,4,9,*] and László Tora[1,2,3,4,*]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France
[2]Centre National de la Recherche Scientifique, UMR7104, 67404 Illkirch, France
[3]Institut National de la Santé et de la Recherche Médicale, U964, 67404 Illkirch, France
[4]Université de Strasbourg, 67404 Illkirch, France
[5]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[6]Université Paris Diderot, Sorbonne Paris Cité, 75205 Paris, France
[7]Molecular Cancer Research and Stem Cell Section, Regenerative Medicine Center and Center for Molecular Medicine, University Medical Center Utrecht c/o Hubrecht Institute, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands
[8]German Cancer Consortium (DKTK) partner site Freiburg, German Cancer Research Center (DKFZ) and Department of Urology, Medical Center-University of Freiburg, 79106 Freiburg, Germany
[9]Lead Contact
*Correspondence: devys@igbmc.fr (D.D.), laszlo@igbmc.fr (L.T.)
http://dx.doi.org/10.1016/j.molcel.2017.08.016

SUMMARY

Prior studies suggested that SAGA and TFIID are alternative factors that promote RNA polymerase II transcription, with about 10% of genes in *S. cerevisiae* dependent on SAGA. We reassessed the role of SAGA by mapping its genome-wide location and role in global transcription in budding yeast. We find that SAGA maps to the UAS elements of most genes, overlapping with Mediator binding and irrespective of previous designations of SAGA- or TFIID-dominated genes. Disruption of SAGA through mutation or rapid subunit depletion reduces transcription from nearly all genes, measured by newly synthesized RNA. We also find that the acetyltransferase Gcn5 synergizes with Spt3 to promote global transcription and that Spt3 functions to stimulate TBP recruitment at all tested genes. Our data demonstrate that SAGA acts as a general cofactor required for essentially all RNA polymerase II transcription and is not consistent with the previous classification of SAGA- and TFIID-dominated genes.

INTRODUCTION

Formation of the transcription preinitiation complex (PIC), containing RNA polymerase II (Pol II) and general transcription factors (GTFs), is a major regulatory step in eukaryotic gene expression. TFIID is a GTF that binds to promoters and is known to nucleate PIC formation. One important class of factors that regulate PIC formation and function are coactivator complexes. Coactivators directly interact with the basal transcription machinery and/or open the chromatin structure at promoter regions through either remodeling activities or post-translational histone modifications (Hahn and Young, 2011; Thomas and Chiang, 2006). Several coactivators were shown to act at specific subsets of genes, whereas others appear to have a more global role in Pol II transcription. For example, the coactivator Mediator has been recognized as a general factor required for essentially all Pol II transcription, interacts directly with Pol II and other basal factors, and plays a key role in PIC stabilization and Pol II activation (Ansari et al., 2009; Holstege et al., 1998; Thompson and Young, 1995).

To determine the specific regulatory network of distinct coactivators, several studies used genome-wide transcriptome analyses to quantify the levels of steady-state mRNA in different mutant strains. Using this approach, a seminal study analyzed the contribution of the SAGA coactivator complex to gene expression in *S. cerevisiae* (Huisinga and Pugh, 2004). The yeast SAGA complex is thought to activate Pol II transcription through both the recruitment of the TATA binding protein (TBP) and via chromatin modifications mediated by its histone acetyltransferase (HAT) and deubiquitinase (DUB) activities (reviewed in Koutelou et al., 2010; Rodríguez-Navarro, 2009; Weake and Workman, 2012). Organization of these activities in distinct structural modules of SAGA allows the inactivation of one specific function without altering others (Lee et al., 2011). Upon deletion of *SPT3*, a TBP-interacting subunit of SAGA, levels of total mRNA were reduced for about 10% of yeast genes (Huisinga and Pugh, 2004). This set of Spt3-regulated genes was compared to those regulated by TBP-associated factor 1 (Taf1), a subunit of the TFIID complex also suggested to deposit TBP at promoters. This led to the distinction between two different gene classes: (1) the SAGA-dominated genes, which are positively regulated by Spt3, but are essentially independent of Taf1, and (2) the larger class (90%) of TFIID-dominated genes, which are more dependent on Taf1 than on Spt3 (Huisinga and Pugh, 2004). However, the inactivation of both Taf1 and Spt3 induced a severe decrease in the steady-state mRNA levels of almost all yeast genes, which was used as an indication that Spt3 function is partially redundant with that of TFIID. Different

Please cite this article in press as: Baptista et al., SAGA Is a General Cofactor for RNA Polymerase II Transcription, Molecular Cell (2017), http://dx.doi.org/10.1016/j.molcel.2017.08.016

gene features were compared in these two gene classes, leading to the conclusion that SAGA-dominated promoters tend to have consensus TATA box, are more stress-regulated/inducible genes, and tend to be more tightly regulated (Basehoar et al., 2004). As a general model, it was proposed that TBP recruitment is primarily dependent on SAGA at TATA-containing promoters, but dominated by TFIID at the TATA-like (or TATA-less) promoters (reviewed in Tora and Timmers, 2010).

Further studies reported different chromatin organization at the two classes of genes, accounting for the higher plasticity in expression of SAGA-dominated genes (Rhee and Pugh, 2012; Tirosh and Barkai, 2008). The TFIID-dominated genes, which have less expression variability, have a large nucleosome-depleted region just upstream of the transcription start site (TSS) and well-positioned flanking nucleosomes, which may play a role in PIC assembly. The presence of a consensus TATA-box is more important for PIC assembly at SAGA-dominated promoters, where less-well-positioned nucleosomes appear to compete with transcription factors. A more recent study systematically analyzed transcriptional changes upon deletion of all non-essential SAGA subunits (Lenstra et al., 2011). This analysis suggested that even fewer genes were SAGA dependent, as steady-state mRNA levels for only ~150 genes were either up- or downregulated in a *spt3*Δ strain. Surprisingly, when compared with genome-wide localization of the same SAGA subunits (Venters et al., 2011), very little correlation was found between SAGA location and SAGA-dominated genes (Lenstra and Holstege, 2012).

Recent observations provide putative explanations for the seemingly conflicting findings on SAGA location and transcriptional effects. Several studies revealed that mRNA buffering through widespread mRNA stabilization is a common response to a global reduction of Pol II transcription (Haimovich et al., 2013; Plaschka et al., 2015; Rodríguez-Molina et al., 2016; Sun et al., 2012). As the steady-state mRNA levels are the net result of both mRNA synthesis and decay, the quantification of total RNA may not appropriately reflect mRNA synthesis rates upon global changes of Pol II transcription. The recruitment of yeast SAGA at a subset of upstream activating sequences (UASs), as determined by ChIP (van Werven et al., 2008; Venters et al., 2011), contrasted with a broad distribution of the HAT or DUB activities of SAGA, either acetylating the promoter or deubiquitinating the transcribed region of almost all Pol II-transcribed genes (Bonnet et al., 2014). This discrepancy could be explained by an inherent difficulty of ChIPing the SAGA complex (H.T.M. Timmers, unpublished data), which can be due to its very dynamic interaction with chromatin (Vosnakis et al., 2017).

To understand these seemingly contradictory observations, we assessed the genome-wide localization of SAGA and its role in transcription using different and independent methodologies, which are not affected by the potential biases indicated above. We show here that SAGA is recruited to the UASs at a majority of yeast genes, similarly to Mediator. In good agreement, SAGA was found to be required for RNA synthesis of essentially all genes transcribed by Pol II. We observed a compensatory increase of the half-life of a majority of mRNAs upon SAGA depletion, explaining the limited changes in steady-state mRNA levels in the different SAGA mutant strains. We further analyzed the relative contributions of the different SAGA activities to Pol II transcription, thereby revealing a synergistic role for Spt3 and the acetyltransferase Gcn5. Our data lead us to propose that SAGA is a general cofactor for PIC recruitment that is required for transcription of the vast majority of Pol II genes.

RESULTS

SAGA Is Recruited to the Regulatory Regions of Both SAGA- and TFIID-Dominated Genes in Yeast

While a gene-specific function for SAGA has been previously reported (Basehoar et al., 2004; Huisinga and Pugh, 2004), the genome-wide localization of SAGA remains unclear (Robert et al., 2004; van Werven et al., 2008; Venters et al., 2011). Here we used chromatin endogenous cleavage coupled with high-throughput sequencing (ChEC-seq), a method orthogonal to ChIP-seq, to generate a high-resolution genome-wide binding pattern for SAGA in *S. cerevisiae* (Grünberg et al., 2016; Grünberg and Zentner, 2017; Zentner et al., 2015). Micrococcal nuclease (MNase) was C-terminally fused to four SAGA-specific subunits: Spt7 is required for the structural integrity and thus functions of SAGA, Spt3 and Spt8 interact with TBP and have been suggested to deposit TBP at promoters of SAGA-dominated genes (Bhaumik and Green, 2001, 2002; Dudley et al., 1999; Laprade et al., 2007; Larschan and Winston, 2001; Mohibullah and Hahn, 2008), and Ubp8 is the catalytic subunit of the SAGA deubiquitination module. Fusion of the MNase to the respective SAGA subunits did not result in a growth phenotype under the conditions tested in this study, suggesting that the tagging did not affect SAGA function (data not shown).

Permeabilized cells were treated with calcium to stimulate MNase activity and promote DNA cleavage in proximity of the tagged SAGA subunits. The cleaved fragment ends were sequenced and mapped to the budding yeast genome (Figures 1 and S1). Recent ChEC-seq mapping of Mediator, as exemplified by the Med8 subunit, was used as a reference for a coactivator complex associated with UASs genome-wide (Grünberg et al., 2016). As the cleavage patterns were almost identical after 5 and 15 min of MNase activation (ρ = 0.9938; Figure S1A), only the 5 min samples were used in the following analyses. We found strong SAGA-MNase-dependent cleavages upstream of an average of ~2,700 genes, showing little preference for genes previously characterized as SAGA- or TFIID-dominated (61% of SAGA- and 49% of TFIID-dominated genes were targeted by SAGA; Xu et al., 2009). At three exemplary, highly expressed, SAGA-dominated (*CDC19*, *ILV5*, and *PDC1*) (Figure 1, left panels) and three highly expressed, TFIID-dominated genes (*EFB1*, *RPS5*, and *YEF3*) (Figure 1, right panels), we observed strong DNA cleavage upstream of the TSSs, with negligible background signal in the coding region using all four MNase fusion proteins. Enrichment of specific cleavages was dependent on fusion of MNase to SAGA subunits, as it was not observed in a strain expressing untethered MNase under control of the Spt3 promoter after 5 and 15 min incubation (Figure S1A).

We next analyzed the localization of SAGA relative to the TSSs. The average profiles for the four SAGA subunits were

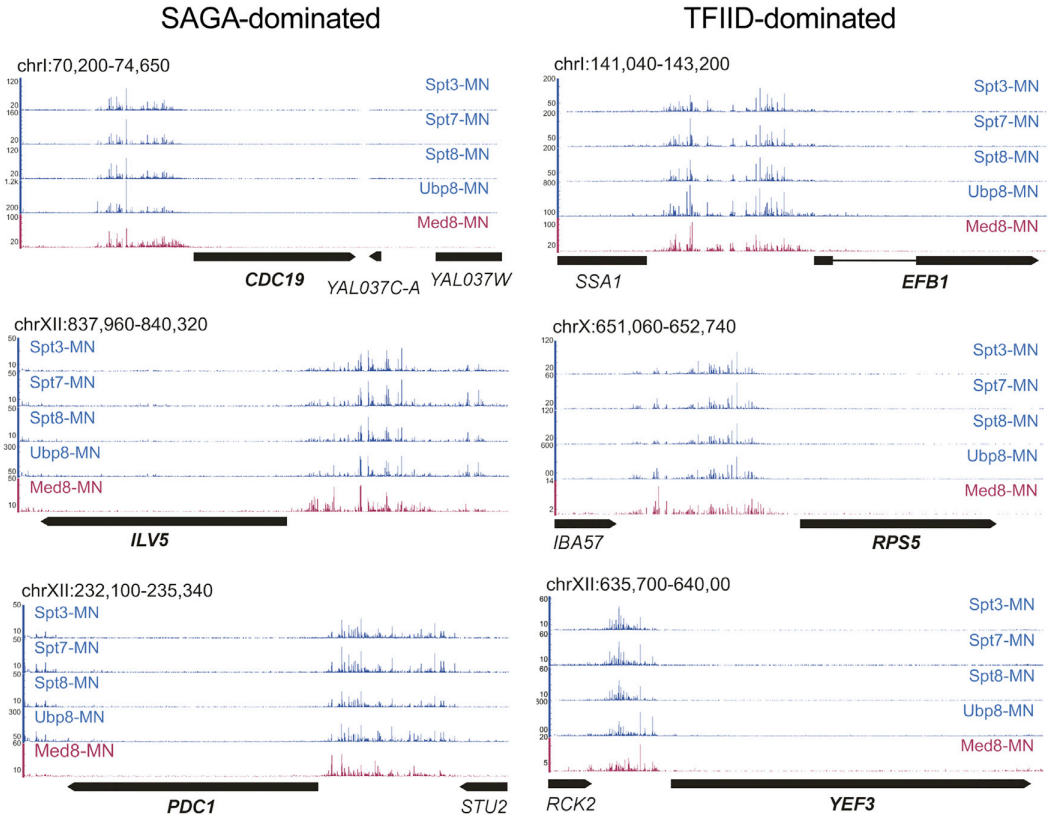


Figure 1. ChEC-Seq Profiling of SAGA-Specific Subunits

Signal tracks showing cleavages induced by Spt3-MNase, Spt7-MNase, Spt8-MNase, and Ubp8-MNase at three representative SAGA-dominated genes (*CDC19*, *ILV5*, and *PDC1*) and three TFID-dominated genes (*EFB1*, *RPS5*, and *YEF3*). Med8-MNase cleavage sites are shown as a reference for the coactivator Mediator binding at UASs (Grünberg et al., 2016). See also Figure S1.

almost identical, suggesting that these analyses reveal the location of the whole SAGA complex (Figure 2A). This was confirmed by pairwise correlations of genes targeted by SAGA-MNase subunits tested, resulting in Spearman's rank correlation coefficients of $\rho > 0.96$ (Figure S1B). When we compared the plots for SAGA (Spt3-MNase) and our previous ChEC-seq data obtained for the Mediator (Med8-MNase) and TFIID (Taf1-MNase) complexes, we found a striking coincidence between SAGA and Mediator position at UASs (Figure 2B) (Grünberg et al., 2016). In agreement with previous ChIP and ChEC-seq data, the peak summit of average SAGA cleavages relative to the TSS at TATA-containing genes was generally more distal (average 286 bp) than at TATA-less genes (average 217 bp) (Grünberg et al., 2016; van Werven et al., 2008; Venters et al., 2011) (Figure 2C). In addition, we found that > 99% of genes bound by SAGA were also targeted by TFIID (Figure S1C) (Grünberg et al., 2016).

Previous mapping of Mediator and expression analysis suggest that occupancy of factors is broadly uncoupled from active gene expression (Grünberg et al., 2016; Lenstra et al., 2011). Like Mediator, levels of SAGA-directed DNA cleavage were higher at UASs of SAGA-dominated genes compared to TFIID-dominated genes, which are generally expressed at significantly higher levels (Churchman and Weissman, 2011). Our data indicate that SAGA, like Mediator, binds to the UASs of many genes, independent of gene classification, and broadly uncoupled from gene activation.

Global Downregulation of Pol II Transcription upon Deletion of SAGA Structural Subunits

The global recruitment of SAGA at the UASs of many Pol II-transcribed genes is in good agreement with the previously reported decrease of Pol II recruitment at most gene promoters in *spt20*Δ yeast strains (Bonnet et al., 2014). To determine whether SAGA acts as a general cofactor for Pol II, we compared the steady-state and newly synthesized mRNA levels in SAGA mutant strains at all genes transcribed by Pol II. Wild-type, *spt7*Δ, and *spt20*Δ strains cultured in rich medium were pulse-labeled with 4-thiouracil (4tU) and mixed in a fixed ratio with labeled *S. pombe* cells for normalization. Purified total (steady-state) RNA and labeled (newly synthesized) RNA were hybridized to Affymetrix microarrays containing probes for *S. cerevisiae* and *S. pombe* transcripts.

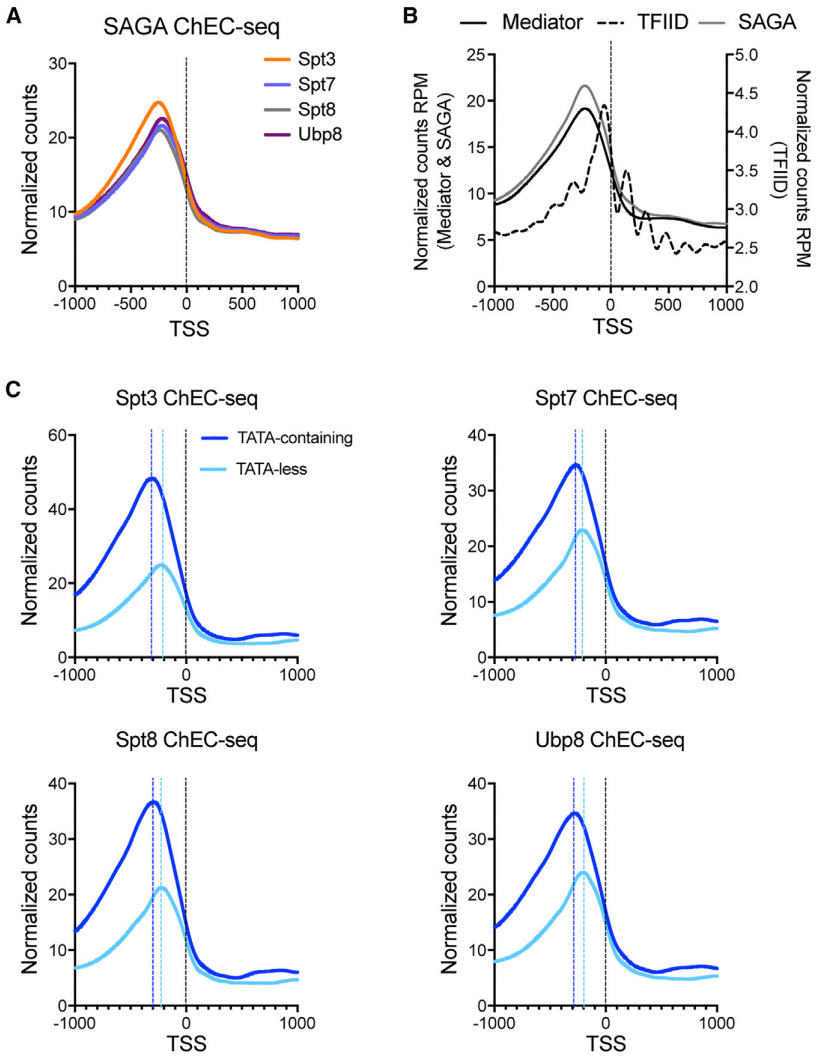


Figure 2. SAGA Associates with UASs of SAGA- and TFIID-Dominated Genes

(A) Plots of average SAGA subunit cleavages relative to the TSSs of all annotated Pol II genes.

(B) Plots of average SAGA (Spt3-MNase, gray line), Mediator (Med8-MNase, black line), and TFIID (Taf1-MNase, dotted line) cleavage at the TSSs of all genes transcribed by Pol II.

(C) Plots of average cleavage of the single subunits at TATA-containing (dark blue line) or TATA-less (light blue line) gene promoters. The dotted lines represent the peak summit to TSS distance at TATA-containing (dark blue) or TATA-less (light blue) genes. See also Figure S1.

When applying a 2-fold change threshold and a p value of 0.05 after normalization to the *S. pombe* spiked-in signal, significant changes for only a very limited number of genes were observed in steady-state mRNA levels in the *spt20*Δ strain (43 genes downregulated and 24 genes upregulated) (Figure 3A). These data are reminiscent of earlier studies suggesting that SAGA regulates a limited number of genes (Lenstra et al., 2011). Importantly, genes with altered steady-state mRNA levels in the SAGA mutant do not constitute a clearly discrete class of genes based on their expression profile, but appear more as the outliers in a Gaussian distribution of gene expression fold changes (Figure 3A). In contrast, the comparison of newly synthesized mRNA levels between *spt20*Δ and wild-type strains revealed a decrease in mRNA synthesis for a vast majority of genes (4,432 genes downregulated and 2 genes upregulated) (Figure 3B). Strikingly different profiles were also observed when comparing levels of total or labeled RNA purified from the *spt7*Δ strain (Figures 3D and 3E). Consistently, the number of downregulated genes and the extent of downregulation dramatically increased when analyzing newly synthesized mRNA as compared to steady-state mRNA, indicating a global decrease of Pol II transcription in the *spt20*Δ and *spt7*Δ strains.

mRNA Levels Are Buffered by Increased mRNA Half-Life in SAGA Mutant Strains

As the analysis of newly synthesized mRNA quantifies mRNA synthesis uncoupled from RNA degradation, the very limited changes in steady-state mRNA levels might be explained by compensatory changes in global mRNA decay. To address this hypothesis, we used comparative dynamic transcriptome analysis (cDTA) (Sun et al., 2012), which allows the calculation of mRNA synthesis and decay rates for all *S. cerevisiae* coding genes. In *spt20*Δ or *spt7*Δ strains, we observed a concomitant decrease in both mRNA synthesis and decay rates when compared to the wild-type strain (Figures 3C and 3F). This compensation was almost complete in the *spt20*Δ strain, with a mean decrease in synthesis of about 3.8-fold and in decay of about 4.1-fold, which accounts for the small number of changes in steady-state mRNA (Figures 3A and 3C). In the *spt7*Δ strain, the decrease in mRNA synthesis was very similar, but the changes in decay rates were more dispersed (Figure 3F). As a result of an imperfect compensation, the steady-state mRNA levels of a higher number of genes were modified in this mutant strain (Figure 3D).

To verify this global mRNA stabilization using an independent method, we compared mRNA decay following transcription inhibition with thiolutin in wild-type and SAGA mutant strains. In the wild-type strain, the calculated mRNA half-lives were very similar to those previously reported (Geisberg et al., 2014). In good agreement with our cDTA measurements, the mRNA half-lives of five randomly selected genes increased (1.5- to 4-fold) in the *spt7*Δ and *spt20*Δ strains (Figure S2). As expected for a gene

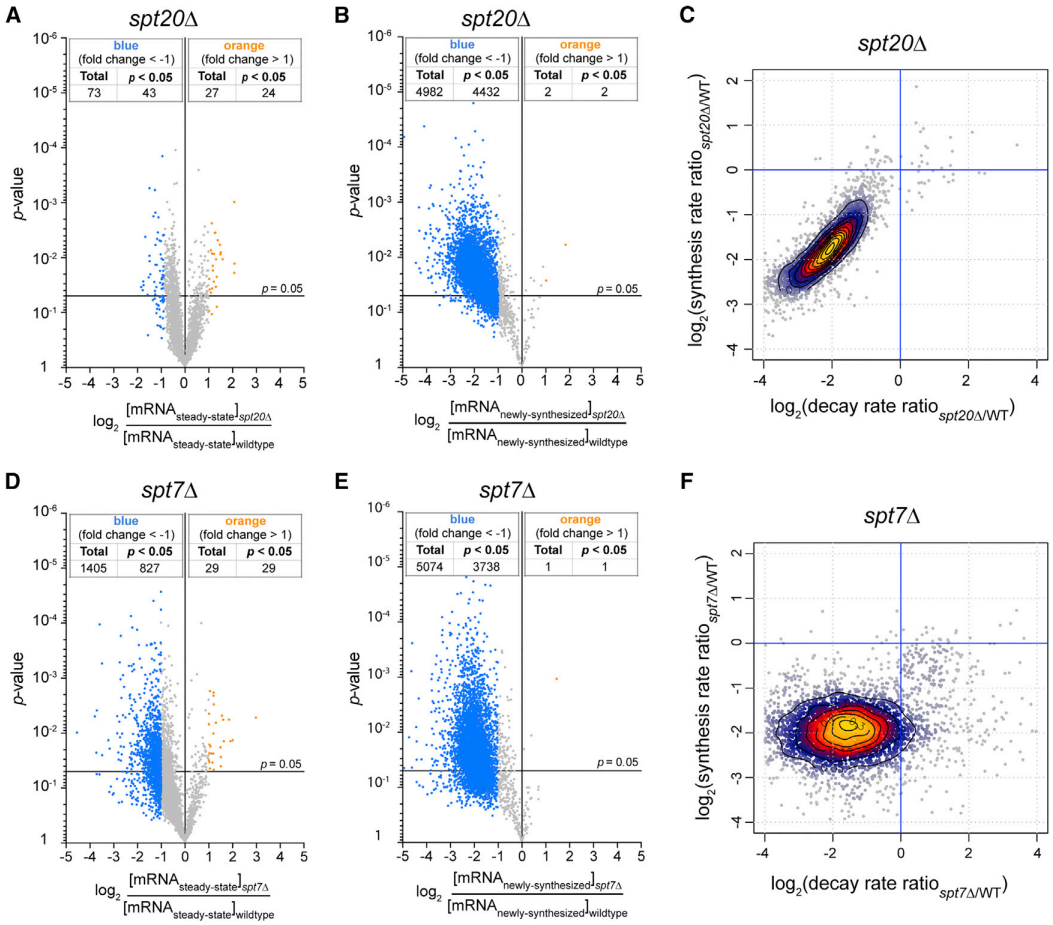


Figure 3. Compensation of an Overall Decrease of Pol II Transcription by a Global Change in mRNA Decay in *SPT20* and *SPT7* Deletion Strains

(A–F) Volcano plots showing fold changes in steady-state mRNA levels (A and D) or newly synthesized mRNA levels (B and E) relative to their significance (p value). Fold changes (FC) were calculated as the log2 of the ratio of the expression value of each gene after normalization to *S. pombe* signal in the *spt20*Δ strain (A and B) or the *spt7*Δ (D and E) strain versus the expression value of the same gene in wild-type *S. cerevisiae*. A total of 5,385 genes were analyzed, and thresholds of 2-fold change (blue dots, more than 2-fold decrease; yellow dots, more than 2-fold increase) and 0.05 p values were considered. cDTA profiles for *spt20*Δ and *spt7*Δ strains are shown in (C) and (F), respectively. For all analyzed genes, changes in synthesis rates were plotted against the changes in mRNA decay rates. Changes were calculated as the log2 of the ratio between mutant and wild-type. 90% of genes are contained within the outer contour. Yellow and red dots correspond to 60% of genes. For each strain, results were obtained from at least two independent biological replicates. See also Figure S2.

transcribed by Pol III and not affected by SAGA inactivation, there was no significant change of the long-lived *scR1* RNA (Figure S2). Our results show that yeast cells buffer a global decrease in Pol II transcription following SAGA depletion by increasing mRNA half-lives, as previously reported for mutations in Pol II and Mediator, or inhibition of the kinase activity in TFIIH (Plaschka et al., 2015; Rodríguez-Molina et al., 2016; Sun et al., 2012).

SAGA Equally Contributes to the Expression of SAGA- and TFIID-Dominated Genes

Our ChEC-seq analysis revealed that the cleavage levels at TATA-containing (or SAGA-dominated) genes were significantly higher than observed at TATA-less (or TFIID-dominated) genes (Figure 2C). We therefore asked whether the expression of these two classes of genes would be differentially controlled by SAGA. For each gene category, changes in synthesis rates between mutant and wild-type strains were plotted, and their distributions in the different gene classes were compared. Inactivation of the SAGA complex through deletion of either *SPT20* or *SPT7* induced a similar decrease in Pol II transcription for either the SAGA- or the TFIID-dominated genes (Figures 4A and 4B). Similarly, no difference was observed based on whether or not these genes contain a TATA box in their promoters (Figures 4C and 4D). We did, however, find that the most highly expressed genes in wild-type cells showed the strongest decrease in mRNA synthesis in both *SPT20*- and *SPT7*-deleted strains (Figures 4E and 4F).

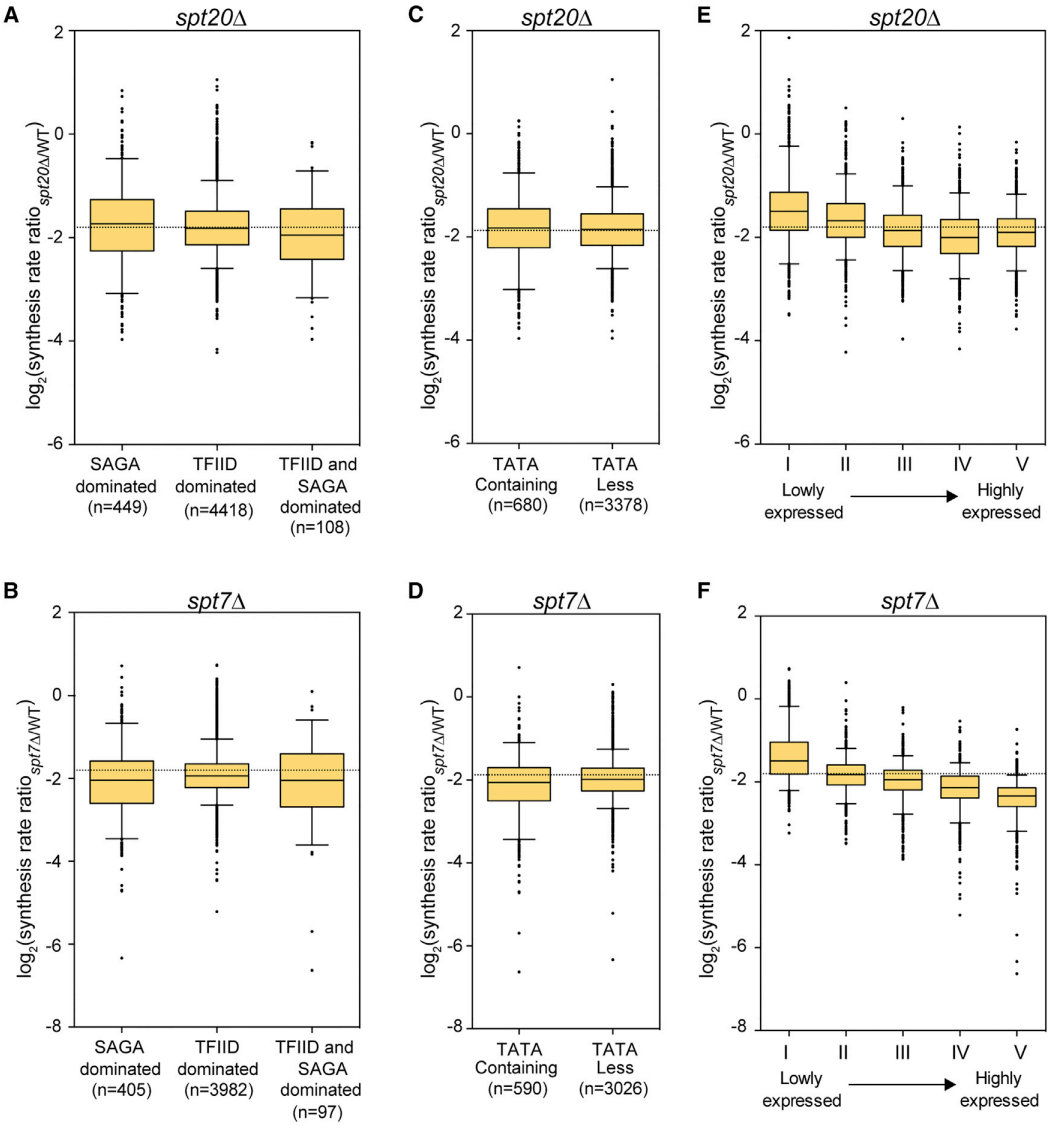


Figure 4. Changes in mRNA Synthesis Rates for Different Classes of Genes in *spt20*Δ and *spt7*Δ

(A–F) Boxplots showing the distribution of changes in mRNA synthesis rates between *spt20*Δ (A, C, and E) or *spt7*Δ (B, D, and F) and wild-type strains. Very similar changes were seen whether genes were described as SAGA- or TFIID-dominated (A and B) or whether their promoters were TATA-containing or TATA-less (C and D). Genes were divided in quintiles according to their expression levels in wild-type cells, and the changes in synthesis rates were plotted for each quintile (E and F). Boxes contain genes between the 25th and the 75th centiles; the line indicates the median and the whiskers correspond to 5th and 95th centiles.

The Global Decrease of Pol II Transcription in SAGA Mutants Is a Primary Event

Growth of *S. cerevisiae* strains deleted for *SPT20* or *SPT7* is impaired in rich medium, which might cause indirect effects on transcription. Indeed, slow growth of *S. cerevisiae* deletion strains has been linked to altered cell-cycle distribution, leading to a common gene expression signature (O'Duibhir et al., 2014). Although our observations of a global decrease in Pol II transcription did not reproduce this signature, a conditional depletion strategy was used to rule out any effect due to the slow growth of the SAGA mutant strains. We used the anchor-away strategy to deplete C-terminally FRB-tagged Spt7 from the nucleus by addition of rapamycin in a strain expressing RPL13A-FKBP12 (Haruki et al., 2008). Nuclear-cytoplasmic fractionation of cells treated or untreated with rapamycin for 30 min demonstrated an efficient nuclear depletion of the fused Spt7 protein, in agreement with the delayed growth of the *SPT7-FRB* strain (Figures S3A and S3B). Viability of the *SPT7-FRB* strain was not affected after 30 or 60 min of rapamycin treatment (Figure S3C). Total and newly synthesized mRNA levels of all

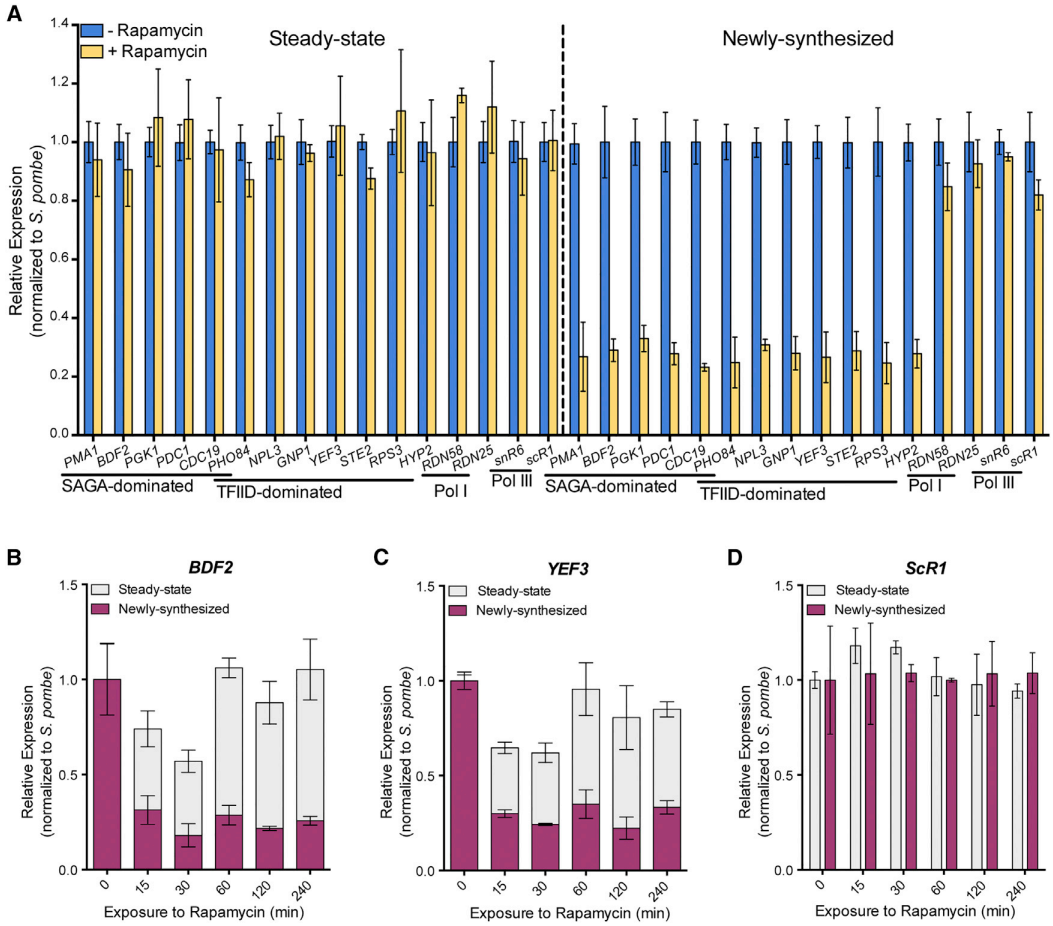


Figure 5. Conditional Nuclear Depletion of Spt7 Decreased Transcription of Both SAGA- and TFIID-Dominated Genes

(A) Spt7 anchor-away strains, untreated or treated with rapamycin for 60 min, were labeled with 4tU. mRNA levels from six SAGA- and six TFIID-dominated genes and RNA levels from four control genes transcribed by Pol I and Pol III were quantified by RT-qPCR. Expression values (mean ± SD of three independent experiments) were normalized to spiked-in *S. pombe* signal and set to 1 in the untreated sample.

(B–D) Time course analysis of changes in steady-state and newly synthesized RNA for a SAGA-dominated (B), a TFIID-dominated (C), and a control gene transcribed by Pol III (D) upon Spt7 nuclear depletion.

See also Figures S3 and S4.

tested genes were unchanged between the parental and the *SPT7-FRB* strain, indicating that the FRB fusion by itself did not induce any detectable phenotype (Figures S3D and S3E).

To determine the effects of Spt7 depletion on transcription, we quantified steady-state and newly synthesized mRNA for a number of selected genes after 60 min of rapamycin treatment and a pulse of RNA labeling using 4tU. Total and labeled RNA were analyzed by RT-qPCR, and mRNA levels for both fractions were normalized to the spiked-in *S. pombe* signal. While the steady-state mRNA levels did not change significantly, the newly synthesized mRNA levels of all tested genes were reduced by 3- to 4-fold in the rapamycin-treated cells when compared to non-treated cells (Figure 5A). The range of decrease in mRNA synthesis was very similar among the five SAGA- and the six TFIID-dominated genes tested. The RNA levels of genes transcribed by Pol I or Pol III were not significantly modified upon nuclear depletion of Spt7 (Figure 5A). Accordingly with the almost complete nuclear depletion of Spt7, the extent of gene expression changes highly resembles that observed in the *spt7*Δ strain.

To better appreciate how Spt7 nuclear depletion affects transcription, we compared steady-state and newly synthesized mRNA levels for a number of selected genes in a time course experiment. Cells were exposed to rapamycin for different periods (0–240 min) and labeled with 4tU for 6 min. Quantification of labeled RNA revealed that mRNA synthesis was strongly reduced as early as 15 min following induction of Spt7 depletion and remained stable over time (Figures 5B, 5C, and S4). Interestingly, steady-state mRNA levels of the same genes were slightly reduced at early time points (15 and 30 min), and they returned to normal levels at later time points. The observed profiles are best explained by an early decrease in mRNA synthesis, whereas efficient buffering of mRNA levels would occur at later time points

Please cite this article in press as: Baptista et al., SAGA Is a General Cofactor for RNA Polymerase II Transcription, Molecular Cell (2017), http://dx.doi.org/10.1016/j.molcel.2017.08.016

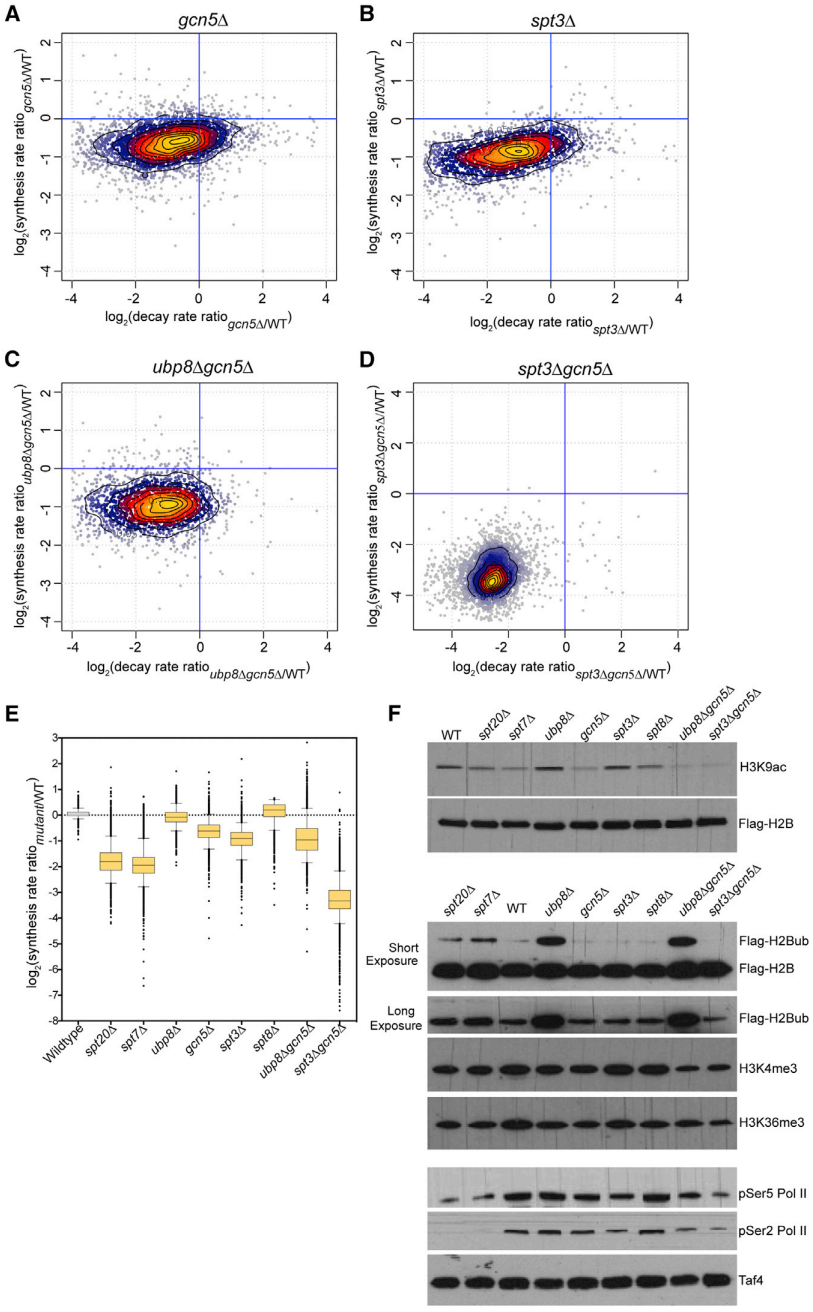


Figure 6. cDTA Analyses of Different SAGA Subunit Deletion Strains

(A–D) Synthesis rates and decay rates were determined for each *S. cerevisiae* gene in *gcn5*Δ (A), *spt3*Δ (B), *ubp8*Δ*gcn5*Δ (C), and *spt3*Δ*gcn5*Δ (D). Changes (calculated as the log2 of the ratio between mutant and wild-type) in synthesis rates were plotted against changes in decay rates.

(E) Boxplots summarizing the extent of changes in mRNA synthesis for all the analyzed deletion strains. For each strain, cDTA data were obtained from at least two independent biological replicates.

(F) Whole-cell extracts from wild-type and the indicated deletion strains were revealed with the antibodies corresponding to histone marks regulated by SAGA (H3K9ac, H2Bub) or associated with active transcription (H3K4me3, H3K36me3) or with antibodies specific to the C-terminal domain of Rpb1 phosphorylated on serine 5 (pSer5 RNA Pol II) or on serine 3 (pSer5 RNA Pol II), as indicated. See also Figures S5, S6, and S7.

Activities of the SAGA Complex Required for Pol II Transcription

The budding yeast SAGA complex is composed of 19 different subunits, which are organized in five functional and structural modules (Han et al., 2014; Lee et al., 2011; Setiaputra et al., 2015). To better understand the role of SAGA in global mRNA synthesis as revealed by the analysis of strains deleted for SAGA structural subunits, we wanted to determine which individual activities of SAGA may play a role in Pol II transcription and what their relative contributions are. The most obvious candidates are the enzymatic activities of SAGA, which are known to regulate histone marks associated with active transcription as well as subunits proposed to recruit TBP at promoters (Koutelou et al., 2010). We thus focused our analysis on Gcn5, the catalytic subunit of the HAT module; Ubp8, the catalytic subunit of the DUB module; and the TBP-interactors Spt3 and Spt8 (Dudley et al., 1999; Laprade et al., 2007; Larschan and Winston, 2001; Mohibullah and Hahn, 2008). Global levels of newly synthesized mRNA were significantly decreased (by an average 1.5-fold) in the *gcn5*Δ strain, suggesting that histone acetylation favors transcription initiation (Figures 6A and 6E). *UBP8* deletion did not affect mRNA synthesis, suggesting that increased H2Bub levels do not affect Pol II transcription by themselves (Figures 6E and S5). From the two subunits interacting with TBP, only Spt3 appeared to play a significant role in Pol II transcription, as the newly synthesized mRNAs were reduced by about (from 60 min on). The RNA levels of *scR1*, a gene transcribed by Pol III, were unchanged in the total and labeled RNA fractions during the same time course experiment (Figure 5D). Thus, our observations indicate that the global downregulation of Pol II transcription upon SAGA inactivation is a primary event, which cannot be explained by indirect effects of *SPT7* or *SPT20* deletions and which seems to be a direct consequence of SAGA acting at the promoters of genes.

2-fold in the *spt3*Δ, and they were unaltered in the *spt8*Δ strain (Figures 6B and S5).

To recapitulate the transcriptional effects seen upon depletion of SAGA structural subunits (Spt7 and Spt20), budding yeast strains deleted for two SAGA-specific subunits, which are located in distinct modules, were generated. Surprisingly, mRNA synthesis was more affected in the double *ubp8*Δ*gcn5*Δ mutant (about 2-fold average decrease) than in the single *GCN5*-deleted strain (Figures 6C and 6E). Most importantly, the cumulative inactivation of TBP-interacting and HAT activities in the *spt3*Δ*gcn5*Δ strain induced a dramatic decrease in mRNA synthesis by, on average, 10-fold (Figures 6D, 6E, and S5) without affecting the overall SAGA assembly (Figure S6). This global loss in Pol II transcription was larger than the cumulative effect of *SPT3* and *GCN5* deletion and even greater than deletion of the structural subunits. Altogether, these results point to an important role for TBP interacting through Spt3 and also for Gcn5 through either acetylation of histones, binding to acetylated histones through the Gcn5 bromodomain, or an alternative mechanism. To further support that these global transcriptional effects are directly related to SAGA function, we compared all our datasets with the previously reported "slow-growth gene expression signature" (O'Duibhir et al., 2014) and did not observe any relevant correlation (Figure S7).

For all deletion strains used in this study, we analyzed the global levels of histone marks directly regulated by SAGA (H3K9ac and H2Bub) or associated with active transcription (H3K4me3 and H3K36me3). As expected, H3K9 acetylation strongly decreased upon *GCN5* deletion and to a lesser extent upon depletion of SAGA structural subunits (Figure 6F). As previously reported, H2Bub levels slightly increased upon *SPT7* or *SPT20* deletion, and they were dramatically increased in either the *ubp8*Δ or the *ubp8*Δ*gcn5*Δ strains (Figure 6F) (Henry et al., 2003). Less pronounced changes were observed for marks enriched at active promoters (H3K4me3) or at transcribed gene regions (H3K36me3), which correlated with global gene expression changes in the corresponding strains (Figure 6F). In addition, the levels of serine 5 and serine 2 phosphorylation on the C-terminal domain of the largest subunit of Pol II (Rpb1) were slightly reduced in *gcn5*Δ, *spt3*Δ, and *ubp8*Δ*gcn5*Δ and strongly decreased in *spt3*Δ*gcn5*Δ, *spt7*Δ, and *spt20*Δ strains (Figure 6F). Importantly, these analyses revealed a strong correlation of alterations in Pol II transcription with Rpb1 phosphorylation levels.

Role of SAGA in the Genome-wide Recruitment of TBP

The distinction of the two proposed classes of SAGA-dominated and TFIID-dominated genes was substantiated by the idea that each complex can independently recruit TBP to the corresponding promoters. In line with this hypothesis, it is predicted that in a *SPT3*-deleted strain, TBP recruitment should be more affected at the SAGA-dominated genes compared to recruitment at the TFIID-dominated genes. Analysis of newly synthesized RNA in a *spt3*Δ strain revealed that mRNA synthesis was significantly reduced for a majority of genes with an average 2-fold decrease in mRNA levels, whereas very few changes were detected in steady-state mRNA levels (Figures 7A and 7B). The changes in synthesis rates between *spt3*Δ and wild-type strains did not discriminate between the different classes of genes (SAGA- versus TFIID-dominated; TATA-containing versus TATA-less promoters) (Figures 7C and 7D). We next investigated whether the SAGA requirement at these different genes was due to different mechanisms. TBP recruitment at selected genes was measured by ChIP in *3HA-TBP* and *spt3*Δ *3HA-TBP* strains. At the promoters from four SAGA- and five TFIID-dominated genes, TBP recruitment was similarly reduced upon *SPT3* deletion, whereas only background levels could be detected at control regions (Figure 7E). These results suggest that SAGA plays an important role in TBP recruitment or stabilization and subsequent Pol II recruitment at almost all active promoters and that this function is independent of the promoter architecture.

DISCUSSION

In this study, we provide a combination of complementary results indicating that SAGA is recruited at most Pol II-transcribed genes in *S. cerevisiae*, where it is plays a critical role in mRNA synthesis. In striking contrast to the current model for SAGA function, in which SAGA predominantly regulates the expression of only a limited subset of genes, we propose that SAGA, like Mediator, acts as a general cofactor for Pol II transcription. Previous observations suggesting a global role for SAGA in Pol II transcription were: (1) SAGA-dependent changes in H3K9 acetylation revealed that SAGA, through its Gcn5 subunit, acetylates most active promoters, and (2) SAGA is required for Pol II recruitment at a vast majority of active promoters (Bonnet et al., 2014). This conclusion is coherently supported by our new findings including: (1) direct localization studies by ChEC-seq indicate recruitment of SAGA to UASs of genes previously categorized as SAGA- or TFIID-dominated, (2) all genes bound by SAGA are also targeted by TFIID, (3) SAGA is required for TBP recruitment at active promoters of both SAGA- and TFIID-dominated genes, and (4) disruption of the SAGA complex led to a global decrease in mRNA synthesis.

Steady-State mRNA Levels Do Not Accurately Reflect SAGA Activity in Pol II Transcription

Earlier studies reporting a gene-specific activity for the SAGA complex were based on the evaluation of steady-state mRNA levels upon deletion of SAGA subunits (Huisinga and Pugh, 2004; Lenstra et al., 2011). By analyzing newly transcribed RNA, we provide evidence that these studies underestimated the role of SAGA on Pol II transcription. In addition, comparison of steady-state with newly synthetized mRNA revealed that major and general perturbations of Pol II transcription can lead to a compensatory mechanism, as reported for other coactivator complexes such as Mediator (Plaschka et al., 2015). This compensation emerges as a way to cope with an abrupt decrease in RNA synthesis that would ultimately lead to a decrease of cellular mRNA. This mechanism is also supported by the increased half-life of mRNAs in Spt7- and Spt20-depleted cells resulting in longer-lived and more stable transcripts. Hence, through a mechanism that remains elusive, a simultaneous decrease in mRNA decay ultimately leads to virtually unchanged RNA levels (Haimovich et al., 2013; Sun et al., 2013). A prime candidate for this would be the CCR4-NOT complex, which is a global regulator of mRNA decay and has been linked to the

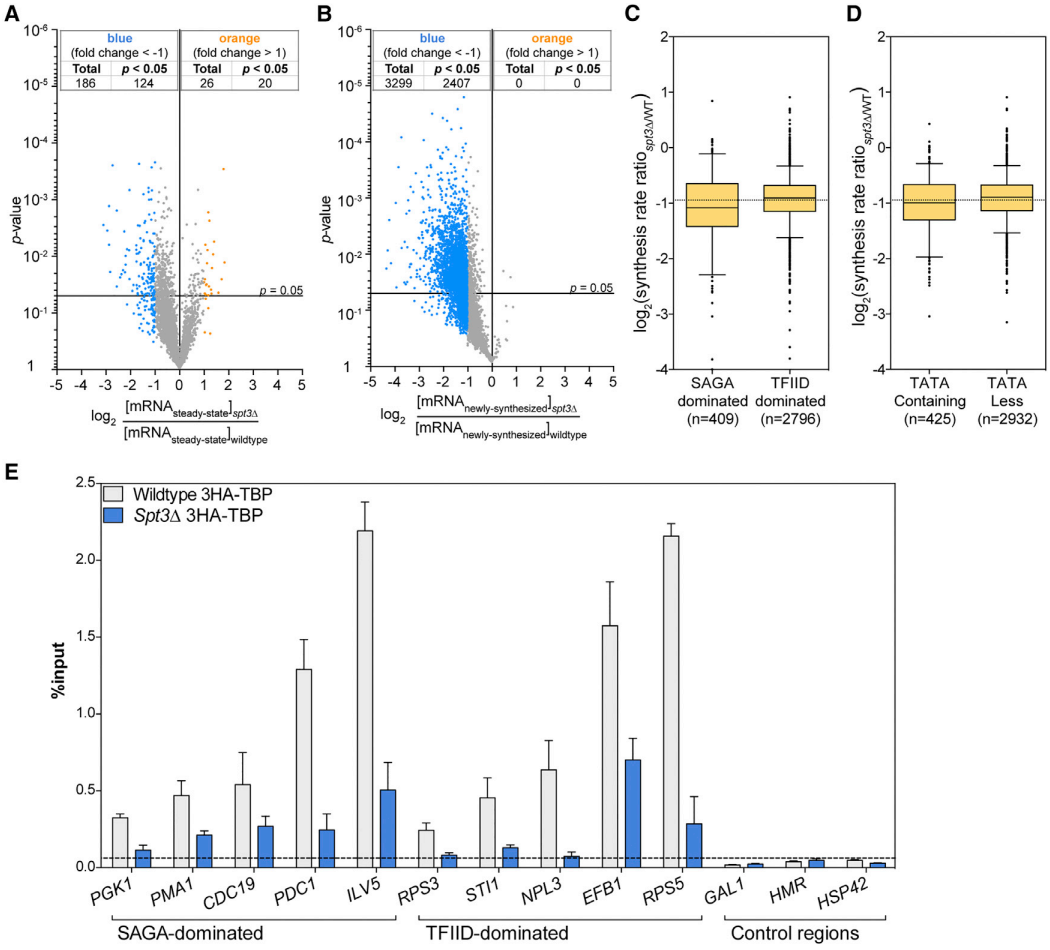


Figure 7. Reduced TBP Occupancy at Promoters of Both SAGA- and TFIID-Dominated Genes in *SPT3* Deletion Strain

(A and B) Volcano plots showing changes in steady-state (A) and newly synthesized mRNA levels between *spt3*Δ and wild-type *S. cerevisiae*.

(C and D) Boxplots representing the distribution of changes in synthesis rates upon *SPT3* deletion for SAGA- versus TFIID-dominated genes (C) and for TATA-containing versus TATA-less genes (D).

(E) TBP enrichment at promoters from five SAGA- and six TFIID-dominated genes as well as at three control regions was determined by HA ChIP in a *spt3*Δ *3HA-TBP* strain and the parental *3HA-TBP* strain and quantified by real-time PCR. The values (mean ± SD of three independent ChIP experiments) are expressed as percentage of input DNA signal.

TFIID, SAGA, and Mediator complexes by genetic means (Collart and Timmers, 2004; Villanyi and Collart, 2015). Importantly, our experiments indicate that the existence of this compensatory mechanism has occluded the full spectrum of SAGA action and that this coactivator has a critical role in optimal Pol II transcription.

SAGA Activities Act in a Synergistic Manner

SAGA is a multifunctional complex with at least five activities: HAT, DUB, TBP interacting, activator binding, and nucleosome binding (reviewed in Koutelou et al., 2010; Rodríguez-Navarro, 2009; Weake and Workman, 2012). As the depletion of different SAGA subunits affected a distinct and relatively small subset of genes, it was suggested that each SAGA subunit makes specific and unique contributions to the function of the complete complex (Lee et al., 2000). Further gene expression analyses of steady-state RNA showed that subunits of a same module share similar transcriptional effects, but each SAGA activity regulated a different set of genes (Lenstra et al., 2011). From the 12 SAGA-specific subunits, we measured the transcriptional effects of 4 subunits affecting HAT, DUB, or TBP-interacting activities as well as two structural subunits. Except for *SPT8* and *UBP8*, we observed a global effect on Pol II transcription in all other mutants, suggesting that most SAGA activities have a general regulatory function at all genes. The decrease in global mRNA synthesis observed in a *GCN5* and *SPT3* double deletion strain was much higher than the sum of any single mutant. Therefore, these subunits having distinct functions (histone acetylation and TBP interacting) and being found in different locations within SAGA (Setiaputra et al., 2015) act in a synergistic manner on Pol

II transcription. This underscores that different activities, which might appear as relatively independent (HAT, DUB and TBP interacting) are joined within a single complex to coordinate their action on Pol II transcription.

Surprisingly, the transcriptional effects seen in the *spt3*Δ*gcn5*Δ strain were even more severe than that observed upon deletion of SAGA structural subunits, which were expected to compromise multiple if not all SAGA functions (Grant et al., 1997; Roberts and Winston, 1997; Sterner et al., 1999). However, recent studies demonstrated that upon depletion of these structural subunits, smaller assemblies or sub-complexes still form (Lee et al., 2011). This is supported by the milder increase of H2Bub levels in *spt7*Δ or *spt20*Δ when compared with *ubp8*Δ, which suggest the formation of a partially active DUB module in the absence of Spt7 or Spt20. Therefore, in all mutant strains analyzed, residual activities of such sub-complexes might fulfill some functions of the whole SAGA complex, although we cannot exclude that they may have negative effects on transcription through sequestration of PIC components. Interestingly, systematic proteomic analyses of deletion strains and our coimmunoprecipitation experiments predict that a SAGA complex lacking only Spt3 and Gcn5 would be assembled in the *spt3*Δ*gcn5*Δ strain (Lee et al., 2011) (Figure S6). Although this incomplete SAGA complex might retain some activities of the wild-type SAGA, mRNA synthesis was dramatically reduced in this strain (by an average 10-fold). This is in agreement with the slower growth of this double-mutant strain when compared to *spt20*Δ cells (Sterner et al., 1999). Hence, further suppression of all SAGA activities may lead to an almost complete suppression of Pol II transcription and cell lethality.

A minor reduction of mRNA synthesis was observed upon deletion of *GCN5*, which is known to be a subunit of both the SAGA and the ADA complexes. The ADA complex is composed of the same four subunits as the SAGA HAT module, with two additional ADA-specific subunits, Ahc1 and Ahc2 (Eberharter et al., 1999; Lee et al., 2011). Depletion of Spt3, which is specific to SAGA on the *gcn5*Δ background, had a much more dramatic effect on transcription, with a more than 10-fold average reduction in mRNA synthesis. This suggests that the role of the ADA complex in Pol II transcription is less dominant than that of SAGA, in agreement with the idea that the different SAGA activities have synergistic roles in Pol II transcription. However, this does not discard the hypothesis that both complexes might act cooperatively and that loss of one of them (namely, SAGA) potentiates a decreased activity of the other (ADA complex).

An Alternative Model for SAGA Function in PIC Assembly

Our study demonstrates that SAGA is recruited to the majority of active promoters where it has a comparable role in transcription independently of the type of gene (TFIID- versus SAGA-dominated; TATA-containing or TATA-less). In this respect, SAGA can be compared to Mediator, a coactivator complex that is required for all Pol II transcription and stimulates PIC formation. Importantly, ChEC-seq analyses revealed that Mediator binds to a majority of UASs, clarifying ambiguous results obtained by ChIP-based methods (Grünberg et al., 2016). Widespread binding of Mediator is consistent with the global and dramatic decrease of mRNA synthesis observed upon the loss of Mediator subunits as demonstrated through the analysis of newly synthesized RNA (Plaschka et al., 2015). The similarities in binding profiles and transcriptional effects of both SAGA and Mediator suggest that the two complexes participate in the assembly and/or stabilization of the PIC at all promoters and should be considered as general cofactors in budding yeast. In addition to the global function of SAGA and Mediator, TFIID was also found at essentially all promoters and was shown to be required for the transcription of both SAGA- and TFIID-dominated genes (Grünberg et al., 2016; Warfield et al., 2017). Therefore, the distinct functions of these three coactivators are needed for Pol II transcription initiation at almost all yeast promoters.

However, as SAGA can contact many different activators and uses several domains to interact with modified histones, SAGA recruitment and dynamics may vary at different gene promoters in yeast. In addition, the different activities of SAGA might have diverse outputs at each gene promoter, depending on many additional factors such as promoter architecture or chromatin environment. Indeed, our analyses of various SAGA mutant strains revealed that, although mRNA synthesis of a vast majority of genes was affected, the extent of changes in Pol II activity varied across the genome. However, the distribution of mRNA synthesis changes appears to indicate a continuum of slightly different sensitivities to SAGA action, from which a distinct group of genes could not be identified. Other gene features such as the presence of a TATA-consensus or a TATA-like element allow a better distinction of gene classes characterized by different mechanisms of transcription regulation (Müller and Tora, 2014; Rando and Winston, 2012). Although SAGA had a comparable overall influence on transcription of these two gene classes, our ChEC-seq profiling revealed that SAGA is differently recruited at TATA-containing versus TATA-less genes. SAGA had a higher occupancy and was recruited more upstream relative to TSSs at TATA-containing genes when compared to TATA-less genes, a profile strikingly similar to that of Mediator (Grünberg et al., 2016). This is in agreement with previous studies indicating distinct chromatin structure and different transcriptional plasticity for these two promoter classes (de Jonge et al., 2017; Kubik et al., 2015; Tirosh et al., 2007). It remains to be determined if SAGA and Mediator have distinct functions in PIC formation at TATA-containing versus TATA-less promoters. Nevertheless, our data do not support the earlier distinction of two different pathways for PIC assembly with TBP delivery at promoters depending on either SAGA or TFIID. In conclusion, our study together with that of the accompanying paper by the Hahn lab (Warfield et al., 2017) demonstrates that in yeast both TFIID and SAGA are required for all mRNA transcription by Pol II.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Yeast strains
  - ChEC-seq

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2017.08.016.

AUTHOR CONTRIBUTIONS

D.D., T.B., S.G., S.H., and L.T. designed the study; T.B. performed all experiments except ChEC-seq analyses; S.G. and N.M. performed ChEC-seq experiments; M.J.E.K. and H.T.M.T. provided anchor-away strains; T.B., D.D., and S.G. analyzed data; D.D., T.B., S.G., S.H., and L.T. wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

We are grateful to Violaine Alunni and Christelle Thibault (IGBMC) for array hybridization. We thank Vincent Géli and Pierre Luciano (CRCM, Marseille, France) for advice and discussion, Gabe Zentner (Indiana University) for advice on ChEC-seq data analysis, and P. Anthony Weil for antibodies. We thank Farrah El Saafin for critically reading the manuscript. T.B. was supported by a Marie Curie-ITN fellowship (EU-FP7 PEOPLE-2013 program, PITN-GA-2013-606806, NR-NET) and the Fondation ARC. This work was supported by NIH grants GM053451 and GM075114 (to S.H.), funds from the Agence Nationale de la Recherche (ANR-15-CE11-0022 SAGA2 to D.D.; ANR-13-BSV8-0021-03 DiscoverIID to L.T.), and the European Research Council Advanced grant (ERC-2013-340551, Birtoaction, to L.T.). This study was also supported by ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02. N.M. is a student in the Magistère de Génétique Graduate Program at Université Paris Diderot, Sorbonne, Paris Cité.

Received: March 29, 2017
Revised: June 28, 2017
Accepted: August 18, 2017
Published: September 14, 2017

REFERENCES

Ansari, S.A., He, Q., and Morse, R.H. (2009). Mediator complex association with constitutively transcribed genes in yeast. Proc. Natl. Acad. Sci. USA *106*, 16734–16739.

Bähler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., 3rd, Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. Yeast *14*, 943–951.

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.

Bhaumik, S.R., and Green, M.R. (2001). SAGA is an essential in vivo target of the yeast acidic activator Gal4p. Genes Dev. *15*, 1935–1945.

Bhaumik, S.R., and Green, M.R. (2002). Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters in vivo. Mol. Cell. Biol. *22*, 7365–7371.

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.

Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D. (1998). Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast *14*, 115–132.

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

Collart, M.A., and Timmers, H.T. (2004). The eukaryotic Ccr4-not complex: a regulatory platform integrating mRNA metabolism with cellular signaling pathways? Prog. Nucleic Acid Res. Mol. Biol. *77*, 289–322.

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.

Dudley, A.M., Rougeulle, C., and Winston, F. (1999). The Spt components of SAGA facilitate TBP binding to a promoter at a post-activator-binding step in vivo. Genes Dev. *13*, 2940–2945.

Eberharter, A., Sterner, D.E., Schieltz, D., Hassan, A., Yates, J.R., 3rd, Berger, S.L., and Workman, J.L. (1999). The ADA complex is a distinct histone acetyltransferase complex in Saccharomyces cerevisiae. Mol. Cell. Biol. *19*, 6621–6631.

Eyboulet, F., Wydau-Dematteis, S., Eychenne, T., Alibert, O., Neil, H., Boschiero, C., Nevers, M.C., Volland, H., Cornu, D., Redeker, V., et al. (2015). Mediator independently orchestrates multiple steps of preinitiation complex assembly in vivo. Nucleic Acids Res. *43*, 9214–9231.

Gardner, R.G., Nelson, Z.W., and Gottschling, D.E. (2005). Ubp10/Dot4p regulates the persistence of ubiquitinated histone H2B: distinct roles in telomeric silencing and general chromatin. Mol. Cell. Biol. *25*, 6123–6139.

Geisberg, J.V., Moqtaderi, Z., Fan, X., Ozsolak, F., and Struhl, K. (2014). Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell *156*, 812–824.

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

Grünberg, S., and Zentner, G.E. (2017). Genome-wide Mapping of Protein-DNA Interactions with ChEC-seq in Saccharomyces cerevisiae. J. Vis. Exp. Published online June 3, 2017. http://dx.doi.org/10.3791/55836.

Grünberg, S., Henikoff, S., Hahn, S., and Zentner, G.E. (2016). Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. EMBO J. *35*, 2435–2446.

Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705–736.

Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millán-Zambrano, G., Barkai, O., Chávez, S., Pérez-Ortín, J.E., Darzacq, X., and Choder, M. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. Cell *153*, 1000–1011.

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

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell *31*, 925–932.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

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

Hirschhorn, J.N., Bortvin, A.L., Ricupero-Hovasse, S.L., and Winston, F. (1995). A new class of histone H2A mutations in Saccharomyces cerevisiae causes specific transcriptional defects in vivo. Mol. Cell. Biol. *15*, 1999–2009.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717–728.

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.

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., and Knop, M. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast *21*, 947–962.

Koutelou, E., Hirsch, C.L., and Dent, S.Y. (2010). Multiple faces of the SAGA complex. Curr. Opin. Cell Biol. *22*, 374–382.

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.L., Rougemont, J., and Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. Mol. Cell *60*, 422–434.

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.

Larschan, E., and Winston, F. (2001). The S. cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4. Genes Dev. *15*, 1946–1956.

Lee, T.I., Causton, H.C., Holstege, F.C., Shen, W.C., Hannett, N., Jennings, E.G., Winston, F., Green, M.R., and Young, R.A. (2000). Redundant roles for the TFIID and SAGA complexes in global transcription. Nature *405*, 701–704.

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

Lenstra, T.L., and Holstege, F.C. (2012). The discrepancy between chromatin factor location and effect. Nucleus *3*, 213–219.

Lenstra, T.L., Benschop, J.J., Kim, T., Schulze, J.M., Brabers, N.A., Margaritis, T., van de Pasch, L.A., van Heesch, S.A., Brok, M.O., Groot Koerkamp, M.J., et al. (2011). The specificity and topology of chromatin interaction pathways in yeast. Mol. Cell *42*, 536–549.

Mohibullah, N., and Hahn, S. (2008). Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. Genes Dev. *22*, 2994–3006.

Müller, F., and Tora, L. (2014). Chromatin and DNA sequences in defining promoters for transcription initiation. Biochim. Biophys. Acta *1839*, 118–128.

O'Duibhir, E., Lijnzaad, P., Benschop, J.J., Lenstra, T.L., van Leenen, D., Groot Koerkamp, M.J., Margaritis, T., Brok, M.O., Kemmeren, P., and Holstege, F.C. (2014). Cell cycle population effects in perturbation studies. Mol. Syst. Biol. *10*, 732.

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., et al. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature *518*, 376–380.

Rando, O.J., and Winston, F. (2012). Chromatin and transcription in yeast. Genetics *190*, 351–387.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.

Robert, F., Pokholok, D.K., Hannett, N.M., Rinaldi, N.J., Chandy, M., Rolfe, A., Workman, J.L., Gifford, D.K., and Young, R.A. (2004). Global position and recruitment of HATs and HDACs in the yeast genome. Mol. Cell *16*, 199–209.

Roberts, S.M., and Winston, F. (1997). Essential functional interactions of SAGA, a Saccharomyces cerevisiae complex of Spt, Ada, and Gcn5 proteins, with the Snf/Swi and Srb/mediator complexes. Genetics *147*, 451–465.

Rodríguez-Molina, J.B., Tseng, S.C., Simonett, S.P., Taunton, J., and Ansari, A.Z. (2016). Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization. Mol. Cell *63*, 433–444.

Rodríguez-Navarro, S. (2009). Insights into SAGA function during gene expression. EMBO Rep. *10*, 843–850.

Schwalb, B., Schulz, D., Sun, M., Zacher, B., Dümcke, S., Martin, D.E., Cramer, P., and Tresch, A. (2012). Measurement of genome-wide RNA synthesis and decay rates with Dynamic Transcriptome Analysis (DTA). Bioinformatics *28*, 884–885.

Setiaputra, D., Ross, J.D., Lu, S., Cheng, D.T., Dong, M.Q., and Yip, C.K. (2015). Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. J. Biol. Chem. *290*, 10057–10070.

Sterner, D.E., Grant, P.A., Roberts, S.M., Duggan, L.J., Belotserkovskaya, R., Pacella, L.A., Winston, F., Workman, J.L., and Berger, S.L. (1999). Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. Mol. Cell. Biol. *19*, 86–98.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res. *22*, 1350–1359.

Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol. Cell *52*, 52–62.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit. Rev. Biochem. Mol. Biol. *41*, 105–178.

Thompson, C.M., and Young, R.A. (1995). General requirement for RNA polymerase II holoenzymes in vivo. Proc. Natl. Acad. Sci. USA *92*, 4587–4590.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Res. *18*, 1084–1091.

Tirosh, I., Berman, J., and Barkai, N. (2007). The pattern and evolution of yeast promoter bendability. Trends Genet. *23*, 318–321.

Tora, L., and Timmers, H.T. (2010). The TATA box regulates TATA-binding protein (TBP) dynamics in vivo. Trends Biochem. Sci. *35*, 309–314.

van Werven, F.J., van Bakel, H., van Teeffelen, H.A., Altelaar, A.F., Koerkamp, M.G., Heck, A.J., Holstege, F.C., and Timmers, H.T. (2008). Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome. Genes Dev. *22*, 2359–2369.

Venters, B.J., Wachi, S., Mavrich, T.N., Andersen, B.E., Jena, P., Sinnamon, A.J., Jain, P., Rolleri, N.S., Jiang, C., Hemeryck-Walsh, C., and Pugh, B.F. (2011). A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. Mol. Cell *41*, 480–492.

Villanyi, Z., and Collart, M.A. (2015). Ccr4-Not is at the core of the eukaryotic gene expression circuitry. Biochem. Soc. Trans. *43*, 1253–1258.

Vosnakis, N., Koch, M., Scheer, E., Kessler, P., Mély, Y., Didier, P., and Tora, L. (2017). Coactivators and general transcription factors have two distinct dynamic populations dependent on transcription. EMBO J. Published online July 19, 2017. http://dx.doi.org/10.15252/embj.201696035.

Please cite this article in press as: Baptista et al., SAGA Is a General Cofactor for RNA Polymerase II Transcription, Molecular Cell (2017), http://dx.doi.org/10.1016/j.molcel.2017.08.016

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*. Published online September 14, 2017. http://dx.doi.org/10.1016/j.molcel.2017.08.014.

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

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Zentner, G.E., Kasinathan, S., Xin, B., Rohs, R., and Henikoff, S. (2015). ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. Nat. Commun. *6*, 8733.

Please cite this article in press as: Baptista et al., SAGA Is a General Cofactor for RNA Polymerase II Transcription, Molecular Cell (2017), http://dx.doi.org/10.1016/j.molcel.2017.08.016

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-Flag M2 antibody	Sigma	Cat# F1804; RRID: AB_262044
Rabbit polyclonal anti-H3K4me3	Abcam	ab8580
Rabbit polyclonal anti-H3K9ac	Abcam	ab4441; RRID: AB_2118292
Rabbit polyclonal anti-H3K36me3	Abcam	ab9050; RRID: AB_306966
Rat monoclonal anti-RNA Pol II CTD phosphoSer2	Active Motif	Cat# 61083; RRID: AB_2687450
Rat monoclonal anti-RNA Pol II CTD phosphoSer5	Active Motif	Cat# 61085; RRID: AB_2687451
Rabbit polyclonal anti-Taf4	P. Anthony Weil	N/A
Rabbit polyclonal anti-Taf5	P. Anthony Weil	N/A
Rabbit polyclonal anti-Taf6	P. Anthony Weil	N/A
Rabbit polyclonal anti-Taf10	P. Anthony Weil	N/A
Mouse monoclonal anti-Taf10	László Tora	2D5
Rabbit polyclonal anti-Ada1	Steven Hahn	5255
Rabbit polyclonal anti-Spt3	Steven Hahn	5113
Rabbit polyclonal anti-Gcn5	P. Anthony Weil	N/A
Rabbit polyclonal anti-Spt7	Fred Winston	N/A
Rabbit polyclonal anti-H3	Abcam	ab1791
Chemicals, Peptides, and Recombinant Proteins		
cOmplete mini, EDTA-free Protease inhibitor cocktail	Roche	Cat# 11836170001
4-Thiouracil	Sigma-Aldrich	Cat# 440736
Rapamycin	Euromedex	Cat# SYN-1185
EZ-Link HPDP Biotin	ThermoFisher	Cat# 21341
Thiolutin	Abcam	ab143556
EZview Red anti-HA Affinity Gel	Sigma-Aldrich	Cat# E6779
Protein A-Sepharose 4B, Fast flow	Sigma-Aldrich	Cat# P9424
Critical Commercial Assays		
RiboPure RNA Purification kit, yeast	ThermoFisher	Cat# AM1926
μMACS Streptavidin kit	Miltenyi Biotec	Cat# 130-074-101
Deposited Data		
Raw and analyzed data (Microarrays)	This paper	GEO: GSE96849
Raw and analyzed data (ChEC-seq)	This paper	GEO: GSE97379
Original images of western blot data	Mendeley Data	http://dx.doi.org/10.17632/gwtwc4ndpj.1
Experimental Models: Organisms/Strains		
S. cerevisiae strain background BY4705 (see Table S1)	Brachmann et al., 1998	N/A
SPT3-3FLAGMNase, BY4705 (see Table S1)	This study	N/A
SPT7-3FLAGMNase, BY4705 (see Table S1)	This study	N/A
SPT8-3FLAGMNase, BY4705 (see Table S1)	This study	N/A
UBP8-3FLAGMNase, BY4705 (see Table S1)	This study	N/A
S. cerevisiae strain background FY406 (see Table S1)	Hirschhorn et al., 1995	N/A
WTyH2B, FY406 (see Table S1)	Bonnet et al., 2014	N/A
ubp8Δ, FY406 (see Table S1)	Bonnet et al., 2014	N/A
gcn5Δ, FY406 (see Table S1)	Bonnet et al., 2014	N/A
spt7Δ, FY406 (see Table S1)	Bonnet et al., 2014	N/A

(*Continued on next page*)

CellPress

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
spt20Δ, FY406 (see Table S1)	Bonnet et al., 2014	N/A
ubp8Δgcn5Δ, FY406 (see Table S1)	This study	N/A
spt3Δ, FY406 (see Table S1)	This study	N/A
spt8Δ, FY406 (see Table S1)	This study	N/A
spt3Δgcn5Δ, FY406 (see Table S1)	This study	N/A
SPT7-FRB, BY4742 (see Table S1)	This study	N/A
S. cerevisiae strain background BY4742 (see Table S1)	Euroscarf	N/A
3HA-TBP, BY4742 (see Table S1)	This study	N/A
spt3Δ, 3HA-TBP, BY4742 (see Table S1)	This study	N/A
Oligonucleotides		
Primers for RT-qPCR, see Table S3	This paper	This paper
Primers for ChIP-qPCR, see Table S3	This paper	This paper
Recombinant DNA		
pGZ108 (see Table S3)	Zentner et al., 2015	N/A
pFA6a–hphNT1(see Table S3)	Janke et al., 2004	N/A
pFA6a-kanMX6 (see Table S3)	Bähler et al., 1998	N/A
YIplac211-3HA-TBP (see Table S3)	Eyboulet et al., 2015	N/A
Software and Algorithms		
GraphPad Prism, version 6.0	GraphPad Software	https://www.graphpad.com/
R	R Core Team	https://www.r-project.org/
RStudio: Integrated Development for R	RStudio Team	https://www.rstudio.com/

Molecular Cell, Volume *68*

Supplemental Information

SAGA Is a General Cofactor for RNA Polymerase II Transcription

Tiago Baptista, Sebastian Grünberg, Nadège Minoungou, Maria J.E. Koster, H.T. Marc Timmers, Steve Hahn, Didier Devys, and László Tora

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead author, Didier Devys (devys@igbmc.fr).

METHOD DETAILS

Yeast strains

Yeast strains for ChEC-seq were generated in the BY4705 background (Brachmann *et al.*, 1998). SAGA subunits were endogenously tagged with 3xFLAG-MNase by transformation with gene-specific PCR products derived from pGZ108 (Zentner *et al.*, 2015). Different yeast strains used in this study were also derived from the previously described FY406 (Hirschhorn *et al.*, 1995). Deletion mutants were generated as described previously (Bahler *et al.*, 1998; Janke *et al.*, 2004) or by transformation with PCR products amplified from genomic DNA of the corresponding strain obtained from the *Saccharomyces* Genome Deletion (SGD) strain database. The HA-TBP strain expressing N-terminal HA-tagged version of TBP was obtained by inserting 3HA epitopes preceded by *URA3* marker using an integrative plasmid (YIplac211-3HA-TBP) (Eyboulet *et al.*, 2015). A complete and descriptive list of the strains and plasmids used in this study can be found in Supplemental Tables S1 and S2, respectively.

ChEC-seq

For ChEC analysis, yeast strains were grown in synthetic media to an OD_{600} of 0.5-0.7. ChEC reactions were performed as described and quenched after 5 min (Grunberg *et al.*, 2016). Sequencing library preparation, alignment, track visualization, and cleavage pattern analysis were performed as described (Zentner *et al.*, 2015; Grunberg *et al.*, 2016). The scripts used for data processing are available at https://github.com/zentnerlab/chec-seq. HOMER (http://homer.salk.edu) (Heinz *et al.*, 2010) was used to generate reads per million (RPM)-normalized average plots. TFIID- and SAGA-dependent (Huisinga and Pugh, 2004) or TATA-containing or TATA-less genes (Rhee and Pugh, 2012) were assigned to previous TSS annotations (Xu *et al.*, 2009).

RNA labeling and newly-transcribed RNA purification

For each yeast strain and biological replicate, 100mL of wild-type and mutant *S. cerevisiae* cells were grown in YPD medium at 30°C to an $OD_{600} \approx 0.8$. Newly-synthesized RNAs were labeled for 6 min by adding freshly-prepared 4-thiouracil (Sigma-Aldrich) until a final concentration of 5mM. In parallel, wild-type *S. pombe* cells were similarly grown in YES medium, at 31°C, and labeled for 6 min to be used as a spike-in across all samples. Upon labeling, cells were immediately pelleted and flash-frozen in liquid N_2 and stored at -80°C until further use. All experiments using deletion strains were performed using at least two independent biological replicates.

Before total RNA extraction, *S. cerevisiae* and *S. pombe* cells were mixed in a ratio of 3:1. Total RNAs were extracted using RiboPure yeast kit (Ambion, Life Technologies) according to the description provided by the manufacturer. Prior to biotinylation, RNA samples were heated for 10 min at 60°C and cooled down on ice for another 5 min. 200µg of total RNA were biotinylated using 200µL of 1mg/mL EZ-link HPDP- Biotin (Pierce) in 100µL of biotinylation buffer (100mM Tris- HCl pH 7.5, 10mM EDTA) and adjusted to a final volume of 1000µL with DEPC-treated RNase-free water (Sigma-Aldrich) for 3 h at room temperature. After chloroform extraction and isopropanol precipitation ($^1/_{10}$ vol 5M NaCl and 2.5 vol isopropanol), purified RNAs were suspended in 100µL of DEPC-treated RNase-free water (Sigma-Aldrich).

Recovered RNA samples were incubated at 65°C for 10 min and allowed to cool down on ice for 5 min. Newly-synthesized biotinylated RNAs were bound to 100µL of µMACS streptavidin microbeads (Miltenyi Biotec) for 90 min at room temperature with gentle shaking. Purification of labeled RNA was then carried out using µMACS streptavidin starting kit (Miltenyi Biotec). Columns were first equilibrated with 1mL of washing buffer (100mM Tris-HCl at pH 7.5, 10mM EDTA, 1M NaCl, 0.1% Tween20). Samples were passed through the columns twice and washed five times with incremental volumes of washing buffer (600, 700, 800, 900, and 1000µL). Ultimately, labeled RNAs were eluted twice with 200µL of 100mM DTT. Following ethanol precipitation (overnight precipitation in $^1/_{10}$ vol of 3M NaOAc, 3 vol of 100% ethanol and 20µg of RNA-grade glycogen), RNAs were washed in ice-cold 70% ethanol and resuspended in 20µL of DEPC-treated RNase-free water (Sigma-Aldrich). Samples were stored at -80°C until further use.

For the anchor-away strain (SPT7-FRB), 200mL of *S. cerevisiae* cells were grown in YPD medium at 30°C to an $OD_{600} \approx 0.8$. At that point, the culture was divided into two equal volumes of 100mL each and to one of those cultures rapamycin was added (+RAPA) until a final concentration of 1μg/mL, for 60 min, to allow nuclear depletion of Spt7. To the other half of the culture, the control samples (-RAPA/minus rapamycin), only the vehicle (DMSO) was added. Newly synthesized RNAs were labeled for 6 min by adding 4-thiouracil (Sigma-Aldrich) until a final concentration of 5mM.

For time-course experiments, 600mL of *S. cerevisiae* were grown and when $OD_{600} \approx 0.8$, 100mL of the culture were collected (timepoint 0) and rapamycin was added up to a final concentration of 1μg/mL. Labeling was performed as described above and aliquots of 100mL were collected, representing 15, 30, 60, 120 and 240 min of nuclear depletion of Spt7. In all cases, cells were collected, counted and frozen in liquid N_2. In parallel, and as described above, wild-type *S. pombe* cells were similarly grown in YES medium, at 31°C, and labeled for 6 min to be used as a spike-in across all samples. All experiments with anchor-away strains were performed using three independent biological replicates.

RT-qPCR

When performing RT-qPCR, cDNA synthesis was performed on 2μg of total RNA or 10μL of labeled RNA using random hexamers and Transcriptor reverse transcriptase (Roche) according to the manufacturer's instructions. Real-time qPCR were performed using SYBR Green I Master (Roche). A list of all of the primers used can be found in Supplemental Table S3. All samples were run in triplicate and using at least two biological replicates. After qPCR, all raw values were corrected for the expression of *S. pombe* tubulin. Finally, results were represented graphically as a relative comparison between the wild-type (set to 1) and mutant samples.

Genome-wide expression analyses

RNA samples were hybridized to GeneChip Yeast Genome 2.0 microarrays following the instructions from the supplier (Affymetrix). Briefly, biotinylated cRNA targets were prepared from 150ng of total RNA using the "MessageAmp™ Premier RNA Amplification Kit" (Ambion), according to the Instruction Manual # 4386269 Revision B, 18 september 2007. Following fragmentation, 4 μg of cRNAs were hybridized for 16 hours at 45°C, 60rpm on

GeneChip® Yeast Genome 2.0 arrays (Affymetrix). The chips were washed and stained in the GeneChip® Fluidics Station 450 (Affymetrix) using the FS450_0003 script and scanned with the GeneChip® Scanner 3000 7G (Affymetrix) at a resolution of 1.56µm. Raw data (.CEL Intensity files) were extracted from the scanned images using the Affymetrix GeneChip® Command Console (AGCC) version 4.1.2. CEL files were further processed with Affymetrix Expression Console software version 1.4.1 to calculate probeset signal intensities, using the Affymetrix statistics-based algorithms MAS-5.0 with default settings and global scaling as normalization method. The trimmed mean target intensity of each chip was arbitrarily set to 100.

All experiments were done using at least two independent biological replicates. Raw data were normalized to *S. pombe* signal and used to calculate fold-changes in total and newly-synthesized RNA levels, as represented in volcano plots. Further calculations of synthesis and decay rates, based on mathematical model as previously described, were performed using a pipeline and R/Bioconductor package publicly available (Schwalb *et al.*, 2012; Sun *et al.*, 2012). TFIID- and SAGA-dependent or TATA-containing or TATA-less genes were defined as previously mentioned (Huisinga and Pugh, 2004; Rhee and Pugh, 2012).

Growth curve and cell viability analyses

For growth curve analysis, the anchor-away (SPT7-FRB) and parental strains were grown overnight and then diluted to $OD_{600} \approx 0.1$ in pre-heated YPD (200mL), before rapamycin addition at a final concentration of 1µg/mL (+ RAPA). To the control samples (-RAPA), only the vehicle (DMSO) was added. Every 30min, OD_{600} was measured until saturation of the culture. For cell viability, strains were grown until log-phase ($OD_{600} \approx 0.7$). After 30 and 60min an aliquot was collected, stained with trypan blue and counted using a neubauer chamber.

Cell fractionation

150 ml cell culture ($OD_{600} \approx 0.5$) were centrifuged and treated with 0.1M Tris, pH 9.4, 10mM DTT for 5 min at room temperature. Cells were pelleted and resuspended in YPD/S (YPD with 1M Sorbitol) and collected by centrifugation. Cells were resuspended in 1mL YPD/S and 750µL of 2M sorbitol. Zymolyaze 100T was added until a final concentration of

0.3mg/mL. Cells were incubated at 30°C for 30min (progress of spheroplasting was checked every 10min). Spheroplasts were washed and resuspended in YPD/S for 30 min at 30°C. Cells were quickly cooled down in ice, centrifuged, washed and resuspended in 1.5 ml lysis buffer (20 mM K-phosphate pH 6.5, 0.5 mM $MgCl_2$, 18% Ficoll, 1xPIC). Spheroplasts were lysed three times with a B dounce and shortly centrifuged at 4500g to remove unbroken cells. The precleared whole cell lysate (Total lysate fraction) was centrifuged at 21000g for 45 min resulting into a crude nuclear pellet (nuclear fraction) and post-nuclear supernatant cytosol fraction (cytoplasmic fraction). 200µl of each fraction was TCA precipitated, washed with ice-cold acetone and resuspended in SDS-sample buffer.

RNA half-life quantification

Wild-type and mutant *S. cerevisiae* cells (two independent biological replicates for each strain) were grown in 300mL YPD medium at 30°C to an OD_{600}≈0.8. A 50mL aliquot wascollected and used as timepoint = 0. Immediately after, thiolutin was added to the cells until a final concentration of 3µg/mL. Aliquots of 50mL were collected after 10, 20, 30, 45 and 60 min exposure to thiolutin. Cells were collected by centrifugation and frozen in liquid N_2. RNA was extracted using RiboPure yeast kit (Ambion, Life Technologies) according to the description provided by the manufacturer. cDNA preparations and RT-qPCR analyses were performed as previously described. Results were represented graphically as a relative comparison between the wild-type (set to 1) and mutant samples.

Whole cell extract and immunoprecipitation

150mL of cultures were grown in YPD at 30°C until OD_{600} ≈ 0.8. Cells were pelleted, washed in ice-cold 1x PBS, centrifuged and the pellet was frozen at 80°C. 500µL of buffer I (40 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.1% Tween, 10% glycerol, 1x PIC) were added to the pellet, and cells were broken by addition of acid-washed glass beads (0.5 mm diameter). Whole-cell extracts were clarified by centrifugation for 5 min at 4°C, and the protein concentration was determined.

Immunoprecipitations were carried out using protein A sepharose beads (Sigma) equilibrated in buffer I (40 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Tween, 10% glycerol, 1x PIC). 2mg of whole cell extract were precleared with 100µL of protein A sepharose beads, and immunoprecipitated using 5µL of antibody (anti-Taf10 and anti-Ada1) and 100µL of

protein A sepharose beads. Beads were washed three times with 1mL of buffer II (40 mM Tris-HCl pH 8.0, 350mM NaCl, 0.1% Tween, 10% glycerol, 1x PIC) and three times with 1mL of buffer I. Bound protein complexes were eluted twice with 50µL 3M glycine pH 2.8 and then neutralized with 6.73µL of Tris-HCl pH9.8. From these eluates, equal volumes of 10µL were separated in a SDS-PAGE and analyzed by Western blot.

Chromatin immunoprecipitation

ChiP was performed as previously reported using three independent biological replicates (Bonnet *et al.*, 2014). Briefly, cell cultures in log-phase (OD≈0.8) were cross-linked with 1% formaldehyde for 15 min (at room temperature, with gentle agitation) and then quenched with 0.125M glycine for 5 min (4°C, with gentle agitation). Cells were harvested, washed twice with ice-cold 1xPBS and resuspended in FA lysis buffer (50mM HEPES pH 7.5, 140mM NaCl, 1mM EDTA, 0.1% sodium deoxycholate, 1% Triton X-100, 0.4mM DTT, protease inhibitor cocktail (Roche)). Cells were lysed by mechanical shearing using acid-washed glass beads for 45 mins at 4°C, with intervals where samples were cooled-down on ice to avoid over-heating. The pellets were collected, washed once with FA lysis buffer and then sonicated in FA lysis buffer using a Covaris E220 machine to achieve an average DNA fragment size of 200bp. The Covaris E220 settings were as following: peak incident power of 140 Watts, duty factor of 5%, 200 bursts per cycle and 600 seconds (10 min) duration. The sonication was performed at 4°C and for six cycles of the described protocol. After each 10 min cycle, the samples would be left in the bath while other samples would be sonicating, to avoid overheating.

For each ChIP reaction, 250µg of chromatin extract were used. The equivalent to 1% of total sonicated chromatin was used as input. Input samples were kept at 4°C until the elution stage, when it was treated the same way the immunoprecipitated samples. After pre-clearing with protein-G sepharose, chromatin extracts were incubated overnight with 50µL EZview Red anti-HA affinity gel beads (Sigma). The resins were washed once with the following buffers at room temperature for 10min, except for the wash with Tris-EDTA buffer that was performed twice under the same conditions: **(i)** FA lysis buffer, **(ii)** FA lysis buffer containing 0.5 M NaCl, **(iii)** LiCl-containing buffer (10mM Tris-HCl pH 7.5, 1mM EDTA, 0.25M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate) and finally **(iv)** Tris-EDTA buffer (10mM Tris-HCl pH 7.5 and 1mM EDTA). All buffers were also complemented with

protease inhibitor cocktail (Roche). Bound chromatins were eluted with 250µL of elution buffer (50mM Tris-HCl pH7.5, 10mM EDTA, 0.1M NaOAc, 1% SDS) for 30 min at 65°C with shaking and then reverse cross-linked with RNase A treatment at 65°C overnight. Immunopurified material was incubated with 20 µg proteinase K for 1 h at 55°C. DNA was extracted once with phenol:chloroform:isoamyl alcohol (25:24:1) and extracted a second time with chloroform only. DNA was precipitated overnight at -20°C ($^1/_{10}$ vol 3M NaOAc, 3 vol 100% ethanol and 20µg of glycogen). DNA were then recovered and resuspended in 60µL of Tris-EDTA buffer (10mM Tris-HCl pH 7.5 and 1mM EDTA).

ChIP-qPCR

For ChIP-qPCR experiments, three independent sets were performed and analyzed by qPCR. qPCR were performed using SYBR Green I Master (Roche) as described above. A list of all of the primers used can be found in Supplemental Table S3. Each sample was run in triplicate and using three biological replicates. Normalization was conducted using the input percent method and results were presented as % of input.

Western blotting

For histone marks analyses, whole-cell extracts were prepared as described (Gardner *et al.*, 2005) with minor changes. Cells from log-phase yeast cultures were harvested by centrifugation and lysed in 400 mL of SUME buffer (8 M urea, 1% SDS, 10 mM MOPS at pH 6.8, 10 mM EDTA, 0.01% bromophenol blue) by mechanical shearing. For RNA Pol II phosphoSer2 and phosphoSer5 analyses, extracts were prepared by treating cell pellets with 0.2M NaOH for 5min and boiling the samples in 2x Laemmli buffer for 5min. Extracts were cleared from debris through centrifugation. Antibodies used in this study were as follows: anti-H3K9ac (Abcam, ab4441), anti-Flag (Sigma, M2), anti-H3K4me3 (Abcam, ab8580), anti-H3K36me3 (Abcam, ab9050), anti-RNA Pol II phosphoSer2 (3E10, Active Motif), anti-RNA Pol II phosphoSer5 (3E8, Active Motif) and anti-Taf4 (gift from P. Anthony Weil).

DATA AND SOFTWARE AVAILABILITY

Data Resources

The accession numbers for the data reported in this paper are GEO: GSE96849 and GSE97379.

Supplemental information

Figure S1. The majority of SAGA-bound UASs is bound by all SAGA-subunits, Related to Figures 1 and 2.

Figure S2. Loss of SAGA leads to increased mRNA half-lives, Related to Figure 3.

Figure S3. FRB strain characterization and validation of Spt7 nuclear depletion, Related to Figure 5.

Figure S4. Nuclear depletion of Spt7 decreased transcription of both SAGA- and TFIID-dominated genes, Related to Figure 5.

Figure S5. cDTA analyses for several SAGA mutants, Related to Figure 6.

Figure S6. SAGA complex characterization upon deletion of one or more subunits, Related to Figure 6.

Figure S7. Transcriptional changes observed in SAGA mutants do not correlate with the slow growth phenotype gene expression signature, Related to Figure 6.

Table S1. Genotypes of the yeast strains used in this study, Related to STAR Methods.

Table S2. List of plasmids used in this study, Related to STAR Methods.

Table S3. List of primers used for RT-qPCR and ChIP-qPCR, Related to STAR Methods.

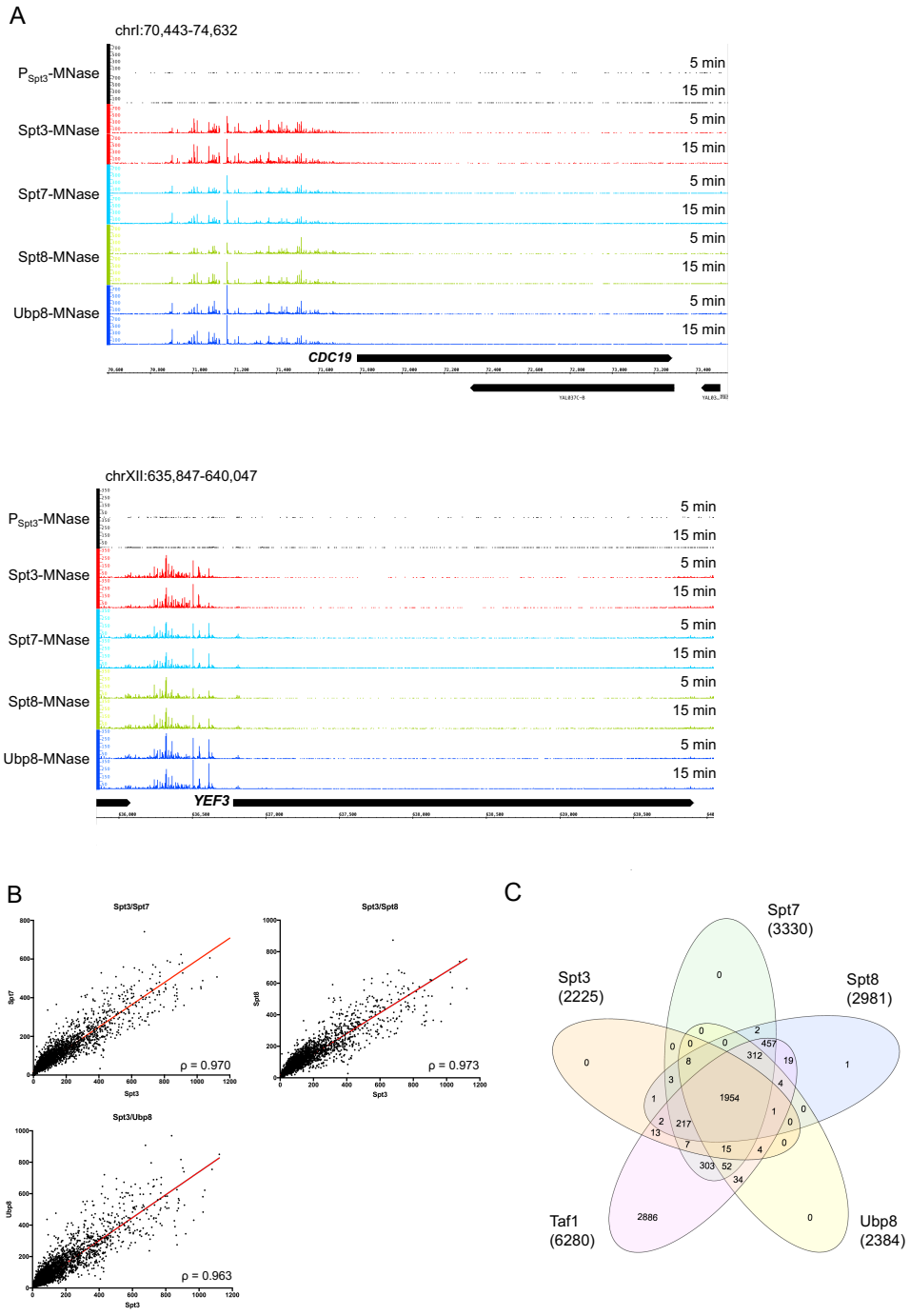
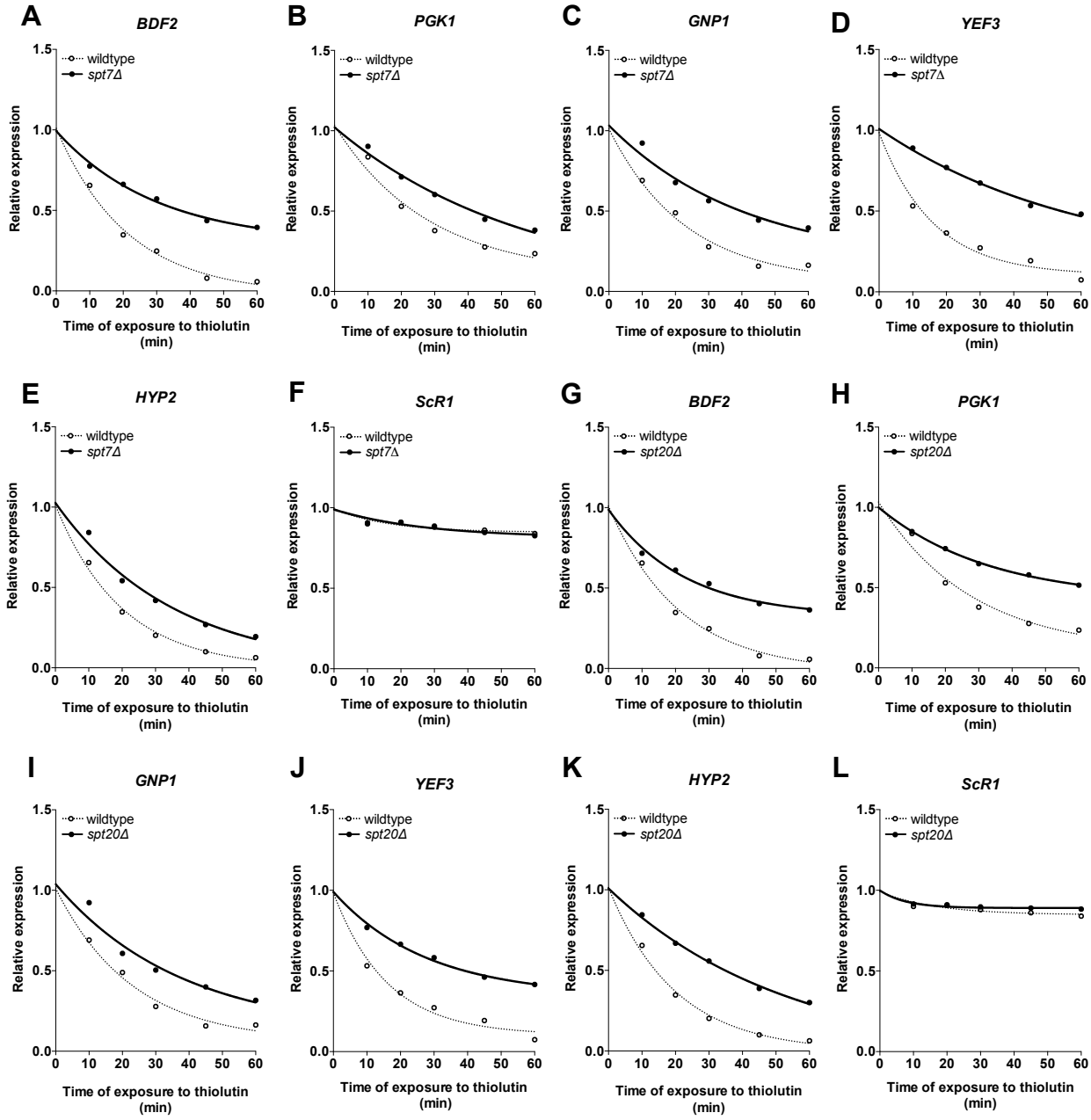


Figure S1. The majority of SAGA-bound UASs is bound by all SAGA-subunits, Related to Figures 1 and 2. (A) ChEC-seq mapping of SAGA at CDC19 (top panel) and YEF3 (bottom panel). Signal tracks showing cleavage of genomic DNA by Spt3- (red), Spt7- (cyan), Spt8- (green), Ubp8-MNase (dark blue), and a PSpt3-MNase control (black) 5 and 15 min after addition of CaCl2. **(B)** Pairwise correlations of SAGA subunits as indicated above the graphs. The Spearman's rank correlation coefficient r for each pairwise comparison is shown. **(C)** More than 99% of genes bound

by SAGA are also bound by TFIID. Genes with at least 5% ChEC-seq signal compared to the gene with the highest signal for the respective SAGA- or Taf1-MNase variants were analyzed.



	Reported $t_{1/2}$ (min)	wildtype $t_{1/2}$ (min)	*spt7Δ* $t_{1/2}$ (min)	*spt20Δ* $t_{1/2}$ (min)
BDF2	15	14.8	34.6	31.2
GNP1	15	15.9	35.9	33.9
HYP2	19	19.8	30.2	37.2
PGK1	23	20.1	30.7	33.4
YEF3	12	11.1	41.5	43.2

Figure S2. Loss of SAGA leads to increased mRNA half-lives, Related to Figure 3. (**A-L**) Measurement of RNA half-life using transcription inhibition with thiolutin in *spt7*Δ **(AF)** or *spt20*Δ

strains **(G-L)**. Comparison between half-lives reported elsewhere and the ones obtained in this experiments.

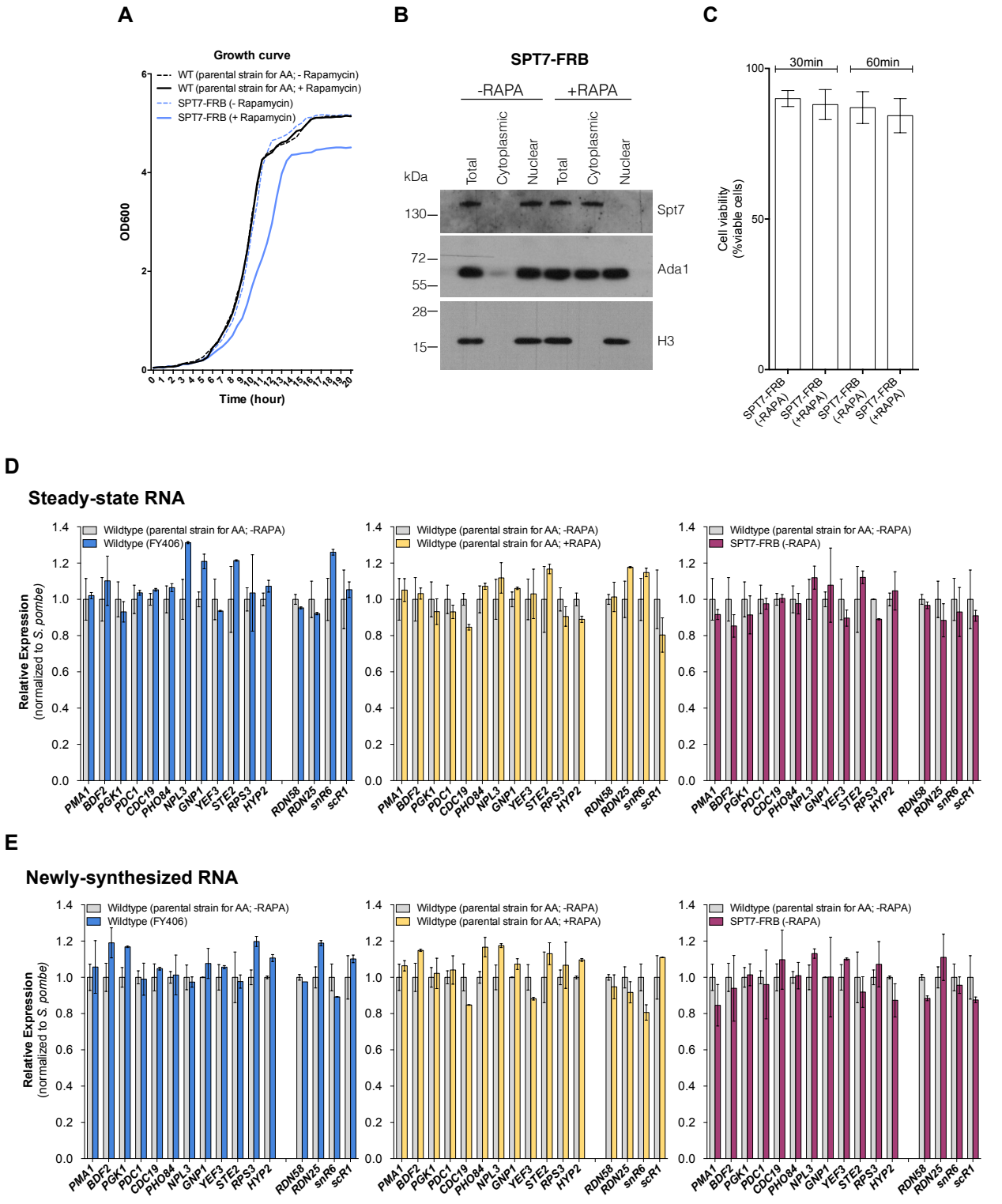


Figure S3. FRB strain characterization and validation of Spt7 nuclear depletion, Related to Figure 5. (A) SPT7-FRB strain does not present a growth phenotype in comparison to the parental strain. Exposure to rapamycin eventually promotes a slower growth phenotype, as observed for the corresponding constitutive deletion strain. **(B)** Cell fraction depicting efficient nuclear depletion of

Spt7 upon exposure to rapamycin. **(C)** Upon 30 min of exposure to rapamycin viability of cells in log-phase is not affected, in comparison to its counterpart with vehicle only. **(D-E)** Fusion of FRB domain to Spt7 does not affect RNA Pol II transcription by itself, both at the steady-state **(D)** or newly-synthesized RNA **(E)** levels. Expression values (mean ± SD of three independent experiments) were normalized to spiked-in *S. pombe* signal.

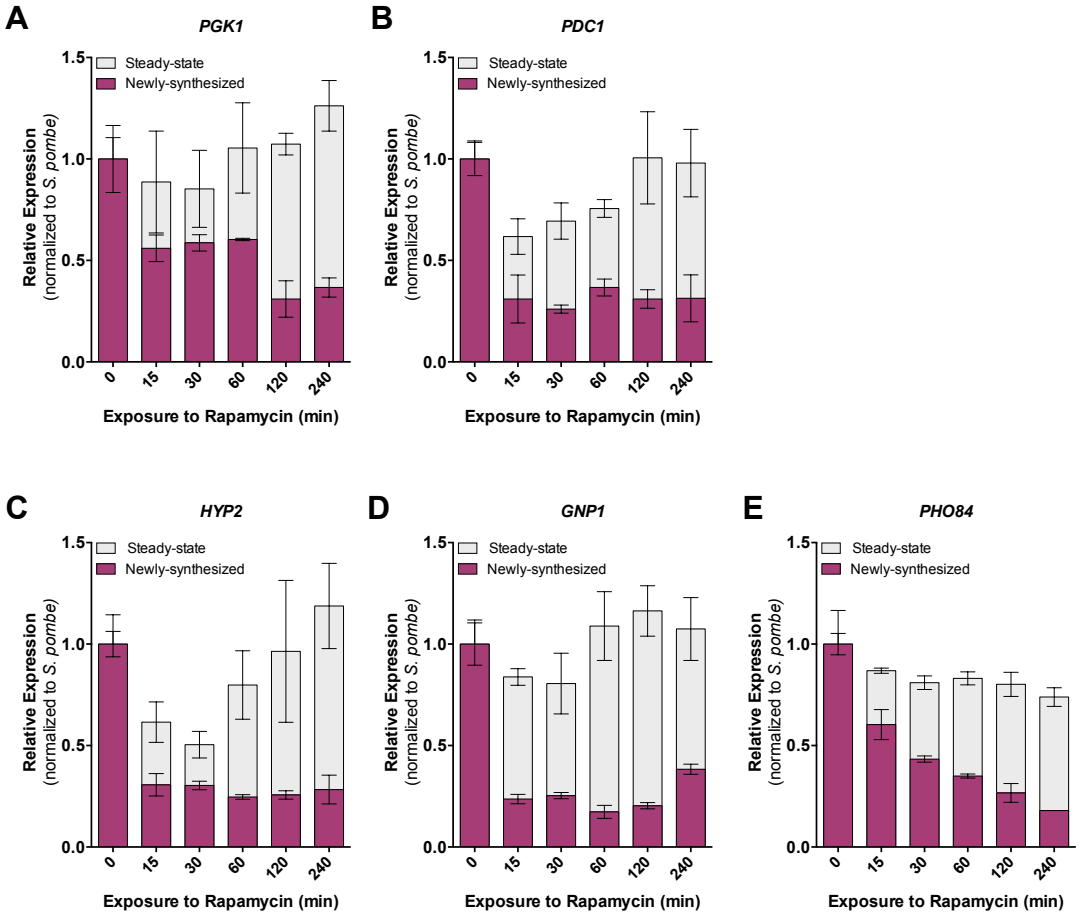


Figure S4. Nuclear depletion of Spt7 decreased transcription of both SAGA- and TFIID-dominated genes, Related to Figure 5. (**A-E**) Time course analysis of changes in steady-state and newly- synthesized RNA for SAGA-dominated genes **(A, B)**, TFIID-dominated genes **(C, D)** and a SAGA- and TFIID-dominated gene **(E)** upon Spt7 nuclear depletion. Expression values (mean ± SD of three independent experiments) were normalized to spiked-in *S. pombe* signal and set to 1 in the untreated sample.

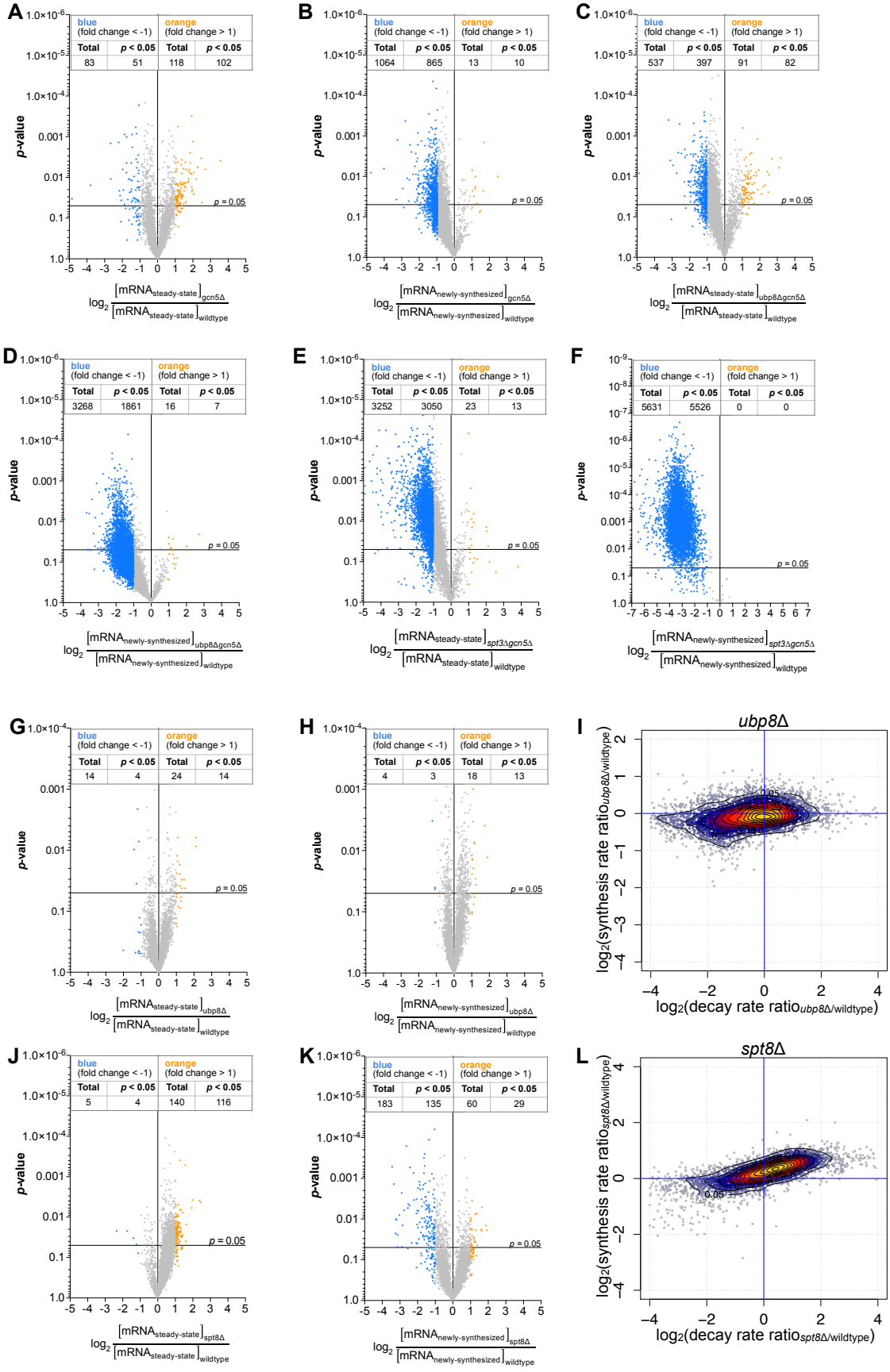

A
blue (fold change < -1) Total 83 p < 0.05 51
orange (fold change > 1) Total 118 p < 0.05 102
p-value
p = 0.05
log2 [mRNAsteady-state]gcn5Δ / [mRNAsteady-state]wildtype
B
blue (fold change < -1) Total 1064 p < 0.05 865
orange (fold change > 1) Total 13 p < 0.05 10
p-value
p = 0.05
log2 [mRNAnewly-synthesized]gcn5Δ / [mRNAnewly-synthesized]wildtype
C
blue (fold change < -1) Total 537 p < 0.05 397
orange (fold change > 1) Total 91 p < 0.05 82
p-value
p = 0.05
log2 [mRNAsteady-state]ubp8Δgcn5Δ / [mRNAsteady-state]wildtype
D
blue (fold change < -1) Total 3268 p < 0.05 1861
orange (fold change > 1) Total 16 p < 0.05 7
p-value
p = 0.05
log2 [mRNAnewly-synthesized]ubp8Δgcn5Δ / [mRNAnewly-synthesized]wildtype
E
blue (fold change < -1) Total 3252 p < 0.05 3050
orange (fold change > 1) Total 23 p < 0.05 13
p-value
p = 0.05
log2 [mRNAsteady-state]spt3Δgcn5Δ / [mRNAsteady-state]wildtype
F
blue (fold change < -1) Total 5631 p < 0.05 5526
orange (fold change > 1) Total 0 p < 0.05 0
p-value
p = 0.05
log2 [mRNAnewly-synthesized]spt3Δgcn5Δ / [mRNAnewly-synthesized]wildtype
G
blue (fold change < -1) Total 14 p < 0.05 4
orange (fold change > 1) Total 24 p < 0.05 14
p-value
p = 0.05
log2 [mRNAsteady-state]ubp8Δ / [mRNAsteady-state]wildtype
H
blue (fold change < -1) Total 4 p < 0.05 3
orange (fold change > 1) Total 18 p < 0.05 13
p-value
p = 0.05
log2 [mRNAnewly-synthesized]ubp8Δ / [mRNAnewly-synthesized]wildtype
I
ubp8Δ
log2(synthesis rate ratioubp8Δ/wildtype)
log2(decay rate ratioubp8Δ/wildtype)
J
blue (fold change < -1) Total 5 p < 0.05 4
orange (fold change > 1) Total 140 p < 0.05 116
p-value
p = 0.05
log2 [mRNAsteady-state]spt8Δ / [mRNAsteady-state]wildtype
K
blue (fold change < -1) Total 183 p < 0.05 135
orange (fold change > 1) Total 60 p < 0.05 29
p-value
p = 0.05
log2 [mRNAnewly-synthesized]spt8Δ / [mRNAnewly-synthesized]wildtype
L
spt8Δ
log2(synthesis rate ratiospt8Δ/wildtype)
log2(decay rate ratiospt8Δ/wildtype)

Figure S5. cDTA analysis for several SAGA mutants, Related to Figure 6. (**A-F, G-H, J-K**) Volcano plots showing changes in steady-state (A, C, E, G and J) and newly-synthesized mRNA levels **(B, D, F, H and I)** between mutant and wild-type *S. cerevisiae* cells relative to their significance (*p*-value). Fold changes (FC) were calculated as the log2 of the ratio of the expression value of each gene after normalization to *S. pombe* signal in the *gcn5*Δ **(A, B)**, *ubp8*Δ*gcn5*Δ **(C, D)**, *spt3*Δ*gcn5*Δ **(E, F)**, *ubp8*Δ **(G, H)** and *spt8*Δ **(J, K)** strains versus the expression value of the same gene in wild-type *S. cerevisiae*. **(I and L)** For all analyzed genes, changes in synthesis rates were plotted against the changes in mRNA decay rates. Changes were calculated as the Log2 of the ratio between *ubp8*Δ **(I)** or *spt8*Δ **(L)** and wild-type. 90% of genes are contained within the outer contour. Yellow and red dots correspond to 60% of genes. For each strain, results were obtained from at least two independent biological replicates.

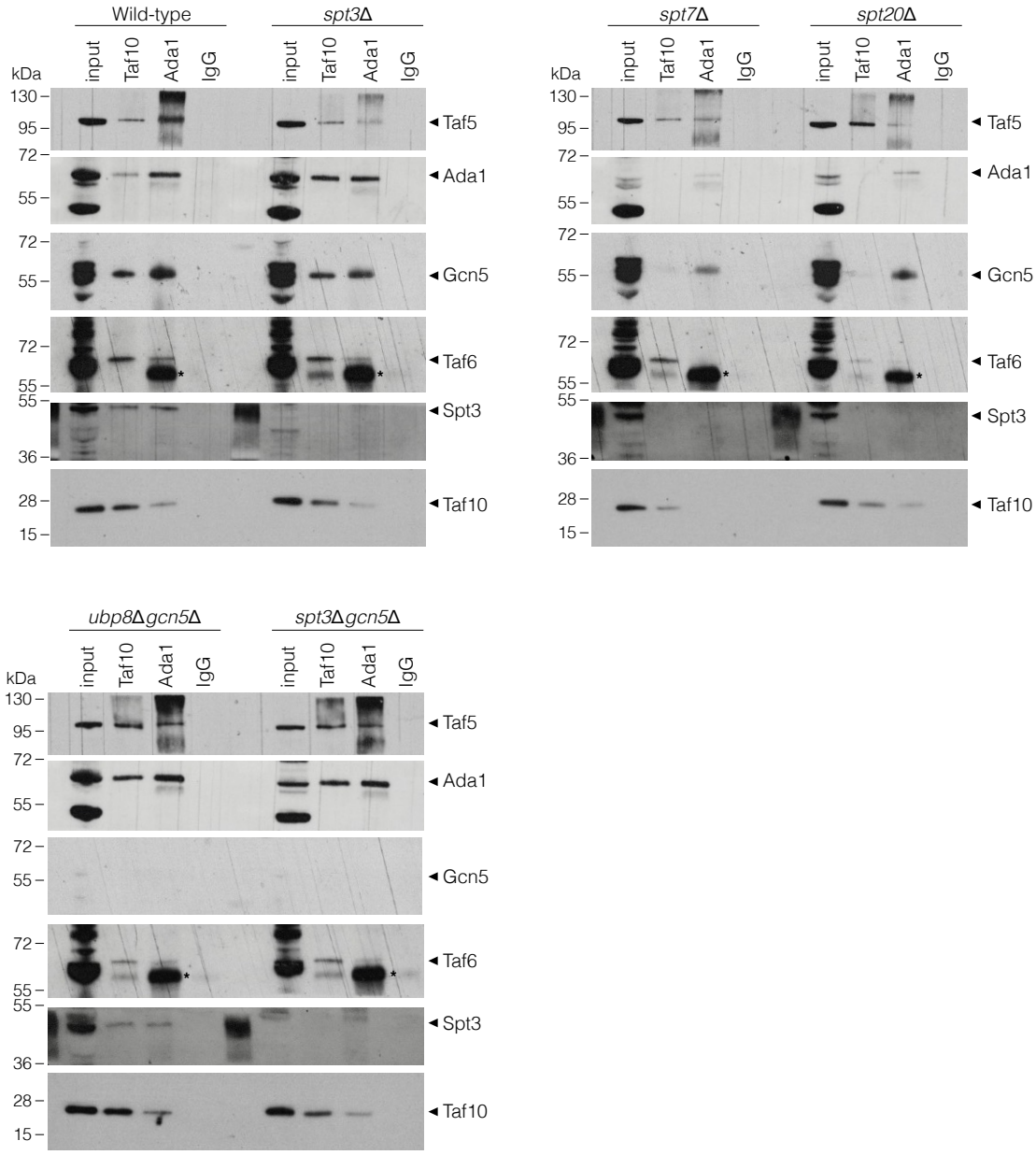
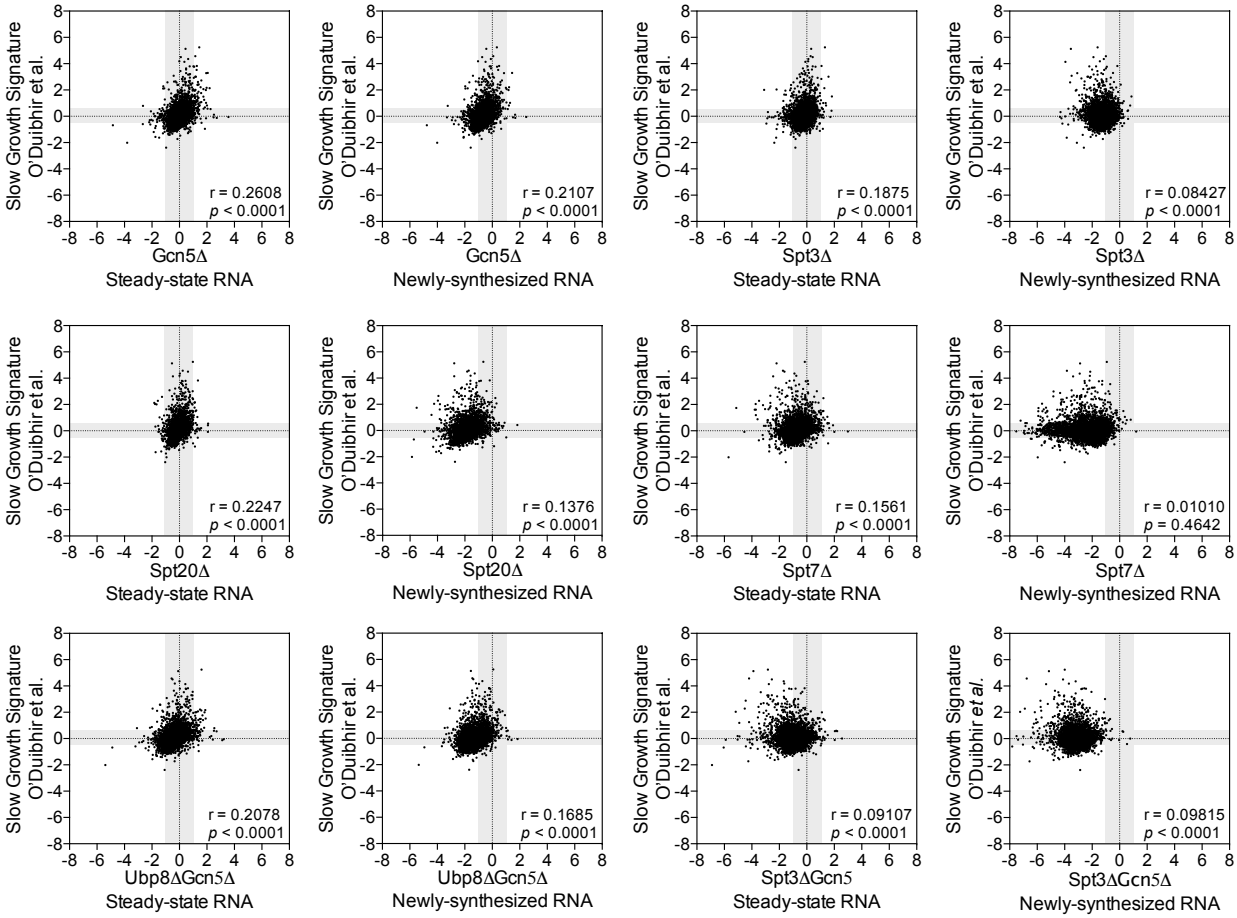


Figure S6. SAGA complex characterization upon deletion of one or more subunits, Related to Figure 6. Upon deletion of *spt3*Δ, *spt7*Δ, *spt20*Δ, *ubp8*Δ*gcn5*Δ and *spt3*Δ*gcn5*Δ SAGA purification was performed by immunoprecipitation of two subunits (Taf10 and Ada1). Eluates were separated by SDS-PAGE and shown are quantitative Western blot analyses of the IP blotted against Taf5, Taf6, Taf10, Ada1, Spt3 and Gcn5.



Steady-state

Wildtype	*gcn5Δ*	*spt3Δ*	*spt7Δ*	*spt20Δ*	*ubp8Δgcn5Δ*	*spt3Δgcn5Δ*
-0.1728	0.2608	0.1875	0.1561	0.2247	0.2078	0.09107

Newly-synthesized

Wildtype	*gcn5Δ*	*spt3Δ*	*spt7Δ*	*spt20Δ*	*ubp8Δgcn5Δ*	*spt3Δgcn5Δ*
-0.1107	0.2107	0.08427	0.0101	0.1376	0.1685	0.09815

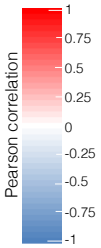


Figure S7. Transcriptional changes observed in SAGA mutants do not correlate with the slow growth phenotype gene expression signature, Related to Figure 6. Transcriptional profiles for SAGA mutants were compared with the slow-growth transcriptional signature obtained elsewhere. The shaded regions on the scatter plots correspond to the threshold applied for each of the studies as a cut-off for either up- or down-regulation. The color code of the tables indicates the degree of correlation between the results obtained in this work and the slow-growth signature (red indicates positive correlation and blue indicates negative correlation).

Table S1. Genotypes of the yeast strains used in this study, Related to STAR Methods.

Name	Genotype	Approach	Source
SGY95	ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 met15Δ0 trp1Δ63 ura3Δ0 SPT3-3FLAGMNase(83-231)-kanMX6	ChEC-seq	This study
SGY96	ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 met15Δ0 trp1Δ63 ura3Δ0 SPT7-3FLAGMNase(83-231)-kanMX6	ChEC-seq	This study
SGY98	ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 met15Δ0 trp1Δ63 ura3Δ0 SPT8-3FLAGMNase(83-231)-kanMX6	ChEC-seq	This study
SGY99	ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 met15Δ0 trp1Δ63 ura3Δ0 UBP8-3FLAGMNase(83-231)-kanMX6	ChEC-seq	This study
FY406	MATa (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63 [pSAB6-(HTA1-HTB1, URA3)]	cDTA	Hirschhorn et.al., 1995
WT yH2B	MATa (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63 [pRS413-(HTA1-Flag-HTB1,CEN, HIS)]	cDTA, mRNA half-life	Bonnet *et al*., 2014
FY406 ubp8Δ- hH2B	MATa ubp8Δ::KANMX4 (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2 128δ his3Δ200 trp1Δ63 [pRS413-(HTA1-Flag-HTB1,CEN, HIS)]	cDTA	Bonnet *et al*., 2014
FY406 gcn5Δ	MATa gcn5Δ::HPH (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63 [pRS413-(HTA1-Flag-HTB1,CEN, HIS)]	cDTA	Bonnet *et al*., 2014
FY406 spt7Δ	MATa spt7Δ::KAN MX4(hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63 [pRS413-(HTA1-Flag-HTB1,CEN, HIS)]	cDTA, mRNA half-life	Bonnet *et al*., 2014
FY406 spt20Δ	MATa spt20Δ::HPH (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63 [pRS413-(HTA1-Flag-HTB1,CEN, HIS)]	cDTA, mRNA half-life	Bonnet *et al*., 2014
FY406 ubp8Δgcn5Δ	MATa ubp8Δ::KANMX6;gcn5Δ::HPH (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63 [pRS413-(HTA1-Flag-HTB1,CEN, HIS)]	cDTA	This study

FY406 spt3Δ	MATa spt3Δ::KANMX6 (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63 [pRS413-(HTA1-Flag-HTB1,CEN, HIS)]	cDTA	This study
FY406 spt8Δ	MATa spt8Δ::KANMX6 (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63 [pRS413-(HTA1-Flag-HTB1,CEN, HIS)]	cDTA	This study
FY406 spt3Δgcn5Δ	MATa spt3Δ::KANMX6;gcn5Δ::HPH (hta1-htb1)Δ::LEU2 (hta2-htb2)Δ::TRP1 leu2Δ1 ura3-52 lys2Δ1 lys2-128δ his3Δ200 trp1Δ63 [pRS413-(HTA1-Flag-HTB1,CEN, HIS)]	cDTA	This study
BY4742 SPT7-FRB	MATα; tor1-1; fpr1del; RPL13A-FKBP12-NAT; MET15; his3-1; leu2; lys2; ura3; SPT7-FRB::Hygro	Ancho-away	This study
BY4742	MATα; his3D1; leu2D0; lys2D0; ura3D0	Parental strain, ChIP	Euroscarf
BY4742 3HA-TBP	BY4742; MATα; ura3Δ0; leu2Δ0; his3Δ1; lys2Δ0; URA3::3HA::SPT15	ChIP	This study
BY4742 spt3Δ	BY4742; MATα; ura3Δ0; leu2Δ0; his3Δ1; lys2Δ0; URA3::3HA::SPT15 spt3::KanMX6	ChIP	This study
BY4742 spt3Δ; 3HA-TBP	BY4742; MATα; ura3Δ0; leu2Δ0; his3Δ1; lys2Δ0; URA3::3HA::SPT15 spt3::KanMX6	ChIP	This study

Table S2. List of plasmids used in this study, Related to STAR Methods.

Plasmid	Description	Source
pGZ108	pFA6a-based vector for C-terminal tagging with 3FLAG-MNase; kanMX6 marker	(Zentner et al., 2015)
pFA6a–hphNT1	Gene deletion cassette: marker pAgTEF-hph-tScCYC1, selectable phenotype: hygromycin resistance.	Janke et al., 2004
pFA6a-kanMX6	Plasmid for yeast gene deletion using the kanMX selectable marker conferring kanamycin resistance.	Bähler et al., 1998
YIplac211-3HA-TBP	Yeast integrative plasmide containing 3HA-TBPΔC for N-terminal tagging of TBP; URA3 marker	Eyboulet et al., 2015

Table S3. List of primers used for RT-qPCR and ChIP-qPCR, Related to STAR Methods.

Gene	Name	Sequence	Approach
PMA1	PMA1_Forward	CTCATCAGCCAACTCAAGAAA	RT-qPCR
	PMA1_Reverse	CGTCATCGTCAGAAGATTCA	
BDF2	BDF2_Forward	CTGAAGAAAATGGAGGTTGAAT	RT-qPCR
	BDF2_Reverse	CTTCCTCTTCCTTTCCTTCG	
PGK1	PGK1_Forward	AGCGTGTCTTCATCAGAG	RT-qPCR
	PGK1_Reverse	TGGCAAAGCAGCAACAA	
PDC1	PDC1_Forward	ATTCACCGACACCGAAG	RT-qPCR
	PDC1_Reverse	TTACGCCGCTGATGGTT	
CDC19	CDC19_Forward	CCAAAGACCAACAACCC	RT-qPCR
	CDC19_Reverse	ATTCGTAAGAACCGTGAGAG	
PHO84	PHO84_Forward	GTGTTGGTTTCTTGACAGATTC	RT-qPCR
	PHO84_Reverse	GCATACTACCGTGCCAG	
NPL3	NPL3_Forward	CACCACCGTCAAGAAGGA	RT-qPCR
	NPL3_Reverse	CAAAGATTTCATTCAACTCGGAT	
GNP1	GNP1_Forward	CGTAATGGGAAACATCGTC	RT-qPCR
	GNP1_Reverse	TGGGCGGAATAATGAGGG	
YEF3	YEF3_Forward	AGAAGTTATCTGTTGCCACTG	RT-qPCR
	YEF3_Reverse	TACCATTCAAGAAAGAAGCGAC	
STE2	STE2_Forward	GACTTACGCTCTCACCG	RT-qPCR
	STE2_Reverse	AGAAGCCACAAGAAGGAC	
RPS3	RPS3_Forward	ATTGTTGAACGGTTTGGC	RT-qPCR
	RPS3_Reverse	CCCTTAGCACCAGATTCCATA	
HYP2	HYP2_Forward	TTGAAACTGCTGACGCT	RT-qPCR
	HYP2_Reverse	TCTTGATGACAACGAAACCG	
RDN58	RDN58_Forward	AACGGATCTCTTGGTTCTCG	RT-qPCR
	RDN58_Reverse	GTGCGTTCAAAGATTCGATG	
RDN25	RDN25_Forward	TGGCAGTCAAGCGTTCATAG	RT-qPCR
	RDN25_Reverse	CGCTTACCGAATTCTGCTTC	
snR6	snR6_Forward	CGAAGTAACCCTTCGTGGAC	RT-qPCR
	SNR6_Reverse	TCATCCTTATGCAGGGGAAC	
scR1	scR1_Forward	CCTTTGGGCAAGGGATAGTT	RT-qPCR
	scR1_Reverse	TTTACGACGGAGGAAAGACG	
S. pombe Tubulin	Sp_Tubulin_F	CCGCTGGTGGAAAGTATGTT	RT-qPCR
	Sp_Tubulin_R	GCCAATTCAGCACCTTCAGT	
PMA1	PMA1_P1_Forward	GATGGTGGGTACCGCTTATG	ChIP-qPCR
	PMA1_P1_Reverse	TTGGTGTTATAGGAAAGAAAGAG	
CDC19	CDC19_P1_Forward	CCTTTCCTTCCCATATGATGC	ChIP-qPCR
	CDC19_P1_Reverse	ACTTTGAAAGGGGACCATGA	
PDC1	PDC1_P1_Forward	CAGCTTATGGTGATGGCACA	ChIP-qPCR
	PDC1_P1_Reverse	ACCCAAATCTGATTGCAAGG	

PGK1	PGK1_P1_Forward	GTTCGTTCGATCGTACTGTT	ChIP-qPCR
	PGK1_P1_Reverse	AAACTAAACCACCCCCTTGG	
ILV5	ILV5_P1_Forward	CACCCAGTATTTTCCCTTTCC	ChIP-qPCR
	ILV5_P1_Reverse	GCGGCTTGAGTTCTCAACAT	
STI1	STI1_P1_Forward	CCAAAAGTCTGCTCCCAAAT	ChIP-qPCR
	STI1_P1_Reverse	TGCAGCGTTACCTTGTTGTT	
RPS3	RPS3_P1_Forward	TCCGTAACATCCATACCTTTCC	ChIP-qPCR
	RPS3_P1_Reverse	TACCACTGCCCATGGGAGAAA	
NPL3	NPL3_P1_Forward	TTTTCTAACGGCCTGTGCTT	ChIP-qPCR
	NPL3_P1_Reverse	GCCACCAATTAGAAGGCTACTC	
EFB1	EFB1_P1_Forward	TCAGCACTGAAGAGTCCAACC	ChIP-qPCR
	EFB1_P1_Reverse	TGACTTGTCAGCCAAAGAAGC	
RPS5	RPS5_P1_Forward	CCAAGAAAAGAGACTAGAAAT	ChIP-qPCR
	RPS5_P1_Reverse	TGGAGTAGCCAAGACGACTG	
YEF3	YEF3_P1_Forward	CTTACGCTCTCTTTCTTTCCT	ChIP-qPCR
	YEF3_P1_Reverse	TTCTAGAACCTTAATGGA	
GAL1	GAL1_P1_Forward	ACATTTCCACACCCTGGAAC	ChIP-qPCR
	GAL1_P1_Reverse	TTCTTCGCGAGAACAATTCA	
HMR	HMR_P1_Forward	ACGATCCCCGTCCAAGTTATG	ChIP-qPCR
	HMR_P1_Reverse	CTTCAAAAGGAGTCTTAATTTCCCTG	
HSP42	HSP42_P1_Forward	GGGAGGCCTCTGTGAAGTTA	ChIP-qPCR
	HSP42_P1_Reverse	GCCTGAACGTGTCCCTATGT	

3. Transcription of nearly all yeast RNA Polymerase II-transcribed genes is dependent on transcription factor TFIID. (L Warfield*, S Ramachandran*, T Baptista *et al.*; Molecular Cell, 2017)

As previously mentioned in previous sections of the present thesis, genes are classified according to their requirements for either SAGA or TFIID as SAGA-dominated and TFIID-dominated genes, respectively. From the data collected in Frank Pugh's laboratory, the class of TFIID-dominated genes was enriched for housekeeping/constitutively expressed genes lacking a TATA consensus sequence on their core promoter. Not only that, but work from the laboratory of Michael Green has shown that some genes do not depend on Tafs for their expression. Specifically, they observed that when thermo-sensitive mutants for several Taf proteins (Taf1, Taf6, Taf9, and Taf12) were shifted to restrictive temperature, the expression of several genes (*RPS5*, *RPS30*, *SSB1*, *ACT1*, and *RPL25*) was totally abrogated, while the expression of others (*ADH1*, *SED1*, *PGK1*, *GAL1*, and *CUP1*) was virtually unaffected. Also, through ChIP-PCR analyses, they have reported that the promoters of those genes were depleted for Tafs (Li *et al.*, 2000). In addition, the group of Kevin Struhl has also addressed this question regarding TFIID dependency of some genes. In general lines, they hypothesized that if TFIID is present at both TAF-dependent and TAF-independent promoters, the Taf/TBP occupancy ratio should be constant at all promoters. Hence, they analyzed the different Taf to TBP ratios in a given number of genes by ChIP-PCR. Importantly, they reported that some genes presented a low Taf/TBP ratio (Taf-depleted genes – *ADH1*, *PGK1*, and *PYK1*), whereas others displayed a higher Taf/TBP ratio (Taf-enriched genes – *RPS8A*, *RPL9A*, *RPL5*, *TRP3*, *ACT1*, and *ETF2*) (Kuras *et al*., 2000). Nevertheless, a recent publication addressed the genome-wide recruitment of Taf1/TFIID using ChEC-seq, reporting that, instead, TFIID is physically recruited to both SAGA- and TFIID-dominated genes (Grunberg *et al*., 2016).

Hence, in line with **aim (e)**, we have re-examined the TFIID dependence of yeast genes and found that, for cells grown in rich media, nearly all RNA Pol II-transcribed genes are strongly dependent on TFIID. The magnitude of TFIID dependence can change in different growth conditions. However, even under different growth conditions, genes in the Taf1-depleted, Taf1-enriched, TATA-containing and TATA-less categories respond similarly to depletion of TFIID-specific subunits such as Taf1, Taf2, Taf7, Taf11, and Taf13.

Altogether, our findings have important implications for the mechanism of gene regulation and the nature of the PIC at both housekeeping and highly regulated genes (Warfield *et al*., 2017).

These results were submitted and accepted for publication in *Molecular Cell* (online publishing: 14th of September 2017; *in print* date: 5th of October 2017). Below are detailed the contributions of every author.

Author's contributions

Transcription of nearly all yeast RNA Polymerase II-transcribed genes is dependent on transcription factor TFIID. L Warfield*, S Ramachandran*, T Baptista, D Devys, L Tora and S Hahn.

*These authors contributed equally

This work as a collaborative effort between Steven Hahn's and our groups. Specifically, the laboratory of Steven Hahn produced all the data obtained through the use of the auxin-induced degradation strains. In our side, I was responsible for the execution of all the experiments that resourced to the use of the anchor-away strains. Additionally, I was in charge of analyzing and interpreting the data acquired in our laboratory. Genome-wide analysis of the transcriptome of the mutant strains was made using the protocol that I established (and optimized) in the laboratory. Having that, I was fully responsible for the execution of the experiments presented in the manuscript (Figure 6 and Supplementary Figures 3 and 4). Additionally, I was involved in the design and conceiving of this work, together with all the other authors of the manuscript. In the end, and once all the data was collected, the manuscript was written with the input of all the authors. Specifically:

Linda Warfield – Conceived and designed experiments. Performed all auxin-degron experiments, 4-tU labeling and RT qPCR analysis of degron-containing cells and TFIID IPs.

Srinivas Ramachandran – Conceived and designed experiments. Performed the computational analysis of all auxin-degron experiments.

Tiago Baptista – Conceived and designed experiments. Performed experiments using anchor-away strains (4-tU labeling, genome wide transcriptional analyses, computational analyses and cell fractionation experiments).

Didier Devys – Conceived and designed experiments.

László Tora – Conceived and designed experiments.

Steven Hahn – Conceived and designed experiments.

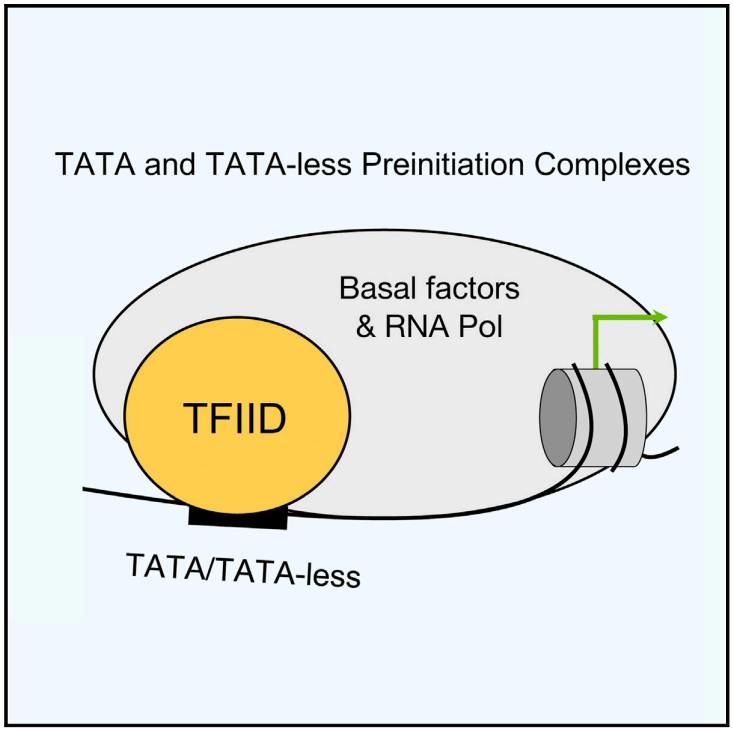
All authors analyzed the results and wrote the manuscript.

Article

Molecular Cell

Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID

Graphical Abstract



Highlights

- Transcription of nearly all yeast mRNAs is strongly TFIID dependent
- Both TATA and TATA-less promoters show similar TFIID dependence
- The magnitude of TFIID dependence can vary with growth conditions
- SAGA and TFIID are not alternative factors but work together at most mRNA genes

Authors

Linda Warfield,
Srinivas Ramachandran,
Tiago Baptista, Didier Devys,
Laszlo Tora, Steven Hahn

Correspondence

shahn@fredhutch.org

In Brief

Early work suggested that expression of yeast mRNA genes is primarily dependent on either transcription factor TFIID or SAGA. Warfield et al. now find that transcription of nearly all these genes is strongly TFIID dependent. This finding has important implications for the mechanism and regulation of transcription at all mRNA genes.

Warfield et al., 2017, Molecular Cell *68*, 1–12
October 5, 2017
http://dx.doi.org/10.1016/j.molcel.2017.08.014

CellPress

Please cite this article in press as: Warfield et al., Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID, Molecular Cell (2017), http://dx.doi.org/10.1016/j.molcel.2017.08.014

Molecular Cell
Article

CellPress

Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID

Linda Warfield,[1,7] Srinivas Ramachandran,[1,2,7] Tiago Baptista,[3,4,5,6] Didier Devys,[3,4,5,6] Laszlo Tora,[3,4,5,6] and Steven Hahn[1,8,*]

[1]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[2]Howard Hughes Medical Institute, Seattle, WA 98109, USA
[3]Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France
[4]UMR7104, Centre National de la Recherche Scientifique, 67404 Illkirch, France
[5]U964, Institut National de la Santé et de la Recherche Médicale, 67404 Illkirch, France
[6]Université de Strasbourg, 67404 Illkirch, Cedex, France
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: shahn@fredhutch.org
http://dx.doi.org/10.1016/j.molcel.2017.08.014

SUMMARY

Previous studies suggested that expression of most yeast mRNAs is dominated by either transcription factor TFIID or SAGA. We re-examined the role of TFIID by rapid depletion of *S. cerevisiae* TFIID subunits and measurement of changes in nascent transcription. We find that transcription of nearly all mRNAs is strongly dependent on TFIID function. Degron-dependent depletion of Taf1, Taf2, Taf7, Taf11, and Taf13 showed similar transcription decreases for genes in the Taf1-depleted, Taf1-enriched, TATA-containing, and TATA-less gene classes. The magnitude of TFIID dependence varies with growth conditions, although this variation is similar genome-wide. Many studies have suggested differences in gene-regulatory mechanisms between TATA and TATA-less genes, and these differences have been attributed in part to differential dependence on SAGA or TFIID. Our work indicates that TFIID participates in expression of nearly all yeast mRNAs and that differences in regulation between these two gene categories is due to other properties.

INTRODUCTION

TFIID and SAGA are two widely used RNA polymerase II (Pol II) transcription factors that together regulate transcription of nearly all Pol II-transcribed genes (Hahn and Young, 2011). Yeast (y) TFIID is comprised of TATA binding protein (TBP) and 14 TBP-associated factors (Tafs) (Berger et al., 2011; Sainsbury et al., 2015). Although transcription at TATA box-containing promoters can be reconstituted with only TBP and other basal transcription factors, the TFIID Tafs are thought to be critical for promoter recognition at the ~80% of eukaryotic promoters lacking a TATA box. TFIID is structurally organized into three lobes with lobe C (Taf1, Taf2, and Taf7) binding several promoter DNA elements found in higher eukaryotes (e.g., Inr, DPE, MTE) (Cianfrocco and Nogales, 2013; Louder et al., 2016; Theisen et al., 2010; Verrijzer et al., 1995). TFIID also binds transcription activators, acetylated nucleosomes, the universally required coactivator Mediator, and probably other components of the basal transcription machinery (Müller and Tora, 2014). These functions allow TFIID to respond to regulatory inputs from both UAS elements and the chromatin state and to act as a platform for assembly of the transcription preinitiation complex (PIC).

SAGA contains 19 subunits and is structurally organized into five modules that function in histone acetylation and deubiquitylation, and in the binding of TBP and transcription activators (Hahn and Young, 2011). Five of the ySAGA subunits are shared with yTFIID, and seven of the SAGA subunits, including four common Tafs, contain histone fold domains that likely form a structural module at the center of SAGA (Bieniossek et al., 2013; Grant et al., 1998; Han et al., 2014; Setiaputra et al., 2015). Although SAGA and TFIID share a set of subunits and both bind activators and TBP, SAGA has no known DNA binding activity or interactions with other components of the basal transcription machinery.

Earlier studies using the yeast *S. cerevisiae* have classified the TFIID and SAGA dependence of genes based on genome-wide mapping of TFIID binding and transcriptional changes observed in strains with mutations in TFIID and SAGA subunits. First, formaldehyde-based ChIP and ChIP-exo methods have mapped the relative ratio of the TFIID subunit Taf1 versus TBP at yeast promoter regions, identifying Taf-enriched versus Taf-depleted genes (Kuras et al., 2000; Li et al., 2000; Rhee and Pugh, 2012). Second, steady-state mRNA levels were measured before and after heat-shock inactivation of temperature-sensitive Taf proteins (Huisinga and Pugh, 2004; Lee et al., 2000; Li et al., 2000; Shen et al., 2003). In parallel, steady-state mRNA levels were measured in strains with mutations in SAGA subunits and in strains with both TFIID and SAGA mutations (Huisinga and Pugh, 2004; Lee et al., 2000; Shen et al., 2003). Combined, these

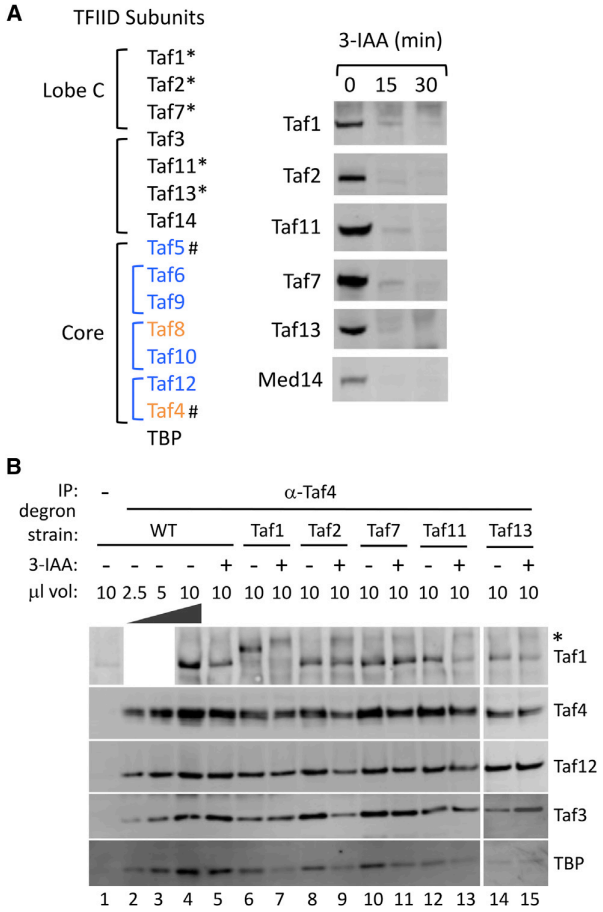


Figure 1. Auxin-Dependent Depletion of Yeast TFIID Subunits

(A) Listed at left are yeast TFIID subunits in lobe C, the shared or similar TFIID/SAGA core subunits, and other auxiliary subunits. Asterisk indicates subunits tagged with the 3xV5-IAA7 degron tag, and hashtag indicates subunits tagged with FRB for anchor-away. Blue indicates subunits shared with SAGA, and orange indicates that SAGA contains a related but SAGA-specific subunit. Shown at right is western analysis of cells treated with 3-IAA for the indicated time. Western blot was probed with anti V5 antibody.

(B) Analysis of TFIID subunits after Taf depletion. After 30 min treatment with 3-IAA, whole-cell extracts were made and immune precipitated with anti-Taf4 antibody. Shown is quantitative western analysis of the IP, probing for Taf1, Taf3, Taf4, Taf12, and TBP. Asterisk indicates a non-specific signal observed with the anti Taf1 antibody.

studies suggested that TFIID and SAGA likely contribute to the expression of nearly all genes but that expression of most genes is dominated by either TFIID or SAGA.

Many studies suggest that there are at least two broad classes of genes: housekeeping genes and highly regulated genes (de Jonge et al., 2017; Eisenberg and Levanon, 2013; Huisinga and Pugh, 2004; Mencía et al., 2002; Ohtsuki et al., 1998; Zabidi and Stark, 2016). These two gene classes differ in their response to transcription activators, chromatin organization, and modifications and in promoter sequence elements. In yeast, promoters lacking a consensus TATA box are typically in the housekeeping class, have the +1 nucleosome adjacent to the site of PIC formation, are enriched in H4 acetylation, and are nearly always classified as Taf1 enriched and TFIID dominated (Basehoar et al., 2004; Huisinga and Pugh, 2004; Rhee and Pugh, 2012). On the other hand, TATA-containing promoters are often highly regulated by stress, are typically modulated by chromatin remodeling and/or covalent histone modifications, and are often classified as Taf1 depleted and SAGA dominated.

Recent results, however, have raised questions about the TFIID and SAGA specificity at these two classes of genes. First, a formaldehyde-independent method for mapping genome-wide binding termed ChEC-seq mapped apparent equal occupancy of TFIID at promoters of both gene classes (Grünberg et al., 2016). Second, mapping SAGA-dependent histone modifications showed that SAGA associates with all transcribed genes (Bonnet et al., 2014). Third, inactivation of SAGA was found to decrease Pol II recruitment and the rate of transcription from several genes previously characterized as TFIID dominated (Bonnet et al., 2014). Finally, it was demonstrated that yeast mRNA levels are largely buffered to major transcriptional perturbations (Haimovich et al., 2013; Munchel et al., 2011; Sun et al., 2013, 2012). For example, inhibition of Pol II activity by several approaches led to a corresponding increase in mRNA stability for most genes. Therefore, using steady-state mRNA levels to determine the role of widely used transcription factors is problematic. Based on these findings, we have re-examined the TFIID dependence of yeast genes and find that, for cells grown in rich media, nearly all Pol II-transcribed genes are strongly dependent on TFIID. The magnitude of TFIID dependence can change in different growth conditions. However, even under different growth conditions, genes in the Taf1-depleted, Taf1-enriched, TATA-containing and TATA-less categories respond similarly to depletion of TFIID-specific subunits such as Taf1, Taf2, Taf7, Taf11, and Taf13. Our findings have important implications for the mechanism of gene regulation and the nature of the PIC at both housekeeping and highly regulated genes.

RESULTS

A Dramatic Decrease in Genome-wide Pol II Transcription upon TFIID Inactivation

Since all TFIID subunits are essential for viability, previous studies in yeast utilized heat shock of temperature-sensitive mutants to inactivate TFIID, followed by steady-state mRNA analysis (Huisinga and Pugh, 2004; Shen et al., 2003). As an alternative approach to TFIID inactivation, we used the auxin-inducible degron system (Nishimura et al., 2009) to rapidly deplete selected TFIID subunits. First, a yeast strain was constructed containing both a triple Flag epitope tag at the C terminus of the Pol II subunit Rpb3 and constitutively expressing the plant-specific F box protein *Os*TIR1. Next, the degron, auxin repressor protein IAA7, was integrated at the C terminus of the TFIID-specific subunits Taf1, Taf2, Taf7, Taf11, and Taf13 (Cianfrocco et al., 2013; Hahn and Young, 2011; Leurent et al., 2004; 2002; Louder et al., 2016) (Figure 1A, left panel). Taf1, Taf11, and Taf13 have been implicated in TFIID-TBP binding (Anandapadamanaban et al., 2013; Lavigne et al., 1999; Shen et al., 2003). IAA7 was also fused to Med14, the Mediator

subunit that connects the head, middle, and tail modules (Plaschka et al., 2016; Tsai et al., 2014), with the reasoning that depletion of this subunit will inactivate Mediator. The control strain, designated as wild-type, was identical except for lacking an IAA7 degron tag. Western blot analyses showed that all of these subunits fused to IAA7 were efficiently and rapidly degraded 30 min after addition of the auxin 3-IAA (Figure 1A, right panel). After auxin addition for 30 min, all strains were ≥88% viable, although prolonged exposure was lethal. Without 3-IAA, the Taf1, Taf2, Taf7, and Taf13 degron-tagged strains grew normally at 25°C, 30°C, and 37°C, while the Taf11-degron strain grew normally at 25°C and 30°C but slowly at 37°C. Analysis of Pol II transcription in the absence of 3-IAA at 30°C (see below) showed that wild-type and all degron strains had highly similar transcription profiles across the 4,808 genes used for genome-wide analysis (Pearson correlation coefficient ≥0.88; Figure S1A).

The integrity of TFIID after Taf depletion was monitored by immune precipitation of TFIID core subunit Taf4 followed by Western analysis of co-precipitated proteins (Figure 1B). Our results show that, upon degradation of Taf1, Taf2, Taf7, Taf11, or Taf13, the tested TFIID subunits (Taf1, Taf12, Taf3, and TBP) still co-precipitated with Taf4, showing no obvious disruption in subunit association of the remaining TFIID complex. We note that ~2-fold lower levels of TBP co-precipitated upon depletion of Taf1 or Taf11, which is expected, since both subunits have been implicated in TBP binding (Anandapadamanaban et al., 2013; Lavigne et al., 1999; Shen et al., 2003).

To monitor genome-wide nascent transcription, we used native Pol II ChIP to analyze the position of Pol II-DNA complexes before and after 3-IAA addition. Exponentially growing cells in rich media were treated with either 500 μM 3-IAA or DMSO for 30 min. Cells were collected by filtration and quick frozen. For a spike-in control to allow comparison of different samples, frozen cells were mixed with 10% by weight of frozen *S. pombe* cells, also containing a Rpb3-triple Flag tag. Cells were broken under conditions to preserve both chromatin and Pol II elongation complexes, treated with Micrococcal nuclease (MNase), and Pol II complexes isolated by IP. Short DNA fragments associated with Pol II were isolated and sequenced. Assays were performed in biological duplicate. The duplicates were highly similar; hence we performed all further analyses using data combined from the duplicates.

Next we analyzed metagene profiles of Pol II elongation complexes aligned relative to the transcription start site (TSS) and overlaid on histone H3 profiles generated from previously published histone H3 ChIP-seq data (Thurtle and Rine, 2014) (Figure 2A). As previously observed under these growth conditions (Churchman and Weissman, 2011), the average level of Pol II-DNA complexes is highest near the TSS and gradually declines toward the 3′ end of genes (Figure 2A, WT). Addition of either 3-IAA or DMSO to cells lacking a degron gave nearly identical profiles. In contrast, depletion of TFIID subunits TAF7 and TAF11 showed dramatic decreases in genome-wide transcription (Figure 2A, Taf7 and Taf11). Similarly, depletion of the Mediator subunit Med14 resulted in a dramatic decrease in genome-wide Pol II transcription, confirming the essential role of Mediator in transcription (Figure 2A, Med14).

Similar TFIID Dependence of Taf-Enriched and Taf-Depleted Genes

To quantitate transcriptional changes at individual genes, the levels of Pol II ChIP (25–140 bp DNA fragments) was averaged over the interval from 1 to 100 bp downstream of the TSS. This analysis showed that transcription from nearly all Pol II-transcribed genes is dependent on TFIID (Figure 2B; Table S3). The median fold changes in Pol II ChIP levels upon 3-IAA addition were −7.8 (Med14), −4.1 (Taf13), −3.7 (Taf11), −3.4 (Taf7), −3.3 (Taf1), and −2.3 (Taf2). We chose 1–100 bp from the TSS since it captured the highest peak of RNAPII downstream of TSS. We also performed the same analysis on tiling windows of 100 bp in the first 1,000 bp downstream of the TSS (or transcription end site [TES], whichever was shorter) and found the fold changes in Pol II upon 3-IAA addition to be similar throughout the gene body (data not shown).

While nearly all genes show decreased transcription upon TFIID inactivation, Figure 2B shows that there are a small number of genes with little or no apparent expression changes upon depletion of individual Tafs. One possibility is that these non-responsive genes are enriched in genes previously characterized as Taf depleted and/or SAGA dominated (Huisinga and Pugh, 2004; Rhee and Pugh, 2012). To test this hypothesis, genes were grouped by the previously defined Taf1-enriched and Taf1-depleted categories and analyzed as above (Figure 3A). While the Taf1-depleted gene class shows slightly more sensitivity to Med14 depletion compared to the Taf1-enriched class, genes in both categories show very similar responses to depletion of Taf13, Taf11, and Taf7. This behavior was also observed for Taf1 and Taf2 (Figure S1B). Previous analyses showed that genes lacking a consensus TATA correlated with Taf1-enriched genes (Basehoar et al., 2004; Rhee and Pugh, 2012). However, in agreement with the above analysis, we found that both TATA-containing and TATA-less genes showed near-identical responses to Taf depletion (Figure 3B). We conclude that, for cells grown in rich media, transcription of nearly all Pol II-transcribed yeast genes is dependent on TFIID.

Transcription of Several Taf1-Independent Genes Is Dependent on Other TFIID Subunits

Early work on yeast TFIID function identified several individual genes with low ChIP ratios of Taf1/TBP and for which steady-state mRNA levels were insensitive to Taf1 inactivation by heat shock (Kuras et al., 2000; Li et al., 2000). We analyzed our genome-wide data and found that several of these genes (*PGK1*, *CDC19*/*PYK1*, *ADH1*) were in fact insensitive to Taf1 depletion (Figure 4). However, all of these genes showed greater than a 2-fold decrease in Pol II ChIP upon degradation of Taf13, showing that most transcription from these genes is TFIID dependent.

To further verify the role of TFIID in transcription, we next examined the effect of Taf1, Taf7, and Med14 depletion on nascent transcription (Bonnet et al., 2014; Duffy et al., 2015; Sun et al., 2013). Thirty minutes after 3-IAA or DMSO addition, RNAs were metabolically labeled for 6 min with 4-thio Uracil and nascent RNAs purified and quantitated by RT-qPCR. Selected genes in each of three previously defined classes were analyzed: Taf1 enriched, TFIID independent, and Taf1

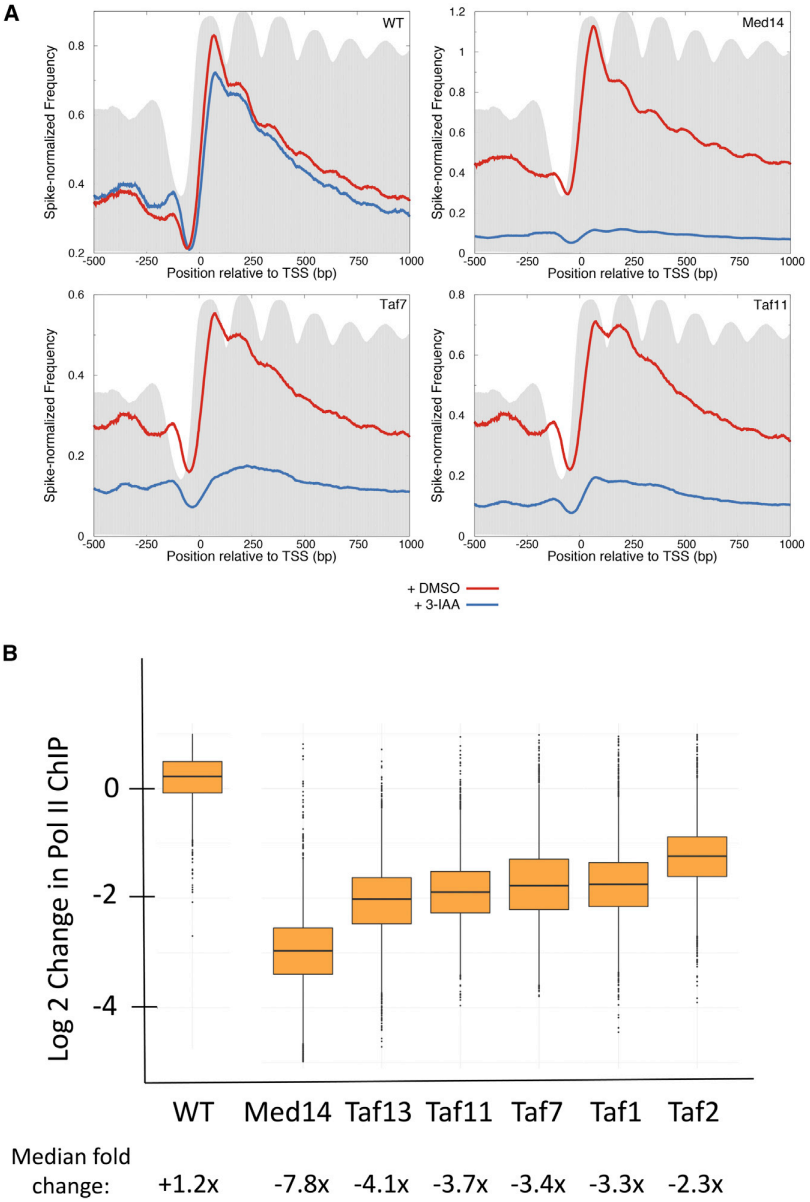


Figure 2. Genome-wide Decreases in Transcription upon Depletion of TFIID and Mediator Subunits

(A) Metagene profile of native Pol II ChIP in wild-type and cells containing the indicated IAA7 degron fusion (4,807 genes). Results are aligned at the major TSS (Park et al., 2014). Blue line, cells treated with 3-IAA for 30 min prior to harvest; red line, cells treated equivalently but with DMSO. Data are averages of biological duplicates with samples normalized by spike-in with *S. pombe* Rpb3-Flag cells. Gray shows metagene profile of histone H3 plotted using published H3 ChIP-seq data (Thurtle and Rine, 2014). The degron fused strains are labeled in the upper right corner of the different panels.

(B) Boxplots show the change in native Pol II ChIP on a log2 scale from 4,807 genes comparing DMSO and 3-IAA addition. Median fold changes in ChIP signals are indicated below. See also Figure S1 and Table S3.

depleted. In agreement with the Pol II ChIP analysis, expression of all genes tested except *PGK1*, *ADH1*, and *CDC19* were highly sensitive to Taf1 and Taf7 depletion, while transcription of all tested genes was dependent on Med14 (Figure S2). From these results we conclude that, while nearly all genes are sensitive to TFIID inactivation, there is a small set of genes that show little change upon inactivation of subunits in TFIID lobe C.

Comparison of Genome-wide Taf Dependence and TFIID Binding

To compare global changes in transcription between the degron strains, we converted the change in Pol II ChIP signals to Z scores. This ensured that the distribution of changes in Pol II ChIP upon addition of 3-IAA was similar across all five Taf degron strains. K-means clustering of the Z scores was used to separate genes into five clusters (Figures 5A and 5B; Table S4). Cluster 1 (with 342 genes) has the lowest decrease in Pol II ChIP after depletion of Taf1, Taf2, Taf7, Taf11, and Taf13, while cluster 5 (with 683 genes) has the largest decrease in Pol II ChIP after 3-IAA addition. Clusters 2, 3, and 4, which make up ~85% of the genes, have a mean Pol II ChIP change within 1 Z score in all five experiments, implying that majority of genes have similar changes in all five degron strains.

To understand the magnitude of changes in transcription in the different clusters, we plotted a heatmap of the changes in Pol II ChIP signals after Taf degradation (Figure 5C) and the mean of these signals for each cluster (Figure 5D). We observed that even for cluster 1, which features the lowest decrease in transcription upon Taf depletion, the mean Pol II ChIP score decreased for all five degron strains. In cluster 2, which featured slightly different Z scores for Taf7 and Taf11 compared with Taf1, Taf2, and Taf13, the log2 fold change in Pol II ChIP was between −1.5 and −2.3 (Figure 5D), indicating that the Z scores were highlighting differences in experiments that translated to minor differences in the Pol II ChIP scores. We also checked for enrichment of Gene Ontology (GO) terms in the different clusters. We found many metabolic processes to be enriched in cluster 1, and RNA processes including tRNA, rRNA, and ncRNA metabolism enriched in clusters 3 and 5. All GO terms with false discovery rate <5% are listed in Table S5.

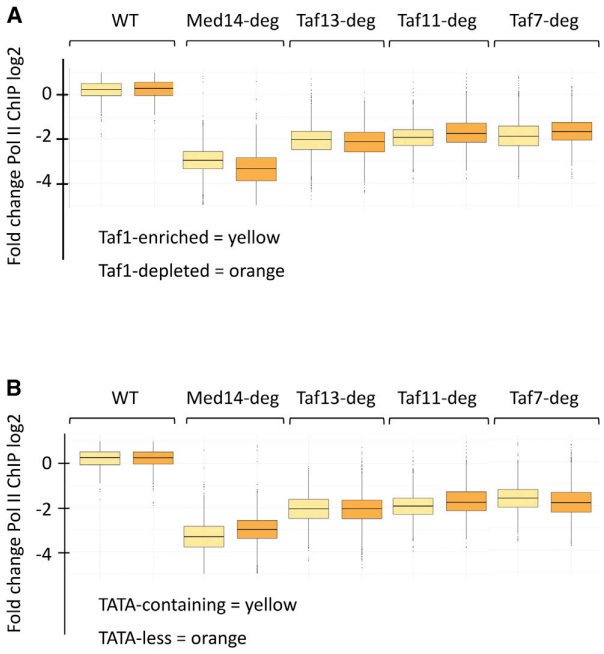


Figure 3. TFIID Functions at Genes Previously Thought to Be Less Dependent on TFIID

(A) Boxplots of data from Figure 3 split into "Taf1-enriched" and "Taf1-depleted" categories (Rhee and Pugh, 2012).

(B) Boxplots of data from Figure 3 split into TATA-containing and TATA-less promoter categories (Rhee and Pugh, 2012).

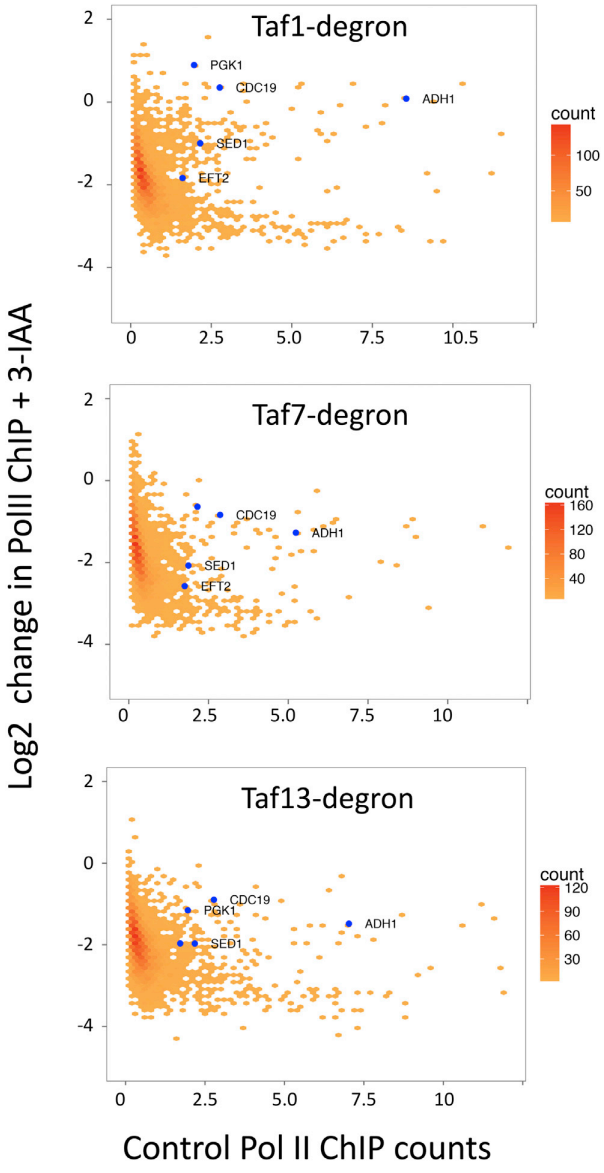


Figure 4. Several Previously Characterized Taf1-Independent Genes Are Dependent on Other TFIID Subunits

Plotted is a hexagonally binned 2D-histogram of Log_2 fold change in Pol II ChIP upon 3-IAA addition versus spike-normalized Pol II ChIP signal (a measure of transcription) averaged over the 1–100 bp from the TSS of each gene (4,807 genes). The blue points are genes used in pre-genome-wide studies to examine TFIID dependence (Kuras et al., 2000; Li et al., 2000). See also Figures S1 and S2 and Table S3.

We next compared the levels of TFIID bound at Pol II promoters to the TFIID dependence clusters. ChEC-seq was recently used to map genome-wide TFIID binding, and we had found that TFIID was localized to nearly all yeast promoters (Grünberg et al., 2016). We next asked whether changes in Pol II ChIP signals observed in clusters 1–5 corresponded to sets of genes with different levels of TFIID. We analyzed the average distribution of Taf1 ChEC-seq signal across the clusters (Figure 5E) and found cluster 1 to be the only outlier in Taf1 distribution, with Taf1 promoter binding being lower for cluster 1 compared to other clusters. It is important to note, however, that genes in cluster 1 have significant levels of bound TFIID (Figure 5E). Cluster 1 features the weakest decrease in Pol II transcription upon degradation of the Tafs, implying that the genes in this cluster depend less on TFIID for transcription compared to the other clusters. We then asked how these clusters compare to previously identified Taf-enriched and Taf-depleted gene lists. We found cluster 1 to have a 1.8-fold enrichment of Taf-depleted genes compared to all genes, indicating some correspondence with the published Taf depleted list. However, only 111 genes out of 870 total Taf-depleted genes were part of cluster 1. Thus, we conclude that, upon Taf depletion, genes in cluster 1 have smaller decreases in Pol II transcription compared to the global average, and this group is enriched for genes that were defined as Taf-depleted in previous studies. Nevertheless, these genes are still TFIID dependent.

Consistent with these findings, the genes with little or mild Taf1 and Taf7 dependence from Figures 4 and S2 (*PGK1*, *CDC19*, and *ADH1*) are found in cluster 1. The heatmap in Figure 5C reveals that genes in cluster 1 seem less sensitive to depletion of Taf1, Taf2, and Taf7, suggesting that genes in cluster 1 are less dependent on TFIID subunits in lobe C. Further analysis showed that

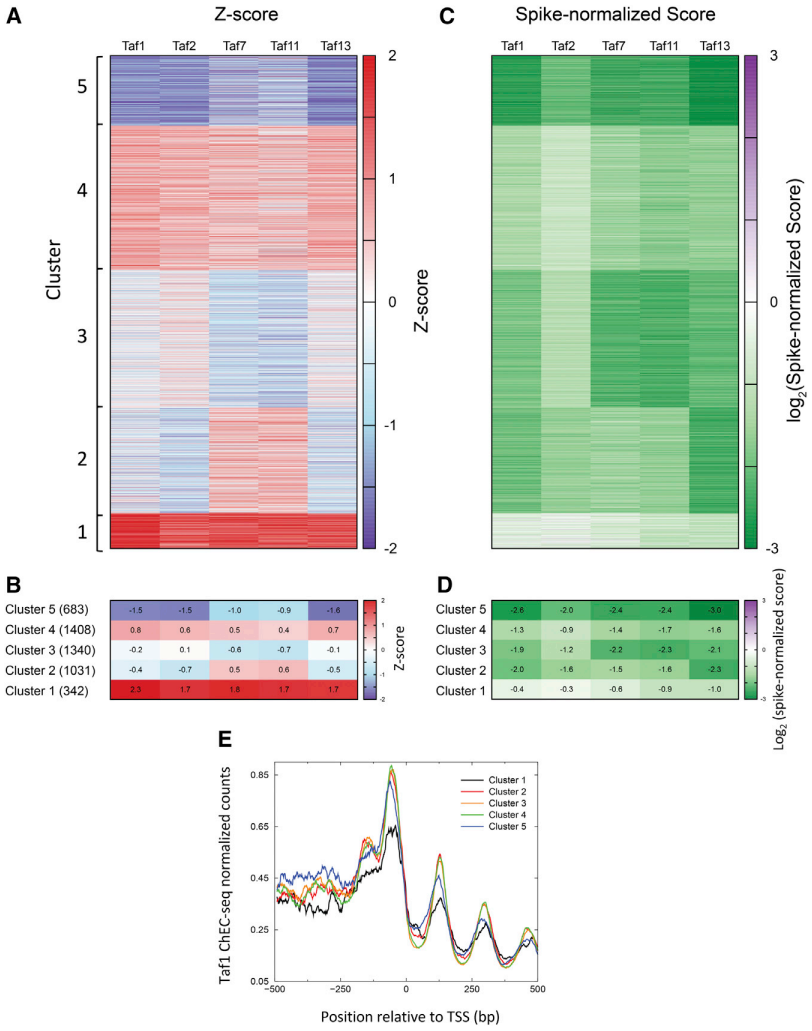


Figure 5. Analysis of Genome-wide Dependence on TFIID Subunits

(A) Heatmap of genes ordered by clusters that were obtained after k-means clustering of the Z scores from Taf-degron experiments.

(B) The mean Z-scores of the five clusters for the five different Taf subunit degrons plotted as a heatmap. The mean values are indicated in each heatmap cell.

(C) Same ordering of genes as (A), but the spike-normalized change in Pol II at the TSS is plotted.

(D) The mean Spike-normalized change of the five clusters for the five different Taf subunit degrons plotted as a heatmap.

(E) Average Taf1 ChEC-seq profiles (Grünberg et al., 2016) plotted for the five clusters shown in (A). See also Table S3, Table S4, and Table S5.

genes with previously characterized strong TFIID dependence (*RPL25* and *RPS5*) are in cluster 5, while most genes previously characterized as Taf1 depleted (e.g., *SSB1*, *VTC1*, *SSH1*, *SED1*, and *ETF2*) are clearly TFIID dependent and found in clusters 2, 3, and 4.

TFIID Subunit Taf4 and TFIID/SAGA Subunit Taf5 Are Required for Genome-wide Pol II Transcription

In an orthogonal approach to test TFIID dependence of Pol II transcription, the TFIID-specific subunit Taf4 and shared TFIID/SAGA subunit Taf5 were depleted using the anchor-away method in which proteins are rapidly exported from the nucleus to the cytoplasm upon rapamycin addition. Taf4 and Taf5 were fused to FRB in a strain expressing RPL13A-FKBP12 to induce anchoring of Taf4-FRB or Taf5-FRB at ribosomes upon rapamycin treatment (Haruki et al., 2008). Both strains grew normally in the absence of rapamycin (Figure S3A), and both steady-state and newly synthesized mRNA levels are similar to wild-type at 16 assayed genes (Figures S3C and S3D). As observed for the Taf-degron strains, nuclear depletion of Taf4 or Taf5 for 30 min was not lethal (Figure S3B), but prolonged exposure was lethal. Fractionation of cells after rapamycin treatment showed efficient change in localization of the tagged subunit from the nucleus to the cytoplasm (Figure S3E). Under these conditions, the other TFIID subunits were stable, but a portion mislocalized to the cytoplasm. For example, upon Taf5 anchor-away, Taf11 was mostly cytoplasmic, while Taf4 stayed mostly nuclear and upon Taf4 anchor-away Taf5 and Taf11 were mostly, but not entirely, nuclear.

Preliminary time course experiments showed that strong transcriptional effects without any detectable growth delay were already observable following 15 min of rapamycin exposure. Thus, anchor-away strains growing in log phase were exposed to rapamycin for 30 min and pulse labeled with 4-thiouracil. Labeled cells were mixed in a fixed ratio with labeled *S. pombe* cells for spike-in normalization. Purified labeled (newly synthesized) and total (steady-state) RNA were hybridized to Affymetrix microarrays containing probes for both *S. cerevisiae* and *S. pombe* transcriptome.

For all yeast genes with a positive signal on microarrays, we measured the ratio of mRNA levels between the *TAF4-FRB* strain treated and untreated with rapamycin. Newly synthesized mRNA levels of a large number of genes (3,423 out of 5,657) were found to be decreased by at least 2-fold, whereas only 38 genes were upregulated by 2-fold (Figure 6A). We used comparative dynamic transcriptome analysis (cDTA) (Sun et al., 2013) to measure changes in both synthesis and decay rates upon nuclear depletion of Taf4 for this gene set. This analysis revealed that, for the vast majority of genes, a decrease in synthesis rate was compensated by a decrease in decay rates (Figure 6B). However, this compensation was sub-optimal, with a median decrease in synthesis rate by 2.7-fold and a median decrease

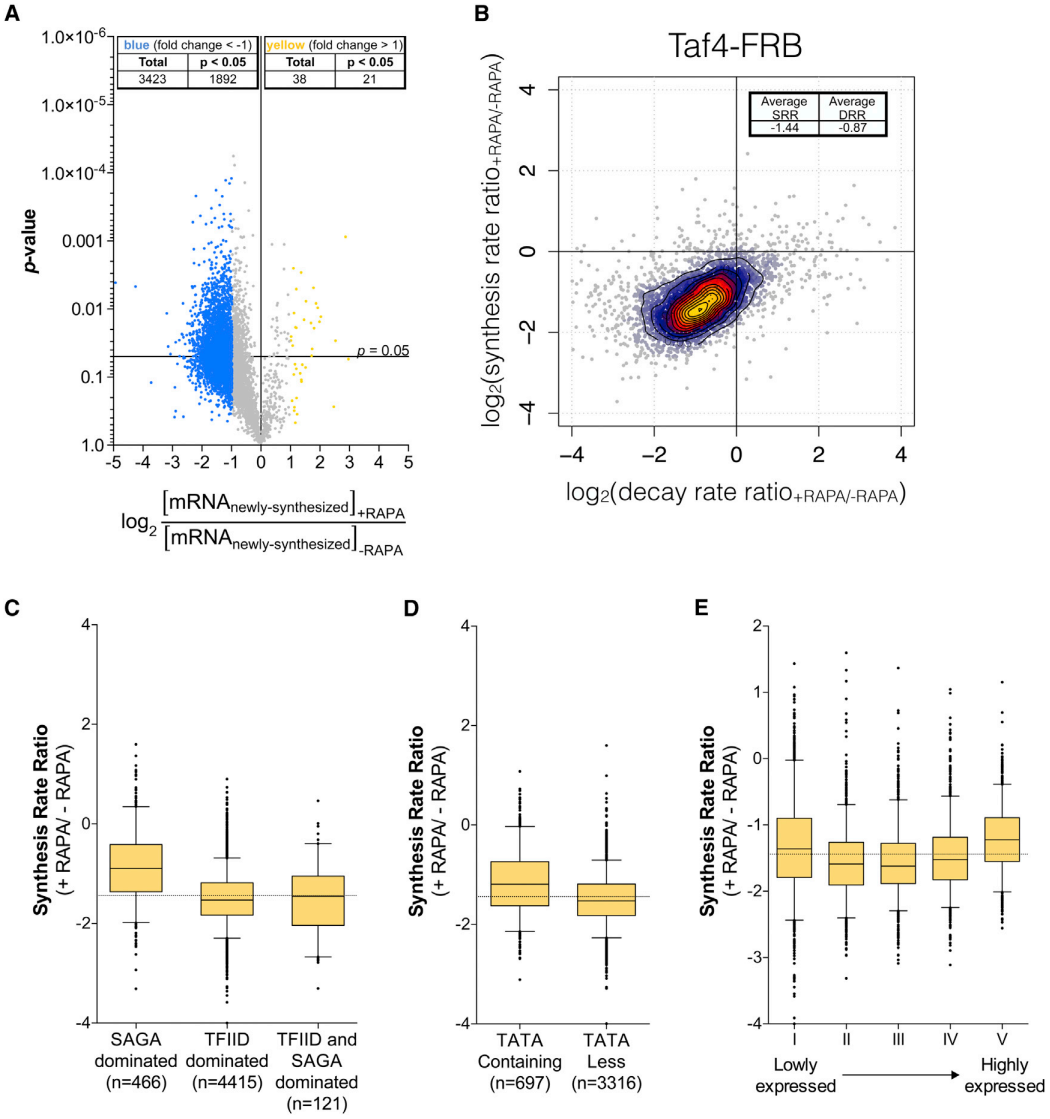


Figure 6. cDTA Analysis of Taf4 Anchor-away Strain

(A) Volcano plot showing fold changes in newly synthesized mRNA levels upon Taf4 nuclear depletion relative to their significance (p value). A total of 5,657 genes were analyzed in two biological replicates, and thresholds of 2-fold change and 0.05 p values were considered (genes upregulated by 2-fold are indicated in yellow and genes downregulated by 2-fold are indicated in blue).

(B) Changes in mRNA synthesis rates for 5,434 genes upon rapamycin treatment in a *TAF4-FRB* strain were plotted against the changes of the decay rates. Log_2 values of the average synthesis rate ratio and decay rate ratio are indicated.

(C–E) Boxplots showing the distribution of changes in synthesis rates upon Taf4 nuclear depletion according to the gene class (C and D) or to the gene expression levels (E). Dotted horizontal lines in (C)–(E) show the median of the synthesis rate ratio. See also Figures S3 and S4.

of decay rate by 1.8-fold, leading to partial buffering of steady-state mRNA levels.

Additionally, we addressed whether changes in synthesis rates for different genes were correlated to their previously proposed dependence on either TFIID or SAGA or the existence of a TATA consensus sequence in their promoters. Although the transcription of all gene classes was affected by the nuclear depletion of Taf4, the overall synthesis rates of the genes previously designated as TFIID dominated (or TATA-less genes) decreased more than that of previously designated SAGA-dominated (or TATA-containing) genes (Figures 6C and 6D). "SAGA-dominated" gene class had a median fold change of −1.86, whereas the "TFIID-dominated" class had a median fold change of −2.88, suggesting that genes from this latter group are more

sensitive to Taf4 depletion. Interestingly, the sensitivity of transcription to Taf4 depletion was not significantly correlated with the level of gene transcription (Figures 6E and S4B).

Nuclear depletion of Taf5 had very similar effects on transcription as Taf4 depletion, but with a broader effect, which might be explained by a combined inactivation of both TFIID and SAGA (see Discussion) (Figures S4C–S4I). Indeed, the newly synthesized mRNA levels of 4,238 genes (~75% of analyzed genes) were decreased by at least 2-fold (Figure S4D). However, as observed above with anchor-away of Taf4, our analysis showed that genes of the "TFIID-dominated" and TATA-less gene classes were more sensitive to Taf5 depletion (see Discussion). Nevertheless, our combined results using several different approaches demonstrate that TFIID is required for transcription of the vast majority of yeast genes regardless of previous gene classification.

Genome-wide TFIID Dependence Can Vary with Growth Conditions

One possible reason for the differences in our findings compared to published studies is a change in cell growth conditions. For example, some previous studies cultured cells in synthetic media, and all earlier studies of yeast Taf dependence involved heat shock of temperature-sensitive strains. To test whether these alternative growth conditions altered TFIID dependence, we repeated the above TFIID depletion experiment with cells grown in synthetic complete glucose media (SC) and with YPD-grown cells that had been subjected to a 30 min heat shock followed by 3-IAA addition. The degron system worked efficiently under these two growth conditions, with <10% Taf11 and Taf13 remaining after 30 min of 3-IAA treatment (Figure S5).

Surprisingly, the Pol II ChIP profile from wild-type cells grown in SC (Figure 7A, left panel, red line) was different from that of YPD-grown cells (left panel, purple line), with the peak of Pol II density over the middle of the coding sequence in SC compared to the peak density of Pol II near the TSS in YPD. This suggests a difference in the rates of initiation and/or elongation in the two growth conditions (Ehrensberger et al., 2013). After Taf depletion in SC, genome-wide Pol II transcription decreased (Figure 7A, middle and right panels), but the magnitude of TFIID dependence was somewhat lower compared to growth in rich media (Figure 7C). For example, the median Taf11 and Taf13-dependence was −3.7- and −4.1-fold in YPD compared with −2.1- and −1.8-fold in SC.

Heat shock of wild-type cells results in a general ~1.3-fold decrease in genome-wide transcription (Figure 7B, left panel, purple line versus red line). Depletion of Taf13 after the 30 min heat shock resulted in a further strong decrease in genome-wide Pol II transcription (Figure 7B, right panel). However, the magnitude of TFIID dependence was again less than that observed for growth at 30°C (Figure 7C). Under heat-shock stress, the median Taf13 dependence was −1.8-fold. However, TFIID dependence in both synthetic media and heat-shock conditions were reflected nearly equally in both the SAGA and TFIID-dominated gene classes (Figure S6). For example, Taf13 depletion in synthetic media and heat-shock conditions had near-equal effects on Taf1-depleted and Taf1-enriched genes as well as genes with or without a consensus TATA box.

Genome-wide comparisons of transcription after Taf-degradation in the three growth conditions are shown in Figure S7. We found that relative Pol II ChIP signals after degradation of individual Tafs are highly correlated between cells grown in YPD and SC at 30°C and in cells with or without heat shock. Therefore, the overall magnitude of TFIID dependence for genome-wide expression can vary with environmental conditions, but the relative dependence on TFIID for expression of individual genes is similar in all tested growth conditions.

DISCUSSION

Previous studies suggested that transcription from most Pol II-transcribed genes can be classified as dominated by either by the transcription factor TFIID or by the transcription factor SAGA. Here we have used orthogonal approaches to deplete TFIID-specific subunits and measure effects on nascent Pol II transcription. In contrast to earlier conclusions, in rich media we found a striking genome-wide dependence on TFIID, with nearly all genes showing greater than 2-fold decreases in transcription upon TFIID inactivation. Importantly, defects in gene expression observed after degron-dependent Taf1, Taf2, Taf7, Taf11, and Taf13 depletion were similar in the previously defined classes of "Taf1-depleted," "Taf1-enriched," TATA-containing, and TATA-less genes. This genome-wide TFIID dependence was observed under three different growth conditions, although the magnitude of TFIID contribution to total Pol II transcription varied with environmental conditions.

Anchor-away depletion of both the TFIID-specific subunit Taf4 and the shared SAGA/TFIID subunit Taf5 also showed decreased transcription from all classes of genes. However, depletion of these two Tafs gave stronger effects at the "TFIID-dominated" and TATA-less gene classes. One possibility is that both of these Tafs are more important for transcription of the TATA-less gene class compared to TATA-containing genes. Alternatively, anchor-away depletion may not be as severe as the degron method, or it may result in some partial TFIID and/or SAGA complexes that differentially function at the two gene classes. We attempted to generate an IAA7-tagged Taf4 strain to compare anchor-away and degron depletion of Taf4. However, we were unable to construct this strain, suggesting that the IAA7 tag is detrimental to Taf4 function. In any case, our combined results clearly show that transcription of nearly all genes of both classes is dependent on TFIID.

Several factors likely contribute to the differences between earlier work and the results reported here. First, formaldehyde crosslinking was used in earlier studies to map the location of TFIID subunits. Many studies have shown that, for most proteins in close contact with DNA, formaldehyde crosslinking is a reliable way to map protein-DNA locations. However, the factors Mediator, TFIID, and SAGA all show striking differences in genome-wide location when comparing formaldehyde-based methods versus the nuclease-based ChEC-seq method and other functional assays. For example, ChEC-seq analysis of TFIID suggested that TFIID occupies both the "Taf1-depleted" and the "Taf1-enriched" classes of genes to the same extent (Grünberg et al., 2016). Similarly, monitoring chromatin modifications and Pol II occupancy suggested that SAGA functions at all

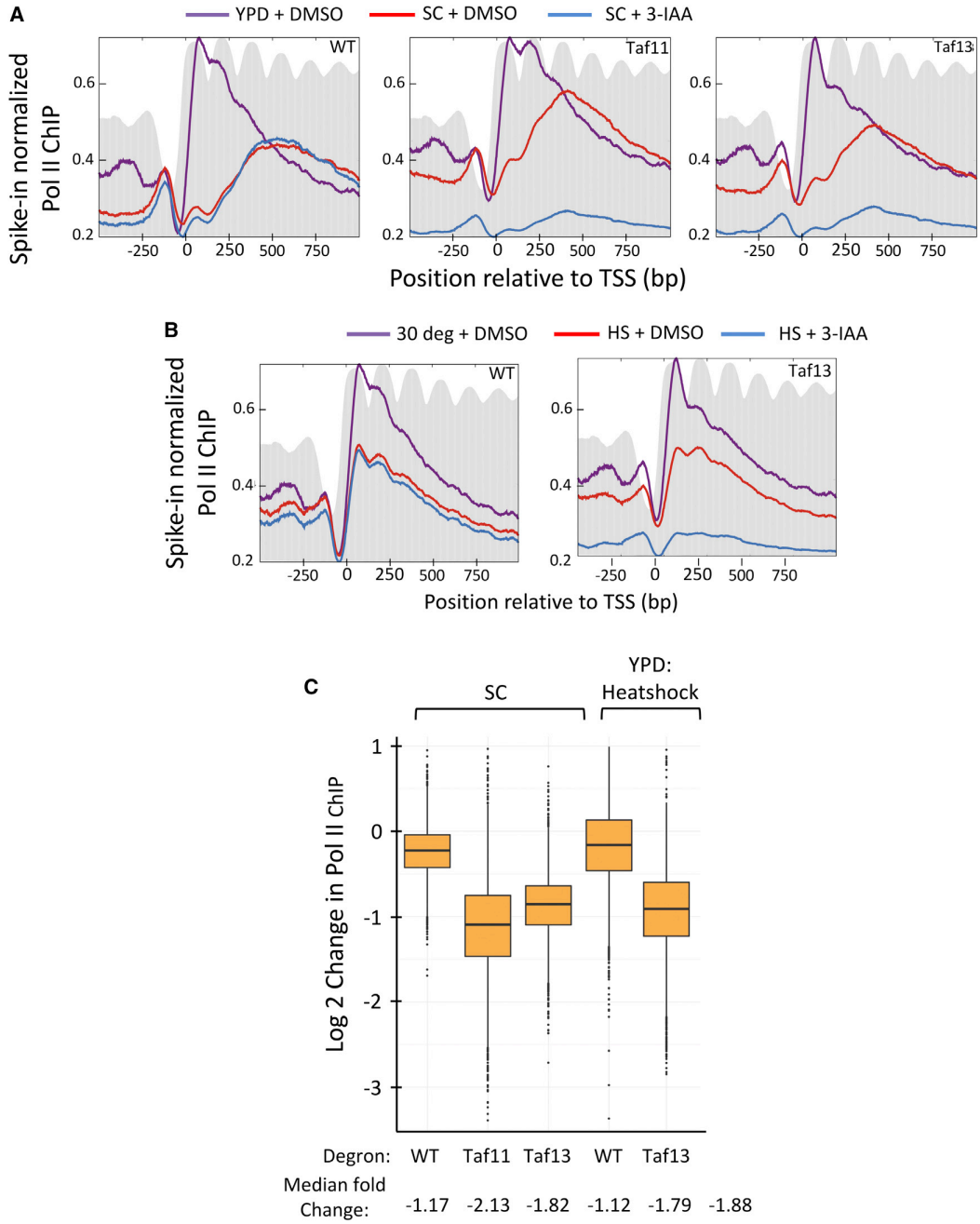


Figure 7. TFIID Dependence of Transcription in Synthetic Media and Heat-Shock Stress

(A) Metagene profile of native Pol II ChIP from cells grown in synthetic complete (SC) media and containing the indicated IAA7 degron fusion. Results are presented as in Figure 2. Purple line is Pol II ChIP data from YPD-grown cells (+DMSO) from Figure 2; red and blue lines are cells grown in SC with or without 3-IAA for 30 min.

(B) Same as in (A), but Pol II ChIP from cells under heat-shock stress. Purple line is Pol II ChIP from YPD-grown cells (+DMSO) in Figure 2. Red and blue lines are from YPD-grown cells heat shocked for 30 min prior to 30 min addition of DMSO or 3-IAA as indicated.

(C) Genome-wide changes in transcription upon TFIID subunit depletion from the data in (A) and (B). Boxplots show the change in native Pol II ChIP on a log2 scale from 4,807 genes comparing DMSO and 3-IAA addition. For heat-shock conditions, the plot indicates changes in Pol II ChIP in heat-shocked WT and heat-shocked degron-containing cells upon addition of 3-IAA. Median fold changes in ChIP signals are indicated below. See also Figures S5, S6, and S7.

expressed genes (Bonnet et al., 2014). While the reasons for these differences in TFIID mapping are not understood, one possible explanation may be the proximity of nucleosomes to PICs at these two promoter types. Genome-wide mapping showed that PICs at TATA-less, "TFIID-dominated" yeast promoters are usually adjacent to the +1 nucleosome, while PICs at TATA-containing promoters are usually located upstream of the +1 nucleosome (Rhee and Pugh, 2012). Perhaps protein-protein crosslinking of TFIID to the adjacent nucleosome contributes to stronger ChIP signals at TATA-less versus TATA-containing PICs. It is important to note that, at most so-called Taf1-depleted genes, low but detectable ChIP signals from Taf1 and other Tafs were detected (Kuras et al., 2000; Rhee and Pugh, 2011). A second difference with earlier work is the measurement of newly synthetized transcription instead of steady-state RNA levels. The cDTA analysis here clearly shows that, although some effects are already visible in total levels of RNA, inactivation of TFIID causes a genome-wide stabilization of steady-state mRNAs, masking the effects of TFIID depletion. Finally, earlier studies used heat shock to inactivate temperature-sensitive TFIID subunits followed by measurement of steady-state mRNA levels. We show here that heat shock causes a general genome-wide decrease in Pol II transcription and that transcription under these repressed conditions is somewhat less dependent on TFIID.

Binding of TFIID to Pol II promoters is one of the first steps on the pathway to PIC formation. TFIID interacts with some transcription activators, binds promoter DNA, and likely interacts with other basal transcription factors. In metazoans, TFIID (including its TBP subunit) interacts with promoter elements including TATA, Inr, MTE, and DPE. These latter interactions are thought to recruit and stabilize TBP to the ~80% of promoters lacking a consensus TATA. Since ChEC-seq mapping showed that TFIID is present at nearly all promoters, the promoter-binding function of TFIID likely explains the viability of yeast with TBP mutations that disrupt specific TBP-TATA binding (Kamenova et al., 2014). Paradoxically, no conserved promoter elements analogous to the Inr, MTE, and DPE have been identified in yeast, so the basis for yeast TFIID promoter recognition is not clear. Lobe C in human TFIID interacts with these downstream promoter elements, so it is expected that lobe C subunits, especially Taf1, should be critical for TFIID function (Louder et al., 2016). However, we found a small number of genes where transcription is less sensitive to Taf1, Taf2, and Taf7 depletions compared to dependence on Taf11 and Taf13 (cluster 1). Thus, the downstream promoter DNA binding function of yeast TFIID may not be as critical as it is in metazoans. Alternatively, other partial TFIID assemblies, such as the core TFIID, the 7TAF, or the 8TAF complexes (Bieniossek et al., 2013; Trowitzsch et al., 2015; Wright et al., 2006), may carry out partial functions at these promoters, which are less sensitive to Taf1 and/or Taf7 depletions.

A popular view is that TFIID and SAGA can replace each other, each functioning independently of the other at different classes of genes. Results shown in the accompanying paper (Baptista et al., 2017, in this issue of *Molecular Cell*) also contradict this view, as yeast SAGA was shown to localize to the UAS regions of nearly all yeast genes and to be important for nearly all Pol II transcription, regardless of TATA presence, or Taf1 enrichment measured by ChIP. Our combined results lead to the view that (1) most active yeast promoters have SAGA localized to UASs and TFIID localized at promoters and that (2) SAGA and TFIID are not alternative factors, but both make independent major contributions to expression of nearly all genes. It is important to note that we found newly synthetized transcription more sensitive to Mediator inactivation compared with TFIID inactivation. Therefore, in yeast there may be a low level of transcription at many genes that is TFIID independent and derives from PICs assembled with TBP, but lacking Tafs. Since the magnitude of TFIID dependence varies with growth conditions, it may be that the level or function of TBP relative to Tafs also varies with environmental conditions. An alternative explanation is that auxin-dependent protein degradation is not complete and that more residual TFIID or partial TFIID complexes (see above) compared with Mediator remain after 3-IAA addition. These scenarios are also plausible when using anchor-away strains, leaving some residual subunits or partial complexes that could replace canonical complexes for transcriptional regulation by holo-TFIID, as previously suggested in metazoans (Indra et al., 2005; Tatarakis et al., 2008; Wright et al., 2006).

Finally, many studies showed at least two broad classes of Pol II-transcribed genes: housekeeping and highly regulated genes. Highly regulated genes are usually TATA containing, are modulated by chromatin rearrangement and/or modifications, are sensitive to levels of transcription activators, and are repressed by the regulators NC2 and Mot1. In contrast, the housekeeping class is generally TATA-less, is associated with high levels of H4 acetylation in promoter regions, is responsive to a different set of activators, and is generally activated (or derepressed) by an NC2/Mot1 interplay. These two gene classes are often characterized as TFIID or SAGA dominated. Our new findings show that these two gene classes are not distinguished by TFIID and SAGA requirements, as both these factors make important contributions to expression of nearly all genes. This suggests that some or all of the other properties described above are likely responsible for the differences in regulation observed between these different gene categories.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Strain Construction
  - Yeast Cell Growth, Harvesting, and Breakage of Frozen Cells for Modified Pol II ChIP-seq
  - Chromatin Extract Preparation and MNase Digestion
  - IPs to Recover Elongation Complexes
  - Preparation of DNA Libraries for NGS
  - Nascent RNA Labeling and Purification for RTqPCR
  - cDNA Synthesis and Quantitative PCR of Nascent RNA
  - cDTA analysis
  - Cell fractionation, growth curve and cell viability

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and five tables and can be found with this article at http://dx.doi.org/10.1016/j.molcel.2017.08.014.

AUTHOR CONTRIBUTIONS

S.H., L.W., S.R., T.B., D.D., and L.T. conceived and designed the experiments; and L.W. performed all auxin-degron experiments, 4 thioU labeling, and RT-qPCR analysis of degron-containing cells and TFIID IPs. S.R. performed the computational analysis of all auxin-degron experiments. T.B. performed experiments using anchor-away strains (4 thioU labeling, genome-wide transcriptional analyses, computational analyses, and cell fractionation experiments). All authors analyzed the results and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Steve Henikoff (FHCRC) for support, advice, and encouragement throughout the course of this work; Matt Miller and Toshi Tsukiyama (FHCRC) for reagents, advice, and assistance with the auxin degron system; Christine Codomo (FHCRC) for advice on construction of DNA sequencing libraries; Erin Duffy and Matt Simon (Yale) for advice on labeling and purification of nascent mRNA; Marc H.T. Timmers (UMC Utrecht, the Netherlands) for anchor-away strains; and Sebastian Grünberg, Rafał Donczew, and Steve Henikoff for comments on the manuscript. Supported by NIH grant GM053451 (S.H.) and the Howard Hughes Medical Institute (to S. Henikoff). T.B. was funded by a Marie Curie Initial Training Networks (EU-FP7 PEOPLE-2013 program, PITN-GA-2013-606806, NR-NET) and Association pour la Recherche sur le Cancer fellowships. This study was also supported by the European Research Council (ERC) Advanced grant (ERC-2013-340551, Birtoaction, to L.T.), ANR grant ANR-15-CE11-0022 (SAGA2), and grant ANR-10-LABX-0030-INRT, 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.

Received: March 29, 2017
Revised: June 28, 2017
Accepted: August 18, 2017
Published: September 14, 2017

REFERENCES

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.

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*. Published online September 14, 2017. http://dx.doi.org/10.1016/j.molcel.2017.08.016.

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.

Berger, I., Blanco, A.G., Boelens, R., Cavarelli, J., Coll, M., Folkers, G.E., Nie, Y., Pogenberg, V., Schultz, P., Wilmanns, M., et al. (2011). Structural insights into transcription complexes. J. Struct. Biol. *175*, 135–146.

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.

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.

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

Cianfrocco, M.A., and Nogales, E. (2013). Regulatory interplay between TFIID's conformational transitions and its modular interaction with core promoter DNA. Transcription *4*, 120–126.

Cianfrocco, M.A., Kassavetis, G.A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J.T., and Nogales, E. (2013). Human TFIID binds to core promoter DNA in a reorganized structural state. Cell *152*, 120–131.

de Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453–1454.

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.

Duffy, E.E., and Simon, M.D. (2016). Enriching s(4) U-RNA using methane thiosulfonate (MTS) chemistry. Curr. Protoc. Chem. Biol. *8*, 234–250.

Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking distinct RNA populations using efficient and reversible covalent chemistry. Mol. Cell *59*, 858–866.

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

Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends Genet. *29*, 569–574.

Grant, P.A., Schieltz, D., Pray-Grant, M.G., Steger, D.J., Reese, J.C., Yates, J.R., 3rd, and Workman, J.L. (1998). A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. Cell *94*, 45–53.

Grünberg, S., Henikoff, S., Hahn, S., and Zentner, G.E. (2016). Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. EMBO J. *35*, 2435–2446.

Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705–736.

Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millán-Zambrano, G., Barkai, O., Chávez, S., Pérez-Ortín, J.E., Darzacq, X., and Choder, M. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. Cell *153*, 1000–1011.

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

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell *31*, 925–932.

Herbig, E., Warfield, L., Fish, L., Fishburn, J., Knutson, B.A., Moorefield, B., Pacheco, D., and Hahn, S. (2010). Mechanism of Mediator recruitment by tandem Gcn4 activation domains and three Gal11 activator-binding domains. Mol Cell Biol. *30*, 2376–2390.

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.

Indra, A.K., Mohan, W.S., 2nd, 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.

Kamenova, I., Warfield, L., and Hahn, S. (2014). Mutations on the DNA binding surface of TBP discriminate between yeast TATA and TATA-less gene transcription. Mol. Cell. Biol. *34*, 2929–2943.

Kuras, L., Kosa, P., Mencia, M., and Struhl, K. (2000). TAF-containing and TAF-independent forms of transcriptionally active TBP in vivo. Science *288*, 1244–1248.

Lavigne, A.-C., Gangloff, Y.-G., Carré, L., Mengus, G., Birck, C., Poch, O., Romier, C., Moras, D., and Davidson, I. (1999). Synergistic transcriptional activation by TATA-binding protein and hTAFII28 requires specific amino acids of the hTAFII28 histone fold. Mol. Cell. Biol. *19*, 5050–5060.

Lee, T.I., Causton, H.C., Holstege, F.C., Shen, W.C., Hannett, N., Jennings, E.G., Winston, F., Green, M.R., and Young, R.A. (2000). Redundant roles for the TFIID and SAGA complexes in global transcription. Nature *405*, 701–704.

Leurent, C., Sanders, S., Ruhlmann, C., Mallouh, V., Weil, P.A., Kirschner, D.B., Tora, L., and Schultz, P. (2002). Mapping histone fold TAFs within yeast TFIID. EMBO J. *21*, 3424–3433.

Leurent, C., Sanders, S.L., Demény, M.A., Garbett, K.A., Ruhlmann, C., Weil, P.A., Tora, L., and Schultz, P. (2004). Mapping key functional sites within yeast TFIID. EMBO J. *23*, 719–727.

Li, X.Y., Bhaumik, S.R., and Green, M.R. (2000). Distinct classes of yeast promoters revealed by differential TAF recruitment. Science *288*, 1242–1244.

Louder, R.K., He, Y., López-Blanco, J.R., Fang, J., Chacón, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature *531*, 604–609.

Mencía, M., Moqtaderi, Z., Geisberg, J.V., Kuras, L., and Struhl, K. (2002). Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. Mol. Cell *9*, 823–833.

Müller, F., and Tora, L. (2014). Chromatin and DNA sequences in defining promoters for transcription initiation. Biochim. Biophys. Acta *1839*, 118–128.

Munchel, S.E., Shultzaberger, R.K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol. Biol. Cell *22*, 2787–2795.

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.

Ohtsuki, S., Levine, M., and Cai, H.N. (1998). Different core promoters possess distinct regulatory activities in the Drosophila embryo. Genes Dev. *12*, 547–556.

Park, D., Morris, A.R., Battenhouse, A., and Iyer, V.R. (2014). Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. Nucleic Acids Res. *42*, 3736–3749.

Plaschka, C., Nozawa, K., and Cramer, P. (2016). Mediator architecture and RNA polymerase II interaction. J. Mol. Biol. *428*, 2569–2574.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408–1419.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.

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.

Schwalb, B., Schulz, D., Sun, M., Zacher, B., Dümcke, S., Martin, D.E., Cramer, P., and Tresch, A. (2012). Measurement of genome-wide RNA synthesis and decay rates with Dynamic Transcriptome Analysis (DTA). Bioinformatics *28*, 884–885.

Setiaputra, D., Ross, J.D., Lu, S., Cheng, D.T., Dong, M.-Q., and Yip, C.K. (2015). Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. J. Biol. Chem. *290*, 10057–10070.

Shen, W.-C., Bhaumik, S.R., Causton, H.C., Simon, I., Zhu, X., Jennings, E.G., Wang, T.-H., Young, R.A., and Green, M.R. (2003). Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly. EMBO J. *22*, 3395–3402.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife *6*, e21856.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res. *22*, 1350–1359.

Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol. Cell *52*, 52–62.

Tatarakis, A., Margaritis, T., Martinez-Jimenez, C.P., Kouskouti, A., Mohan, W.S., 2nd, 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.

Theisen, J.W.M., Lim, C.Y., and Kadonaga, J.T. (2010). Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. Mol. Cell. Biol. *30*, 3471–3479.

Thurtle, D.M., and Rine, J. (2014). The molecular topography of silenced chromatin in Saccharomyces cerevisiae. Genes Dev. *28*, 245–258.

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.

Tsai, K.-L., Tomomori-Sato, C., Sato, S., Conaway, R.C., Conaway, J.W., and Asturias, F.J. (2014). Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. Cell *157*, 1430–1444.

Verrijzer, C.P., Chen, J.L., Yokomori, K., and Tjian, R. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. Cell *81*, 1115–1125.

Wright, K.J., Marr, M.T., 2nd, and Tjian, R. (2006). TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. Proc. Natl. Acad. Sci. USA *103*, 12347–12352.

Zabidi, M.A., and Stark, A. (2016). Regulatory enhancer-core-promoter communication via transcription factors and cofactors. Trends Genet. *32*, 801–814.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-V5	Invitrogen	Cat # 46-0705
Anti-FLAG M2 Magnetic Beads	Sigma-Aldrich	Cat # M8823
Rabbit polyclonal anti-Taf1	Hahn lab	Rabbit # 3506
Rabbit polyclonal anti-Taf4	Hahn lab	Rabbit # 4022
Rabbit polyclonal anti-Taf12	Hahn lab	Rabbit # 3434
Rabbit polyclonal anti-Taf3	Hahn lab	Rabbit # 4045
Rabbit polyclonal anti-TBP	Hahn lab	Rabbit # 7370
Rabbit polyclonal anti-H3	Abcam	Cat # ab1791
Rabbit polyclonal anti-Taf4	P. Anthony Weil	N/A
Rabbit polyclonal anti-Taf5	P. Anthony Weil	N/A
Rabbit polyclonal anti-Taf11	P. Anthony Weil	N/A
Chemicals, Peptides, and Recombinant Proteins		
3-Indoleacetic acid (3-IAA)	Sigma-Aldrich	I3750; CAS: 87-51-4
Micrococcal Nuclease (MNase)	Worthington Biochemical Corporation	Cat # NFCB
cOmplete Mini, EDTA-free Protease inhibitor cocktail tablets	Roche	Cat # 11836170001
4-thiouracil	Sigma-Aldrich	Cat # 440736; CAS: 591-28-6
MTSEA biotin-XX	Biotium	Cat # 90066
Dynabeads MyOne Streptavidin C1	Invitrogen, ThermoFisher	Cat # 65001
AMPureXP beads	Beckman Coulter	Cat # A63881
EZ-Link HPDP Biotin	ThermoFisher	Cat # 21341
Rapamycin	Euromedex	Cat # SYN-1185
Dyanabeads Protein G	Thermo Fisher Scientific	Cat # 10004D
0.5 mm zirconia disruption beads	Research Products International	Cat # 9834
Critical Commercial Assays		
RiboPure yeast kit	Ambion, Life Technologies	Cat # AM1926
RNeasy MinElute Cleanup Kit	QIAGEN	Cat # 742040
μMACS Streptavidin Kit	Miltenyi Biotec	Cat# 130-074-101
Deposited Data		
All sequencing datasets have been uploaded in GEO under accession GSE97081		
Western blot original images	Mendeley Data file	http://dx.doi.org/10.17632/gwtwc4ndpj.1
Experimental Models: Organisms/Strains		
S. pombe: Strain background: 972h	Laboratory of Gerald Smith	ATCC: 24843
S. cerevisiae: Strain background: BY4705	Laboratory of Dan Gottschling	ATCC: 200869
Yeast strains, see Table S2	This paper	N/A
Sequence-Based Reagents		
Primers for RT-qPCR, see Table S1	This paper	N/A
Software and Algorithms		
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	N/A
Shell Scripts for Sequence Alignment	This paper	
Perl Scripts for analysis	This paper	
R Scripts for analysis	This paper	

Molecular Cell, Volume *68*

Supplemental Information

Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID

Linda Warfield, Srinivas Ramachandran, Tiago Baptista, Didier Devys, Laszlo Tora, and Steven Hahn

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author, Dr. Steven Hahn (shahn@fredhutch.org).

METHOD DETAILS

Strain Construction

S. cerevisiae and *S. pombe* strains were constructed using standard yeast methods. Proteins were chromosomally tagged by high efficiency yeast transformation and homologous recombination of PCR-amplified DNA. Plasmid pFA6a-3V5-IAA7-KanMX6 (from the Karstein Weiss laboratory, UC Berkeley - a gift from M. Miller, FredHutch) was used as the template for generating the IAA7 degron tags. The C-terminal tag contains three copies of the V5 epitope tag followed by the IAA7 degron (32 kDa total).

Yeast Cell Growth, Harvesting, and Breakage of Frozen Cells for Modified Pol II ChIP-seq

For each yeast strain and condition, two liters of *S. cerevisiae* were grown at 30 deg in YPD (+ 100 µl Antifoam 204) to an absorbance of ~0.7-1.0. Cells were then induced for protein degradation by the addition of 10 ml 200X 3-IAA (500 µM final working concentration) dissolved in DMSO, or 10 ml DMSO for the no 3-IAA controls, for 30 minutes at 30 deg. Cells were rapidly harvested by vacuum filtration through a 90 mm diameter 0.45 micron nitrocellulose filter, followed by a 100 ml wash with the cell culture media. The cell paste was scraped off with a pre-chilled spatula and plunged into liquid nitrogen to flash freeze cells. For spike-in normalization, an 8 liter batch of S. pombe was grown in YE (+ Antifoam 204) to a similar absorbance and harvested as described above. 2.5 g *S. cerevisiae* were combined with 0.25 g S. pombe in a mortar pre-chilled on dry ice and liquid nitrogen. Frozen cells were lysed with a 4.5" OD mortar and large pestle. Cells were ground for 1 min, liquid nitrogen was then added to the cells and ground for 1 min. The grinding and addition of liquid nitrogen was repeated for a total of 6 minutes. The frozen powdered ground cells were transferred to a cold conical tube and stored at -80 deg.

For the minimal media growth conditions, two liters of *S. cerevisiae* were grown at 30 deg as above, except in synthetic complete media containing 2% glucose and lacking

isoleucine and valine (+ 100 µl Antifoam 204). For heat shock conditions, two liters of *S. cerevisiae* were grown at 30 deg in YPD as above, except when cells reached an absorbance of ~0.7-1.0 two liters of YPD prewarmed to 44 deg was added and cells were transferred to a 37 deg C shaker for 30 min before the addition of 3-IAA for 30 min. Cells were then harvested and processed as described above.

Chromatin Extract Preparation and MNase Digestion

0.5 g frozen cells were added to a pre-chilled 15 ml conical tube and kept on dry ice until all cell preps were weighed. Tubes containing frozen cells were placed on ice for 5 min to slowly warm. 2.5 ml cold lysis buffer + protease inhibitors (50 mM HEPES, 150 mM NaCl, 0.5 mM EDTA, 0.05% TritonX-100, pH 7.4, cOmplete Mini protease inhibitor tablet (Roche)) was added to the cells and the cells were allowed to resuspend with occasional mixing on ice for ~ 5-10 min. These resuspended, lysed cells were incubated in a 37 deg water bath for 5 min, then 2.5 mM $CaCl_2$ and 5 µl 100 U/µl MNase were added and mixed by inversion. Chromatin was MNase digested for 5 min in a 37 deg C water bath, then the reaction was stopped by addition of EGTA to a final concentration of 12 mM. Reactions were incubated on ice for at least 10 minutes while other samples were processed. Extracts were spun at 3K x g for 2 min at 4 deg., then the supernatant transferred to a new tube and spun at 17K x g for 15 min at 4 deg. Supernatant was used immediately for IP.

IPs to Recover Elongation Complexes

1 ml of MNase-treated chromatin extract from above was incubated with 100 µl magnetic anti-Flag beads (prewashed 3 times in lysis buffer + 5 mg/ml BSA + 1.5 mM EDTA + cOmplete Mini protease inhibitor tablet Roche) overnight at 4 deg with rotation. Beads were washed 4 times with 1.5 ml lysis buffer + 1.5 mM EDTA and resuspended in 400 µl same buffer. DNA was eluted by the following: add 20 µg RNaseA and additional NaCl and EDTA so that the final concentrations increase by 100 mM (NaCl) and 10 mM (EDTA). Samples were incubated at 37 deg for 10 min with mixing. 0.5% SDS (final concentration) and 80 µg Proteinase K (Invitrogen) were added and samples incubated at 65 deg for 20 min with mixing. Supernatant was phenol/CHCl3/IAA extracted and DNA precipitated with 30 µg glycogen (Roche), 1/10 vol 3 M NaOAc, and 0.7 vol isopropanol. DNA was resuspended in 0.1X TE and quantitated by PicoGreen assay (Thermo Fisher).

Preparation of DNA Libraries for NGS

DNA isolated from IPs was prepared for sequencing as previously described (Skene and Henikoff, 2017), except for the following changes. End Repair and A-tailing was performed in a 20 ml volume containing up to 50 ng DNA and the following final concentrations: 1X T4 DNA ligase buffer (NEB #B0202S), 0.5 mM each dNTP, 0.25 mM ATP, 2.5% PEG 4000, 2.5 U T4 PNK (NEB #M0201S), 0.05 U T4 DNA polymerase (Invitrogen #18005025), and 0.05 U Taq DNA polymerase and reactions were incubated at 12 deg 15 min, 37 deg 15 min, 72 deg 20 min, then put on ice and immediately used in adapter ligation reactions. Adapter Ligation was performed in a 50 ml volume containing 100:1 adapter:insert molar ratios, 1X Rapid DNA ligase buffer (Enzymatics #B101L) and 3000 U DNA ligase (Enzymatics #L6030-HC-L) and reactions were incubated at 20 deg for 15 min. A single post-ligation cleanup was performed using 0.4X vol AMPure XP reagent (Beckman Coulter #A63881) and DNA eluted in 40 ul 0.1X TE. Library Enrichment was performed in a 25 µl volume containing 16.25 µl DNA from previous step and the following final concentrations: 1X KAPA buffer, 0.3 mM each dNTP, 0.5 mM each P5 and P7 PCR primer, and 0.5 U KAPA HS HIFI polymerase (#KK2502). DNA was amplified with the following program: 98 deg 45 sec, [98 deg 15 sec, ramp to 60 deg @ 1 deg/sec, 60 deg 20 sec, ramp to 98 deg @ 1 deg/sec] 8-10X, 72 deg 1 min. A post-PCR cleanup was performed using 1.4X vol AMPure XP reagent and DNA was eluted into 40 ml 0.1X TE. Libraries were sequenced on the Illumina HiSeq2500 platform using 25 bp paired-ends at the Fred Hutchinson Cancer Research Center Genomics Shared Resources facility.

Nascent RNA Labeling and Purification for RTqPCR

S. cerevisiae and *S. pombe* cells were grown and newly synthesized RNAs were labeled as previously described (Bonnet et al., 2014), except that *S. cerevisiae* strains at an A600 of 1.0 were treated with 500 µM 3-IAA or DMSO for 30 minutes immediately prior to RNA labeling. Labeled *S. cerevisiae* and *S. pombe* cells were mixed in a 3:1 ratio before total RNA was extracted using the RiboPure yeast kit (Ambion, Life Technologies). RNA was then biotinylated essentially as described (Duffy et al., 2015; Duffy and Simon, 2016) using 40 µg total RNA and 4 µg MTSEA biotin-XX (Biotium). Unreacted MTS-biotin was removed by phenol/CHCl3 extraction and RNA was precipitated by isopropanol precipitation. Biotinylated RNA was purified also as described (Duffy and Simon, 2016) using MyOne Streptavidin C1 Dynabeads (Invitrogen) and RNAs eluted into 50 µl

streptavidin elution buffer (100 mM DTT, 20 mM HEPES 7.4, 1 mM EDTA, 100 mM NaCl, 0.05% Tween-20). At this point, 10% input RNA (4 µg) was diluted into 50 µl streptavidin elution buffer and processed the same as the labeled RNA samples. 50 µl each input and purified RNA was adjusted to 100 µl with nuclease-free water and purified on RNeasy columns (Qiagen) using the modified protocol as described (Duffy and Simon, 2016). RNAs were eluted into 14 µl nuclease-free water.

cDNA Synthesis and Quantitative PCR of Nascent RNA

Eleven microliters RNA was used to generate cDNA using Transcriptor (Roche), random hexamer primer, and the manufacturer's instructions. cDNA was used either undiluted, 1/20, or 1/100 for quantitative PCR (qPCR) depending on the gene analyzed. Gene specific qPCR was performed as previously described (Herbig et al., 2010) except that primers were designed near the 5' end of the gene and reactions were run on the QuantStudio5 Real-Time System (ABI). Relative amounts of transcript were normalized to S. pombe tubulin transcripts.

cDTA analysis

Comparative dynamic transcriptome analysis (cDTA) is an approach that allows analysis of newly-synthesized RNA in order to determine transcriptional defects at much higher sensitivity than those methods solemnly relying on evaluation of steady-state levels of RNA. This method relies on a short, non-perturbing metabolic labeling of newly-synthesized RNAs with 4-thiouracil for 6 min. After microarray hybridization or RNA sequencing, it is possible to determine synthesis and decays rates simultaneously for every transcript.

For each yeast strain and condition, 200mL of *S. cerevisiae* cells were grown in YPD medium at 30°C to an $OD_{600} \approx 0.8$. At that point, the culture was divided into two equal volumes of 100mL each and to one of those cultures rapamycin was added (+RAPA) until a final concentration of 1µg/mL, for 30 min, to allow nuclear depletion of either Taf4 or Taf5. To the other half of the culture, the control samples (-RAPA/minus rapamycin), only the vehicle (DMSO) was added. Newly synthesized RNAs were labeled for 6 min by adding 4-thiouracil (Sigma-Aldrich) until a final concentration of 5mM. In parallel, wild-type *S. pombe* cells were similarly grown in YES medium, at 31°C, and labeled for 6 min to be used as a spike-in across all samples. Cells were immediately pelleted and flash-frozen in liquid N_2 and stored at -80°C until further use. Immediately before freezing the cells, a small aliquot was

collected for cell counting purposes. *S. cerevisiae* and *S. pombe* cells were mixed in a ratio of 3:1. Total RNAs were extracted using RiboPure yeast kit (Ambion, Life Technologies) according to the description provided by the manufacturer.

Prior to biotinylation, RNA samples were heated for 10 min at 60°C and cooled down on ice for another 5 min. Subsequently, RNA biotinylation was carried out on 200µg of total RNA using 200µL of 1mg/mL EZ-link HPDP- Biotin (Pierce) in 100µL of biotinylation buffer (100mM Tris- HCl at pH 7.5, 10mM EDTA) and adjusted to a final volume of 1000µL with DEPC-treated RNase-free water (Sigma-Aldrich) for 3 h at room temperature, protected from light and gentle agitation. After chloroform extraction and isopropanol precipitation (1/10 vol 5M NaCl and 2.5 vol isopropanol), purified RNAs were suspended in 100µL of DEPC-treated RNase-free water (Sigma-Aldrich).

Recovered RNA samples were incubated at 65°C for 10 min and allowed to cool down on ice for 5 min. Newly synthesized biotinylated RNAs were bound to 100µL of µMACS streptavidin microbeads (Miltenyi Biotec) for 90 min at room temperature with gentle shaking. Purification of labeled RNA was then carried out using µMACS streptavidin starting kit (Miltenyi Biotec). Columns were first equilibrated with 1mL of washing buffer (100mM Tris-HCl at pH 7.5, 10mM EDTA, 1M NaCl, 0.1% Tween20). Samples were passed through the columns twice and washed five times with increasing volumes of washing buffer (600, 700, 800, 900, and 1000µL). Ultimately, labeled RNAs were eluted twice with 200µL of 100mM DTT. Following ethanol precipitation (overnight precipitation in 1/10 vol of 3M NaOAc, 3 vol of 100% ethanol and 20µg of RNA-grade glycogen), RNAs were washed in ice-cold 70% ethanol and resuspended in 20µL of DEPC-treated RNase-free water (Sigma-Aldrich). The quality of the samples was ascertained and finally hybridized to GeneChip Yeast Genome 2.0 microarrays following the instructions from the supplier (Affymetrix). All experiments were done using two independent biological replicates.

Raw data were normalized to S. pombe signal and used to calculate fold-changes in total and newly-synthesized RNA levels, as represented in volcano plots. Further calculations of synthesis and decay rates, based on mathematical model as described in Sun et al., were performed using a pipeline and R/Bioconductor package publicly available (Schwalb et al., 2012).

Cell fractionation, growth curve and cell viability

150 ml cell culture ($OD_{600} \approx 0.5$) was harvested and treated with 0.1M Tris, pH 9.4, 10mM DTT for 5 min at room temperature. Cells were pelleted and resuspended in YPD/S (YPD with 1M Sorbitol) and collected by centrifugation. Cells were resuspended in 1mL YPD/S and 750µL 2M sorbitol. Zymolyaze 100T was added until a final concentration of 0.3mg/mL. Cells were incubated at 30°C for 30min (progress of spheroplasting every was checked every 10min). After this treatment, spheroplasts were washed and resuspended in room temperature YPD/S for 30 min at 30°C. Cells were quickly cooled down in ice, centrifuged, washed and resuspended in 1.5 ml lysis buffer (20 mM K-phosphate pH 6.5, 0.5 mM MgCl2, 18% Ficoll, 1XPIC). Spheroplasts were lysed three times with a B dounce and shortly centrifuged at 4500g to remove unbroken cells. The precleared whole cell lysate (Total lysate fraction) was centrifuged at 21000g for 45 min resulting into a crude nuclear pellet (nuclear fraction) and post-nuclear supernatant cytosol fraction (cytoplasmic fraction). 200µl of each fraction was TCA precipitated, washed with ice-cold acetone and resuspended in SDS-sample buffer.

For growth curve analysis, the anchor-away (TAF4-FRB and TAF5-FRB) and parental strains were grown overnight and then diluted to $OD_{600} \approx 0.1$ in pre-heated YPD (200mL). To the culture, rapamycin was added until a final concentration of 1µg/mL (+ RAPA). To the control samples (-RAPA), only the vehicle (DMSO) was added. Every 30min, OD_{600} was measured until saturation of the culture.

For cell viability, strains were grown until log-phase (OD600 ≈ 0.7). To the culture, rapamycin was added until a final concentration of 1µg/mL (+ RAPA). To the control samples (-RAPA), only the vehicle (DMSO) was added. After 30min an aliquot was collected, stained with methylene blue and counted using a neubauer chamber.

TFIID Immunoprecipitations and Western Blot Analysis

For each yeast strain, 400 ml of *S. cerevisiae* was grown at 30 deg in YPD to an absorbance of ~0.7-1.0. Cultures were then split into two 200 ml flasks and were treated by addition either 1 ml of 200X 3-indoleacetic acid (3-IAA) (500 µM final working concentration) dissolved in DMSO, or 1 ml DMSO for the no 3-IAA controls, for 30 minutes at 30 deg. Cells were collected by centrifugation, washed with 25 ml IP buffer (40 mM HEPES, 150 mM NaCl, 10% glycerol, 0.1% Tween-20) supplemented with 1 mM

dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 0.31 mg/ml benzamidine, 0.3 μg leupeptin, 1.4 μg/ml pepstatin, and 2 μg/ml chymostatin, and then resuspended in ~750 μl of the same buffer. Cells were transferred to a 2 ml tube containing 1.5 ml 0.5 mm zirconia beads and small scale whole cell extracts were prepared in a mini bead beater-96 (BioSpec products) by shaking 5 times for 3 min each with 1 min chill in ice water between each lysis step. Lysates were clarified by a 3000 x g spin for 5 min followed by a high speed spin in a microfuge at 20K x g for 15 min. Concentrations of extracts were determined by Bradford assays. For IP experiments, 5 mg of extracts were incubated at 4°C overnight with 30 μl Protein G Dynabeads which had been conjugated and crosslinked with α-Taf4 antibody or no antibody as a control. For antibody conjugation, Protein G Dynabeads were washed twice with PBS, then resuspended in PBS keeping the beads at 30 mg/ml, and either 10 μl rabbit polyclonal α-Taf4 or 10 μl PBS then rotated at room temp for 45 min. Beads were washed 3X in PBS, 1X in 0.2M triethanolamine pH 8.2, then resuspended in 0.2M triethanolamine pH 8.2 containing 25 mM dimethyl pimelimidate (DMP), rotated at room temp for 30 min, then washed 1X in PBS. The DMP incubation and PBS wash were repeated for a total of 3 times. Beads were washed 1X in 0.1M ethanolamine and rotated for 5 min, then washed 1X in PBS. Beads were washed in 1M glycine pH 3.0 and rotated for 10 min at room temp and then repeated. Beads were finally washed 3X in IP buffer supplemented with 1 mM DTT and protease inhibitors listed above. Following the overnight IP step, beads were washed 3X with IP buffer, then proteins were eluted by incubating beads in 15 μl 1 M glycine pH 2.5 for 10 min at room temperature. The elution step was repeated and elutions combined. 10 μl 1M Tris pH 8.5 was added to the combined eluate. Eluted proteins were separated by SDS-PAGE and analyzed by Western blot using rabbit polyclonal antibodies. Protein signals were quantified using Odyssey scanner software (Li-Cor) by generating a standard curve using a titration from WT IP. Each protein analyzed was normalized to the amount of the Taf4 protein obtained by IP.

Yeast Viability Following 3-IAA Treatment

For each yeast strain, 15 ml of *S. cerevisiae* were grown at 30 deg in YPD to an absorbance of ~0.5-0.7. Cultures were split into two 5 ml cultures and were then induced for protein degradation by the addition of 25 μl 200X 3-IAA (500 μM final working

concentration) dissolved in DMSO, or 25 μl DMSO for the no 3-IAA controls, for 30 minutes at 30 deg. Cultures were diluted in H_2O and plated to YPD plates for recovery at 30 deg. Colonies were counted after 3 days, and the percent viability was calculated for cells treated with 3-IAA versus DMSO only treatment. Each strain was grown and assayed in duplicate.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of ChIP-Seq Data

Paired-end reads were independently aligned to *S. cerevisiae* reference genome (sacCer3) and S. pombe reference genome (ASM294v2.20). Coverage at each base-pair of the *S. cerevisiae* reference genome was calculated as number of reads of given length (25-140 bp for the analyses presented here) that mapped at that position normalized by the factor N:

$$N = \frac{10{,}000}{Total\ number\ of\ reads\ that\ mapped\ to\ S\ pombe\ genome}$$

Here 10,000 is an arbitrarily chosen number. The normalized coverage at each base-pair from each replicate was averaged when combining multiple replicates. TSS positions were obtained from (Park et al., 2014). Only those genes were considered which had no overlap in with other genes in the TSS - 50 bp to TSS + 50% of gene length interval, which resulted in a list of 4808 genes. For quantifying changes in Pol II levels upon 3-IAA addition to various strains, first, total normalized signal per bp in the interval TSS to TSS+100 bp was calculated for each gene in the +DMSO and +3-IAA datasets. Then, a log ratio of +3-IAA to +DMSO was calculated as the net effect of depletion of different factors in the degron strains. For nucleosome profiles shown in Fig. 2, 140-155 bp fragments from published H3 ChIP-seq paired-end data (SRA Accession: SRR1105662) was used and coverage at each position of a gene was normalized to the average reads over TSS ± 1000 bp interval of that gene.

For determining clusters of genes that change concordantly due to Taf degradation across the 5 Taf degron experiments (Taf1, Taf2, Taf7, Taf11, and Taf13), we first converted the log2 changes in spike-normalized Pol2 enrichment to Z-scores using the mean and standard deviation calculated by fitting a Gaussian function to the log2 distributions. We then performed k-means clustering on the Z-scores across five distributions using Cluster 3.0 (de Hoon et al., 2004) using Euclidean distance as the similarity metric and organizing the genes

into 5 clusters. Published Taf1 ChEC-seq data (GEO: GSM2149798) (Grünberg et al., 2016) was used in Figure 4E.

DATA AND SOFTWARE AVAILABILITY

All sequencing datasets have been uploaded in GEO under accession GSE97081. Original Western blot images have been deposited at mendeley.com: http://dx.doi.org/10.17632/gwtwc4ndpj.1

Supplemental information

Figure S1. (related to Figures 1- 4). (A) Shown is the Pearson correlation coefficient of Pol II ChIP score between WT + DMSO, WT + 3-IAA and all Degron strains + DMSO. We observe high correlation across the 4808 gene list that we used for all our analysis, indicating that adding the degron tag did not affect transcription in the absence of 3-IAA. (B) Genome-wide changes in transcription upon Taf1 and Taf2 depletion. Box plots show the change in native Pol II ChIP on a log2 scale from 4807 genes comparing DMSO and 3-IAA addition. WT and Med14-degron data from Fig 3 are shown for comparison.

Figure S2. (related to Figures 2-5). Analysis of changes in nascent mRNA levels upon Mediator and TFIID-depletion.

Figure S3. (related to Figure 6) Impact and depletion efficiency in Taf4 and Taf5 anchor-away strains.

Figure S4. (related to Figure 6). cDTA analyses of Taf4 and Taf5 anchor-away.

Figure S5. (related to Figure 7). Degron efficiency in alternative growth conditions. Western blot of degron efficiency in cells grown under heat shock stress and in synthetic complete media.

Figure S6. (related to Figure 7). Nearly all genes are sensitive to TFIID depletion in cells grown in synthetic media and undergoing heat shock stress.

Figure S7. (related to Fig 7) Comparison of genome-wide transcription after Taf-depletion in different growth conditions.

Table S1. Primers used (related to Method Details: cDNA Synthesis and Quantitative PCR of Nascent RNA).

Table S2. Yeast Strains (related to Method Details: Strain Construction).

Table S3. Summary of Spike-normalized Pol II ChIP changes after 3-IAA addition (Related

to Figures 2-5, 7, S1, S6).

Table S4. TFIID-dependent Gene Clusters (Related to Fig 5).

Table S5. TFIID-dependent GO terms analysis (Related to Fig 5).

Supplemental information

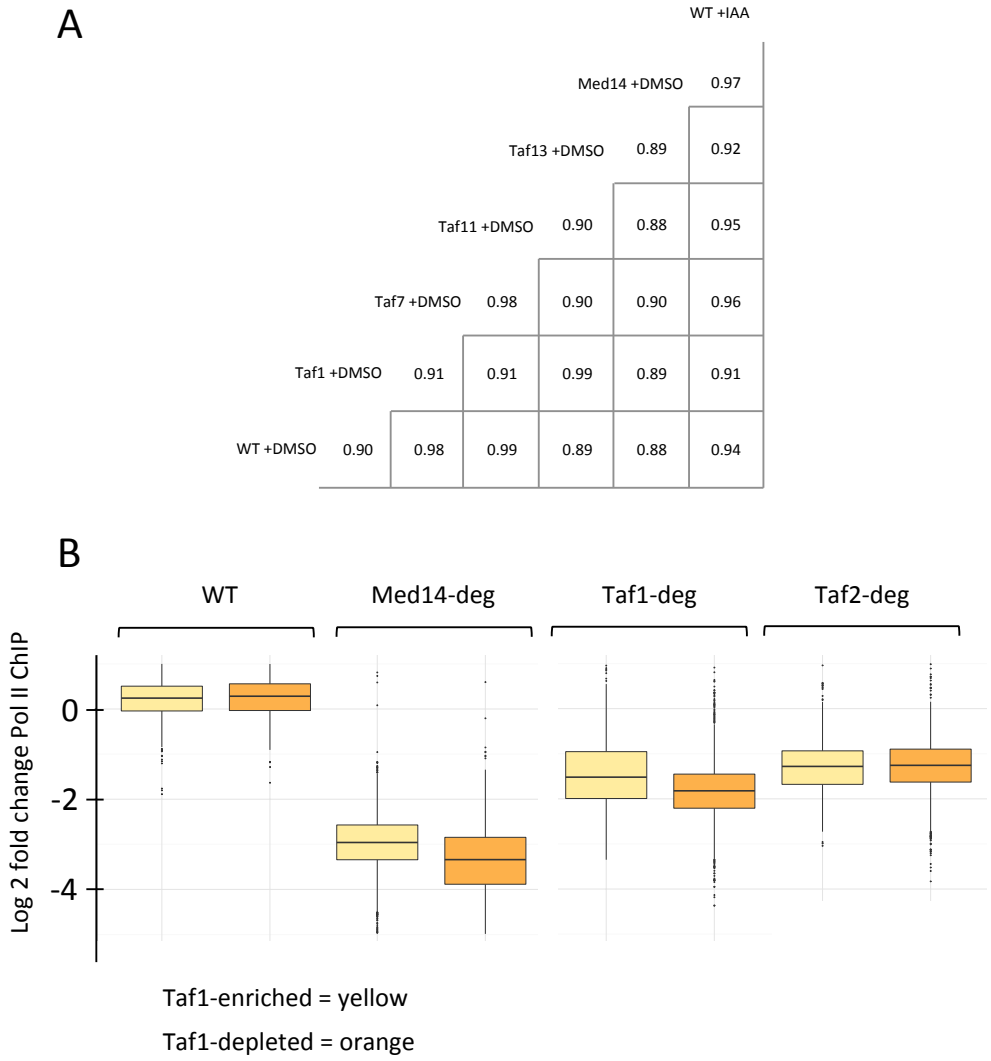


Figure S1 (related to Figures 1- 4). (**A**) Shown is the Pearson correlation coefficient of Pol II ChIP score between WT + DMSO, WT + 3-IAA and all Degron strains + DMSO. We observe high correlation across the 4808 gene list that we used for all our analysis, indicating that adding the degron tag did not affect transcription in the absence of 3-IAA. (**B**) Genome-wide changes in transcription upon Taf1 and Taf2 depletion. Box plots show the change in native Pol II ChIP on a log2 scale from 4807 genes comparing DMSO and 3-IAA addition. WT and Med14-degron data from Fig 3 are shown for comparison.

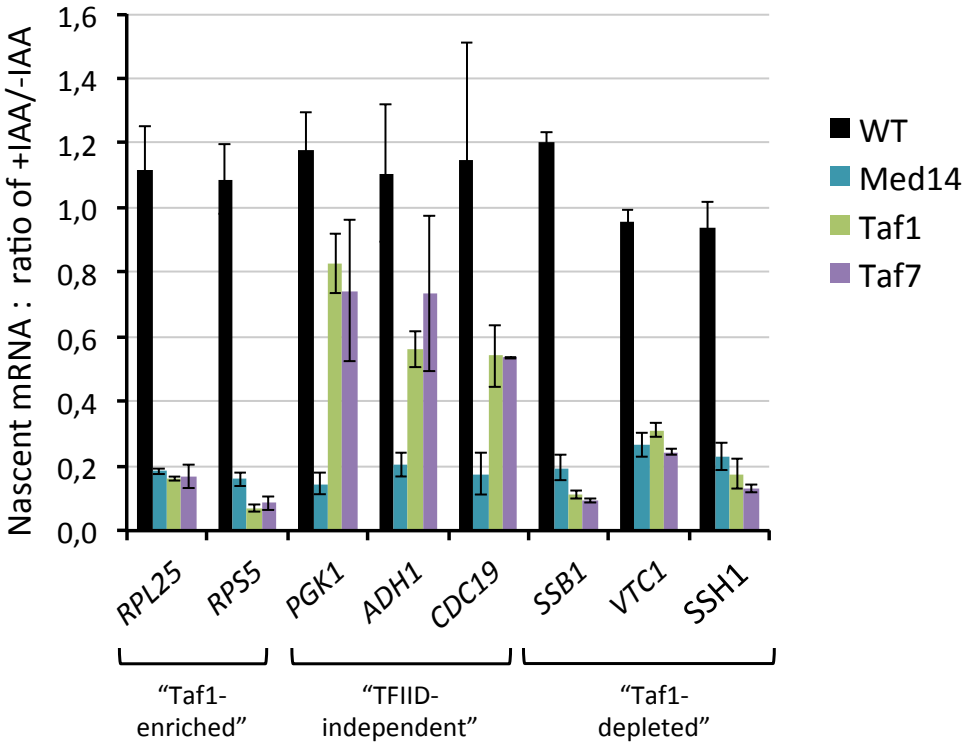


Figure S2 (related to Figures 2-5). Analysis of changes in nascent mRNA levels upon Mediator and TFIID-depletion. RT qPCR analysis of 4-thio Uracil labeled mRNAs purified from cells treated with either DMSO or 3-IAA in the indicated degron-tagged strains. Samples were normalized by spike-in of labeled *S. pombe* cells before mRNA isolation. Genes are grouped by previously published categories (Kuras et al., 2000; Li et al., 2000; Rhee and Pugh, 2012).

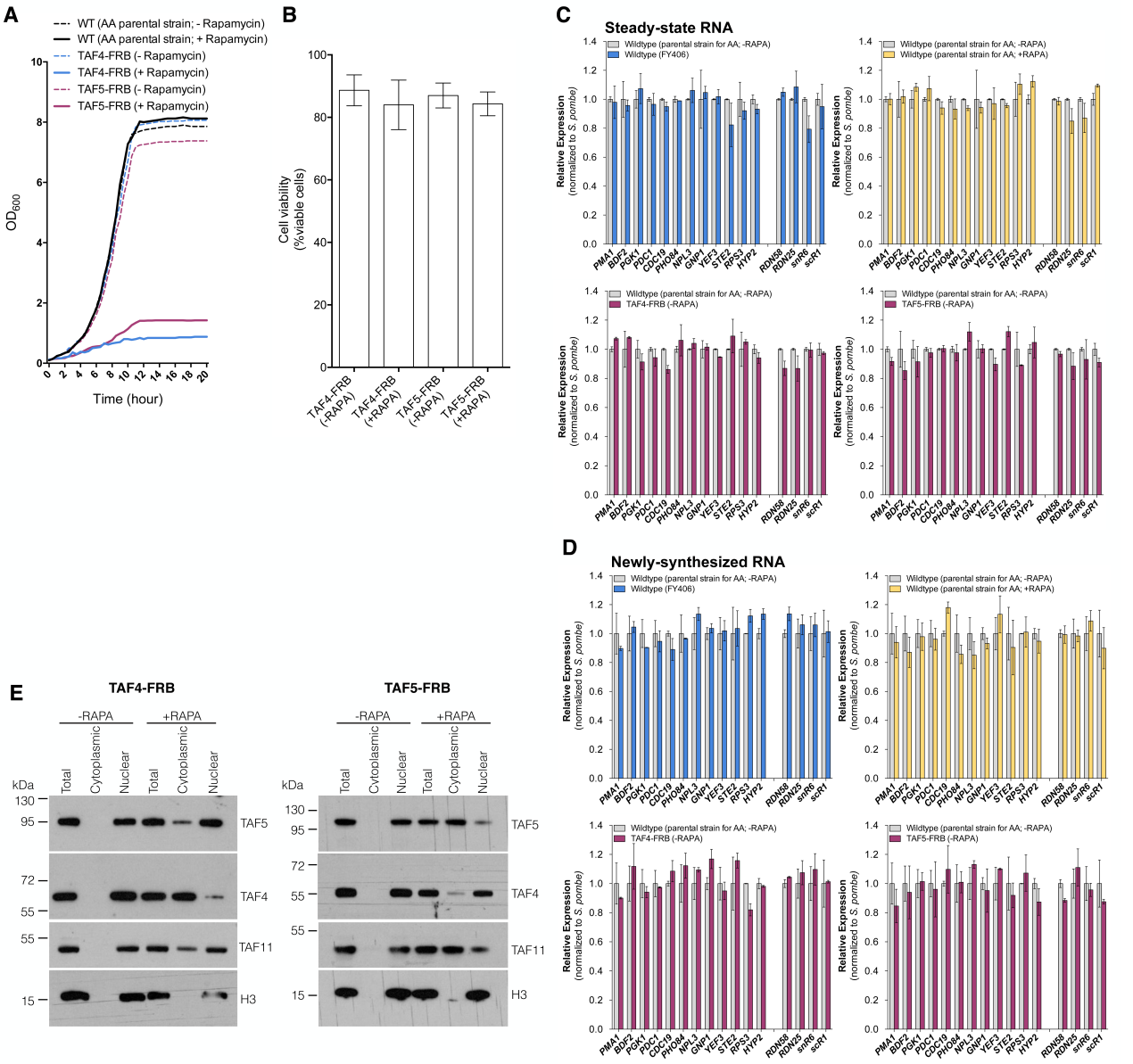


Figure S3. (related to Figure 6) Impact and depletion efficiency in Taf4 and Taf5 anchor-away strains. **(A)** Addition of FRB domain to either Taf4 or Taf5 does not affect growth of these strains in the absence of rapamycin, while long term exposure to rapamycin leads to lethality. **(B)** After 30 min of exposure to rapamycin, viability of cells in log-phase is not affected, in comparison to its counterpart with vehicle only. **(C-D)** Fusion of FRB domain to either Taf4 or Taf5 does not affect RNA Pol II transcription, both at the steady-state (**C**) or newly-synthesized RNA (**D**) levels. **(E)** Cell fractionation depicting efficient nuclear depletion of both Taf4 and Taf5 upon exposure to rapamycin.

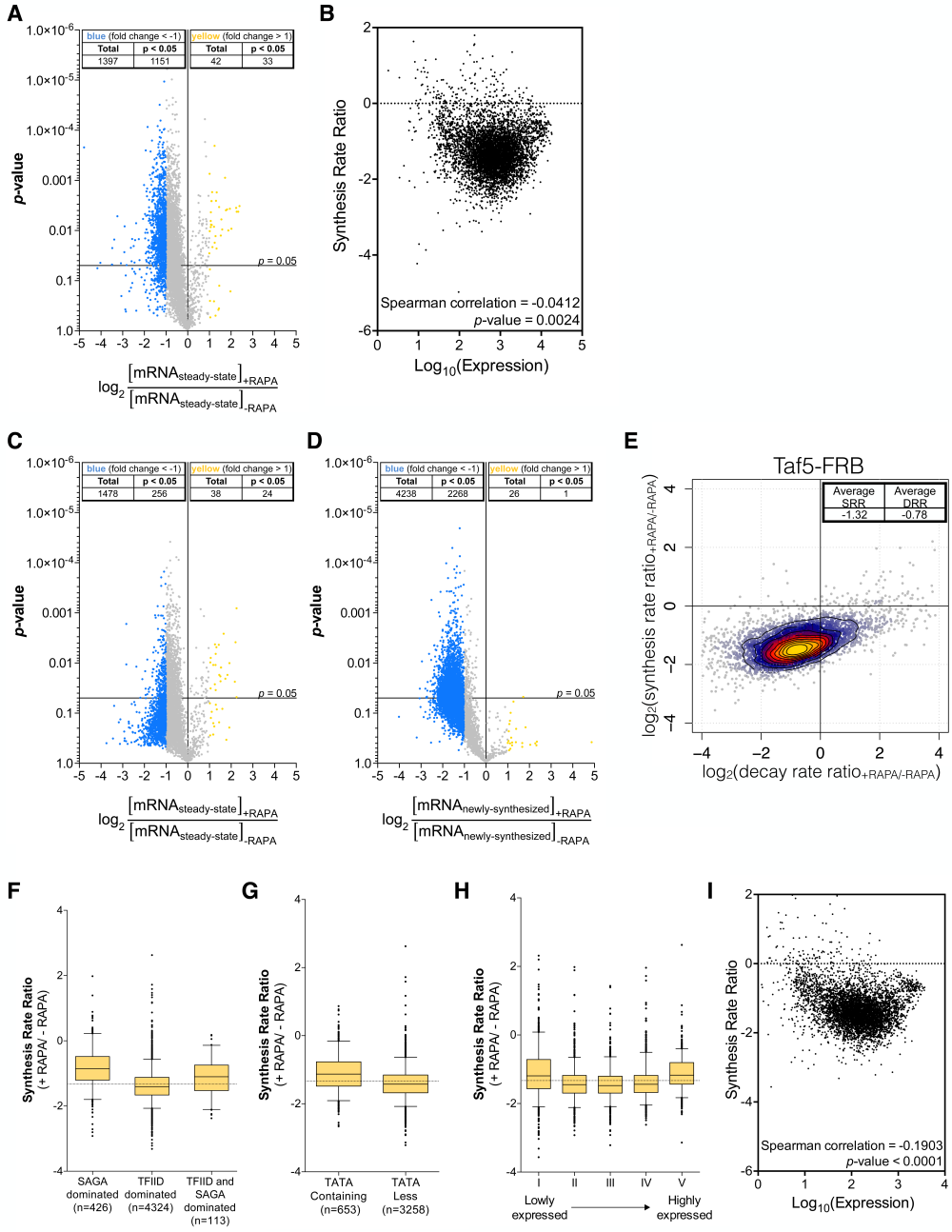


Figure S4 (related to Figure 6). cDTA analyses of Taf4 and Taf5 anchor-away. (**A**) Volcano plot showing changes in steady-state mRNA levels for Taf4 depletion by anchor-away. (**B**) Scatter plot showing a lack of correlation between changes in synthesis rates upon Taf4 depletion and gene expression levels in the wild-type strain. (**C-I**) Nuclear depletion of Taf5 reduced the transcription of a large number of genes. Volcano plots showing changes in steady-state (**C**) and newly-synthesized (**D**) mRNA levels in a *TAF5-FRB* strain treated or untreated with rapamycin. Genome wide changes in mRNA synthesis rates and decay rates upon Taf5 nuclear depletion (**E**). Changes in synthesis rates

according to the gene category were analyzed as in Fig. 6 (**F-I**). All results were obtained from two independent biological replicates.

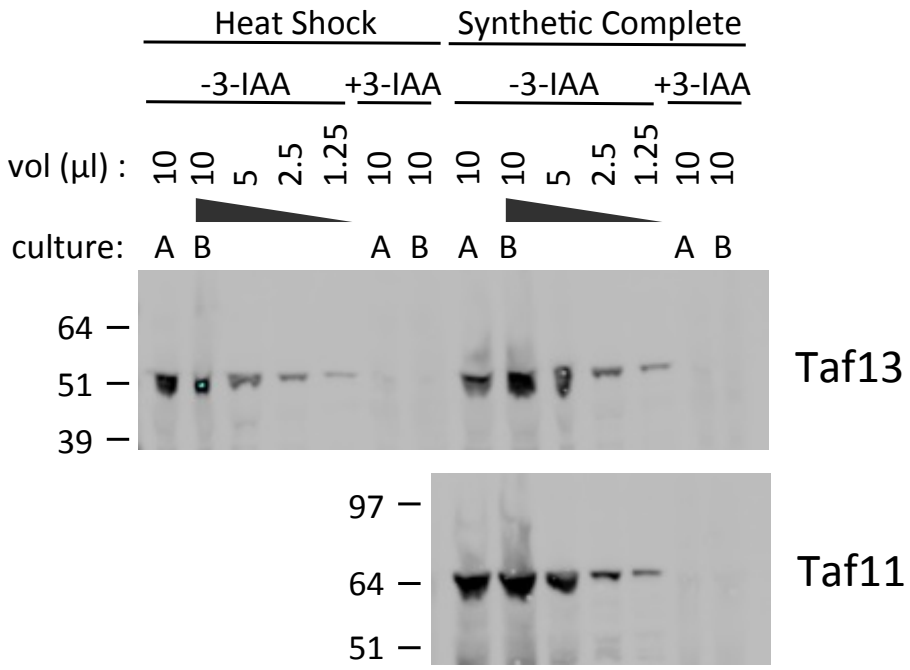


Figure S5 (related to Figure 7). Degron efficiency in alternative growth conditions. Western blot of degron efficiency in cells grown under heat shock stress and in synthetic complete media. Experiment was done in biological duplicate (A and B cultures). Variable amounts of the -3-IAA "B" samples were loaded to allow quantitation of degron efficiency. The positions of molecular weight markers are indicated.

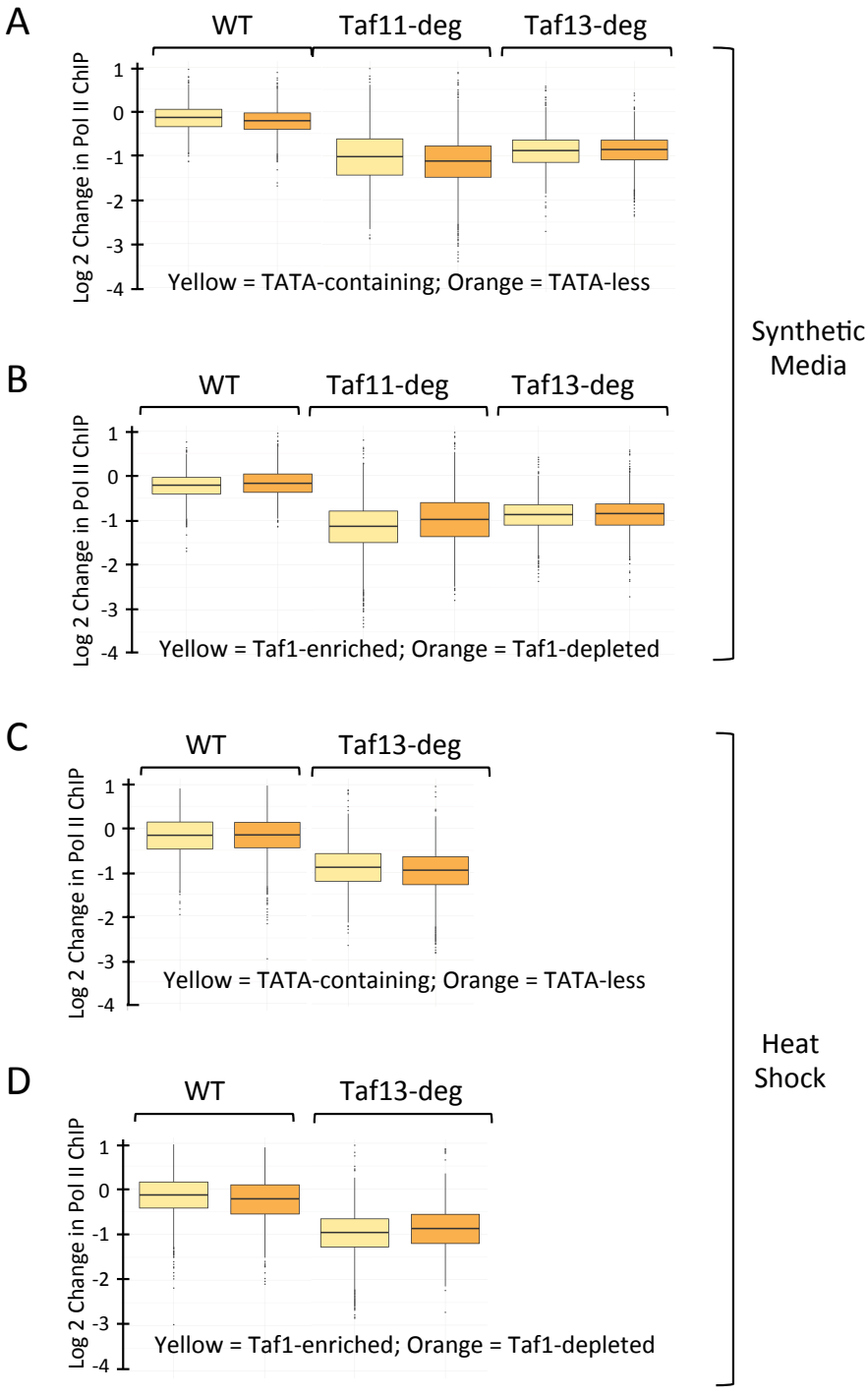


Figure S6 (related to Figure 7). Nearly all genes are sensitive to TFIID depletion in cells grown in synthetic media and undergoing heat shock stress. (**A and B**) Data from cells grown in synthetic media from Fig 7A were split into TATA-containing, TATA-less, "Taf1-enriched" and "Taf1-depleted" gene categories and plotted as in Fig 4. (**C and D**) same as A and B except data from cells under heat shock stress from data in Fig 7B.

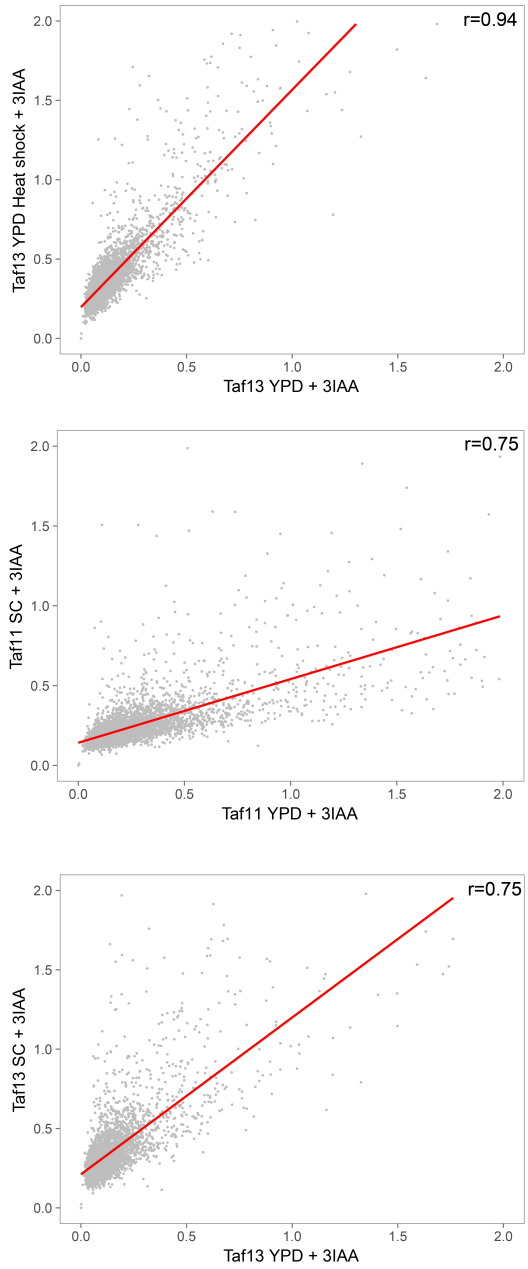


Figure S7 (related to Fig 7) Comparison of genome-wide transcription after Taf-depletion in different growth conditions. Shown in the scatterplot are the spike-normalized Pol II ChIP signals after 3-IAA addition to the indicated degron strains. Pol II ChIP signals from synthetic complete media (SC) and heat shock in YPD are compared to strains grown in YPD at 30 deg.

Table S1. Primers used

Gene	Primer Name	Sequence
RPS5	RPS5_F	GTCGTCTTGGCTACTCCAATTC
	RPS5_R	GTCAACCAAAGAAGCATCCTTAAC
RPL25	RPL25_F	GCTCCATCTGGTATGTGAACTG
	RPL25_R	GCACTATTCTTGTCGTCGGATAG
PGK1	PGK1_F	AGCGTGTCTTCATCAGAG
	PGK1_R	TGGCAAAGCAGCAACAA
ADH1	ADH1_F	CTTCTACGAATCCCACGGTAAG
	ADH1_R	GTGTGACAGACACCAGAGTATT
CDC19	CDC19_F	CAAAGACCAACAACCCAGAAAC
	CDC19_R	GGTATTCGTAAGAACCGTGAGAG
SSB1	SSB1_F	CGTCATTACTGTCCCAGCTTAC
	SSB1_R	GGCAGTAGGTTCGTTGATGATA
SSH1	SSH1_F	CCCAAAGCTACCACACCTAAT
	SSH1_R	ACCCACTAGAAATGTTGGGAAA
VTC1	VTC1_F	TGCCAATGAGCGTACCTTT
	VTC1_R	TGCACTGACCCTACCTATCT
S. pombe tubulin (NDA3)	Tubulin_F	CCGCTGGTGGAAAGTATGTT
	Tubulin_R	GCCAATTCAGCACCTTCAGT

Table S2. Yeast strains

Strain	Background	Description
SHY1058	972h-	*S. pombe*: RPB3-3xFlag::KanMX
SHY1035	BY4705	*S. cerevisiae*: *mat alpha delta ade2::hisG his3 delta 200 leu2 delta 0 lys2 delta 0 met15 delta 0 trp1 delta 63 ura3 delta 0 RPB3-3X Flag::NatMX*
SHY1036	SHY1035	*S. cerevisiae*: *RPB3-3x Flag::NatMX, pGPD1-OSTIR::HIS3 (constitutively expressed OSTIR integrated at HIS3)*
SHY1039	SHY1036	*S. cerevisiae*: *TAF1-3xV5 IAA7::KanMX, RPB3-3x Flag::NatMX, pGPD1-OSTIR::HIS3*
SHY1041	SHY1036	*S. cerevisiae*: *TAF7-3xV5 IAA7::KanMX, RPB3-3x Flag::NatMX, pGPD1-OSTIR::HIS3*
SHY1042	SHY1036	*S. cerevisiae*: *TAF11-3xV5 IAA7::KanMX, RPB3-3x Flag::NatMX, pGPD1-OSTIR::HIS3*
SHY1043	SHY1036	*S. cerevisiae*: *TAF13-3xV5 IAA7::KanMX, RPB3-3x Flag::NatMX, pGPD1-OSTIR::HIS3*
SHY1054	SHY1036	*S. cerevisiae*: *MED14-3xV5 IAA7::KanMX, RPB3-3x Flag::NatMX, pGPD1-OSTIR::HIS3*
MJE110	W303	*S. cerevisiae: MATalpha tor1-1; fpr1::NAT; RPL13A-2xFKB12::TRP1; TAF4-FRB::kanMX6*
MJE112	W303	*S. cerevisiae: MATalpha tor1-1; fpr1::NAT; RPL13A-2xFKB12::TRP1; TAF4-FRB::kanMX6*

4. The genome-wide distributed H2B monoubiquitination plays an impactful, but not global, role in transcription by RNA Polymerase II

Due to the fact that coactivators are recruited to chromatin by specific activators, those complexes were suggested to participate in gene expression control of a small set of genes. Particularly, SAGA was initially thought to be required for the expression of only 10% of the genes in budding yeast, but currently we acknowledge its broader role on RNA Pol II transcription. In fact, in the past, chromatin-modifying enzymes (SAGA and other complexes) have all been implicated to be fundamental for the regulation of the expression of TATA-containing genes (Basehoar *et al.*, 2004; Huisinga and Pugh, 2004; Tirosh and Barkai, 2008). Nevertheless, do coactivators or chromatin modifiers/remodelers have such gene-specific activities or, instead, do they have a broader spectrum of action like SAGA?

For example, considering the genome-wide distribution of H2Bub, deposited by the E2 ubiquitin-conjugating enzyme Rad6 and the E3ubiquitin-ligase Bre1, it would be expected that Rad6 and Bre1 would have a great impact on transcription regulation. First, not only is H2Bub associated with active elongation, but also H2B ubiquitination is required for methylation of H3K4 by Set1/COMPASS and of H3K79 by Dot1 (Dover *et al.*, 2002; Shahbazian *et al.*, 2005). Second, active deposition of ubiquitin onto H2B by Rad6/Bre1 and subsequent removal by SAGA/Ubp8 are intimately associated with proficient transcription (Henry *et al.*, 2003; Kao *et al.*, 2004). Nevertheless, studies performed with steady-state levels of mRNA failed to report a genome-wide role of these ubiquitinating enzymes on transcription (Lenstra *et al.*, 2011). Equally modest effects were observed for mammalian cells through knockdown of RNF20 (RNF20/40 are the human ortholog of Bre1) (Minsky *et al.*, 2008). Instead, only a couple of hundred of genes were affected, either up- or down-regulated.

To better understand the implication of loss of H2Bub, I investigated whether deletion of *BRE1* would have a broader role in transcription in yeast. As a control, I performed a similar analysis in a strain where lysine 123 in histone H2B was mutated to an arginine (*K123R*), hence being non-ubiquitinable. Newly-transcribed RNA was metabolically labeled with 4-tU and steady-state and newly-synthesized RNA were analyzed by hybridization onto chips containing probes for the whole budding yeast transcriptome. Analyses were performed

as described in the two papers part of this thesis (Baptista *et al.*, 2017; Warfield *et al*., 2017) (**Figure 35**). Interestingly, both mutants displayed virtually no changes on transcription when analyzing steady-state levels of RNA (**Figure 35A** and **D**). However, when newly-transcribed RNAs were quantified, changes on transcription were more visible (**Figure 35B** and **E**). In fact, transcriptional effects were more pronounced in the *K123R* mutant than in the *bre1*Δ deletion strain. Through cDTA analysis, we can also conclude that the decrease in the synthesis rates in the mutants is very discreet, just like the compensatory decrease in decay rates (**Figure 35C** and **F**). Altogether, and while preliminary, these results suggest that active deposition of ubiquitin in H2B does not have a great impact on nascent transcription, just like its accumulation upon *UBP8* deletion.

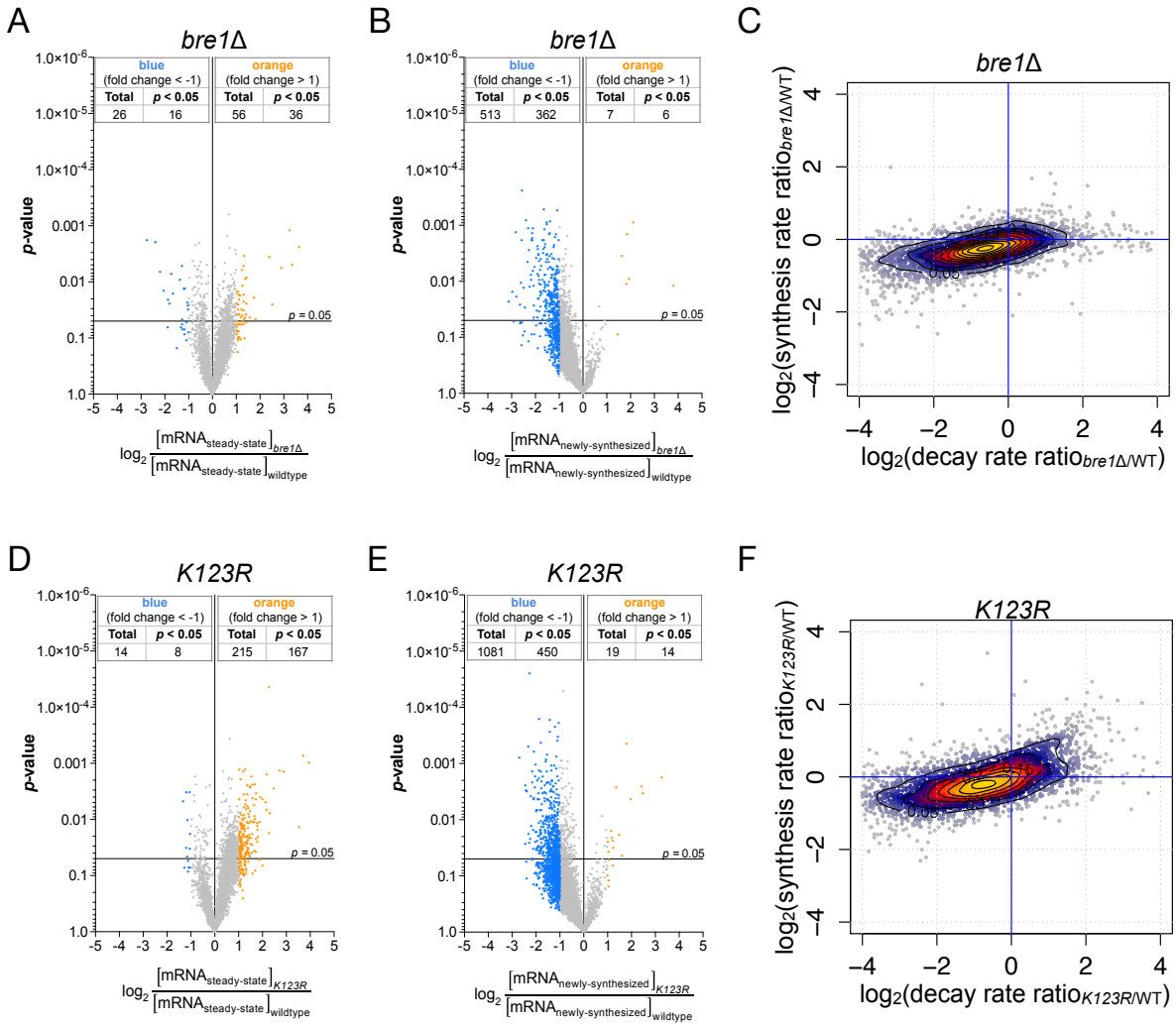


Figure 35 – Analysis of steady-state and nascent RNA Pol II transcription in *BRE1* deletion strain and *K123R* mutant strain. Volcano plots showing fold changes in steady-state mRNA levels **(A,D)** or newly-synthesized mRNA levels **(B,E)** relative to their significance (*p*-value). Fold changes (FC) were calculated as the Log2 of the ratio of the expression value of each gene after normalization to *S. pombe* signal in the *bre1*Δ strain **(A,B)** or the *K123R* **(D,E)** strain versus the expression value of the same gene in wild-type *S. cerevisiae*. cDTA profiles for *bre1*Δ **(C)** and *K123R* **(F)** strains. For all analyzed genes, changes in synthesis rates were plotted against the changes in mRNA decay rates. Changes were calculated as the Log2 of the ratio between mutants and wild-type.

GENERAL DISCUSSION & PERSPECTIVES

General discussion and perspectives

1. A never-ending SAGA: coactivators and general transcription factors joining forces to drive transcription

In the results presented in this thesis, I provide a combination of complementary results indicating that SAGA is recruited at most Pol II-transcribed genes in *S. cerevisiae*, where it plays a critical role in mRNA synthesis. Similar evidence were also found for the transcription factor TFIID. Altogether, the results demonstrated that:

(i) SAGA acetylates H3K9 on most active promoters and deubiquitinates H2B on the transcribed region of expressed genes (Bonnet *et al.*, 2014);

(ii) SAGA is required for normal Pol II recruitment at a vast majority of active promoters (Bonnet *et al.*, 2014);

(iii) SAGA broadly binds to UASs of genes previously categorized as SAGA- or TFIID- dominated (Baptista *et al*., 2017);

(iv) SAGA is required for TBP recruitment at active promoters of both SAGA and TFIID-dominated genes (Baptista *et al*., 2017);

(v) Gcn5 acetyltransferase synergizes with Spt3 (TBP-binding) to promote overall RNA Pol II transcription (Baptista *et al*., 2017);

(vi) Disruption of the SAGA or TFIID complexes led to a global decrease in mRNA synthesis (Baptista *et al*., 2017; Warfield *et al*., 2017);

In remarkable contrast to the previous model for SAGA function, in which SAGA predominantly regulates the expression of only a limited subset of genes, we propose that SAGA and TFIID act as general co-factors for RNA Pol II transcription. Importantly, a more detailed discussion can be found in the *Results* section, where the results are specifically discussed in light of the current knowledge. However, the discussion section below aims to present the results in a more general context and to reveal questions that are still left unanswered.

1.1. SAGA and TFIID: delineating genome-wide RNA Pol II transcription in budding yeast

As the results have shown, there is a genome-wide requirement for SAGA in order for RNA Pol II to function optimally. Nevertheless, transcriptional effects upon SAGA loss are not necessarily equal across the whole genome. In fact, while SAGA-/TFIID-dominated and TATA-containing/-less genes all decrease their expression to the same extent (on average), it is possible to observe that some genes are more affected than others. As depicted in Baptista *et al.*, 2017 (Figure 4 and Figure 6E), while the great majority of the genes are down-regulated to the same extent within the different classes, a dispersion is visible. This clearly indicates that SAGA is required for the expression of all genes, but each gene's requirement is slightly different. The same is true for TFIID (Warfield *et al*., 2017 – Figure 2B, Figure 3 and Figure 6). Importantly, these specifications prove that, while all the genome requires SAGA and TFIID (and Mediator) for optimal transcription, how genes are influenced vary accordingly. Remarkably, the fact that expression is affected differently from gene-to-gene might be due to more than SAGA itself. For instance, promoter architecture and requirements for other coactivators might justify why SAGA activities have slightly different outputs genome-wide. Specifically, to understand whether SAGA coordinates its activities with other coactivators (other than TFIID and Mediator), similar studies could be performed (discussed below).

1.1.1. SAGA in the mammalian system: conservation of genome-wide control of RNA Pol II transcription?

The usage of *S. cerevisiae* or *S. pombe* represents a powerful tool for the understanding of molecular biology events, namely due to their relatively easy manipulation. Nevertheless, while highly conserved complexes from yeast to human might indicate similar activities in the two organisms, evolution allowed the establishment of several differences in the transcriptional environment and the epigenetic landscape. Also, the cellular complexity increased dramatically from yeast to mammals. In the process of evolution, several gene duplications led to the extension of the protein repertoire in mammalian cells, namely when it comes to the SAGA complex. While the SAGA complex is highly conserved in structure and modularity from yeast to mammals, gene duplication led to the diversification of the mHAT of the complex (Gcn5 *versus* PCAF) and diversification of HAT-containing complexes (SAGA *versus* ATAC) (Spedale *et al.*, 2012b). Interestingly, in mammals, the mDUB also

present different flavors: ATXN7 (ySgf11) has two paralogs (ATXN7L1 and ATXN7L2) (Helmlinger *et al.*, 2004) and two other ubiquitin-specific proteases, USP27X and USP51, have been shown to interact with components with the mDUB in a SAGA-independent manner and to compete with the activity of USP22 (yUbp8) for H2B deubiquitination (Atanassov *et al.*, 2016). Altogether, while the knowledge of SAGA in yeast is incredibly rich, the same is not true for mammals, specifically for the poorly characterized ATAC complex. Hence, several questions still need to be addressed: Has the genome-wide role of SAGA been preserved in more complex, multi-cellular organisms? Are SAGA and ATAC HAT activities independent or redundant? If different, what is the molecular basis that distinguishes the activity of GCN5/PCAF within SAGA or ATAC? What are the roles of SAGA and/or ATAC in cell differentiation and development? And the different mDUB? Do they have specific or redundant activities?

In order to verify if SAGA holds a genome-wide impact on RNA Pol II transcription, similar approaches that have been applied in yeast can be used in mammalian cells. For that, cell lines depleted for different subunits of the SAGA complex need to be obtained. Gratefully, genetic manipulation in mammalian cells is much more attainable in our days, through the usage of the CRISPR-Cas9 technology (Doudna and Charpentier, 2014; Hsu *et al.*, 2014). Once the cells are available, one can perform metabolic labeling of newly-transcribed RNA using 4-sU and ascertain the impact on transcription. Also, using the same cellular model, one can evaluate the role of the SAGA complex during cellular differentiation, through the usage of mouse embryonic stem cells lacking SAGA components. Similarly, the same methods can be used for the study of the ATAC complex. Additionally, something that is still not well explored is the role of the SAGA complex in embryonic development. While some studies have made a rough characterization of SAGA subunits' depletion in mouse, the molecular basis behind the phenotypes, all leading to lethality, are poorly identified. This work is currently being put into practice in our lab. Similarly, equivalent studies would be of high relevance to understand the role of Mediator in multicellular eukaryotes and comprehend whether its function is conserved through evolution.

1.2. SAGA recruitment to chromatin

Our studies demonstrate that SAGA and TFIID have comparable roles in transcription independently of the classification of genes (TFIID- *versus* SAGA-dominated; TATA-containing *versus* TATA-less). In addition, ChEC-seq experiments have shown that SAGA is recruited to both TFIID-/SAGA-dominated and TATA-containing/-less genes (Baptista *et al.* 2017 – Figure 1 and Figure 2). However, the genome-wide binding profile of SAGA has its peculiarities. First, the average binding profile of SAGA is virtually the same as that of Mediator, and clearly distinct from TFIID. This indicates that both SAGA and Mediator not only contribute to the expression of the vast majority of RNA Pol II genes, but also bind to the same regulatory regions with equivalent extent. This highlights that both SAGA and Mediator are general cofactors for expression of class II genes. Second, levels of SAGA-directed DNA cleavage – via the fused MNase – were higher at UASs of SAGA-dominated genes compared to TFIID-dominated genes, just as observed for Mediator, but not for TFIID (Grunberg *et al.*, 2016). While one could argue that this is due to higher expression levels of SAGA-dominated genes, this is not true (Churchman and Weissman, 2011; Grunberg *et al.*, 2016). Hence, this increased binding is uncoupled from transcription. Third, the peak summit of average SAGA cleavages relative to the TSS at TATA-containing genes was generally more distal (average 286 bp) than at TATA-less genes (average 217 bp) raising the question why do SAGA and Mediator bind more frequently and distally at TATA-containing genes? While we did not explore this question, one can argue that this occurs due to inherent characteristics of the two different classes of genes, such as promoter architecture. For instance, SAGA-dominated genes/TATA-containing genes have been reported to have a specific promoter architecture that allows for more transcriptional plasticity and more transcriptional noise (Tirosh and Barkai, 2008; Kubik *et al.*, 2015; Kubik *et al.*, 2017). In order to better understand the phenomenon, the following different questions should be addressed, as we can see below.

1.2.1. How do SAGA, Mediator and TFIID cooperate?

From our results, we could observe that SAGA, just like Mediator, binds to UASs, while TFIID binds to promoters. Also, all these three coactivators have a primary role in maintaining normal RNA Pol II transcription. Moreover, as previously mentioned, the binding of SAGA and Mediator differs depending on the existence of a TATA consensus

sequence or not. While one can attribute these differences to general promoter architecture, it would be of high interest to gain more insights regarding this last part, namely how these three complexes coordinate their activities and how they contribute to PIC assembly in different chromatin contexts. Also, to better comprehend their function and how the different activities of the complexes are coordinated, it is essential to gain insight into their 3D structure.

First, it would be important to know the high-resolution structure of these complexes. While one of the main problems with SAGA and Mediator is, for instance, their high-flexibility, some alternative methods could be used to acquire a good structure of both coactivators. One of these strategies is to use an approach like the one used by Cynthia Wolberger's group. In Morgan *et al*. they describe a high-resolution structure of the mDUB of SAGA, through its immobilization/stabilization using a nucleosome containing a non-cleavable monoubiquitinated histone H2B (H2Bub). In this case, the stability of interaction between the module and its substrate was dramatically increased and it was possible to crystallize the purified complex and resolve the structure with a resolution of 3.9Å (Morgan *et al.*, 2016). Though challenging, with this method one might expect that nucleosomes containing non-cleavable H2Bub would capture a specific conformation of SAGA or might stabilize such conformation. Hence cryo-electron microscopy should provide a much better resolution structure of the whole SAGA complex. Additionally, crosslinking studies like the ones performed in the laboratory of Steven Hahn might help elucidate the structure of Mediator and TFIID, similarly to what has been done for SAGA (Han *et al.*, 2014).

Second, *in vitro* studies could be used to understand the interplay between SAGA, TFIID, and Mediator in PIC assembly. How do these complexes interact within the context of the PIC? How do SAGA and Mediator contribute to PIC assembly? In recent years, there have been collective efforts to understand how the PIC is assembled and how its different components interact among each other. In fact, a partial RNA Pol II PIC containing RNA Pol II, TBP, TFIIB, and TFIIF was captured with a core Mediator. However, the tail module of Mediator, the CTD of Rpb1 (RNA Pol II) and other PIC components are missing in this structure (Plaschka *et al.*, 2015; Louder *et al.*, 2016; Plaschka *et al.*, 2016; Nozawa *et al.*, 2017). Also, how does SAGA associate with the PIC and how does this coactivator fit within the current knowledge that we have of PIC structure? Considering this, it is important to include SAGA within structural studies done on the PIC, similarly to Mediator. Also, SAGA and TFIID are both capable of interacting with TBP, although with different strengths. While

SAGA interaction with TBP seems to be relatively weak, it is necessary for RNA Pol II transcription, as presented in this thesis (Baptista *et al*., 2017). So how does SAGA coordinate TBP recruitment with TFIID? Is SAGA responsible for the recruitment of TBP *per se*, or is its main role is to stabilize TBP at promoters, counteracting the activity of Mot1 and NC2 and facilitating PIC assembly?

Third, the current PIC structures were obtained through the assembling of the different components onto naked DNA templates. Since *in vivo* genomic DNA is found within the context of chromatin, wrapped and compacted by nucleosomes, it would be of high interest to perform similar structural studies in this environment. Also, the DNA template used in the structural studies is a TATA-containing DNA stretch. However, our and other results suggest that the PIC structure with SAGA and Mediator is different depending on the existence of a TATA consensus sequence (Rhee and Pugh, 2011; Grunberg *et al.*, 2016). Thus, structural studies performed in a TATA-less DNA template would be extremely important.

Fourth, it is necessary to understand how SAGA (or Mediator) transmits regulatory signals directly to the transcription machinery. As indicated above, both SAGA and Mediator bind to UASs. Still, they have a relevant role in PIC formation, that occur downstream of the UAS, on the core promoter. Also, both SAGA and Mediator were not found to bind at promoters using ChEC-seq, as opposed to TFIID. Thus, how information is transmitted from one region to the other is still unclear. Most likely, this occurs through looping, as observed in high eukaryotes. Nevertheless, together with structural information discussed above, the use of high-resolution chromosome conformation capture might offer answers that will help understand how these two coactivators regulate transcription.

1.2.2. Recruitment of SAGA through Tra1

While the work presented here has allowed a better comprehension of SAGA recruitment, there are still a lot of open questions, particularly regarding the mechanisms through which SAGA is recruited to the UASs of a vast majority of class II genes and whether these different paths might be indicative of how SAGA is differently recruited to distinct sets of genes. Briefly, coactivators have been described to be recruited to chromatin through (i) interactions with gene-specific activators, (ii) interaction with general transcription factors or (iii) through chromatin interacting domains. In *S. cerevisiae*, since

SAGA has been implicated in the regulation of a small subset of genes, it was previously assumed that its preferable mechanism of recruitment to its binding sites would be through interaction with specific activators. Nevertheless, our current knowledge discloses that SAGA is recruited to regulatory regions genome-wide and is responsible for the regulation of essentially all RNA Pol II transcription. Therefore, its recruitment should rely on more general or many different gene-specific mechanisms. Hence, one should aim to understand whether the interaction with activators is indeed the preferable way for SAGA to be recruited or if instead, the complex relies on more generic mechanisms.

In budding yeast, the activator-interacting subunit of SAGA is Tra1. However, Tra1 is a subunit of two distinct co-activators (SAGA and NuA4). Importantly, this subunit is essential for cell survival. In fact, lethality through *Tra1* deletion is likely due to impact on NuA4, since other subunits of this coactivator are also essential (like the acetyltransferase Esa1) and none of the SAGA-specific subunits is required for viability (Allard *et al*., 1999). Additionally, all studies aiming towards the understanding of Tra1 function were not able to unveil which of the two complexes mediates a given effect. Nevertheless, in the fission yeast *S. pombe*, a gene duplication event led to the evolution of two paralogue Tra proteins: Tra1, non-essential and exclusively integrated within SAGA, and Tra2, essential and exclusively integrated into NuA4 (Helmlinger *et al*., 2011). Hence, to try to understand whether SAGA only relies on Tra1 to be recruited to chromatin, we could use a *tra1*Δ *S. pombe* strain and define the transcriptional effects (newly-transcribed and steady-state mRNA) upon loss of this SAGA subunit. Alternatively, another way to tackle this question is by looking at how the recruitment itself is affected upon loss of Tra1 using, for example, ChEC-seq as a method of choice. Two options are possible:

- The same cellular system can be used (S. *pombe* strain deleted for *Tra1* where MNase is fused to a SAGA-specific subunit);
- Or *Tra1* can be conditionally depleted in *S. cerevisiae* (Auxin-inducible-degradation or anchor-away system), in a background where the MNase is fused to a SAGA specific subunit.

Interestingly, this question has been recently addressed in our laboratory. We have analyzed a *S. pombe* mutant strain deleted for Tra1. Previous results have indicated that only a small fraction of genes (32 genes) was dysregulated upon Tra1 loss. Hence, we tested whether the analysis of newly-transcribed RNA would disclose a broader effect on RNA Pol

II transcription. Representative genes, designated "Tra1-dependent" (if steady-state transcription was affected by Tra1 deletion) and "Tra1-independent" (if steady-state transcription was not affected by Tra1 deletion), were selected and tested by RT-qPCR. Importantly, these preliminary results indicated that only the "Tra1-dependent genes" are down-regulated upon Tra1 deletion, both steady-state and newly-transcribed mRNA. Hence, by itself, Tra1 does not explain the global recruitment of SAGA. Nevertheless, this study should be extended, since only a limited number of genes were investigated. Interestingly, from the genes previously identified to be regulated by Tra1 in *S. pombe*, a big proportion are stress-induced genes (Helmlinger *et al.*, 2011). Indeed, Tra1 has been extensively reported in *S. cerevisiae* to interact with stress-response activators (Brown *et al.*, 2001; Huisinga and Pugh, 2004; Breitkreutz *et al.*, 2010). From our results, loss of Tra1 impairs the expression of heat-shock response genes upon heat-shock. Thus, in *S. pombe*, recruitment of SAGA to chromatin through Tra1 might only be pivotal for stress-response genes. However, this should be explored in more detail and for more stresses other than heat-shock.

1.2.3. Recruitment of SAGA through other mechanisms

Altogether, our observations indicate that SAGA is recruited by specific activators (via Tra1) to stress-response genes. On the other hand, "Tra1-independent" genes are not affected by the loss of Tra1. Nevertheless, our results in *S. cerevisiae* indicate that SAGA is globally recruited to the transcribed genome and that is essential for the expression of a vast majority of RNA Pol II genes. While we are not aware whether SAGA inactivation promotes a global decrease in RNA Pol II transcription in fission yeast, the extremely severe phenotype observed in a strain deleted for Spt7 supports that idea. Considering that, SAGA has to be recruited to chromatin through other mechanisms other than interaction with activators.

Other than the possibility of being recruited by Tra1 through interaction with activators, it has been hypothesized that SAGA can be recruited to chromatin through other subunits. For instance, Taf12, a subunit of both SAGA and TFIID, is able to interact with acidic activators like Gcn4 and Gal4, similarly to Tra1 in SAGA (Reeves and Hahn, 2005; Herbig *et al.*, 2010). While it is not known whether this activity of Taf12 is independent of Tra1, one can also imagine that these subunits are redundant. Besides, SAGA can be recruited to chromatin through interaction with histones. In fact, several SAGA subunits possess domains known to recognize and bind specific modification in histones. One example

is the double-Tudor domain of Sgf29, able to interact with H3K4me2/3, a modification catalyzed by Set1 in histones surrounding the promoter of actively transcribed genes (Vermeulen *et al.*, 2010; Bian *et al.*, 2011). Also, Spt7 and Gcn5 each have a bromodomain, capable of interacting with acetylated nucleosomes assisting in the retention of SAGA on chromatin in the absence of activators (Hassan *et al.*, 2002). Hence, in order to have a broader picture of the mechanism behind SAGA recruitment, one plausible procedure is to use ChEC-seq in SAGA mutants lacking one or more of these domains. Specifically, the usage of ChEC-seq is advisable over the more conventional ChIP-seq, due to the reasons described in the *Introduction*.

2. SAGA: nascent RNAs *versus* steady state-transcription

Earlier predictions of the contribution that SAGA has in RNA Pol II transcription suggested that this coactivator complex plays a limited and specialized function. Nevertheless, past studies focused on the quantification of steady-state levels of RNA. As we could observe in the *Introduction* section, steady-state levels do not always reflect the real state of transcription in a specific condition. Due to the interplay between mRNA synthesis and decay, several studies reported that measurement of nascent transcription better represents the role of a factor in transcription (Sun *et al*., 2012; Rodriguez-Molina *et al*., 2016; Baptista *et al*, 2017). In fact, through analyses of newly-synthesized RNA, we have shown that previous reports underestimated the role of the SAGA coactivator complex in the expression of class II genes. Additionally, these studies also disclosed that global perturbations in transcription (or RNA degradation), could lead to a compensatory mechanism. This compensation emerges as a way for the cells to cope with these dramatic alterations: dramatic decrease in the synthesis rate (or degradation) of transcripts is buffered by a simultaneous decrease in the decay (or synthesis) rates, ultimately buffering the levels of RNA. This has been shown for RNA Pol II subunits (Sun *et al.*, 2012), the general coactivator Mediator (Plaschka *et al.*, 2015), the GTF TFIIH (Helenius *et al.*, 2011; Rodriguez-Molina *et al.*, 2016), and elements of the mRNA degradation machinery (Sun *et al.*, 2012; Haimovich *et al.*, 2013b; Sun *et al.*, 2013). Specifically, in our SAGA mutants, such compensatory events were universally observed, thus supporting why previous studies have only disclosed a minor impact on transcription. Additionally, we were also able to prove that transcripts are stabilized in *spt7*Δ and *spt20*Δ mutant strains, as indicated by their longer

half-lives, thus backing up this compensatory mechanism. Nevertheless, this compensation is not always "perfect". For instance, in what we have observed in our TFIID mutants (anchor-away strains for TAF4 and TAF5), decrease in transcription was only accompanied by a sub-optimal decrease in the decay rates, thus resulting in only partial buffering on the transcripts. All considered, a lot of questions are still left unanswered, such as the nature of the compensatory mechanism, how is it triggered and why is it not equally efficient in all the contexts.

2.1 Xrn1 and Ccr4-Not complex as putative mediators of compensatory response

Although the mechanisms behind the compensation observed in SAGA/TFIID mutants remains elusive, one of the primary candidates could be the Ccr4-Not complex. First, as covered in the Introduction section, Ccr4-Not is a deadenylation complex involved in the very first steps of cytoplasmic mRNA degradation (Parker, 2012). Second, it has been reported as being a direct regulator of transcription (Liu *et al.*, 1998; Collart, 2003; Collart and Timmers, 2004; Kruk *et al.*, 2011). Third, past studies have reported a direct and/or genetic interaction between Ccr4-Not subunits and SAGA, Mediator and TFIID (Collart and Timmers, 2004; Villanyi and Collart, 2015). Sun and colleagues postulated that the effectors of such response/feedback loop should be transcriptional inhibitors or degradation enhancers. Hence, since Ccr4-Not complex is a factor that positively acts on transcription, the authors excluded such positive factor as a potential mediator of this compensatory mechanism, in their context (Sun *et al.*, 2012). The same group, together with an independent study, has reported that Xrn1 was the factor involved in this process (Haimovich *et al.*, 2013b; Sun *et al.*, 2013). However, the hypothesis of the Ccr4-Not complex being a mediator in this process should not be excluded prematurely, since it was also shown that a *pop2*Δ strain, a subunit of the Ccr4-Not complex, also revealed incapacity to perfectly buffer mRNA levels (Sun *et al.*, 2013). Hence, both Xrn1 and Ccr4-Not complex are putative candidates to regulate the mRNA compensatory response.

Also, while Ccr4-Not was excluded as being associated with a global participation in a compensatory response to global RNA Pol II transcription abrogation, it cannot be excluded as being involved in this process in a context-dependent manner, such as in the context of SAGA loss. If this scenario is true, one would expect that the absence of either Xrn1 or Pop2, in a background where SAGA subunits are deleted, would lead to lethality, since steady-state

levels of mRNA would not be sustained. Hence, in order to successfully validate this hypothesis, one would have to delete components of Ccr4-Not complex or Xrn1 in a genetic background that allows conditional depletion of Spt7 or Spt20, like the anchor-away strain used in one of the publications (Baptista *et al*, 2017).

2.2. Codon optimality: a universal way of mRNA stability?

The genetic code is degenerate, with 61 nucleotide triplet codons encoding 20 different amino acids (Crick, 1963). As a consequence, all amino acids, excluding methionine and tryptophan, are encoded by two or more "synonymous" codons. The term "codon optimality", which should not be mistaken with codon bias (or codon usage bias, refers to differences in the frequency of occurrence of synonymous codons in coding DNA), describes the differential decoding efficiency for the 61 different codons. At a minimum, codon optimality depends on the abundance of all tRNAs that can recognize a codon, and the nature of the base-pairing interaction between the codon and the tRNAs that recognize it (dos Reis *et al.*, 2004; Bicknell and Ricci, 2017). Briefly, optimal codons, which are recognized by abundant tRNAs and efficiently translated are correlated with mRNA stability, while non-optimal codons, which are recognized by less abundant tRNAs, are correlated with RNA instability (Presnyak *et al.*, 2015). Hence, the higher the content in optimal codons, the more stable the transcript is. During translation, the ribosome slows down when it encounters a non-optimal codon, leading to the destabilization and degradation of the mRNA. While initially the mechanism behind this observation was unknown, recently, the group of Jeff Coller reported that Dhh1, a DEAD box helicase, is involved in the slowing-down of the ribosome, to promote degradation of non-optimal genes (Radhakrishnan *et al.*, 2016). In fact, this mechanism has been proposed to regulate the overall levels of mRNA within the cell.

Now, imagining a situation in which transcription is highly impaired, how could codon optimality play a role on the buffering mechanism? In a situation where RNA Pol II transcription abruptly decreases, prior to compensation, levels of cytoplasmic mRNA decrease, as we could observe in our results, for instance (Baptista *et al*., 2017 – Figure 5 and Figure S4). In this scenario, we can observe:

(i) An overall decrease in the steady-state levels of all mRNA, but tRNA levels are unmodified, resulting in fewer mRNAs being translated into proteins;

(ii) Codons that were non-optimal now become sub-optimal/optimal, due to the relative increase of the amounts of both optimal and non-optimal tRNAs available;

(iii) The ribosome rapidly scans through the transcript, reads the codon and efficiently translates it;

(iv) Transcripts are stabilized through this regulatory mechanism and the degradation of the mRNAs decreases;

(v) When the steady-state mRNAs have again reached stable levels, the non-optimal tRNAs, which were optimal for a short period, go back to their initial non-optimal state.

As a whole, this exciting yet provocative hypothesis offers a great explanation to what we observe in our SAGA and, to a lesser extent, TFIID mutants. To explore this idea several steps should be taken. There are two distinct hypotheses, the first one being that increased tRNAs abundance might occur as a consequence of decreased mRNA levels. For instance, one could determine the total amount of a given (or all) tRNA(s) and measure the ratio of this tRNA relative to the abundance of the corresponding codon. One could thus confirm whether in this situation (loss of SAGA, for example) changes in this ratio occur thus explaining how non-optimal codons would shift to a sub-optimal/optimal stage. The second hypothesis refers to up-regulation of tRNAs that "read" non-optimal codons to maintain the stability of the diminished amount of transcripts. Interestingly, it has been shown that in a breast cancer cell line, overexpression of two rare tRNAs (Arg-CCG and Glu-UUC) triggered stabilization of the correspondent mRNAs containing those codons (Gingold *et al.*, 2014). Eventually, one can speculate that coordinated expression of mRNAs and tRNAs might act synergistically to control transcript levels. So, instead of increased relative abundance of mRNAs as a response to decreased transcription (first hypothesis), we can observe, instead, an increased expression of tRNAs to stabilize the less abundant transcripts. This slightly different conclusion indicates that increased optimality is a primary event/response, and not a consequence of decreased levels of RNA, as the first hypothesis suggests.

In our SAGA mutants, while most of the transcripts are compensated with an equivalent decrease in transcription and decay, some escapees are visible (Baptista *et al.*, 2017 – Figure 3, Figure 6 and Figure S5). These transcripts are either overcompensated (decrease in decay rates far exceed the decrease in the synthesis rates - overstabilization) or undercompensated (decrease in decay rates is not enough to match the decrease in the

synthesis rates - understabilization). Actually, one can find the explanation for this in light of the codon optimality hypothesis: genes that are overcompensated could be those containing a high percentage of optimal codons; genes undercompensated could be those enriched for non-optimal/sub-optimal codons. In order to validate if this is the case, we could simply select those two distinct groups of genes and determine their composition in terms of optimal/non-optimal codons. After that, and comparing with the composition of genes that are "perfectly" buffered, we can determinate whether there is a bias, or not, for codon optimality.

Nevertheless, there are some experimental obstacles and missing links in this hypothesis, considering our specific context. Indeed, there is one major experimental hurdle: quantification of tRNAs abundance. While most of the studies indirectly quantify tRNA abundance relying on the number of copies of a given tRNA in the genome, this was proven to be misleading (dos Reis *et al.*, 2004; Parmley and Huynen, 2009). Also, tRNAs are heavily modified and present a complex secondary structure, which was proven to alter the sequencing efficiency. Nevertheless, in recent years new approaches for tRNA sequencing have been developed and can now be applied in several and specific contexts (Cozen *et al.*, 2015; Zheng *et al.*, 2015). Besides experimental difficulties, there is a potential flaw in this hypothesis when it comes to our observations in SAGA mutants, but also in the previous reports for other complexes: codon optimality just answers to changes in mRNA stability in a situation when transcription is globally affected. However, as we could notice for degradation complexes, such as Ccr4-Not complex, a similar mechanism of RNA buffering is observed (Sun *et al.*, 2012). If one would expect a universal mechanism for increased RNA stability in a situation of defective synthesis or decay, one could expect the mechanism to work in both directions: to decrease decay when transcription is affected or decrease synthesis when decay is impaired. In regards to that, the hypotheses that Ccr4-Not or Xrn1 are involved in this mechanism are more intuitive. Nevertheless, one cannot exclude that the mechanism regulating compensation after transcription abrogation is seemingly different from that regulating buffering upon decay impairment. Hence, there could be two different mechanisms to answer to the two distinct situations: one to compensate changes in transcription and another independent one to compensate changes in decay.

3. Taking a detour on this SAGA: role of other coactivators on transcription by RNA Pol II

Due to the fact that coactivators are recruited to chromatin by specific activators, those complexes were suggested to participate in gene expression control of a small set of genes. Particularly, SAGA was initially thought to be required for the expression of only 10% of the genes in budding yeast, but currently we acknowledge its broader role on RNA Pol II transcription. In fact, in the past, chromatin-modifying enzymes (SAGA and other complexes) have all been implicated to be fundamental for the regulation of the expression of TATA-containing genes (Basehoar *et al.*, 2004; Huisinga and Pugh, 2004; Tirosh and Barkai, 2008). Nevertheless, do coactivators or chromatin modifiers/remodelers have such gene-specific activities or, instead, do they have a broader spectrum of action like SAGA?

Altogether, and while preliminary, these results suggest that active deposition of ubiquitin in H2B does not have a great impact on nascent transcription, just like its accumulation upon *UBP8* deletion. Yet, this mark is highly associated with active transcription. In the *bre1*Δ and the *K123R* mutant strains, the levels of H3K4me3 and H3K79me3 are undetectable and, in good agreement, H3K4me3, deposited by Set1/COMPASS complex, was found to have little to no effects on nascent transcription (Howe *et al.*, 2017). While this kind of study should be extended to other chromatin-modifying/-remodeling complexes, this lack of association of histone marks with transcription is highly interesting. If histone modifications found within transcribed regions are not associated with transcription, what is their function? Eventually, they can be involved in epigenetic memory, like it has been reported for several organisms, or chromatin bookmarking, to facilitate recognition of transcribed *loci* after cell division (Zaidi *et al.*, 2011; D'Urso and Brickner, 2014; Allis and Jenuwein, 2016; Liu *et al.*, 2017). Alternatively, a specific combination of histone marks might be required for efficient transcription explaining why changes in a single mark might have little effects on transcription. In agreement with this hypothesis, we observed that RNA pol II transcription was more decreased in the *gcn5*Δ*ubp8*Δ strain than in the single *gcn5*Δ mutant, although *UBP8* deletion alone does not have any measurable transcriptional defect.

As discussed above, recruitment of coactivators, like SAGA and Mediator, occur genome-wide. However, the pattern of binding is different depending on whether there is a

TATA-box or not. While above we discussed alternatives to understand this question, such as structural-based analyses of PIC-Mediator or PIC-SAGA complex, they are mainly *in vitro*. While *in vitro* experiments are a tremendously rich source of information, ideally one could study the differential recruitment of SAGA and Mediator (and other coactivators, chromatin remodelers/modifiers) in an *in vivo* system. Here, differential recruitment does not relate to recruitment itself, just like it has been addressed in our and other past works, but instead refers to the determinants that lead to a different binding pattern of SAGA/Mediator at TATA-containing/TATA-less genes. Using synthetic chromosomes, it is possible to engineer a stretch of DNA containing one gene (or more) together with its regulatory regions. By changing the composition of the regulatory region and promoter (such as swapping a TATA-box for a TATA-less promoter, insert a large NFR or remove it, etc…), one can understand how these alterations would impact on SAGA and Mediator recruitment and disclose the determinants of differential binding of coactivators complexes.

Since coactivators are highly conserved in distinct organisms and assuming that their function is conserved throughout evolution, one could engineer a synthetic chromosome containing a gene (and its regulatory region) belonging to *S. pombe* and transform it onto *S. cerevisiae*. Like this, we could consider this sequence as naïve, since it does not belong to that organism, and likely this could result in a less biased studying of the mechanism. In fact, this approach does not need to be restricted to SAGA and Mediator. Upon evaluation of other coactivators on transcription (analysis of nascent transcription upon disruption of the complexes and an evaluation of genome-wide recruitment through ChEC-seq), one can aim to better understand the complex network of coactivators in gene regulation. In fact, these analyses should be expanded to corepressors, considering that they potentially act in a genome-wide fashion. Alternatively, instead of using synthetic chromosomes, it is also possible to use engineered mini-chromosomes. These mini-chromosomes are chromatinized, with well-established nucleosome positioning, efficiently and easily purified from budding yeast and can be used for several downstream applications, such as mass-spectrometry or sequencing. Moreover, this system can also be used for *in vitro* studies (Unnikrishnan *et al.*, 2010; Unnikrishnan *et al.*, 2012).

CONCLUSIONS

Conclusions

During my thesis, I have explored the differential requirement of both SAGA and TFIID for RNA Pol II transcription in *S. cerevisiae*. Using a plethora of experimental approaches, my collaborators and I have particularly dissected the mechanisms of PIC assembly and transcription initiation by RNA Pol II, thus clarifying mixed messages from prior studies.

Through analysis of chromatin marks and their dependency on the SAGA complex, we have characterized the broad enzymatic activities of the coactivators throughout the whole RNA Pol II-transcribed genome. Additionally, we have also shown that the integrity of the SAGA complex is essential for optimal RNA Pol II. Moreover, via ChEC-seq, SAGA recruitment we have undoubtedly elucidated that SAGA is recruited to the regulatory regions of genes in a genome-wide manner.

Additionally, by studying nascent transcription, instead of only steady-state levels of mRNA, we have disclosed that upon SAGA loss (via deletions of its structural subunits), mRNA synthesis rate is highly impaired. Not only that, but decay rates of mRNAs are simultaneously decreased leading to the buffering of total levels of mRNA. This observation justifies why past results have underestimated the role of the SAGA complex in RNA Pol II transcription. Importantly, we have also reported that transcription control by SAGA is achieved by the synergistic activities of its Gcn5 and Spt3 subunits, responsible for histone acetylation and TBP-binding, respectively. Finally, through the analysis of the transcriptional complex TFIID, we have disclosed that disruption of the GTF leads to the overall decrease in RNA Pol II recruitment at transcribed genes and impaired nascent transcription.

Altogether, these results challenge prior models of PIC assembly that stated that TBP delivery at promoters would either depend on SAGA or TFIID, thus leading to the nomenclature of SAGA-dominated and TFIID-dominated genes. Here, I show that differential gene regulation, largely attributed to either SAGA or TFIID dominancy, is not appropriate, but instead depends on other features of gene promoters, thus finally clarifying a longstanding conundrum.

BIBLIOGRAPHY

Bibliography

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

Akoulitchev, S., Makela, T.P., Weinberg, R.A., and Reinberg, D. (1995). Requirement for TFIIH kinase activity in transcription by RNA polymerase II. Nature *377*, 557-560.

Alekseev, S., Nagy, Z., Sandoz, J., Weiss, A., Egly, J.M., Le May, N., and Coin, F. (2017). Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression. Molecular cell *65*, 504-514 e504.

Allard, S., Utley, R.T., Savard, J., Clarke, A., Grant, P., Brandl, C.J., Pillus, L., Workman, J.L., and Cote, J. (1999). NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. The EMBO journal *18*, 5108-5119.

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nature reviews Genetics *17*, 487-500.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999). The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. Genes & development *13*, 2148-2158.

Anderson, J.S., and Parker, R.P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. The EMBO journal *17*, 1497-1506.

Andrau, J.C., van de Pasch, L., Lijnzaad, P., Bijma, T., Koerkamp, M.G., van de Peppel, J., Werner, M., and Holstege, F.C. (2006). Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. Molecular cell *22*, 179-192.

Ansari, S.A., He, Q., and Morse, R.H. (2009). Mediator complex association with constitutively transcribed genes in yeast. Proceedings of the National Academy of Sciences of the United States of America *106*, 16734-16739.

Ansari, S.A., Ganapathi, M., Benschop, J.J., Holstege, F.C., Wade, J.T., and Morse, R.H. (2012). Distinct role of Mediator tail module in regulation of SAGA-dependent, TATA-containing genes in yeast. The EMBO journal *31*, 44-57.

Armache, K.J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America *100*, 6964-6968.

Atanassov, B.S., Evrard, Y.A., Multani, A.S., Zhang, Z., Tora, L., Devys, D., Chang, S., and Dent, S.Y. (2009). Gcn5 and SAGA regulate shelterin protein turnover and telomere maintenance. Molecular cell *35*, 352-364.

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. Molecular cell *62*, 558-571.

Auble, D.T., and Hahn, S. (1993). An ATP-dependent inhibitor of TBP binding to DNA. Genes & development *7*, 844-856.

Auble, D.T., Wang, D., Post, K.W., and Hahn, S. (1997). Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. Molecular and cellular biology *17*, 4842-4851.

Babour, A., Shen, Q., Dos-Santos, J., Murray, S., Gay, A., Challal, D., Fasken, M., Palancade, B., Corbett, A., Libri, D.*, et al.* (2016). The Chromatin Remodeler ISW1 Is a Quality Control Factor that Surveys Nuclear mRNP Biogenesis. Cell *167*, 1201-1214 e1215.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L.*, et al.* (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Molecular cell *32*, 878-887.

Baek, H.J., Kang, Y.K., and Roeder, R.G. (2006). Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. The Journal of biological chemistry *281*, 15172-15181.

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. Molecular cell.

Barlev, N.A., Emelyanov, A.V., Castagnino, P., Zegerman, P., Bannister, A.J., Sepulveda, M.A., Robert, F., Tora, L., Kouzarides, T., Birshtein, B.K.*, et al.* (2003). A novel human Ada2 homologue functions with Gcn5 or Brg1 to coactivate transcription. Molecular and cellular biology *23*, 6944-6957.

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.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends in genetics : TIG *32*, 29-41.

Belotserkovskaya, R., Sterner, D.E., Deng, M., Sayre, M.H., Lieberman, P.M., and Berger, S.L. (2000). Inhibition of TATA-binding protein function by SAGA subunits Spt3 and Spt8 at Gcn4-activated promoters. Molecular and cellular biology *20*, 634-647.

Berger, S.L., Cress, W.D., Cress, A., Triezenberg, S.J., and Guarente, L. (1990). Selective inhibition of activated but not basal transcription by the acidic activation domain of VP16: evidence for transcriptional adaptors. Cell *61*, 1199-1208.

Berger, S.L., Pina, B., Silverman, N., Marcus, G.A., Agapite, J., Regier, J.L., Triezenberg, S.J., and Guarente, L. (1992). Genetic isolation of ADA2: a potential transcriptional adaptor required for function of certain acidic activation domains. Cell *70*, 251-265.

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

Bhaumik, S.R., and Green, M.R. (2001). SAGA is an essential in vivo target of the yeast acidic activator Gal4p. Genes & development *15*, 1935-1945.

Bian, C., Xu, C., Ruan, J., Lee, K.K., Burke, T.L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B.O.*, et al.* (2011). Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. The EMBO journal *30*, 2829-2842.

Bicknell, A.A., and Ricci, E.P. (2017). When mRNA translation meets decay. Biochemical Society transactions *45*, 339-351.

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.

Bleichenbacher, M., Tan, S., and Richmond, T.J. (2003). Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. Journal of molecular biology *332*, 783-793.

Boeck, R., Tarun, S., Jr., Rieger, M., Deardorff, J.A., Muller-Auer, S., and Sachs, A.B. (1996). The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. The Journal of biological chemistry *271*, 432-438.

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

Boeing, S., Rigault, C., Heidemann, M., Eick, D., and Meisterernst, M. (2010). RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. The Journal of biological chemistry *285*, 188-196.

Bonnet, J., Wang, Y.H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W., Timmers, H.T., Tora, L., Devys, D.*, et al.* (2010). The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. EMBO reports *11*, 612-618.

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

Borde, V., Robine, N., Lin, W., Bonfils, S., Geli, V., and Nicolas, A. (2009). Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. The EMBO journal *28*, 99-111.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes & development *20*, 1123-1136.

Brand, M., Leurent, C., Mallouh, V., Tora, L., and Schultz, P. (1999a). Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. Science *286*, 2151-2153.

Brand, M., Yamamoto, K., Staub, A., and Tora, L. (1999b). Identification of TATA-binding protein-free TAFII-containing complex subunits suggests a role in nucleosome acetylation and signal transduction. The Journal of biological chemistry *274*, 18285-18289.

Braun, K.A., and Young, E.T. (2014). Coupling mRNA synthesis and decay. Molecular and cellular biology *34*, 4078-4087.

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A., and Choder, M. (2011). Promoter elements regulate cytoplasmic mRNA decay. Cell *147*, 1473-1483.

Bregman, D.B., Halaban, R., van Gool, A.J., Henning, K.A., Friedberg, E.C., and Warren, S.L. (1996). UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. Proceedings of the National Academy of Sciences of the United States of America *93*, 11586-11590.

Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.Y., Breitkreutz, B.J., Stark, C., Liu, G.*, et al.* (2010). A global protein kinase and phosphatase interaction network in yeast. Science *328*, 1043-1046.

Brown, C.E., and Sachs, A.B. (1998). Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. Molecular and cellular biology *18*, 6548-6559.

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 *292*, 2333-2337.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.

Bryant, G.O., and Ptashne, M. (2003). Independent recruitment in vivo by Gal4 of two complexes required for transcription. Molecular cell *11*, 1301-1309.

Bucheli, M., and Sweder, K. (2004). In UV-irradiated Saccharomyces cerevisiae, overexpression of Swi2/Snf2 family member Rad26 increases transcription-coupled repair and repair of the non-transcribed strand. Molecular microbiology *52*, 1653-1663.

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

Burke, T.L., Miller, J.L., and Grant, P.A. (2013). Direct inhibition of Gcn5 protein catalytic activity by polyglutamine-expanded ataxin-7. The Journal of biological chemistry *288*, 34266-34275.

Burley, S.K., and Roeder, R.G. (1996). Biochemistry and structural biology of transcription factor IID (TFIID). Annual review of biochemistry *65*, 769-799.

Bushnell, D.A., and Kornberg, R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. Proceedings of the National Academy of Sciences of the United States of America *100*, 6969-6973.

Cabart, P., Ujvari, A., Pal, M., and Luse, D.S. (2011). Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. Proceedings of the National Academy of Sciences of the United States of America *108*, 15786-15791.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193-198.

Carey, M., Li, B., and Workman, J.L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. Molecular cell *24*, 481-487.

Carroll, J.S., Munchel, S.E., and Weis, K. (2011). The DExD/H box ATPase Dhh1 functions in translational repression, mRNA decay, and processing body dynamics. The Journal of cell biology *194*, 527-537.

Carrozza, M.J., John, S., Sil, A.K., Hopper, J.E., and Workman, J.L. (2002). Gal80 confers specificity on HAT complex interactions with activators. The Journal of biological chemistry *277*, 24648-24652.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P.*, et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Castells-Roca, L., Garcia-Martinez, J., Moreno, J., Herrero, E., Belli, G., and Perez-Ortin, J.E. (2011). Heat shock response in yeast involves changes in both transcription rates and mRNA stabilities. PloS one *6*, e17272.

Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., and Young, R.A. (2001). Remodeling of yeast genome expression in response to environmental changes. Molecular biology of the cell *12*, 323-337.

Cereghino, G.P., Atencio, D.P., Saghbini, M., Beiner, J., and Scheffler, I.E. (1995). Glucose-dependent turnover of the mRNAs encoding succinate dehydrogenase peptides in Saccharomyces cerevisiae: sequence elements in the 5' untranslated region of the Ip mRNA play a dominant role. Molecular biology of the cell *6*, 1125-1143.

Cereghino, G.P., and Scheffler, I.E. (1996). Genetic analysis of glucose regulation in saccharomyces cerevisiae: control of transcription versus mRNA turnover. The EMBO journal *15*, 363-374.

Chambers, R.S., Wang, B.Q., Burton, Z.F., and Dahmus, M.E. (1995). The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II

and by the general transcription factors IIF and IIB. The Journal of biological chemistry *270*, 14962-14969.

Chapman, R.D., Palancade, B., Lang, A., Bensaude, O., and Eick, D. (2004). The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability. Nucleic acids research *32*, 35-44.

Chen, C.Y., Chang, C.C., Yen, C.F., Chiu, M.T., and Chang, W.H. (2009). Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. Proceedings of the National Academy of Sciences of the United States of America *106*, 127-132.

Cheng, B., and Price, D.H. (2007). Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. The Journal of biological chemistry *282*, 21901-21912.

Chereji, R.V., Ocampo, J., and Clark, D.J. (2017). MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers. Molecular cell *65*, 565-577 e563.

Cheung, A.C., and Cramer, P. (2012). A movie of RNA polymerase II transcription. Cell *149*, 1431-1437.

Chi, Y., Huddleston, M.J., Zhang, X., Young, R.A., Annan, R.S., Carr, S.A., and Deshaies, R.J. (2001). Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. Genes & development *15*, 1078-1092.

Chiang, C.M., and Roeder, R.G. (1995). Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. Science *267*, 531-536.

Chicca, J.J., 2nd, Auble, D.T., and Pugh, B.F. (1998). Cloning and biochemical characterization of TAF-172, a human homolog of yeast Mot1. Molecular and cellular biology *18*, 1701-1710.

Cho, E.J., and Buratowski, S. (1999). Evidence that transcription factor IIB is required for a post-assembly step in transcription initiation. The Journal of biological chemistry *274*, 25807-25813.

Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes & development *15*, 3319-3329.

Chowdhury, A., Mukhopadhyay, J., and Tharun, S. (2007). The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. Rna *13*, 998-1016.

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

Ciurciu, A., Komonyi, O., and Boros, I.M. (2008). Loss of ATAC-specific acetylation of histone H4 at Lys12 reduces binding of JIL-1 to chromatin and phosphorylation of histone H3 at Ser10. Journal of cell science *121*, 3366-3372.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annual review of biochemistry *78*, 273-304.

Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R., and Boothroyd, J.C. (2005). Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nature biotechnology *23*, 232-237.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A., and Johnston, M. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science *301*, 71-76.

Coleman, R.A., Taggart, A.K., Burma, S., Chicca, J.J., 2nd, and Pugh, B.F. (1999). TFIIA regulates TBP and TFIID dimers. Molecular cell *4*, 451-457.

Collart, M.A. (1996). The NOT, SPT3, and MOT1 genes functionally interact to regulate transcription at core promoters. Molecular and cellular biology *16*, 6668-6676.

Collart, M.A. (2003). Global control of gene expression in yeast by the Ccr4-Not complex. Gene *313*, 1-16.

Collart, M.A., and Timmers, H.T. (2004). The eukaryotic Ccr4-not complex: a regulatory platform integrating mRNA metabolism with cellular signaling pathways? Progress in nucleic acid research and molecular biology *77*, 289-322.

Coller, J., and Parker, R. (2004). Eukaryotic mRNA decapping. Annual review of biochemistry *73*, 861-890.

Coller, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. Cell *122*, 875-886.

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

Conaway, R.C., Garrett, K.P., Hanley, J.P., and Conaway, J.W. (1991). Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors alpha and beta gamma promote entry of polymerase into the preinitiation complex. Proceedings of the National Academy of Sciences of the United States of America *88*, 6205-6209.

Cozen, A.E., Quartley, E., Holmes, A.D., Hrabeta-Robinson, E., Phizicky, E.M., and Lowe, T.M. (2015). ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. Nature methods *12*, 879-884.

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

Crick, F.H. (1963). On the genetic code. Science *139*, 461-464.

D'Urso, A., and Brickner, J.H. (2014). Mechanisms of epigenetic memory. Trends in genetics : TIG *30*, 230-236.

Dahan, O., Gingold, H., and Pilpel, Y. (2011). Regulatory mechanisms and networks couple the different phases of gene expression. Trends in genetics : TIG *27*, 316-322.

Das, S., Sarkar, D., and Das, B. (2017). The interplay between transcription and mRNA degradation in Saccharomyces cerevisiae. Microbial cell *4*, 212-228.

Daugeron, M.C., Mauxion, F., and Seraphin, B. (2001). The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. Nucleic acids research *29*, 2448-2455.

Davidson, I. (2003). The genetics of TBP and TBP-related factors. Trends in biochemical sciences *28*, 391-398.

Deluen, C., James, N., Maillet, L., Molinete, M., Theiler, G., Lemaire, M., Paquet, N., and Collart, M.A. (2002). The Ccr4-not complex and yTAF1 (yTaf(II)130p/yTaf(II)145p) show physical and functional interactions. Molecular and cellular biology *22*, 6735-6749.

Denis, C.L., and Chen, J. (2003). The CCR4-NOT complex plays diverse roles in mRNA metabolism. Progress in nucleic acid research and molecular biology *73*, 221-250.

Deshmukh, M.V., Jones, B.N., Quang-Dang, D.U., Flinders, J., Floor, S.N., Kim, C., Jemielity, J., Kalek, M., Darzynkiewicz, E., and Gross, J.D. (2008). mRNA decapping is promoted by an RNA-binding channel in Dcp2. Molecular cell *29*, 324-336.

Dolken, L., Ruzsics, Z., Radle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P*., et al.* (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. Rna *14*, 1959-1972.

Dori-Bachash, M., Shalem, O., Manor, Y.S., Pilpel, Y., and Tirosh, I. (2012). Widespread promoter-mediated coordination of transcription and mRNA degradation. Genome biology *13*, R114.

dos Reis, M., Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. Nucleic acids research *32*, 5036-5044.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002). Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. The Journal of biological chemistry *277*, 28368-28371.

Dunckley, T., and Parker, R. (1999). The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. The EMBO journal *18*, 5411-5422.

Dunn, T., Gable, K., and Beeler, T. (1994). Regulation of cellular Ca2+ by yeast vacuoles. The Journal of biological chemistry *269*, 7273-7278.

Durand, A., Bonnet, J., Fournier, M., Chavant, V., and Schultz, P. (2014). Mapping the deubiquitination module within the SAGA complex. Structure *22*, 1553-1559.

Eberharter, A., Sterner, D.E., Schieltz, D., Hassan, A., Yates, J.R., 3rd, Berger, S.L., and Workman, J.L. (1999). The ADA complex is a distinct histone acetyltransferase complex in Saccharomyces cerevisiae. Molecular and cellular biology *19*, 6621-6631.

Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. The Journal of biological chemistry *266*, 71-75.

Eisenmann, D.M., Dollard, C., and Winston, F. (1989). SPT15, the gene encoding the yeast TATA binding factor TFIID, is required for normal transcription initiation in vivo. Cell *58*, 1183-1191.

Eisenmann, D.M., Arndt, K.M., Ricupero, S.L., Rooney, J.W., and Winston, F. (1992). SPT3 interacts with TFIID to allow normal transcription in Saccharomyces cerevisiae. Genes & development *6*, 1319-1331.

Enssle, J., Kugler, W., Hentze, M.W., and Kulozik, A.E. (1993). Determination of mRNA fate by different RNA polymerase II promoters. Proceedings of the National Academy of Sciences of the United States of America *90*, 10091-10095.

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., and Werner, M. (2008). Mediator-dependent recruitment of TFIIH modules in preinitiation complex. Molecular cell *31*, 337-346.

Eyboulet, F., Wydau-Dematteis, S., Eychenne, T., Alibert, O., Neil, H., Boschiero, C., Nevers, M.C., Volland, H., Cornu, D., Redeker, V.*, et al.* (2015). Mediator independently orchestrates multiple steps of preinitiation complex assembly in vivo. Nucleic acids research *43*, 9214-9231.

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

Fan, X., Moqtaderi, Z., Jin, Y., Zhang, Y., Liu, X.S., and Struhl, K. (2010). Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. Proceedings of the National Academy of Sciences of the United States of America *107*, 17945-17950.

Farago, M., Nahari, T., Hammel, C., Cole, C.N., and Choder, M. (2003). Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. Molecular biology of the cell *14*, 2744-2755.

Fassler, J.S., and Winston, F. (1988). Isolation and analysis of a novel class of suppressor of Ty insertion mutations in Saccharomyces cerevisiae. Genetics *118*, 203-212.

Feaver, W.J., Gileadi, O., Li, Y., and Kornberg, R.D. (1991). CTD kinase associated with yeast RNA polymerase II initiation factor b. Cell *67*, 1223-1230.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J.*, et al.* (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212-224.

Fishburn, J., Mohibullah, N., and Hahn, S. (2005). Function of a eukaryotic transcription activator during the transcription cycle. Molecular cell *18*, 369-378.

Fishburn, J., and Hahn, S. (2012). Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. Molecular and cellular biology *32*, 12-25.

Fishburn, J., Tomko, E., Galburt, E., and Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. Proceedings of the National Academy of Sciences of the United States of America *112*, 3961-3966.

Friedel, C.C., and Dolken, L. (2009). Metabolic tagging and purification of nascent RNA: implications for transcriptomics. Molecular bioSystems *5*, 1271-1278.

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.

Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C., and Faye, G. (2003). Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. The EMBO journal *22*, 1588-1598.

Garcia-Martinez, J., Delgado-Ramos, L., Ayala, G., Pelechano, V., Medina, D.A., Carrasco, F., Gonzalez, R., Andres-Leon, E., Steinmetz, L., Warringer, J.*, et al.* (2016). The cellular growth rate controls overall mRNA turnover, and modulates either transcription or degradation rates of particular gene regulons. Nucleic acids research *44*, 3643-3658.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Molecular biology of the cell *11*, 4241-4257.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M.*, et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. Science *272*, 830-836.

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

Gillette, T.G., Gonzalez, F., Delahodde, A., Johnston, S.A., and Kodadek, T. (2004). Physical and functional association of RNA polymerase II and the proteasome. Proceedings of the National Academy of Sciences of the United States of America *101*, 5904-5909.

Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sorensen, K.D.*, et al.* (2014). A dual program for translation regulation in cellular proliferation and differentiation. Cell *158*, 1281-1292.

Ginsburg, D.S., Govind, C.K., and Hinnebusch, A.G. (2009). NuA4 lysine acetyltransferase Esa1 is targeted to coding regions and stimulates transcription elongation with Gcn5. Molecular and cellular biology *29*, 6473-6487.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Molecular and cellular biology *29*, 5455-5464.

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

Goldstrohm, A.C., Seay, D.J., Hook, B.A., and Wickens, M. (2007). PUF protein-mediated deadenylation is catalyzed by Ccr4p. The Journal of biological chemistry *282*, 109-114.

Goler-Baron, V., Selitrennik, M., Barkai, O., Haimovich, G., Lotan, R., and Choder, M. (2008). Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. Genes & development *22*, 2022-2027.

Gonzalez, D., Hamidi, N., Del Sol, R., Benschop, J.J., Nancy, T., Li, C., Francis, L., Tzouros, M., Krijgsveld, J., Holstege, F.C.*, et al.* (2014). Suppression of Mediator is regulated by Cdk8-dependent Grr1 turnover of the Med3 coactivator. Proceedings of the National Academy of Sciences of the United States of America *111*, 2500-2505.

Goodrich, J.A., Hoey, T., Thut, C.J., Admon, A., and Tjian, R. (1993). Drosophila TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. Cell *75*, 519-530.

Goppelt, A., and Meisterernst, M. (1996). Characterization of the basal inhibitor of class II transcription NC2 from Saccharomyces cerevisiae. Nucleic acids research *24*, 4450-4455.

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

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

Grant, P.A., Schieltz, D., Pray-Grant, M.G., Steger, D.J., Reese, J.C., Yates, J.R., 3rd, and Workman, J.L. (1998a). A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. Cell *94*, 45-53.

Grant, P.A., Schieltz, D., Pray-Grant, M.G., Yates, J.R., 3rd, and Workman, J.L. (1998b). The ATM-related cofactor Tra1 is a component of the purified SAGA complex. Molecular cell *2*, 863-867.

Grant, P.A., Eberharter, A., John, S., Cook, R.G., Turner, B.M., and Workman, J.L. (1999). Expanded lysine acetylation specificity of Gcn5 in native complexes. The Journal of biological chemistry *274*, 5895-5900.

Grienberger, C., and Konnerth, A. (2012). Imaging calcium in neurons. Neuron *73*, 862-885.

Grunberg, S., and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. Trends in biochemical sciences *38*, 603-611.

Grunberg, S., Henikoff, S., Hahn, S., and Zentner, G.E. (2016). Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. The EMBO journal *35*, 2435-2446.

Grunberg, S., and Zentner, G.E. (2017). Genome-wide Mapping of Protein-DNA Interactions with ChEC-seq in Saccharomyces cerevisiae. Journal of visualized experiments : JoVE.

Guelman, S., Suganuma, T., Florens, L., Weake, V., Swanson, S.K., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2006). The essential gene wda encodes a WD40 repeat subunit of Drosophila SAGA required for histone H3 acetylation. Molecular and cellular biology *26*, 7178-7189.

Guelman, S., Kozuka, K., Mao, Y., Pham, V., Solloway, M.J., Wang, J., Wu, J., Lill, J.R., and Zha, J. (2009). The double-histone-acetyltransferase complex ATAC is essential for mammalian development. Molecular and cellular biology *29*, 1176-1188.

Hahn, S., Buratowski, S., Sharp, P.A., and Guarente, L. (1989). Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identical to the SPT15 suppressor of Ty element insertions. Cell *58*, 1173-1181.

Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705-736.

Hahn, S., and Buratowski, S. (2016). Structural biology: Snapshots of transcription initiation. Nature *533*, 331-332.

Haimovich, G., Choder, M., Singer, R.H., and Trcek, T. (2013a). The fate of the messenger is pre-determined: a new model for regulation of gene expression. Biochimica et biophysica acta *1829*, 643-653.

Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millan-Zambrano, G., Barkai, O., Chavez, S., Perez-Ortin, J.E., Darzacq, X., and Choder, M. (2013b). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. Cell *153*, 1000-1011.

Hamada, M., Huang, Y., Lowe, T.M., and Maraia, R.J. (2001). Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. Molecular and cellular biology *21*, 6870-6881.

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

Hantsche, M., and Cramer, P. (2017). Conserved RNA polymerase II initiation complex structure. Current opinion in structural biology *47*, 17-22.

Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., and Workman, J.L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369-379.

Hatfield, L., Beelman, C.A., Stevens, A., and Parker, R. (1996). Mutations in trans-acting factors affecting mRNA decapping in Saccharomyces cerevisiae. Molecular and cellular biology *16*, 5830-5838.

Hayakawa, T., and Nakayama, J. (2011). Physiological roles of class I HDAC complex and histone demethylase. Journal of biomedicine & biotechnology *2011*, 129383.

He, Q., Johnston, J., and Zeitlinger, J. (2015). ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. Nature biotechnology *33*, 395-401.

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

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

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

Helenius, K., Yang, Y., Tselykh, T.V., Pessa, H.K., Frilander, M.J., and Makela, T.P. (2011). Requirement of TFIIH kinase subunit Mat1 for RNA Pol II C-terminal domain Ser5 phosphorylation, transcription and mRNA turnover. Nucleic acids research *39*, 5025-5035.

Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, L., Potier, N., Van-Dorsselaer, A., Wurtz, J.M*., et al.* (2004). Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. Human molecular genetics *13*, 1257-1265.

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

Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S., and Young, R.A. (1998). Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. Molecular cell *2*, 43-53.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proceedings of the National Academy of Sciences of the United States of America *108*, 18318-18323.

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

Henry, N.L., Sayre, M.H., and Kornberg, R.D. (1992). Purification and characterization of yeast RNA polymerase II general initiation factor g. The Journal of biological chemistry *267*, 23388-23392.

Henry, N.L., Campbell, A.M., Feaver, W.J., Poon, D., Weil, P.A., and Kornberg, R.D. (1994). TFIIF-TAF-RNA polymerase II connection. Genes & development *8*, 2868-2878.

Herbig, E., Warfield, L., Fish, L., Fishburn, J., Knutson, B.A., Moorefield, B., Pacheco, D., and Hahn, S. (2010). Mechanism of Mediator recruitment by tandem Gcn4 activation domains and three Gal11 activator-binding domains. Molecular and cellular biology *30*, 2376-2390.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717-728.

Horiuchi, J., Silverman, N., Pina, B., Marcus, G.A., and Guarente, L. (1997). ADA1, a novel component of the ADA/GCN5 complex, has broader effects than GCN5, ADA2, or ADA3. Molecular and cellular biology *17*, 3220-3228.

Howe, F.S., Fischl, H., Murray, S.C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? BioEssays : news and reviews in molecular, cellular and developmental biology *39*, 1-12.

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

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Huisinga, K.L., and Pugh, B.F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Molecular cell *13*, 573-585.

Imbalzano, A.N., Zaret, K.S., and Kingston, R.E. (1994). Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. The Journal of biological chemistry *269*, 8280-8286.

Ingvarsdottir, K., Krogan, N.J., Emre, N.C., Wyce, A., Thompson, N.J., Emili, A., Hughes, T.R., Greenblatt, J.F., and Berger, S.L. (2005). H2B ubiquitin protease Ubp8 and Sgf11 constitute a discrete functional module within the Saccharomyces cerevisiae SAGA complex. Molecular and cellular biology *25*, 1162-1172.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.

Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. Nature structural & molecular biology *21*, 449-455.

Jiang, F., Frey, B.R., Evans, M.L., Friel, J.C., and Hopper, J.E. (2009). Gene activation by dissociation of an inhibitor from a transcriptional activation domain. Molecular and cellular biology *29*, 5604-5610.

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

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

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

Johnson, K.M., Wang, J., Smallwood, A., Arayata, C., and Carey, M. (2002). TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. Genes & development *16*, 1852-1863.

Johnson, K.M., and Carey, M. (2003). Assembly of a mediator/TFIID/TFIIA complex bypasses the need for an activator. Current biology : CB *13*, 772-777.

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Molecular cell *20*, 971-978.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental biology *339*, 225-229.

Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R.G., Meisterernst, M., and Burley, S.K. (2001). Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. Cell *106*, 71-81.

Kamenova, I., Warfield, L., and Hahn, S. (2014). Mutations on the DNA binding surface of TBP discriminate between yeast TATA and TATA-less gene transcription. Molecular and cellular biology *34*, 2929-2943.

Kang, J.J., Auble, D.T., Ranish, J.A., and Hahn, S. (1995). Analysis of the yeast transcription factor TFIIA: distinct functional regions and a polymerase II-specific role in basal and activated transcription. Molecular and cellular biology *15*, 1234-1243.

Kao, C.F., Hillyer, C., Tsukuda, T., Henry, K., Berger, S., and Osley, M.A. (2004). Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. Genes & development *18*, 184-195.

Kasinathan, S., Orsi, G.A., Zentner, G.E., Ahmad, K., and Henikoff, S. (2014). High-resolution mapping of transcription factor binding sites on native chromatin. Nature methods *11*, 203-209.

Kassem, S., Villanyi, Z., and Collart, M.A. (2017). Not5-dependent co-translational assembly of Ada2 and Spt20 is essential for functional integrity of SAGA. Nucleic acids research *45*, 1186-1199.

Kasten, M., Szerlong, H., Erdjument-Bromage, H., Tempst, P., Werner, M., and Cairns, B.R. (2004). Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. The EMBO journal *23*, 1348-1359.

Keleher, C.A., Redd, M.J., Schultz, J., Carlson, M., and Johnson, A.D. (1992). Ssn6-Tup1 is a general repressor of transcription in yeast. Cell *68*, 709-719.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241-254.

Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. The Journal of biological chemistry *268*, 10416-10424.

Kent, N.A., Adams, S., Moorhouse, A., and Paszkiewicz, K. (2011). Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. Nucleic acids research *39*, e26.

Khaperskyy, D.A., Ammerman, M.L., Majovski, R.C., and Ponticelli, A.S. (2008). Functions of Saccharomyces cerevisiae TFIIF during transcription start site utilization. Molecular and cellular biology *28*, 3757-3766.

Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell *77*, 599-608.

Kimura, M., Suzuki, H., and Ishihama, A. (2002). Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in Schizosaccharomyces pombe involves direct interaction between Fcp1 and the Rpb4 subunit of pol II. Molecular and cellular biology *22*, 1577-1588.

Kirschner, D.B., vom Baur, E., Thibault, C., Sanders, S.L., Gangloff, Y.G., Davidson, I., Weil, P.A., and Tora, L. (2002). Distinct mutations in yeast TAF(II)25 differentially affect the composition of TFIID and SAGA complexes as well as global gene expression patterns. Molecular and cellular biology *22*, 3178-3193.

Knutson, B.A., and Hahn, S. (2011). Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. Molecular and cellular biology *31*, 818-831.

Kohler, A., Pascual-Garcia, P., Llopis, A., Zapater, M., Posas, F., Hurt, E., and Rodriguez-Navarro, S. (2006). The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5

acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11. Molecular biology of the cell *17*, 4228-4236.

Kohler, A., Schneider, M., Cabal, G.G., Nehrbass, U., and Hurt, E. (2008). Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. Nature cell biology *10*, 707-715.

Kohler, A., Zimmerman, E., Schneider, M., Hurt, E., and Zheng, N. (2010). Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. Cell *141*, 606-617.

Kokubo, T., Swanson, M.J., Nishikawa, J.I., Hinnebusch, A.G., and Nakatani, Y. (1998). The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. Molecular and cellular biology *18*, 1003-1012.

Koleske, A.J., and Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. Nature *368*, 466-469.

Kolodziej, P.A., Woychik, N., Liao, S.M., and Young, R.A. (1990). RNA polymerase II subunit composition, stoichiometry, and phosphorylation. Molecular and cellular biology *10*, 1915-1920.

Kostrewa, D., Zeller, M.E., Armache, K.J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature *462*, 323-330.

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

Krebs, A.R., Demmers, J., Karmodiya, K., Chang, N.C., Chang, A.C., and Tora, L. (2010). ATAC and Mediator coactivators form a stable complex and regulate a set of non-coding RNA genes. EMBO reports *11*, 541-547.

Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Molecular cell *44*, 410-423.

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

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M.*, et al.* (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Molecular cell *11*, 721-729.

Kruk, J.A., Dutta, A., Fu, J., Gilmour, D.S., and Reese, J.C. (2011). The multifunctional Ccr4-Not complex directly promotes transcription elongation. Genes & development *25*, 581-593.

Kuang, Z., Cai, L., Zhang, X., Ji, H., Tu, B.P., and Boeke, J.D. (2014). High-temporal-resolution view of transcription and chromatin states across distinct metabolic states in budding yeast. Nature structural & molecular biology *21*, 854-863.

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.L., Rougemont, J., and Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. Molecular cell *60*, 422-434.

Kubik, S., Bruzzone, M.J., and Shore, D. (2017). Establishing nucleosome architecture and stability at promoters: Roles of pioneer transcription factors and the RSC chromatin remodeler. BioEssays : news and reviews in molecular, cellular and developmental biology *39*.

Kuras, L., and Struhl, K. (1999). Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. Nature *399*, 609-613.

Kuras, L., Kosa, P., Mencia, M., and Struhl, K. (2000). TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo. Science *288*, 1244-1248.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes & development *16*, 2893-2905.

Ladurner, A.G., Inouye, C., Jain, R., and Tjian, R. (2003). Bromodomains mediate an acetyl-histone encoded antisilencing function at heterochromatin boundaries. Molecular cell *11*, 365-376.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes & development *12*, 34-44.

Lang, G., Bonnet, J., Umlauf, D., Karmodiya, K., Koffler, J., Stierle, M., Devys, D., and Tora, L. (2011). The tightly controlled deubiquitination activity of the human SAGA complex differentially modifies distinct gene regulatory elements. Molecular and cellular biology *31*, 3734-3744.

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.

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

Lariviere, L., Plaschka, C., Seizl, M., Wenzeck, L., Kurth, F., and Cramer, P. (2012). Structure of the Mediator head module. Nature *492*, 448-451.

Larschan, E., and Winston, F. (2001). The S. cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4. Genes & development *15*, 1946-1956.

Lebreton, A., Tomecki, R., Dziembowski, A., and Seraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature *456*, 993-996.

Lee, K.K., Florens, L., Swanson, S.K., Washburn, M.P., and Workman, J.L. (2005). The deubiquitylation activity of Ubp8 is dependent upon Sgf11 and its association with the SAGA complex. Molecular and cellular biology *25*, 1173-1182.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nature reviews Molecular cell biology *8*, 284-295.

Lee, K.K., Swanson, S.K., Florens, L., Washburn, M.P., and Workman, J.L. (2009). Yeast Sgf73/Ataxin-7 serves to anchor the deubiquitination module into both SAGA and Slik(SALSA) HAT complexes. Epigenetics & chromatin *2*, 2.

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

Lee, T.I., Causton, H.C., Holstege, F.C., Shen, W.C., Hannett, N., Jennings, E.G., Winston, F., Green, M.R., and Young, R.A. (2000). Redundant roles for the TFIID and SAGA complexes in global transcription. Nature *405*, 701-704.

Lenstra, T.L., Benschop, J.J., Kim, T., Schulze, J.M., Brabers, N.A., Margaritis, T., van de Pasch, L.A., van Heesch, S.A., Brok, M.O., Groot Koerkamp, M.J.*, et al.* (2011). The specificity and topology of chromatin interaction pathways in yeast. Molecular cell *42*, 536-549.

Lenstra, T.L., and Holstege, F.C. (2012). The discrepancy between chromatin factor location and effect. Nucleus *3*, 213-219.

Li, X.Y., Virbasius, A., Zhu, X., and Green, M.R. (1999). Enhancement of TBP binding by activators and general transcription factors. Nature *399*, 605-609.

Li, X.Y., Bhaumik, S.R., and Green, M.R. (2000). Distinct classes of yeast promoters revealed by differential TAF recruitment. Science *288*, 1242-1244.

Liao, S.M., Zhang, J., Jeffery, D.A., Koleske, A.J., Thompson, C.M., Chao, D.M., Viljoen, M., van Vuuren, H.J., and Young, R.A. (1995). A kinase-cyclin pair in the RNA polymerase II holoenzyme. Nature *374*, 193-196.

Liu, H.Y., Badarinarayana, V., Audino, D.C., Rappsilber, J., Mann, M., and Denis, C.L. (1998). The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. The EMBO journal *17*, 1096-1106.

Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell *127*, 1223-1237.

Liu, Y., Pelham-Webb, B., Di Giammartino, D.C., Li, J., Kim, D., Kita, K., Saiz, N., Garg, V., Doane, A., Giannakakou, P.*, et al.* (2017). Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. Cell reports *19*, 1283-1293.

Lombardo, A., Cereghino, G.P., and Scheffler, I.E. (1992). Control of mRNA turnover as a mechanism of glucose repression in Saccharomyces cerevisiae. Molecular and cellular biology *12*, 2941-2948.

Lorch, Y., Maier-Davis, B., and Kornberg, R.D. (2014). Role of DNA sequence in chromatin remodeling and the formation of nucleosome-free regions. Genes & development *28*, 2492-2497.

Lotan, R., Bar-On, V.G., Harel-Sharvit, L., Duek, L., Melamed, D., and Choder, M. (2005). The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. Genes & development *19*, 3004-3016.

Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G., and Choder, M. (2007). The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. The Journal of cell biology *178*, 1133-1143.

Louder, R.K., He, Y., Lopez-Blanco, J.R., Fang, J., Chacon, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature *531*, 604-609.

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

Madison, J.M., and Winston, F. (1997). Evidence that Spt3 functionally interacts with Mot1, TFIIA, and TATA-binding protein to confer promoter-specific transcriptional control in Saccharomyces cerevisiae. Molecular and cellular biology *17*, 287-295.

Malave, T.M., and Dent, S.Y. (2006). Transcriptional repression by Tup1-Ssn6. Biochemistry and cell biology = Biochimie et biologie cellulaire *84*, 437-443.

Marcus, G.A., Silverman, N., Berger, S.L., Horiuchi, J., and Guarente, L. (1994). Functional similarity and physical association between GCN5 and ADA2: putative transcriptional adaptors. The EMBO journal *13*, 4807-4815.

Marcus, G.A., Horiuchi, J., Silverman, N., and Guarente, L. (1996). ADA5/SPT20 links the ADA and SPT genes, which are involved in yeast transcription. Molecular and cellular biology *16*, 3197-3205.

Margaritis, T., Oreal, V., Brabers, N., Maestroni, L., Vitaliano-Prunier, A., Benschop, J.J., van Hooff, S., van Leenen, D., Dargemont, C., Geli, V.*, et al.* (2012). Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. PLoS genetics *8*, e1002952.

Marr, M.T., 2nd, Isogai, Y., Wright, K.J., and Tjian, R. (2006). Coactivator cross-talk specifies transcriptional output. Genes & development *20*, 1458-1469.

Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. The Journal of biological chemistry *270*, 12335-12338.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. The Journal of biological chemistry *271*, 27176-27183.

Martinez, E., Kundu, T.K., Fu, J., and Roeder, R.G. (1998). A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID. The Journal of biological chemistry *273*, 23781-23785.

Matangkasombut, O., and Buratowski, S. (2003). Different sensitivities of bromodomain factors 1 and 2 to histone H4 acetylation. Molecular cell *11*, 353-363.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008a). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome research *18*, 1073-1083.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C.*, et al.* (2008b). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

McCullough, C.E., and Marmorstein, R. (2016). Molecular Basis for Histone Acetyltransferase Regulation by Binding Partners, Associated Domains, and Autoacetylation. ACS chemical biology *11*, 632-642.

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.

McMahon, S.J., Pray-Grant, M.G., Schieltz, D., Yates, J.R., 3rd, and Grant, P.A. (2005). Polyglutamine-expanded spinocerebellar ataxia-7 protein disrupts normal SAGA and SLIK histone acetyltransferase activity. Proceedings of the National Academy of Sciences of the United States of America *102*, 8478-8482.

Melvin, W.T., Milne, H.B., Slater, A.A., Allen, H.J., and Keir, H.M. (1978). Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. European journal of biochemistry *92*, 373-379.

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dumcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dolken, L.*, et al.* (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Molecular systems biology *7*, 458.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proceedings of the National Academy of Sciences of the United States of America *98*, 12902-12907.

Milligan, L., Decourty, L., Saveanu, C., Rappsilber, J., Ceulemans, H., Jacquier, A., and Tollervey, D. (2008). A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. Molecular and cellular biology *28*, 5446-5457.

Minsky, N., Shema, E., Field, Y., Schuster, M., Segal, E., and Oren, M. (2008). Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nature cell biology *10*, 483-488.

Mitchell, P., and Tollervey, D. (2003). An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. Molecular cell *11*, 1405-1413.

Mohibullah, N., and Hahn, S. (2008). Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. Genes & development *22*, 2994-3006.

Molina-Navarro, M.M., Martinez-Jimenez, C.P., and Rodriguez-Navarro, S. (2011). Transcriptional elongation and mRNA export are coregulated processes. Genetics research international *2011*, 652461.

Morgan, M.T., Haj-Yahya, M., Ringel, A.E., Bandi, P., Brik, A., and Wolberger, C. (2016). Structural basis for histone H2B deubiquitination by the SAGA DUB module. Science *351*, 725-728.

Moyle-Heyrman, G., Viswanathan, R., Widom, J., and Auble, D.T. (2012). Two-step mechanism for modifier of transcription 1 (Mot1) enzyme-catalyzed displacement of TATA-binding protein (TBP) from DNA. The Journal of biological chemistry *287*, 9002-9012.

Muller, F., Zaucker, A., and Tora, L. (2010). Developmental regulation of transcription initiation: more than just changing the actors. Current opinion in genetics & development *20*, 533-540.

Muller, F., and Tora, L. (2014). Chromatin and DNA sequences in defining promoters for transcription initiation. Biochimica et biophysica acta *1839*, 118-128.

Munchel, S.E., Shultzaberger, R.K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Molecular biology of the cell *22*, 2787-2795.

Murakami, K., Tsai, K.L., Kalisman, N., Bushnell, D.A., Asturias, F.J., and Kornberg, R.D. (2015). Structure of an RNA polymerase II preinitiation complex. Proceedings of the National Academy of Sciences of the United States of America *112*, 13543-13548.

Muratoglu, S., Georgieva, S., 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. Molecular and cellular biology *23*, 306-321.

Nagy, Z., Riss, A., Romier, C., le Guezennec, X., Dongre, A.R., Orpinell, M., Han, J., Stunnenberg, H., and Tora, L. (2009). The human SPT20-containing SAGA complex plays a direct role in the regulation of endoplasmic reticulum stress-induced genes. Molecular and cellular biology *29*, 1649-1660.

Nagy, Z., Riss, A., Fujiyama, S., Krebs, A., Orpinell, M., Jansen, P., Cohen, A., Stunnenberg, H.G., Kato, S., and Tora, L. (2010). The metazoan ATAC and SAGA

coactivator HAT complexes regulate different sets of inducible target genes. Cellular and molecular life sciences : CMLS *67*, 611-628.

Nair, D., Kim, Y., and Myers, L.C. (2005). Mediator and TFIIH govern carboxyl-terminal domain-dependent transcription in yeast extracts. The Journal of biological chemistry *280*, 33739-33748.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Molecular cell *11*, 709-719.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature *377*, 119-128.

Nissan, T., Rajyaguru, P., She, M., Song, H., and Parker, R. (2010). Decapping activators in Saccharomyces cerevisiae act by multiple mechanisms. Molecular cell *39*, 773-783.

Nourani, A., Utley, R.T., Allard, S., and Cote, J. (2004). Recruitment of the NuA4 complex poises the PHO5 promoter for chromatin remodeling and activation. The EMBO journal *23*, 2597-2607.

Nozawa, K., Schneider, T.R., and Cramer, P. (2017). Core Mediator structure at 3.4 A extends model of transcription initiation complex. Nature *545*, 248-251.

Ogryzko, V.V., Kotani, T., Zhang, X., Schiltz, R.L., Howard, T., Yang, X.J., Howard, B.H., Qin, J., and Nakatani, Y. (1998). Histone-like TAFs within the PCAF histone acetylase complex. Cell *94*, 35-44.

Okamoto, T., Yamamoto, S., Watanabe, Y., Ohta, T., Hanaoka, F., Roeder, R.G., and Ohkuma, Y. (1998). Analysis of the role of TFIIE in transcriptional regulation through structure-function studies of the TFIIEbeta subunit. The Journal of biological chemistry *273*, 19866-19876.

Orlicky, S.M., Tran, P.T., Sayre, M.H., and Edwards, A.M. (2001). Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. The Journal of biological chemistry *276*, 10097-10102.

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

Ossipow, V., Tassan, J.P., Nigg, E.A., and Schibler, U. (1995). A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. Cell *83*, 137-146.

Ozsolak, F., Song, J.S., Liu, X.S., and Fisher, D.E. (2007). High-throughput mapping of the chromatin structure of human promoters. Nature biotechnology *25*, 244-248.

Palhan, V.B., Chen, S., Peng, G.H., Tjernberg, A., Gamper, A.M., Fan, Y., Chait, B.T., La Spada, A.R., and Roeder, R.G. (2005). Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. Proceedings of the National Academy of Sciences of the United States of America *102*, 8472-8477.

Pan, G., and Greenblatt, J. (1994). Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. The Journal of biological chemistry *269*, 30101-30104.

Pardee, T.S., Bangur, C.S., and Ponticelli, A.S. (1998). The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. The Journal of biological chemistry *273*, 17859-17864.

Park, D., Lee, Y., Bhupindersingh, G., and Iyer, V.R. (2013). Widespread misinterpretable ChIP-seq bias in yeast. PloS one *8*, e83506.

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. Molecular cell *25*, 635-646.

Parker, R. (2012). RNA degradation in Saccharomyces cerevisae. Genetics *191*, 671-702.

Parmley, J.L., and Huynen, M.A. (2009). Clustering of codons with rare cognate tRNAs in human genes suggests an extra level of expression regulation. PLoS genetics *5*, e1000548.

Parthun, M.R., Widom, J., and Gottschling, D.E. (1996). The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. Cell *87*, 85-94.

Parthun, M.R. (2007). Hat1: the emerging cellular roles of a type B histone acetyltransferase. Oncogene *26*, 5319-5328.

Paul, E., Zhu, Z.I., Landsman, D., and Morse, R.H. (2015). Genome-wide association of mediator and RNA polymerase II in wild-type and mediator mutant yeast. Molecular and cellular biology *35*, 331-342.

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

Peng, J., Zhu, Y., Milton, J.T., and Price, D.H. (1998b). Identification of multiple cyclin subunits of human P-TEFb. Genes & development *12*, 755-762.

Pereira, L.A., van der Knaap, J.A., van den Boom, V., van den Heuvel, F.A., and Timmers, H.T. (2001). TAF(II)170 interacts with the concave surface of TATA-binding protein to inhibit its DNA binding activity. Molecular and cellular biology *21*, 7523-7534.

Pereira, L.A., Klejman, M.P., and Timmers, H.T. (2003). Roles for BTAF1 and Mot1p in dynamics of TATA-binding protein and regulation of RNA polymerase II transcription. Gene *315*, 1-13.

Pina, B., Berger, S., Marcus, G.A., Silverman, N., Agapite, J., and Guarente, L. (1993). ADA3: a gene, identified by resistance to GAL4-VP16, with properties similar to and different from those of ADA2. Molecular and cellular biology *13*, 5981-5989.

Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F.*, et al.* (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature *518*, 376-380.

Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. Nature *533*, 353-358.

Platt, A., and Reece, R.J. (1998). The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex. The EMBO journal *17*, 4086-4091.

Pollard, K.J., and Peterson, C.L. (1997). Role for ADA/GCN5 products in antagonizing chromatin-mediated transcriptional repression. Molecular and cellular biology *17*, 6212-6222.

Pray-Grant, M.G., Schieltz, D., McMahon, S.J., Wood, J.M., Kennedy, E.L., Cook, R.G., Workman, J.L., Yates, J.R., 3rd, and Grant, P.A. (2002). The novel SLIK histone acetyltransferase complex functions in the yeast retrograde response pathway. Molecular and cellular biology *22*, 8774-8786.

Prelich, G. (2002). RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. Eukaryotic cell *1*, 153-162.

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R.*, et al.* (2015). Codon optimality is a major determinant of mRNA stability. Cell *160*, 1111-1124.

Prieto, S., de la Cruz, B.J., and Scheffler, I.E. (2000). Glucose-regulated turnover of mRNA and the influence of poly(A) tail length on half-life. The Journal of biological chemistry *275*, 14155-14166.

Pugh, B.F., and Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes & development *5*, 1935-1945.

Qiu, H., Hu, C., and Hinnebusch, A.G. (2009). Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. Molecular cell *33*, 752-762.

Radhakrishnan, A., Chen, Y.H., Martin, S., Alhusaini, N., Green, R., and Coller, J. (2016). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. Cell *167*, 122-132 e129.

Raices, M., and D'Angelo, M.A. (2017). Nuclear pore complexes and regulation of gene expression. Current opinion in cell biology *46*, 26-32.

Raithatha, S., Su, T.C., Lourenco, P., Goto, S., and Sadowski, I. (2012). Cdk8 regulates stability of the transcription factor Phd1 to control pseudohyphal differentiation of Saccharomyces cerevisiae. Molecular and cellular biology *32*, 664-674.

Ramirez, C.V., Vilela, C., Berthelot, K., and McCarthy, J.E. (2002). Modulation of eukaryotic mRNA stability via the cap-binding translation complex eIF4F. Journal of molecular biology *318*, 951-962.

Rando, O.J., and Winston, F. (2012). Chromatin and transcription in yeast. Genetics *190*, 351-387.

Ranish, J.A., Yudkovsky, N., and Hahn, S. (1999). Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. Genes & development *13*, 49-63.

Reeves, W.M., and Hahn, S. (2005). Targets of the Gal4 transcription activator in functional transcription complexes. Molecular and cellular biology *25*, 9092-9102.

Reid, J.L., Iyer, V.R., Brown, P.O., and Struhl, K. (2000). Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. Molecular cell *6*, 1297-1307.

Reinberg, D., Horikoshi, M., and Roeder, R.G. (1987). Factors involved in specific transcription in mammalian RNA polymerase II. Functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. The Journal of biological chemistry *262*, 3322-3330.

Reinberg, D., and Roeder, R.G. (1987). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE. The Journal of biological chemistry *262*, 3310-3321.

Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H., and Price, D.H. (2001). A highly purified RNA polymerase II elongation control system. The Journal of biological chemistry *276*, 42601-42609.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408-1419.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295-301.

Robert, F., Pokholok, D.K., Hannett, N.M., Rinaldi, N.J., Chandy, M., Rolfe, A., Workman, J.L., Gifford, D.K., and Young, R.A. (2004). Global position and recruitment of HATs and HDACs in the yeast genome. Molecular cell *16*, 199-209.

Roberts, S.M., and Winston, F. (1996). SPT20/ADA5 encodes a novel protein functionally related to the TATA-binding protein and important for transcription in Saccharomyces cerevisiae. Molecular and cellular biology *16*, 3206-3213.

Roberts, S.M., and Winston, F. (1997). Essential functional interactions of SAGA, a Saccharomyces cerevisiae complex of Spt, Ada, and Gcn5 proteins, with the Snf/Swi and Srb/mediator complexes. Genetics *147*, 451-465.

Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.J., Burlingame, A.L., and Kornberg, R.D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell *166*, 1411-1422 e1416.

Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L., and Buratowski, S. (2000). Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. Molecular and cellular biology *20*, 104-112.

Rodriguez-Molina, J.B., Tseng, S.C., Simonett, S.P., Taunton, J., and Ansari, A.Z. (2016). Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization. Molecular cell *63*, 433-444.

Rodriguez-Navarro, S. (2009). Insights into SAGA function during gene expression. EMBO reports *10*, 843-850.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. The EMBO journal *20*, 7137-7148.

Romero-Santacreu, L., Moreno, J., Perez-Ortin, J.E., and Alepuz, P. (2009). Specific and global regulation of mRNA stability during osmotic stress in Saccharomyces cerevisiae. Rna *15*, 1110-1120.

Rougeulle, C., Chaumeil, J., Sarma, K., Allis, C.D., Reinberg, D., Avner, P., and Heard, E. (2004). Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. Molecular and cellular biology *24*, 5475-5484.

Runner, V.M., Podolny, V., and Buratowski, S. (2008). The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. Molecular and cellular biology *28*, 1883-1891.

Ruppert, S., and Tjian, R. (1995). Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. Genes & development *9*, 2747-2755.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nature reviews Molecular cell biology *16*, 129-143.

Saleh, A., Schieltz, D., Ting, N., McMahon, S.B., Litchfield, D.W., Yates, J.R., 3rd, Lees-Miller, S.P., Cole, M.D., and Brandl, C.J. (1998). Tra1p is a component of the yeast Ada.Spt transcriptional regulatory complexes. The Journal of biological chemistry *273*, 26559-26565.

Sanders, S.L., Jennings, J., Canutescu, A., Link, A.J., and Weil, P.A. (2002). Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. Molecular and cellular biology *22*, 4723-4738.

Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G., Sanchez-Rotunno, M., Arraiano, C.M., and van Hoof, A. (2009). The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. Nature structural & molecular biology *16*, 56-62.

Schmid, M., Durussel, T., and Laemmli, U.K. (2004). ChIC and ChEC; genomic mapping of chromatin proteins. Molecular cell *16*, 147-157.

Schmid, M., Arib, G., Laemmli, C., Nishikawa, J., Durussel, T., and Laemmli, U.K. (2006). Nup-PI: the nucleopore-promoter interaction of genes in yeast. Molecular cell *21*, 379-391.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Schuller, R., Forne, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Molecular cell *61*, 305-314.

Schulz, D., Pirkl, N., Lehmann, E., and Cramer, P. (2014). Rpb4 subunit functions mainly in mRNA synthesis by RNA polymerase II. The Journal of biological chemistry *289*, 17446-17452.

Seizl, M., Hartmann, H., Hoeg, F., Kurth, F., Martin, D.E., Soding, J., and Cramer, P. (2011). A conserved GA element in TATA-less RNA polymerase II promoters. PloS one *6*, e27595.

Selitrennik, M., Duek, L., Lotan, R., and Choder, M. (2006). Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways. Eukaryotic cell *5*, 2092-2103.

Sellam, A., Askew, C., Epp, E., Lavoie, H., Whiteway, M., and Nantel, A. (2009). Genome-wide mapping of the coactivator Ada2p yields insight into the functional roles of SAGA/ADA complex in Candida albicans. Molecular biology of the cell *20*, 2389-2400.

Sermwittayawong, D., and Tan, S. (2006). SAGA binds TBP via its Spt8 subunit in competition with DNA: implications for TBP recruitment. The EMBO journal *25*, 3791-3800.

Setiaputra, D., Ross, J.D., Lu, S., Cheng, D.T., Dong, M.Q., and Yip, C.K. (2015). Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. The Journal of biological chemistry *290*, 10057-10070.

Shahbazian, M.D., Zhang, K., and Grunstein, M. (2005). Histone H2B ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1. Molecular cell *19*, 271-277.

Shalem, O., Dahan, O., Levo, M., Martinez, M.R., Furman, I., Segal, E., and Pilpel, Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. Molecular systems biology *4*, 223.

Shivaswamy, S., and Iyer, V.R. (2008). Stress-dependent dynamics of global chromatin remodeling in yeast: dual role for SWI/SNF in the heat shock stress response. Molecular and cellular biology *28*, 2221-2234.

Singer, V.L., Wobbe, C.R., and Struhl, K. (1990). A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes & development *4*, 636-645.

Skene, P.J., Hernandez, A.E., Groudine, M., and Henikoff, S. (2014). The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. eLife *3*, e02042.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annual review of biochemistry *72*, 449-479.

Smith, E., Lin, C., and Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. Genes & development *25*, 661-672.

Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C., and Werner, M. (2011). Direct interaction of RNA polymerase II and mediator required for transcription in vivo. Science *331*, 1451-1454.

Spedale, G., Mischerikow, N., Heck, A.J., Timmers, H.T., and Pijnappel, W.W. (2010). Identification of Pep4p as the protease responsible for formation of the SAGA-related SLIK protein complex. The Journal of biological chemistry *285*, 22793-22799.

Spedale, G., Meddens, C.A., Koster, M.J., Ko, C.W., van Hooff, S.R., Holstege, F.C., Timmers, H.T., and Pijnappel, W.W. (2012a). Tight cooperation between Mot1p and NC2beta in regulating genome-wide transcription, repression of transcription following heat shock induction and genetic interaction with SAGA. Nucleic acids research *40*, 996-1008.

Spedale, G., Timmers, H.T., and Pijnappel, W.W. (2012b). ATAC-king the complexity of SAGA during evolution. Genes & development *26*, 527-541.

Sterner, D.E., Grant, P.A., Roberts, S.M., Duggan, L.J., Belotserkovskaya, R., Pacella, L.A., Winston, F., Workman, J.L., and Berger, S.L. (1999). Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. Molecular and cellular biology *19*, 86-98.

Sterner, D.E., Belotserkovskaya, R., and Berger, S.L. (2002). SALSA, a variant of yeast SAGA, contains truncated Spt7, which correlates with activated transcription. Proceedings of the National Academy of Sciences of the United States of America *99*, 11622-11627.

Stevens, A. (2001). 5'-exoribonuclease 1: Xrn1. Methods in enzymology *342*, 251-259.

Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell *98*, 1-4.

Suganuma, T., Gutierrez, J.L., Li, B., Florens, L., Swanson, S.K., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2008). ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. Nature structural & molecular biology *15*, 364-372.

Suganuma, T., Mushegian, A., Swanson, S.K., Abmayr, S.M., Florens, L., Washburn, M.P., and Workman, J.L. (2010). The ATAC acetyltransferase complex coordinates MAP kinases to regulate JNK target genes. Cell *142*, 726-736.

Sugihara, F., Kasahara, K., and Kokubo, T. (2011). Highly redundant function of multiple AT-rich sequences as core promoter elements in the TATA-less RPS5 promoter of Saccharomyces cerevisiae. Nucleic acids research *39*, 59-75.

Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Molecular cell *61*, 297-304.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Lariviere, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome research *22*, 1350-1359.

Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Molecular cell *52*, 52-62.

Sun, X., Zhang, Y., Cho, H., Rickert, P., Lees, E., Lane, W., and Reinberg, D. (1998). NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. Molecular cell *2*, 213-222.

Swanson, M.J., Qiu, H., Sumibcay, L., Krueger, A., Kim, S.J., Natarajan, K., Yoon, S., and Hinnebusch, A.G. (2003). A multiplicity of coactivators is required by Gcn4p at individual promoters in vivo. Molecular and cellular biology *23*, 2800-2820.

Swisher, K.D., and Parker, R. (2010). Localization to, and effects of Pbp1, Pbp4, Lsm12, Dhh1, and Pab1 on stress granules in Saccharomyces cerevisiae. PloS one *5*, e10006.

Synowsky, S.A., van Wijk, M., Raijmakers, R., and Heck, A.J. (2009). Comparative multiplexed mass spectrometric analyses of endogenously expressed yeast nuclear and cytoplasmic exosomes. Journal of molecular biology *385*, 1300-1313.

Tan, S., Hunziker, Y., Sargent, D.F., and Richmond, T.J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. Nature *381*, 127-151.

Teytelman, L., Thurtle, D.M., Rine, J., and van Oudenaarden, A. (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. Proceedings of the National Academy of Sciences of the United States of America *110*, 18602-18607.

Tharun, S., and Parker, R. (2001). Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. Molecular cell *8*, 1075-1083.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Critical reviews in biochemistry and molecular biology *41*, 105-178.

Timmers, H.T., Meyers, R.E., and Sharp, P.A. (1992). Composition of transcription factor B-TFIID. Proceedings of the National Academy of Sciences of the United States of America *89*, 8140-8144.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome research *18*, 1084-1091.

Travis, G.H., Colavito-Shepanski, M., and Grunstein, M. (1984). Extensive purification and characterization of chromatin-bound histone acetyltransferase from Saccharomyces cerevisiae. The Journal of biological chemistry *259*, 14406-14412.

Trcek, T., Larson, D.R., Moldon, A., Query, C.C., and Singer, R.H. (2011). Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. Cell *147*, 1484-1497.

Trivedi, A., Young, L.S., Ouyang, C., Johnson, D.L., and Sprague, K.U. (1999). A TATA element is required for tRNA promoter activity and confers TATA-binding protein responsiveness in Drosophila Schneider-2 cells. The Journal of biological chemistry *274*, 11369-11375.

Tucker, M., and Parker, R. (2000). Mechanisms and control of mRNA decapping in Saccharomyces cerevisiae. Annual review of biochemistry *69*, 571-595.

Tucker, M., Valencia-Sanchez, M.A., Staples, R.R., Chen, J., Denis, C.L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. Cell *104*, 377-386.

Tucker, M., Staples, R.R., Valencia-Sanchez, M.A., Muhlrad, D., and Parker, R. (2002). Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in Saccharomyces cerevisiae. The EMBO journal *21*, 1427-1436.

Ujvari, A., and Luse, D.S. (2006). RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. Nature structural & molecular biology *13*, 49-54.

Unnikrishnan, A., Gafken, P.R., and Tsukiyama, T. (2010). Dynamic changes in histone acetylation regulate origins of DNA replication. Nature structural & molecular biology *17*, 430-437.

Unnikrishnan, A., Akiyoshi, B., Biggins, S., and Tsukiyama, T. (2012). An efficient purification system for native minichromosome from Saccharomyces cerevisiae. Methods in molecular biology *833*, 115-123.

van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Ne, P., Loeillet, S.*, et al.* (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114-117.

Van Dyke, M.W., Sawadogo, M., and Roeder, R.G. (1989). Stability of transcription complexes on class II genes. Molecular and cellular biology *9*, 342-344.

van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell *109*, 745-756.

van Oevelen, C.J., van Teeffelen, H.A., and Timmers, H.T. (2005). Differential requirement of SAGA subunits for Mot1p and Taf1p recruitment in gene activation. Molecular and cellular biology *25*, 4863-4872.

van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nature biotechnology *18*, 424-428.

van Werven, F.J., van Bakel, H., van Teeffelen, H.A., Altelaar, A.F., Koerkamp, M.G., Heck, A.J., Holstege, F.C., and Timmers, H.T. (2008). Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome. Genes & development *22*, 2359-2369.

Venkatasubrahmanyam, S., Hwang, W.W., Meneghini, M.D., Tong, A.H., and Madhani, H.D. (2007). Genome-wide, as opposed to local, antisilencing is mediated redundantly by the euchromatic factors Set1 and H2A.Z. Proceedings of the National Academy of Sciences of the United States of America *104*, 16609-16614.

Venters, B.J., Wachi, S., Mavrich, T.N., Andersen, B.E., Jena, P., Sinnamon, A.J., Jain, P., Rolleri, N.S., Jiang, C., Hemeryck-Walsh, C*., et al.* (2011). A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. Molecular cell *41*, 480-492.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58-69.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G*., et al.* (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967-980.

Verrijzer, C.P., and Tjian, R. (1996). TAFs mediate transcriptional activation and promoter selectivity. Trends in biochemical sciences *21*, 338-342.

Vethantham, V., Yang, Y., Bowman, C., Asp, P., Lee, J.H., Skalnik, D.G., and Dynlacht, B.D. (2012). Dynamic loss of H2B ubiquitylation without corresponding changes in H3K4 trimethylation during myogenic differentiation. Molecular and cellular biology *32*, 1044-1055.

Villanyi, Z., and Collart, M.A. (2015). Ccr4-Not is at the core of the eukaryotic gene expression circuitry. Biochemical Society transactions *43*, 1253-1258.

Viswanathan, P., Ohn, T., Chiang, Y.C., Chen, J., and Denis, C.L. (2004). Mouse CAF1 can function as a processive deadenylase/3'-5'-exonuclease in vitro but in yeast the deadenylase function of CAF1 is not required for mRNA poly(A) removal. The Journal of biological chemistry *279*, 23988-23995.

Vitaliano-Prunier, A., Menant, A., Hobeika, M., Geli, V., Gwizdek, C., and Dargemont, C. (2008). Ubiquitylation of the COMPASS component Swd2 links H2B ubiquitylation to H3K4 trimethylation. Nature cell biology *10*, 1365-1371.

Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nature protocols *2*, 1467-1478.

Wang, Y.L., Faiola, F., Xu, M., Pan, S., and Martinez, E. (2008). Human ATAC Is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. The Journal of biological chemistry *283*, 33808-33815.

Warfield, L., Ranish, J.A., and Hahn, S. (2004). Positive and negative functions of the SAGA complex mediated through interaction of Spt8 with TBP and the N-terminal domain of TFIIA. Genes & development *18*, 1022-1034.

Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., and Hahn, S. (2017). Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. Molecular cell.

Weake, V.M., Swanson, S.K., Mushegian, A., Florens, L., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2009). A novel histone fold domain-containing protein that replaces TAF6 in Drosophila SAGA is required for SAGA-dependent gene expression. Genes & development *23*, 2818-2823.

Weake, V.M., Dyer, J.O., Seidel, C., Box, A., Swanson, S.K., Peak, A., Florens, L., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2011). Post-transcription initiation function of the ubiquitous SAGA complex in tissue-specific gene activation. Genes & development *25*, 1499-1509.

Winston, F., Chaleff, D.T., Valent, B., and Fink, G.R. (1984a). Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics *107*, 179-197.

Winston, F., Durbin, K.J., and Fink, G.R. (1984b). The SPT3 gene is required for normal transcription of Ty elements in S. cerevisiae. Cell *39*, 675-682.

Winston, F., Dollard, C., Malone, E.A., Clare, J., Kapakos, J.G., Farabaugh, P., and Minehart, P.L. (1987). Three genes are required for trans-activation of Ty transcription in yeast. Genetics *115*, 649-656.

Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Molecular cell *54*, 601-612.

Wu, C.H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., and Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes & development *17*, 1402-1414.

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

Wu, P.Y., Ruhlmann, C., Winston, F., and Schultz, P. (2004). Molecular architecture of the S. cerevisiae SAGA complex. Molecular cell *15*, 199-208.

Wu, S.Y., and Chiang, C.M. (1998). Properties of PC4 and an RNA polymerase II complex in directing activated and basal transcription in vitro. The Journal of biological chemistry *273*, 12492-12498.

Wu, S.Y., Thomas, M.C., Hou, S.Y., Likhite, V., and Chiang, C.M. (1999). Isolation of mouse TFIID and functional characterization of TBP and TFIID in mediating estrogen receptor and chromatin transcription. The Journal of biological chemistry *274*, 23480-23490.

Wu, S.Y., and Chiang, C.M. (2001). TATA-binding protein-associated factors enhance the recruitment of RNA polymerase II by transcriptional activators. The Journal of biological chemistry *276*, 34235-34243.

Wu, Y., Reece, R.J., and Ptashne, M. (1996). Quantitation of putative activator-target affinities predicts transcriptional activating potentials. The EMBO journal *15*, 3951-3963.

Xi, Y., Yao, J., Chen, R., Li, W., and He, X. (2011). Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. Genome research *21*, 718-724.

Yamaguchi, Y., Narita, T., Inukai, N., Wada, T., and Handa, H. (2001). SPT genes: key players in the regulation of transcription, chromatin structure and other cellular processes. Journal of biochemistry *129*, 185-191.

Yan, Q., Moreland, R.J., Conaway, J.W., and Conaway, R.C. (1999). Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. The Journal of biological chemistry *274*, 35668-35675.

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.

Yokomori, K., Admon, A., Goodrich, J.A., Chen, J.L., and Tjian, R. (1993). Drosophila TFIIA-L is processed into two subunits that are associated with the TBP/TAF complex. Genes & development *7*, 2235-2245.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626-630.

Yukawa, Y., Dieci, G., Alzapiedi, M., Hiraga, A., Hirai, K., Yamamoto, Y.Y., and Sugiura, M. (2011). A common sequence motif involved in selection of transcription start sites of Arabidopsis and budding yeast tRNA genes. Genomics *97*, 166-172.

Zaidi, S.K., Young, D.W., Montecino, M., van Wijnen, A.J., Stein, J.L., Lian, J.B., and Stein, G.S. (2011). Bookmarking the genome: maintenance of epigenetic information. The Journal of biological chemistry *286*, 18355-18361.

Zentner, G.E., and Henikoff, S. (2014). High-resolution digital profiling of the epigenome. Nature reviews Genetics *15*, 814-827.

Zentner, G.E., Kasinathan, S., Xin, B., Rohs, R., and Henikoff, S. (2015). ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. Nature communications *6*, 8733.

Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodriguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012). Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. The Journal of biological chemistry *287*, 8541-8551.

Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M., and Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. Nature methods *12*, 835-837.

Zurita, M., and Cruz-Becerra, G. (2016). TFIIH: New Discoveries Regarding its Mechanisms and Impact on Cancer Treatment. Journal of Cancer *7*, 2258-2265.

ANNEXES

Annexes

Publication submitted in Elife.

Architecture of TAF11/TAF13/TBP complex suggests novel regulatory state in General Transcription Factor TFIID function

Kapil Gupta [1, 2], Aleksandra A. Watson[3], **Tiago Baptista**[4-7], Elisabeth Scheer[4-7], Anna L. Chambers[1], Christine Koehler[8], Juan Zou[9], Ima Obong-Ebong[10], Eaazhisai Kandiah[2,11], Arturo Temblador[2], Adam Round[2], Eric Forest[11], Petr Man[12], Christoph Bieniossek[2], Ernest D. Laue[3], Edward A. Lemke[8], Juri Rappsilber[9], Carol V. Robinson[10], Didier Devys[4-7], Laszlo Tora[4-7*] and Imre Berge[1*]

[1]The School of Biochemistry, Biomedical Sciences and BrisSynBio Centre, University of Bristol, Tankard's Close, Bristol BS8 1TD, United Kingdom

[2]European Molecular Biology Laboratory, 71 Avenue des Martyrs, 38000 Grenoble, France

[3]Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom

[4]Institut de Génétique et de Biologie Moléculaire et Cellulaire IGBMC, Illkirch, France

[5]Centre National de la Recherche Scientifique, UMR7104, Illkirch, France

[6]Institut National de la Santé et de la Recherche Médicale, U964, Illkirch, France

[7]Université de Strasbourg, Illkirch, France

8European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

[9]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Max Born Crescent, Edinburgh, EH9 3BF, United Kingdom and Chair of Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany

[10]Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, United Kingdom

[11]Institut de Biologie Structurale IBS, 71 Avenue des Martyrs, 38042 Grenoble, France

[12]BioCeV - Institute of Microbiology, The Czech Academy of Sciences, Prumyslova 595, 252 50 Vestec and Faculty of Science, Charles University, Hlavova 8, 128 43 Prague, Czech Republic

[#]Corresponding authors: Laszlo Tora (laszlo.tora@igbmc.fr) & Imre Berger (imre.berger@bristol.ac.uk)

Abstract

General transcription factor TFIID is a key component of RNA polymerase II transcription initiation. Human TFIID is a megadalton-sized complex comprising TATA-binding protein (TBP) and 13 TBP-associated factors (TAFs). TBP binds to core promoter DNA, recognizing the TATA-box. We identified a ternary complex formed by TBP and the histone fold (HF) domain-containing TFIID subunits TAF11 and TAF13. We demonstrate that TAF11/TAF13 competes for TBP binding with TATA-box DNA, and also with the N-terminal domain of TAF1 previously implicated in TATA-box mimicry. In an integrative approach combining crystal coordinates, biochemical analyses and data from cross-linking mass-spectrometry (CLMS), we determine the architecture of the TAF11/TAF13/TBP complex, revealing TAF11/TAF13 interaction with the DNA binding surface of TBP. We identify a highly conserved C-terminal TBP-binding domain (CTID) in TAF13 which is essential for supporting cell growth. Our results thus have implications for cellular TFIID assembly and suggest a novel regulatory state for TFIID function.

Tiago FIDALGO BAPTISTA

Functional study of the coactivator SAGA: role in RNA Polymerase II transcription

Résumé

Des études antérieures suggèrent que les complexes SAGA et TFIID sont des facteurs jouant un rôle complémentaire dans la transcription par l'ARN polymérase II. Chez S. cerevisiae, environ 10% des gènes semblent dépendants de SAGA alors que TFIID aurait un rôle dominant sur 90% du génome. De nouvelles approches m'ont permis de montrer que SAGA est recruté sur les régions régulatrices d'une majorité de gènes, indépendamment de leur classification. Des analyses d'ARN nouvellement-synthétisés ont également démontré que l'inactivation des complexes SAGA ou TFIID, par mutation ou déplétion inductible de leurs sous-unités, altèrent la transcription de pratiquement tous les gènes par l'ARN polymérase II. L'acétyltransférase Gcn5 et la sous-unité Spt3 agissent de façon synergique au sein du complexe SAGA pour stimuler le recrutement de TBP et la transcription par l'ARN polymérase II. Ces données indiquent que les complexes SAGA et TFIID agissent comme des cofacteurs généraux, étant nécessaires pour la synthèse de quasiment tous les ARNm et ayant des effets équivalents sur les gènes précédemment décrits comme dominés par SAGA ou par TFIID.

Mots clés: *S. cerevisiae* ◆ SAGA ◆ TFIID ◆ Transcription ◆ ARN Polymérase II

Summary

Prior studies suggested that SAGA and TFIID are alternative factors that promote RNA polymerase II (RNA Pol II) transcription with about 10% of genes in *S. cerevisiae* dependent on SAGA. The remainder 90% of the genome would be regulated by TFIID. We reassessed the role of SAGA by mapping its genome-wide location and role in global transcription in budding yeast. We observed that SAGA maps to regulatory elements of most genes, irrespective of previous designations of SAGA- or TFIID-dominated genes. Additionally, disruption of either SAGA or TFIID through mutation or rapid subunit depletion reduces transcription from nearly all genes, measured by newly-synthesized RNA or RNA Pol II chromatin immunoprecipitation. We also found that the acetyltransferase Gcn5 synergizes with Spt3 to promote global transcription and that Spt3 functions to stimulate TBP recruitment at all tested genes. Our data demonstrate that both SAGA and TFIID act as general cofactors required for essentially all RNA Pol II transcription and is not consistent with the previous classification of SAGA- and TFIID-dominated genes.

Keywords: *S. cerevisiae* ◆ SAGA ◆ TFIID ◆ Transcription ◆ RNA Polymerase II