Analysis of the composition and the function of oocyte-specific TBP2-containing transcription machinery during mouse oogenesis

Changwei Yu

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE 414 - Sciences de la Vie et de la Santé

IGBMC - CNRS UMR 7104 - Inserm U 1258 - Université de Strasbourg

THÈSE présentée par :

Changwei YU

soutenue le : 13 **décembre 2018**

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

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

Analysis of the composition and the function of oocyte-specific TBP2-containing transcription machinery during mouse oogenesis

THÈSE dirigée par :

M. László TORA	Directeur de recherche, IGBMC, Université de Strasbourg
M. Stephane VINCENT	Chargé de recherche INSERM, IGBMC, Université de Strasbourg

RAPPORTEURS :

M. Ferenc MUELLER	Professor, Institute of Cancer and Genomic Sciences, University of Birmingham
M. Antoine H.F.M. PETERS	Directeur de recherche, Friedrich Miescher Institute for Biomedical Research, University of Basel

AUTRES MEMBRES DU JURY :

Mme. Petra HAJKOVA	Professor, Institute of Clinical Sciences, Imperial College London
M. Irwin DAVIDSON	Directeur de recherche, IGBMC, Université de Strasbourg

Acknowledgements

First of all, I would like to thank Prof. Dr. Petra Hajkova, Dr. Irwin Davidson, Dr. Antoine H.F.M. Peters and Prof. Dr. Ferenc Mueller for accepting to be members of my thesis committee. Thank you for your time and effort to read and evaluate my thesis.

I would like thank Laszlo for being my supervisor and for giving me the opportunity to work in such an excellent scientific environment, and thank Stephane for being my co-supervisor. Thank both of you for your patience, guidance, trust, encouragement and full support over the last four years. I could always come to you if I had questions or problems and you were always there to help me, I am really grateful to have both of you as my supervisors, thank you!

I would like to thank both the past and current members of the Tora lab. Paul, Sascha and Pooja that started around the same time as me, thank you for the help, great discussions and companionship over the years and for being the best “module-mates”. I want to thank Tiago and Farrah (as well as Paul and Sascha), for the crazy moments we have spent together, especially those Halloween nights and Pokémon hunting. I want to thank Ivanka, Federica, and Nikolaos for the wonderful moments we shared inside and outside of the lab. I thank Vincent, Veronique, Gizem and Fang for the fun time (game night, laser tag and canoeing, etc.) we spend together, thank Kenny and Emma for bring me home when I got drunk after the BBQ party. These are the best memories of my PhD. I would like to thank Didier for the scientific discussions and supportive comments, especially during the lab meetings. I want to thank Matthieu and Eli for the help in the lab, especially Eli, she is amazing to keep everything (including common antibodies, regents, documents and working area) super organized.

I also would like to thank everyone that had important technical and/or scientific input on my work. I want to thank all the IGBMC facilities, especially, Tao, Bernard and Christelle of the Genomeast platform for the kind and continuous help and discussions; Mustapha from the antibodies facility for the antibodies; Betty of the cell culture facility for cells and medium; Claudine of Flow cytometry facility for the help with FCAS; the histology platform; and the animal facility for the animal care taking. I want to thank our collaborators Dr. Maria Elena Torres-Padilla and Mate Borsos for the exciting oocyte DamID experiments, Prof. Dr. Robert Schneider and Dr. Igor Kukhtevich for designing and making the microfluidic devices for us. I also want to thank Dr. Dinah S. Singer and Anne Gegonne for the gift of the Taf7flox mouse line, Henk Stunnenberg for TFIIA antibodies, and Dr. Steven Henikoff for the Protein A-micrococcal nuclease (pA-MN) fusion protein and spike-in DNA. Moreover, I would like to

thank my mid-thesis committee members, Dr. Maria Elena Torres-Padilla and Dr. Gavin Kelsey, for their kind guidance and helpful advice. I also would like to thank the ERC for financing of my work.

Last but never least, I want to thank everyone outside of my scientific life who helped and supported me over the years. I want to thank all my friends inside and outside of IGBMC. Especially Zhirong, Guoqiang, Xieyang, Wenjin and Ruicheng for the help, guidance and joyful moments we shared together. A very special thanks to my family who are always there for me, especially my wife, for all her love and constant support!

Table of Contents

Abstract

Mammalian oocytes go through a differentiation process, during which the synthesis and accumulation of RNAs and proteins are essential for oocyte growth, maturation, fertilization and early embryogenesis. Although some crucial transcription factors associated with this developmental program have already been identified, little is known about the nature and function of the transcriptional machinery that is involved in RNA polymerase II (Pol II) transcription initiation during oogenesis.

In somatic cells, the Pol II general transcription factor (GTF) TFIID, is the first to bind to gene promoters in order to nucleate the pre-initiation complex (PIC) also composed of TFIIA, -IIB, -IIE, -IIF, -IIH and Pol II. In metazoans, TFIID is composed of the TATA-binding protein (TBP) and 13 TBP-associated factors (TAFs). During oocyte growth TBP is replaced by a vertebrate specific TBP-type protein, TBP2 (also called TRF3 or TBPL2) and *Tbp2*$^{-/-}$ females are sterile due to defect in the differentiation of the secondary follicles.

In this study, we aimed to understand whether and how TBP2 is controlling transcription initiation during oogenesis. First, to identify the genes that are regulated by TBP2, we carried out RNA-seq analyses from wild-type and *Tbp2*$^{-/-}$ growing oocytes from primary (post-natal day 7; P7) and secondary follicles (P14). These analyses show a main decrease in the expression of the most abundantly expressed genes as well as specific down-regulation of the expression of the MaLR (mammalian apparent LTR retrotransposons)-type endogenous retroviral elements. Second, to identify the nature of the complex associated with TBP2 in the oocytes, we carried out immunoprecipitation followed by mass spectrometry. We demonstrate that, in the oocytes, TBP2 associates with TFIIA, but does not assemble into a TFIID-like complex. Third, in order to identify the genes directly regulated by TBP2 during oocyte growth, we performed a TBP2-Dam-ID-seq on P7 and P14 oocytes. As a complementary approach, we also applied TBP2 uliCUT&RUN in oocytes. In addition, we performed TBP2 ChIP-seq in TBP2 overexpressing NIH3T3 cells. Our data showed that TBP2 binds to gene promoters genome-widely, and that the TBP2 average binding profile shifts slightly downstream of the TSS , suggesting that TBP2 might be involved in a different TSS usage. Forth, since oocytes have abundant storage of transcripts, in order to tear apart transcription initiation from steady state RNAs, we are currently carrying out nascent transcripts analyses using SLAM-seq and mouse TU-tagging approaches.

Altogether, our data show that a specific TBP2-TFIIA-containing transcription machinery, different from canonical TFIID, drives transcription in mouse growing oocytes.

Résumé

Introduction

Chez les mammifères, la maturation des ovocytes passe par plusieurs étapes successives de différenciation. Ce processus implique la synthèse et l'accumulation d'ARN et de protéines nécessaires à la fertilité et au développement embryonnaire précoce. Bien que de nombreux gènes nécessaires à ce programme de développement ont déjà été indentifiés, peu de choses sont encore connues sur la nature et la fonction de la machinerie transcriptionnelle présente au cours de l'ovogenèse.

La régulation de l'initiation de la transcription par l'ARN polymérase II (Pol II) est cruciale pour le développement embryonnaire. Lors de l'initiation de la transcription, la Pol II et les facteurs généraux de la transcription (GTF) TFIIA,-B,-D,-E,-F,-H forment le complexe de préinitiation (PIC) au niveau des promoteurs. Dans les cellules somatiques, le facteur général TFIID est le premier à initier l'assemblage du PIC au niveau des promoteurs pour le recrutement de la Pol II. Chez les métazoaires, TFIID est composé de la protéine *TATA box Binding Protein* (TBP) associé à 13 TBP-*Associated Factors* (TAFs)(Tora, 2002; Muller et al., 2004). Il existe trois protéines paralogues de TBP chez les métazoaires. TRF1 est spécifique des insectes. La protéine *TBP-like factor* (TLF/TBPL1/TRF2/TRP) a été identifiée chez plusieurs espèces de métazoaires, et peut interagir avec TFIIA et TFIIB mais ne peut pas lier la boite TATA. Enfin, TLF est essentielle à la spermatogenèse chez la souris(Zhang et al., 2001b). La dernière est la protéine TBP2 (aussi appelée TRF3 ou TBPL2) qui est spécifique des vertébrés.

Dans leurs parties C terminales contenant le domaine de liaison à l'ADN, TBP et TBP2 partagent 92% de similarités au niveau de la séquence protéique. TBP2 est capable de lier la boite TATA et interagit avec TFIIA et TFIIB. De plus, TBP2 est capable de médier la transcription *in vitro*(Bartfai et al., 2004; Jallow et al., 2004; Deato et al., 2008). Des études antérieures menées dans le laboratoire, ont montré que TBP est remplacée par TBP2 au cours de la croissance ovocytaire et que la délétion de *Tbp2* bloque le

développement des ovocytes à leur terme, entrainant la stérilité des femelles(Gazdag et al., 2007). Ces données suggèrent donc que TBP2 joue un rôle crucial dans la régulation transcriptionnelle dans les ovocytes, contrôlant ainsi leur croissance et leur maturation.

Dans cette étude, nous avons caractérisé la machinerie transcriptionnelle contenant TBP2 puis nous avons analysé sa fonction dans la régulation transcriptionnelle au cours de l'ovogenèse.

Objectifs, stratégie et résultats

Le but de ce projet est de comprendre pourquoi spécifiquement dans les ovocytes, TBP est remplacée par TBP2. Ce projet s'intéresse en particulier au rôle et à la fonction de TBP2 et de la machinerie transcriptionnelle associée pour la transcription dans les ovocytes.

1. Identification de complexes similaires à TFIID contenant TBP2

Des études antérieures menées dans le laboratoire ont montré l'existence d'une machinerie basale de la transcription spécifique aux ovocytes. Afin de caractériser cette machinerie, nous avons utilisé la technique d'immunoprécipitation couplée à la spectrométrie de masse.

Dans un premier temps, nous avons généré et testé 10 anticorps polyclonaux de lapin dirigés contre la protéine TBP2 murine afin de sélectionner les meilleurs anticorps à utiliser pour les techniques de western-blot, d'immuno-précipitation (IP), d'immuno-fluorescence (IF) et d'immuno-précipitation de la chromatine (ChIP).

Dans une lignée de cellules 3T3 surexprimant TBP2 (lignée II10), j'ai d'abord recherché les protéines de TFIID potentiellement associées à TBP2. Il se trouve que TBP2 est retrouvé dans un complexe similaire à TFIID mais TBP2, contrairement à TBP, n'interagit pas avec TFIIB, BTAF1 et les sous-unités des machineries transcriptionnelles associées avec l'ARN Pol I et Pol III. En revanche, d'après les analyses

d'immunoprécipitation couplées à la chromatographie d'exclusion, TBP2 interagit fortement avec TFIIA au sein d'un même complexe.

Des analyses similaires dans les ovocytes, avec l'optimisation du protocole d'immunoprécipitation (*microIP*) à partir de 500 fois moins de matériel de départ, ont montré que TBP2 interagit avec TFIIA mais avec aucune sous-unité de TFIID. L'interaction entre TBP2 et TFIIA est maintenue même après déplétion de toute sous-unité de TFIID, par immunoprécipitation de TAF7 et de TAF10. Cela suggère donc que TBP2 et TFIIA forment également un complexe stable *in vivo*.

2. Identification des gènes régulés par TBP2 dans les ovocytes

Afin d'identifier les gènes régulés par TBP2 dans les ovocytes, j'ai comparé les profils d'enrichissement de TBP2 au niveau de la chromatine, obtenus avec les techniques de DamID et Cut&Run, avec les niveaux d'expression des gènes, obtenus par mRNA-seq et SLAM-seq.

i. RNA-seq et SLAM-seq dans les ovocytes

Les ARN collectés à partir d'ovocytes contrôles ou mutants pour *Tbp2* chez des souris âgées de 7 ou de 4 jours ont été séquencés par RNA-seq. Les analyses ont révélé que les niveaux d'expression de plus d'un millier de gènes sont diminués de plus de deux fois à 7 et 14 jours après la naissance. Parmi ces gènes, de nombreux rétrovirus endogènes sont présents, dont notamment les MaLR. Afin de confirmer ces résultats, j'ai analysé la transcription naissante grâce à la technique de SLAM-seq. Les résultats sont en cours d'analyse.

ii. DamID-seq et Cut&Run pour TBP2

Les techniques de DamID-seq et de ChIP-seq pour l'analyse du profil d'enrichissement de TBP2 dans les cellules surexprimant TBP2 ont montré que TBP2 lie les séquences appartenant à plus de 4000 gènes. Grâce à la collaboration avec Màté

Borsos du laboratoire de Maria Elena Torres Padilla, nous avons procédé à l'injection de l'ARNm codant la protéine fusion Dam-TBP2 dans des ovocytes collectés à partir de femelles âgées de 7 ou 14 jours. L'analyse des résultats sont en cours, tout comme ceux obtenus avec la technique de Cut&Run.

3. Importance du remplacement de TBP par TBP2 pour la transcription dans les ovocytes

Dans le but de comprendre l'importance de TBP2 pour la transcription dans les ovocytes lorsque TBP est absente, nous avons utilisé un système permettant de forcer l'expression de TBP dans les ovocytes. La séquence codante de TBP a été insérée au niveau du locus de *Tbp2* afin d'exprimer TBP à la place de TBP2 dans les ovocytes. Cette expérience a été infructueuse, les souris génétiquement modifiées avec cette construction étant stériles, indiquant que la régulation de l'expression de *Tbp2* est complexe.

Conclusions

Dans cette étude, nous avons analysé la fonction de TBP2 pour l'initiation de la transcription au cours de l'ovogenèse. Premièrement, nous avons démontré ici que TBP2 n'est pas présent au sein du complexe TFIID *in vivo*. Deuxièmement, TBP2 régule l'expression de très nombreux gènes en contactant physiquement de nombreux gènes. Afin de confirmer ces résultats, nous avons aussi employé les techniques de Dam-ID-seq et de Cut&Run dont les résultats sont en cours d'analyse. Par ailleurs, afin de distinguer la contribution de TBP2 pour l'initiation de la transcription et la stabilité des transcrits, liée à la grande quantité de transcrits contenue dans les ovocytes, nous avons analysé la transcription naissante en utilisant la technique de SLAM-seq. Les résultats obtenus ici montrent l'existence d'une machinerie transcriptionnelle spécifique dans les ovocytes, différente de TFIID pour la transcription.

Bibliographie:

1. Orphanides, G. et al. *Genes Dev*, 1996. 10, 2657-83.
2. Tora, L. *Genes Dev*, 2002. 16, 673-5.
3. Muller, F. & Tora, L. *EMBO J*, 2004. 23, 2-8.
4. Zhang, D. et al. *Science*, 2001. 292, 1153-5.
5. Bartfai, R. et al. *Current Biology*, 2004. 14, 593-598.
6. Jallow, Z. et al. *Proc Natl Acad Sci U S A*, 2004. 101, 13525-30.
7. Deato, M.D.E. et al. *Molecular Cell*, 2008. 32, 96-105.
8. Gazdag, E. et al. *Reproduction*, 2007. 134, 51-62.

Publications :

Sarina Ravens, Changwei Yu, Tao Ye, Matthieu Stierle and Laszlo Tora. Tip60 complex binds to active Pol II promoters and a subset of enhancers and co‐regulates the c‐Myc network in mouseembryonic stem cells. Epigenetics Chromatin. 2015 Nov 6; 8:45

Congrès :

Changwei YU, Máté BORSOS, Tao YE, Maria-Elena Torres-Padilla, Stéphane D. VINCENT, Lászlo TORA. An oocyte-specific TBP2-containing transcription machinery drives maternal mRNA expression during mouse oogenesis. Poster presentation. Stem Cell Conference Basel 2018 « Stem Cell Dynamics Throughout Life: From Development to the Adult ». 29-31 August 2018, Basel, Switzerland

Changwei YU, Máté BORSOS, Tao YE, Maria-Elena Torres-Padilla, Stéphane D. VINCENT, Lászlo TORA. An oocyte-specific TBP2-containing transcription machinery drives maternal mRNA expression in the mouse. Poster presentation. CSHL Epigenetics & Chromatin. September 11 - 15, 2018, Cold Spring Harbor, New York, USA

Table of Figures

Figures of introduction

Table of Tables

Introduction

In eukaryotic cells, the nuclear genome is organized into chromatin which comprises of DNA, RNA and associated proteins (Mondal et al., 2010; Yadav et al., 2018). Transcription is the process of decoding the genetic information from DNA into RNA by RNA polymerase (RNA Pol) enzymes. Typically, the eukaryotic nuclear genome is transcribed by three ubiquitous multisubunit complexes: RNA Pol I, II and III (except in higher plants also by two extra non-essential RNA polymerases, Pol IVa and Pol IVb (Pikaard et al., 2008)).

The three polymerases were first described after their chromatographic separation (Roeder et al., 1969, 1970). Subsequent studies of their sensitivity to the toxin α-amanitin revealed that each RNA Pol is responsible for the synthesis of different classes of RNA (Seifart et al., 1969; Kedinger et al., 1970; Lindell et al., 1970; Zylber et al., 1971; Weinmann et al., 1974). RNA Pol I synthesizes the 47S ribosomal RNA (rRNA) precursor, which could be processed into three mature rRNAs (28S, 18S and 5.8S), while 5S rRNA, transfer RNAs (tRNAs), 7SL RNA, U6 small nuclear RNA (snRNA) and a few other small stable RNAs are synthesized by RNA Pol III. Together the activities of RNA Pol I and Pol III transcription dominate over 80% of total RNA synthesis in growing cells (Paule et al., 2000). Finally, RNA Pol II synthesizes precursors of messenger RNAs (mRNAs), most snRNAs, microRNAs and long noncoding RNAs.

Gene expression is regulated primarily at the level of transcription and underlies all life processes. How transcription is regulated in a time-precise and cell type-specific manner is thus a central question in biology. Due to the fact that all protein-coding genes are transcribed by RNA Pol II, regulation of RNA Pol II transcription is one of the most important steps in controlling of cell identity, growth, differentiation, development, homeostasis and pathologies.

The focus of my introduction mainly concerns the mechanisms of RNA Pol II-Mediated transcription. In the first chapter, I will describe chromatin organization, plasticity, dynamics during development and their impacts on transcription. Secondly, I will detail how RNA Pol II transcription is tightly controlled at different levels with special emphasis on the formation of the pre-initiation complex (PIC). Thirdly, I will explore the cell-type-specific transcription machinery, and particularly I will highlight the diversification of the basal transcription machinery. Finally, I will present advances in transcription regulation during oogenesis.

1. Chromatin dynamics and transcription regulation

1.1 The hierarchy of chromatin organization

Eukaryotic cells compact their nuclear DNA in the nucleus through hierarchical levels of chromatin organization, which has profound effects on gene regulation and activity, by modulating the accessibility of DNA to the regulatory factors and elements.

The basic unit of chromatin is the nucleosome, which is composed of a core particle with 147 base pairs (bp) of DNA wrapped around a histone octamer and a segment of 20 to 90 bp linker DNA that can bind to histone H1 (Olins et al., 2003; Richmond et al., 2003; Szerlong et al., 2011). The histone octamer consists of two copies of histones H2A, H2B, H3 and H4. Non-condensed nucleosomes can form “beads-on-a-string” structure (Olins et al., 1976). The chromatin filaments further coil to reach higher-level structures (***Figure 1-1***).

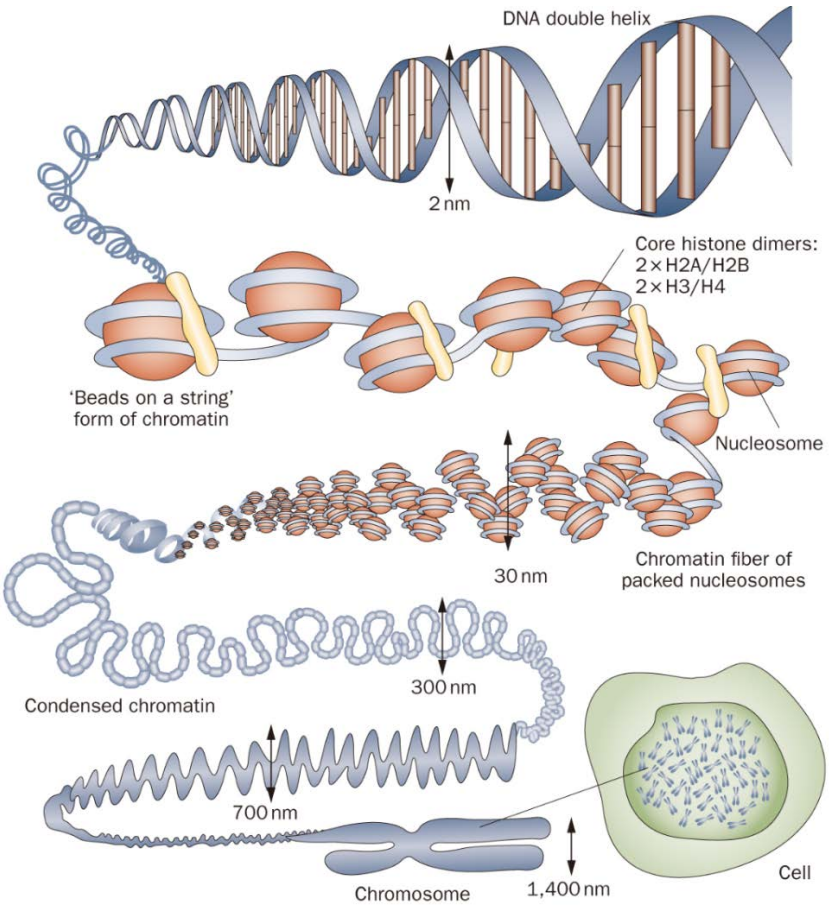


Figure 1-1: The hierarchical model of chromatin organization. DNA is wrapped around nucleosomes to form chromatin chains and fibers, which can undergo further compaction to fit into the nucleus. From (Tonna et al., 2010).

The long-standing textbook model of chromatin compaction is that 11 manometer (nm) nucleosome chains fold into 30nm chromatin fibers, subsequently into 300 nm to 700 nm fibers,

and ultimately mitotic chromosomes (Woodcock et al., 2010; Tonna et al., 2010) (***Figure 1-1***). However, this hierarchical model has been challenged by the findings that chromatin is a flexible and disordered 5 to 24 nm-diameter chain that is packed together at different 3D concentration densities in interphase nuclei and mitotic chromosomes (Fussner et al., 2012; Ou et al., 2017).

Traditionally, chromatin is divided into two structurally and functionally distinguishable territories: heterochromatin and euchromatin (Babu et al., 1987; Huisinga et al., 2006; Tamaru, 2010). Heterochromatin was originally defined as the chromatin that remains condensed and deeply stained at interphase. It typically refers to the highly condensed, gene-poor and less transcriptionally active regions of chromatin, and it has been further subdivided into two subsets, the permanently condensed constitutive heterochromatin and facultative heterochromatin, which may change its state of condensation during development (Babu et al., 1987; Wegel et al., 2005). The euchromatin is less condensed, gene-rich, and transcription-prone. The chromatin can also be classified into five principal types defined by unique combinations of proteins (Filion et al., 2010).

Further refinement of genomic approaches for mapping chromatin properties and chromatin interactions facilitated the generation of numerous high-resolution genome-wide maps, which have provided us a better understanding of chromatin spatial organization (***Figure 1-2***) with chromatin loops, topologically associating domains (TADs), A/B compartments and chromosome territories (Bickmore et al., 2013; Gibcus et al., 2013; Ea et al., 2015; Pombo et al., 2015; Sexton et al., 2015; Bonev et al., 2016; Schmitt et al., 2016; Dekker et al., 2017).

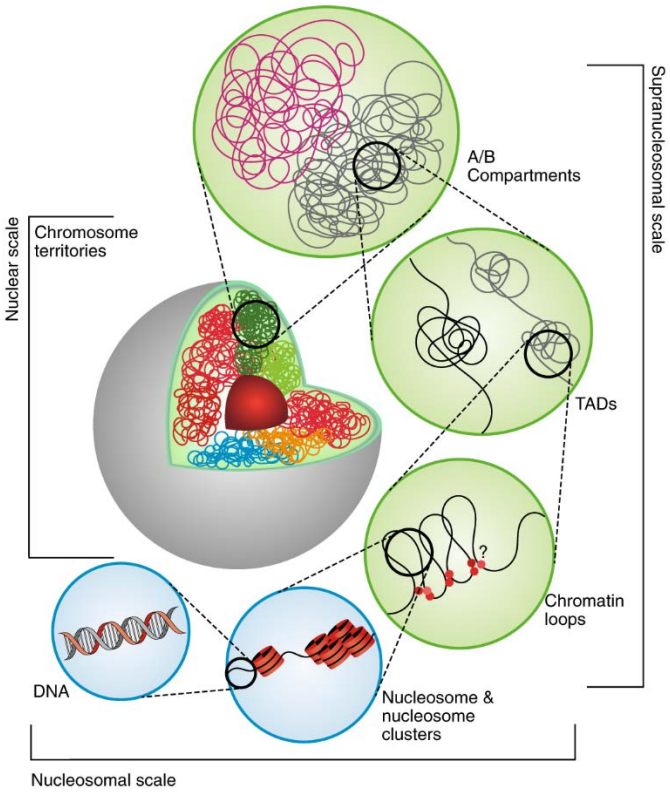


Figure 1-2: 3D genome organization. Nucleosome contacts form clusters and fibers, which could further engage in longer distance chromatin loops. Chromatin looping gives rise to TADs. Associations among TADs form A/B compartments, which generally reflect euchromatin and heterochromatin, respectively. Coalescence of A/B compartments in the same chromosome makes up chromosome territories. From (Dogan et al., 2018).

1.2 Transcription-relevant epigenetic regulations

Epigenetics was defined as “a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence” (Berger et al., 2009). Epigenetic regulations, including DNA methylation, histone post-translational modifications (PTMs), chromatin remodeling and histone variants can alter or even reshape the chromatin landscape locally at the individual gene level as well as globally across the epigenome, therefore modulating the genome accessibility and activity.

1.2.1 DNA methylation

Many eukaryotic genomes contain DNA methylation such as 5-methylcytosine (5mC) and N6-methyladenine (6mA) at different abundance levels (Li et al., 2014; Luo et al., 2017). In mammals, the major form is 5mC, which is mainly found in CpG dinucleotides and generally associated with transcriptional silencing (Zemach et al., 2010; Schubeler, 2015). 5mC can change the functional state of regulatory regions and thus play important roles in various biological processes including genomic imprinting, X chromosome inactivation, repetitive elements silencing and transcription regulation (Bird, 2002; Jones, 2012; Neri et al., 2017).

The mammalian DNA methylation pattern is established during embryonic development by *de novo* DNA methyltransferases (DNMTs) DNMT3A and DNMT3B, in combination with the cofactor DNMT3L (Okano et al., 1998; Okano et al., 1999; Bourc'his et al., 2001). Of note, a rodent DNMT3C methylates the promoters of young retrotransposon during spermatogenesis (Barau et al., 2016). 5mC at the CpG dinucleotides can be maintained during cell division by DNMT1, which recognizes and completes hemi-methylated CpG sites with its functional partner (Hermann et al., 2004; Bostick et al., 2007; Sharif et al., 2007; Song et al., 2011).

5mC is reversible (***Figure 1-3***), and it can be removed either by passive demethylation through imperfect maintenance (Chen et al., 2003a) or by active DNA demethylation through ten-eleven translocation (TET) family of proteins (Tahiliani et al., 2009; Wu et al., 2017).

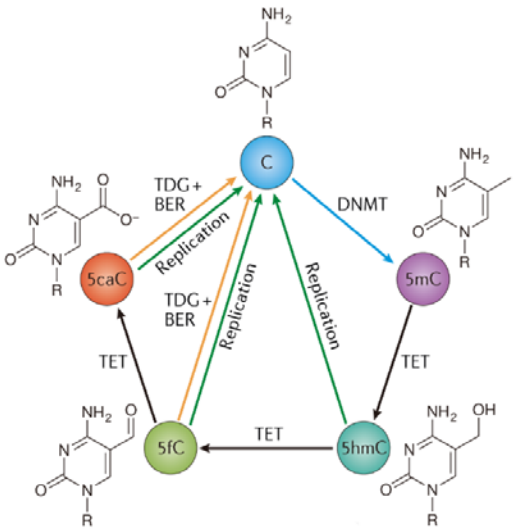


Figure 1-3: Cycle of active DNA demethylation.
DNMTs convert cytosine to 5mC. TET proteins convert 5mC back to cytosine by iterative oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). 5fC or 5caC can be efficiently removed by thymine DNA glycosylase (TDG) coupled with base excision repair (BER). From (Wu et al., 2017).

Whole-genome analysis of DNA methylation have revealed the dynamics of 5mC at different gene regulatory regions as well as in different cell types and developmental stages, and this indicates that 5mC is more than a repressive regulator of promoter activity, and that it has broader regulatory roles in development and disease (Ziller et al., 2013; Schultz et al., 2015; Luo et al., 2018).

1.2.2 Histone modifications

Both the protruding N-terminal tails and the globular domains of histones can carry post-translational modifications (PTMs) that include acetylation, methylation, phosphorylation, ubiquitinylation, citrullination, SUMOylation, ADP ribosylation, deamination, propionylation and butyrylation (Kouzarides, 2007; Kebede et al., 2015).

Histone PTMs can affect the chromatin structure and function as platforms for the recruitment of specific effector proteins (Bannister et al., 2011). Thus, histone PTMs play important roles at different levels of transcriptional regulation, from chromatin architecture to specific loci regulation through the recruitment of transcriptional regulators. The PTMs that have been reported to be involved in transcription regulation are listed in (***Table 1-1***).

Table 1-1. Overview of the transcription related histone PTMs. Adapted from (Lawrence et al., 2016).

Histone		Modifications	Roles	References
H2A	Histone Tail	H2AK4/5ac	Mitosis; chromatin assembly	(Barber et al., 2004)
		H2AK4/5ac	Transcriptional activation	(Fusauchi et al., 1984)
		H2AK7ac	Transcriptional activation	(Suka et al., 2001)
		H2AK119P	Spermatogenesis	(Baarends et al., 2007)
		H2AK119ub	Transcriptional repression	(Wang et al., 2004)
	Globular Domain	H2AQ105	Enriched over RNA Pol I transcribed genes	(Tessarz et al., 2014)
		H2AK119ub	Linked to Polycomb-mediated gene silencing	(Wang et al., 2004)
H2B	Histone Tail	H2BS33P	Transcriptional activation	(Maile et al., 2004)
		H2BK5ac	Transcriptional activation	(Golebiowski et al., 2005)
		H2BK11/12ac	Transcriptional activation	(Suka et al., 2001)
		H2BK15/16ac	Transcriptional activation	(Suka et al., 2001)
		H2BK20ac	Transcriptional activation	(Golebiowski et al., 2005)
		H2BK120ub	Spermatogenesis / meiosis	(Baarends et al., 2007)
		H2BK123ub	Transcriptional activation	(Robzyk et al., 2000)
		H3K4me1	Enhancer mark	(Calo et al., 2013)

H3	Histone Tail	H3K4me2	Linked to active transcription	(Liang et al., 2004)
		H3K4me3	Transcription elongation; active chromatin	(Santos-Rosa et al., 2002)
		H3K9me1/2/3	Transcriptional repression	(Wang et al., 2008)
		H3R17me1/2	Transcriptional activation	(Bauer et al., 2002)
		H3K27me2	Transcriptional repression	(Wang et al., 2008)
		H3K27me3	Transcriptional silencing; bivalent genes	(Bernstein et al., 2006)
		H3K36me3	Transcriptional elongation	(Bell et al., 2007)
		H3K4ac	Transcriptional activation	(Strahl et al., 1999)
		H3K9ac	Histone deposition; transcriptional activation	(Suka et al., 2001)
		H3K14ac	Transcriptional activation	(Agalioti et al., 2002)
		H1K18ac	Transcriptional activation; DNA repair / replication	(Wang et al., 2008)
		H3K23ac	Transcriptional activation	(Tsai et al., 2010)
		H3K27ac	Active enhancer mark	(Calo et al., 2013)
		H3S10P	Mitosis; meiosis; transcriptional activation	(Lee et al., 2008)
		H3S28P	Mitosis; transcriptional activation	(Lau et al., 2011)
	Globular Domain	H3Y41	Prevents HP1α binding to chromatin	(Dawson et al., 2009)
		H3R42me2a	Positive transcriptional effects	(Casadio et al., 2013)
		H3K56me3	Conserved heterochromatin mark	(Jack et al., 2013)
		H3K64ac	Facilitates transcription	(Di Cerbo et al., 2014)
		H3K64me3	Enriched at pericentric heterochromatin	(Daujat et al., 2009)
		H3K79me1/2/3	Active chromatin; Telomeric silencing	(Lawrence et al., 2016)
		H3T118	Transcriptional activation; Nucleosome remodeling	(North et al., 2011)
		H3K122	Transcriptional activation; Nucleosome eviction	(Tropberger et al., 2013)
H4	Histone Tail	H4R3me	Transcriptional activation	(Wang et al., 2001a)
		H4K20me1	Transcriptional silencing	(Nishioka et al., 2002)
		H4K20me3	Heterochromatin	(Schotta et al., 2004)
		H4K5ac	Histone deposition; transcriptional activation	(Suka et al., 2001)
		H4K8ac	Transcriptional activation and elongation	(Suka et al., 2001)
		H4K12ac	Telomeric silencing; transcriptional activation	(Suka et al., 2001)
		H4K16ac	Transcriptional activation; DNA repair	(Akhtar et al., 2000)
	Globular Domain	H4S47P	H3.3 deposition regulation	(Zhang et al., 2013b)
		H4K91ac	Enriched at TSS of active and poised genes	(Wang et al., 2008)

Histone PTMs can be reversible by histone-modifying enzymes including a rich repertoire of writers, readers and erasers (Allis et al., 2016). For instance, histone acetyltransferases (Marmorstein et al., 2014) and deacetylases (Seto et al., 2014) for histone acetylation, and histone methyltransferases and demethylases for histone methylation (Greer et al., 2012; Hyun et al., 2017). Thus the dynamics of histone PTMs adds another layer of regulation for accurate gene expression.

1.2.3 Chromatin remodelers

ATP-dependent chromatin-remodeling complexes (remodelers) modulate the chromatin structure and accessibility through loading, moving, destabilizing, ejecting or restructuring the nucleosomes. They are thereby essential regulators of all chromosomal processes, including transcription.

Phylogenetically, transcription-relevant chromatin remodelers can be classified into four subfamilies: imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), switch/sucrose non-fermentable (SWI/SNF) and INO80. Functionally, chromatin remodelers can be classified as assembly remodelers, access remodelers and editing remodelers, according to their specialty or preference to conduct function of nucleosome assembly and organization, chromatin access or nucleosome editing (histone exchange), respectively (Flaus et al., 2006; Clapier et al., 2009; Hargreaves et al., 2011; Clapier et al., 2017).

Most ISWI subfamily chromatin remodelers assemble nucleosomes and ensure the proper density and spacing of nucleosomes to limit chromatin accessibility, thus contribute to gene repression, except the NURF complex, which has accessory subunits to promote transcription (Xiao et al., 2001; Clapier et al., 2017). CHD subfamily remodelers are involved in all three general remodeling functions (Lusser et al., 2005; Murawska et al., 2011; Konev et al., 2007). SWI/SNF subfamily remodelers typically facilitate chromatin access as they can slide nucleosomes, evict histone dimers or eject octamers, which usually promote gene expression by exposing the binding sites for transcription factors (Boeger et al., 2004). INO80 subfamily remodelers mainly carry out nucleosome editing through replication-independent nucleosome histone exchange to remove a particular histone and replace it with either a canonical or a variant histone (Mizuguchi et al., 2004; Goldberg et al., 2010).

Together, chromatin remodelers ensure the proper density and spacing of nucleosomes on the chromatin, while still ensure rapid access to regulatory factors to specific loci for other genome activities. Chromatin remodelers provide an important mechanism for chromatin structure and DNA accessibility modulation, thus adding another layer of control for transcription regulation.

1.2.4 Histone variants and chaperones

Histone variants are non-allelic isoforms of their corresponding canonical histones counterparts with differences in their primary sequence, expression level and expression timing.

In human, eight H2A variants, two testis-specific H2B variants and six H3 variants have been identified (***Figure 1-4***) (Buschbeck et al., 2017). The variants show different genomic distribution and physiological roles. H2A.Z.1, H3.3 and CENP-A are essential for mouse embryonic development as knockout mice are embryonic lethal, while macroH2A and H2A.X are not essential for development but KO mice display defects (***Table 1-2***).

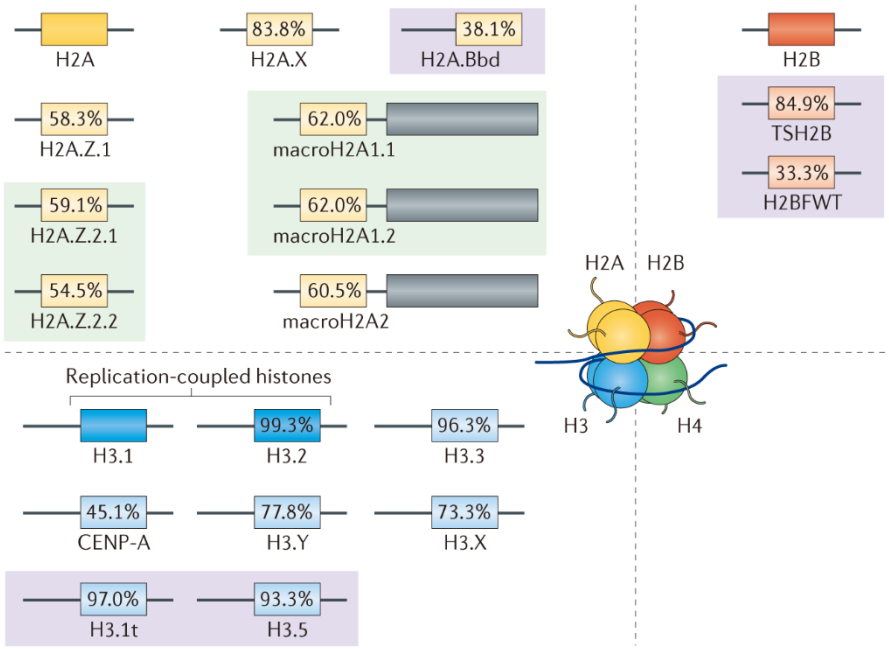


Figure 1-4: Histone variants of human core histones. Light purple and green represent testis-specific variants and alternative splice isoforms, respectively. From (Buschbeck et al., 2017).

Histone chaperones are a group of proteins that bind to histones and regulate nucleosome assembly (Burgess et al., 2013). They escort histones throughout their cellular life and play a crucial role in histones folding, transport, oligomerization, assembly into nucleosome and turnover on the chromatin (Gurard-Levin et al., 2014; Hammond et al., 2017). Thus histone chaperones are important actors in all chromosomal process, including transcription.

In fact, the deposition and function of histone variants are strongly linked to specialized chaperones (***Table1-3***) and remodelers (Clapier et al., 2017). For example, H3.3 turnover is governed by the histone regulator A (HIRA) complex in transcribed regions and promoters, as

well as by DAXX–ATRX in heterochromatic regions (Drane et al., 2010; Goldberg et al., 2010). HIRA facilitates H3.3–H4 deposition at promoters and in gene bodies by interacting with transcription factors and RNA Pol II, respectively (Banaszynski et al., 2013; Pchelintsev et al., 2013; Soni et al., 2014).

Table 1-2: Function of histone variants. Adapted from (Buschbeck et al., 2017).

Histone Variant	Distribution	Function
H2A.X	Genome-wide	DNA damage response (Rogakou et al., 1998), knockout (KO) mice are viable but show male infertility (Celeste et al., 2002).
H2A.Z.1	Regulatory regions (promoters and enhancers) and heterochromatin	Facilitates the binding of regulatory complexes and increases nucleosome dynamics (Hu et al., 2013a), KO mice are embryonic lethal (Faast et al., 2001).
H2A.Z.2		Facilitates the binding of regulatory complexes (Vardabasso et al., 2015).
MacroH2A1	Facultative and constitutive heterochromatin	Gene repression and signal-induced gene activation (Changolkar et al., 2007), KO mice are viable but show metabolic defects (Boulard et al., 2010).
MacroH2A2		Gene repression and signal-induced gene activation, macroH2A KO mice show reduced growth (Pehrson et al., 2014); MacroH2A2 KO zebrafishes show malformations (Buschbeck et al., 2009).
H2A.Bbd	Euchromatin (In testis and brain)	Associates with sites of active transcription and replication and increases nucleosome dynamics (Tolstorukov et al., 2012; Sansoni et al., 2014).
TH2A and TH2B	(Testis and oocytes)	Paternal genome activation in zygote (Shinagawa et al., 2014).
H3.3	*cis*-regulatory elements, gene bodies and repetitive, heterochromatic sequences	Gene activation, retroviral element silencing, genome integrity and the establishment of heterochromatin (Goldberg et al., 2010; Banaszynski et al., 2013; Elsasser et al., 2015; Jang et al., 2015; Nashun et al., 2015), double KO of *H3f3a* and *H3f3b* are early embryonic lethal, *H3f3a* knockouts are viable but show male subfertility, *H3f3b* knockouts are growth-deficient and died at birth. *H3f3b* heterozygotes also show males infertility (Tang et al., 2015a).
CENP-A	Centromere	Kinetochore attachment, chromosome segregation (Lacoste et al., 2014), KO mice are lethal at E3.5–E8.5 in mice (Howman et al., 2000).

Histone variants, together with histone chaperones, serve as another basis for epigenetic regulation by profoundly changing nucleosome properties and punctuating the chromatin, which could alter chromatin structure, nucleosome stability, DNA accessibility, and, ultimately, transcription.

Table 1-3: Histone chaperones involved in histone deposition. Adapted from (Gurard-Levin et al., 2014) and (Hammond et al., 2017).

Histone chaperones	Histone preference	Complex(s)	Main function(s)
ASF1A/B	H3.1-, H3.2-, H3.3-H4	Multiple	Histone donor for CAF-1 and HIRA (Tyler et al., 1999; Munakata et al., 2000)
CHAF1A / p150	H3.1-H4	CAF1 (RBAP48 can also form into HDAC, NuRF, NuRD, and PRC2)	Deposition factor coupled to DNA synthesis: replication, DNA repair (Smith et al., 1989; Volk et al., 2015)
CHAF1B / p60	H3.1-H4		
RBAP48	H3.1-, H3.2-, H3.3-H4		
RBAP46	H3-H4	HAT, HDAC, NuRF, NuRD, PRC2	Interchangeable with RBAP48 in their complexes, apart from HAT1 (RBAP46) and CAF1 (RBAP48).
DAXX	H3.3-H4	DAXX-ATRX	Deposition factor independent of DNA synthesis (Drane et al., 2010; Goldberg et al., 2010)
ATRX	ND		
HIRA	H3.3-H4	HIRA	Deposition factor independent of DNA synthesis (Ray-Gallet et al., 2002; Ricketts et al., 2015; Nashun et al., 2015)
UBN1	H3.3-H4		
CABIN1	H3.3-H4		
HJURP	CENP‑A-H4 (Cse4‑H4)		Deposition factor, centromere maintenance (Dunleavy et al., 2009; Foltz et al., 2009)
MCM2	CENP−A-, H3.1-, H3.2-, H3.3-H4	MCM2-7 complex	Symmetric inheritance of modified histones during DNA replication (Huang et al., 2015; Petryk et al., 2018)
s/tNASP	H3.1-, H3.2-, H3.3-H4, H1	HAT	Protects H3–H4 from degradation (Cook et al., 2011)
SPT2	H3-H4	ND	H3/H4 tetramer maintenance during transcription (Chen et al., 2015)
SPT6	H3-H4	ND	Nucleosome reassembly during gene transcription (Bortvin et al., 1996; Dronamraju et al., 2018)
SPT16	H2A-H2B, H3-H4	FACT	Replication-coupled and-independent histone deposition, H2A–H2B eviction during transcription (Belotserkovskaya et al., 2003; Kemble et al., 2015; Tsunaka et al., 2016; Yang et al., 2016)
SSRP1	H2A-H2B, H3-H4		
ANP32E	H2A.Z-H2B	P400-TIP60	H2A.Z-H2B eviction (Mao et al., 2014; Obri et al., 2014)
YL1 / VPS72	H2A.Z-H2B	SRCAP/SWR‑C, P400-TIP60	H2A.Z-H2B deposition (Liang et al., 2016; Latrick et al., 2016)
NAP1L1‑6	H2A-, H2A.Z-H2B, H3-H4, H1	Nuclear import importin 9	Nuclear transport, replication, transcription (Zlatanova et al., 2007; Kuryan et al., 2012; Aguilar-Gurrieri et al., 2016)
Nucleolin	H2A-H2B, H1	SWAP	Transcription elongation (Angelov et al., 2006)

DNA and histone modifiers, histone variants, histone chaperones and chromatin remodelers, together with methylated DNA binding proteins and histone PTMs readers, endow chromatin with plasticity and establish a sophisticated regulatory network to modulate chromatin structure and properties (***Figure 1-5***). The chromatin dynamics could ultimately affect transcription by altering the accessibility of DNA to transcription factors and transcription machinery.

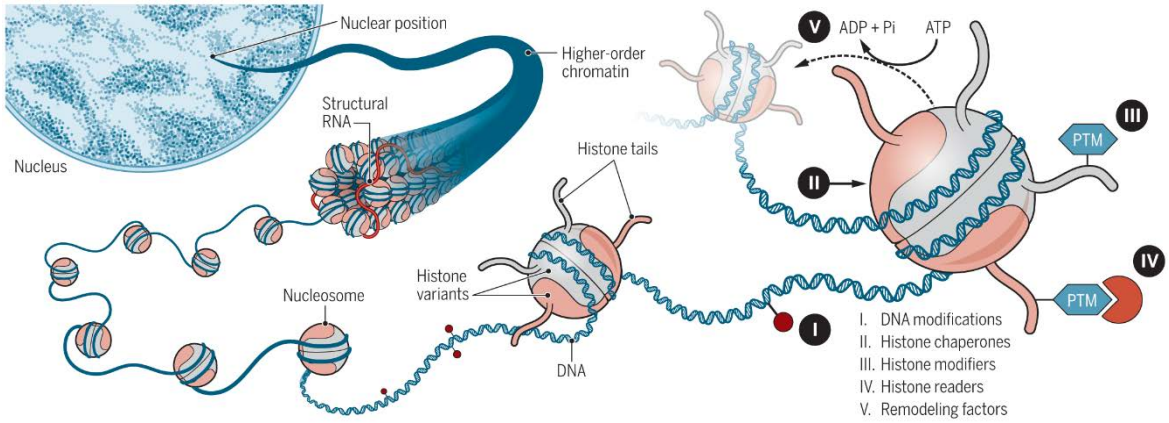


Figure 1-5: Chromatin plasticity. Variations by dynamic combinations of DNA methylation, histone variants incorporation and histone PTMs, together with the modifiers, readers and remodelers, enable chromatin plasticity. From (Yadav et al., 2018).

1.3 Chromatin dynamics during mouse germ cell development

In metazoan, germ cells represents a continuous cellular link that ensure the perpetuation of the genetic and epigenetic information across generations. After primordial germ cells (PGCs) specification and migration, germ cells acquire the competencies to create totipotency during spermatogenesis and oogenesis through elaborate transcription regulation, along with morphology changes and chromatin dynamics. Here I mainly focus on cytological aspects of germ cell development and the chromatin dynamics.

1.3.1 Germ cell development

1.3.1.1 PGCs specification and migration

In mammals, PGCs, the precursors for both sperm and oocytes, are specified from a small population of proximal-posterior epiblast cells during embryonic development (Hayashi et al., 2007). The specification of PGCs is induced by signals from extra-embryonic tissues (***Figure 1-6***), other than asymmetric inheritance of maternal cytoplasmic factors (Lawson et al., 1994; Saitou et al., 2003; Ohinata et al., 2009).

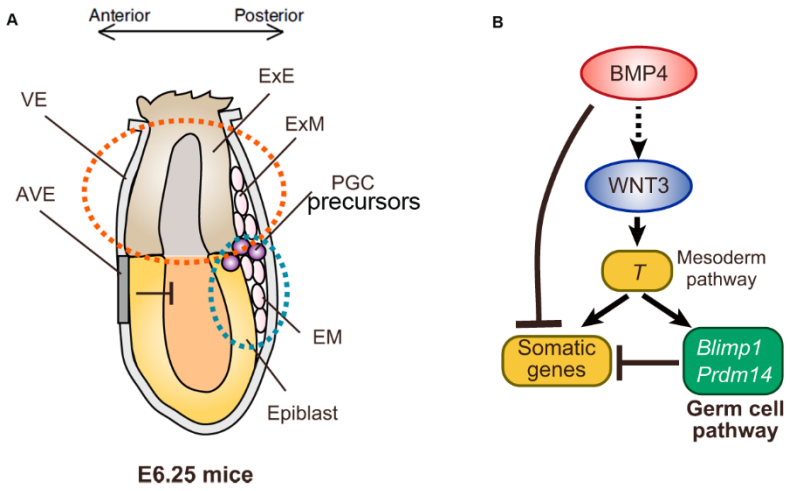


Figure 1-6: Proposed mechanisms for PGC specification in mice. (A) Schematics of mouse embryo at the stage of germ cell specification. Red and blue dotted circles show the BMP and the WNT signals, respectively. VE, visceral endoderm; AVE, anterior visceral endoderm; ExE, extra-embryonic ectoderm; ExM, extra-embryonic mesoderm; EM, embryonic mesoderm. (B) Proposed model for mouse PGC specification (Aramaki et al., 2013). Adapted from (Saitou et al., 2016)

In response to bone morphogenetic protein (BMP) and WNT signals, the earliest known marker of PGC precursors, PR domain containing protein 1 (*Prdm1/Blimp1*) (Vincent et al., 2005), is initially expressed in approximately six cells in the proximal-posterior epiblast at embryonic day 6.25 (E6.25) (Ohinata et al., 2005). Expression of another key regulators for

PGC specification, *Prdm14*, is also induced by the signals and can be detected at ~ E6.5 in PRDM1-positive cells (Of note, *Prdm14* is transiently expressed in the inner cell mass (ICM) cells of blastocysts and silenced by E5.5 (Yamaji et al., 2008)). PRDM1 and PRDM14 further induce the expression of *AP2γ* at around E7.25 (Weber et al., 2010; Magnusdottir et al., 2013). Together, these three transcription factors form a core specification network that is necessary and sufficient for PGC induction (Magnusdottir et al., 2013; Tang et al., 2016).

PRDM1-positive PGC precursors form a cluster of ~20 cells at E6.75 (Lawson et al., 1994; Kurimoto et al., 2008). These cells first become identifiable as PGCs as a cluster of ~40 alkaline phosphatase (AP)-positive cells at the base of incipient allantois at ~E7.25, and expression of *Dppa3*/*Stella* also marks the establishment of PGCs (***Figure 1-7***) (Ginsburg et al., 1990; Saitou et al., 2002; Sato et al., 2002; Saitou et al., 2012b). PGCs migrate to the developing hindgut endoderm at ~E7.75, and the majority of migrating PGCs are arrested at the G2 phase between E7.75 and E8.75. They continue to migrate through the hindgut endoderm and mesentery individually while keep proliferating, and finally colonize the genital ridges at ~E10.5 with ~1,000 cells (Tam et al., 1981; Molyneaux et al., 2001; Seki et al., 2007; Saitou et al., 2016).

PGCs exit migratory state and initiate sexual differentiation while continue to proliferate between E10.5 and E 12.5 (Lesch et al., 2012). In the male genital ridge, the PGCs are committed to spermatogenesis at ~E12.5 and gradually enter into mitotic arrest at ~E13.5 until after birth (McLaren, 2003; Western et al., 2008). Whereas, ~12,000 female PGCs are committed to oogenesis at ~E13.5 and subsequently enter the first meiotic prophase in the mouse (Pepling et al., 2001; McLaren, 2003). This marks the end of PGC stage of germ cell development.

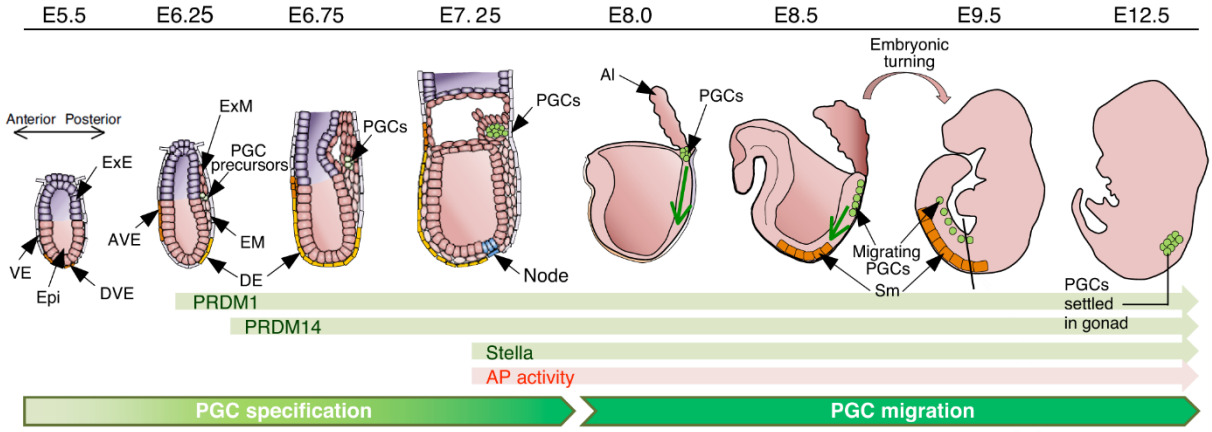


Figure 1-7: Scheme of mouse PGCs specification and migration. Epi, epiblast; DVE, distal visceral endoderm; Sm, somite; Al, allantois. From (Saitou et al., 2012a)

During PGC specification and migration, genome-wide epigenetic reprogramming occurs, including DNA demethylation and histone modification changes, which will be described later

together with spermatogenesis and oogenesis (see ***section 1.3.2***).

1.3.1.2 Spermatogenesis

Spermatogenesis is the developmental process during which diploid spermatogonia generate haploid spermatozoa through mitotic expansions, meiotic reduction divisions and spermiogenesis (***Figure 1-8***) (Griswold, 2016; Chen et al., 2018b).

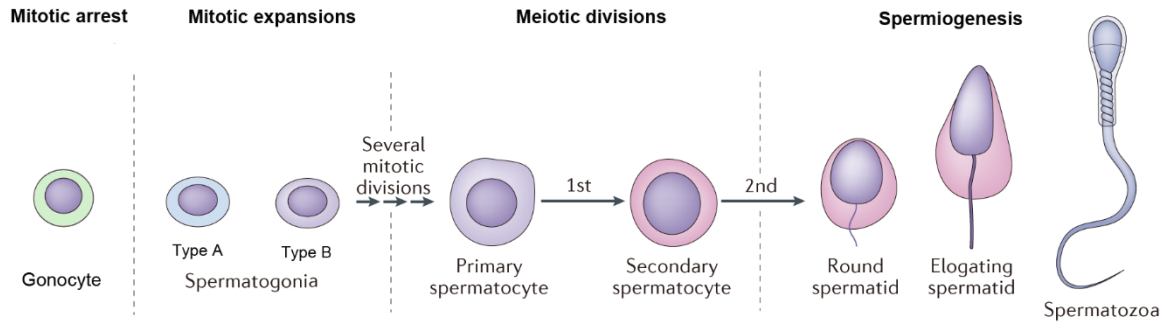


Figure 1-8: Overview of mouse spermatogenesis. Adapted from (Schagdarsurengin et al., 2016)

In mouse embryonic testis, mitotically arrested germ cells lost their alkaline phosphatase expression at ~14.5 and differentiate into gonocytes (also called pro-spermatogonia) (Culty, 2009). At around postnatal day (P) 5, gonocytes resume active proliferation, and some of them are recruited as spermatogonial stem cells (SSCs), which are the singly isolated predominant stem cells during spermatogenesis (Yoshida, 2010; Saitou et al., 2012b). The SSCs can divide to form a pair of cells connected by an intercellular bridge termed A paired spermatogonia (Apr), and Apr can further divide to form A aligned cell syncytia of 4, 8 or 16 cells. This pool of undifferentiated spermatogonia differentiate into A1 differentiating spermatogonia without division and subsequently undergo 5 divisions to form B spermatogonia, which gives rise to two preleptotene spermatocytes through another mitosis (Spradling et al., 2011; Griswold, 2016).

Meiosis starts in preleptotene spermatocytes. Following DNA replication, synapsis and crossing over occur in the prolonged meiotic prophase I, which can be subdivided into leptotene, zygotene, pachytene and diplotene stages. After meiotic prophase I, spermatocytes undergo two rounds of chromosome segregation to produce haploid round spermatids. These round spermatids further undergo dramatic morphological and cytological changes to form mature spermatozoa through the process of spermiogenesis (Jan et al., 2012; Chen et al., 2018b).

Concomitant with cytological changes, epigenetic changes including paternal imprint acquisition, histones replacement by histone variants, transition proteins and protamines, also

occur at different stages of spermatogenesis.

1.3.1.3 Oogenesis

Oogenesis begins when female PGCs colonize the genital ridge, proliferate as oogonia and form germline cysts from ~E10.5 (Pepling, 2006). At ~E13.5, unlike male PGCs entering mitotic arrest, most female germ cells within germline cysts begin to enter meiosis and subsequently arrest in the diplotene stage of meiotic prophase I gradually from ~E17.5, and it takes several days until all oocytes are in diplotene arrest (***Figure 1-9***) (Pepling et al., 2001; Dutta et al., 2016).

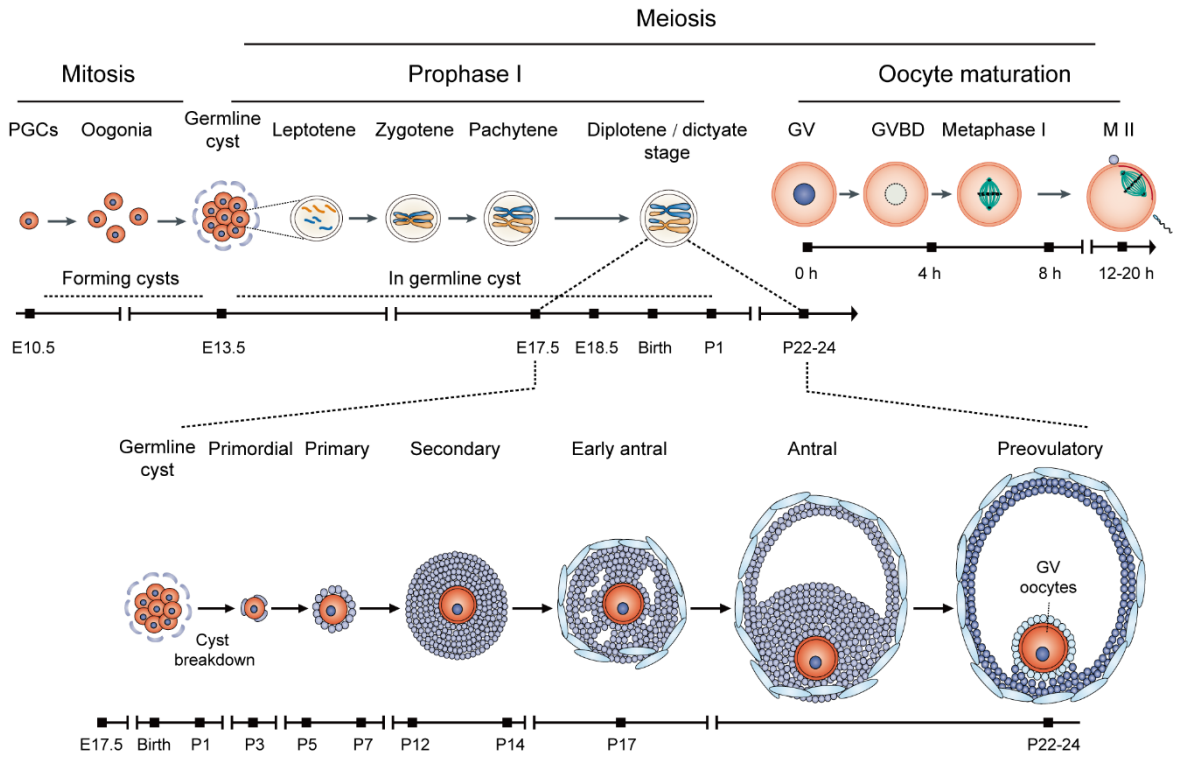


Figure 1-9: Schematic representation of oogenesis. Adapted from (Pepling, 2006; Racki et al., 2006; Sasaki et al., 2008; Li et al., 2013b).

After birth, germline cysts break apart as individual oocyte enclosed with several pre-granulosa cells to form dormant primordial follicles, and this process is accompanied by extensive germ cell loss, of which only the minority undergo apoptosis (Pepling et al., 2001). It is generally thought that this pool of oocytes in primordial follicles are finite and non-renewable after birth (Telfer et al., 2005; Albertini et al., 2015), although this theory has been challenged by studies showing the existence of oogonial stem cells (OSCs) in adult ovary (Johnson et al., 2004; White et al., 2012).

The majority of oocytes enclosed in primordial follicles are maintained in dormancy during reproductive life (Kim, 2012). These resting oocytes surrounded by flat granulosa cells in

primordial follicles are small in size, with a diameter of 12-20 μm, and can be routinely obtained from ovaries of P1 to P3 mice (Pedersen et al., 1968; Mangia et al., 1975).

Upon induction, a limited number of dormant primordial follicles are recruited into the growing pool and undergo further development progressively in an asynchronous manner. Activated primordial follicles first develop into primary follicles, in which granulosa cells increase in number and form single-layered cuboidal granulosa cells, followed by dramatic growth of oocytes and the initiation of zona pellucida formation. Growing oocytes in primary follicles with a diameter of 30 to 60 μm can be obtained from P5 to P7 (Mangia et al., 1975; Sanchez et al., 2012).

Secondary follicle formation proceeds with the generation of multiple layers of granulosa cells around the oocyte. In secondary follicles, the zona pellucida is completely formed, and the diameter of the oocytes is around 60 to 75 μm, and can be obtained at P12 to P14 (Mangia et al., 1975). Theca cell layer forms during the transition from the preantral to early antral stage (Orisaka et al., 2009), followed by the formation of antrum and cumulus oophorus in antral stage. In the preovulatory follicle, germinal vesicle (GV) oocyte and cumulus cells reside in the fluid-filled antrum (Racki et al., 2006; Paulini et al., 2014).

Upon hormonal signaling, oocyte enter the maturation stage with the germinal vesicle breakdown (GVBD), followed by meiosis I spindle assembly and chromosome migration. Subsequently, meiosis II starts and the mature oocyte arrests in metaphase II, which will be bypassed after fertilization (Li et al., 2013b).

1.3.2 Chromatin dynamic during germ cell development

Global epigenetic reprogramming occurs during germ cell development (***Figure 1-10***), including changes in DNA methylation, histone modifications, histone variants incorporation and chromatin organization.

1.3.2.1 Chromatin dynamics in PGCs

Mouse PGCs are specified from epiblast, whose DNA is hypermethylated. PGCs undergo genome-wide DNA demethylation as they migrate and colonize the genital ridge, which results in the erasure of genomic imprinting and X-chromosome reactivation. By ~E13.5, CpG methylation levels drop from ~70% in the epiblast to 14% and 7% in male and female PGCs, respectively (Seisenberger et al., 2012; Kobayashi et al., 2013). Of note, some repetitive elements, such as IAPs and LTR-ERV1, show resistance to this global methylation erasure (Hajkova et al., 2002; Guibert et al., 2012).

Along with DNA methylation, reorganization of repressive histone marks also occurs. Apart from widespread depletion of H3K9me2, H3K27me3 and H2A/H4R3me2s are enriched in PGCs, whereas H3K9me3 is retained relatively constant at centromeric heterochromatin (Seki et al., 2005; Ancelin et al., 2006). Moreover, a non-canonical form of H3K27me3 with broad distribution is acquired in PGCs (Sachs et al., 2013; Ng et al., 2013; Hammoud et al., 2014; Zheng et al., 2016; Hanna et al., 2018a).

Subsequently, DNA methylation re-establishment and chromatin changes occurred in a sex-specific manner.

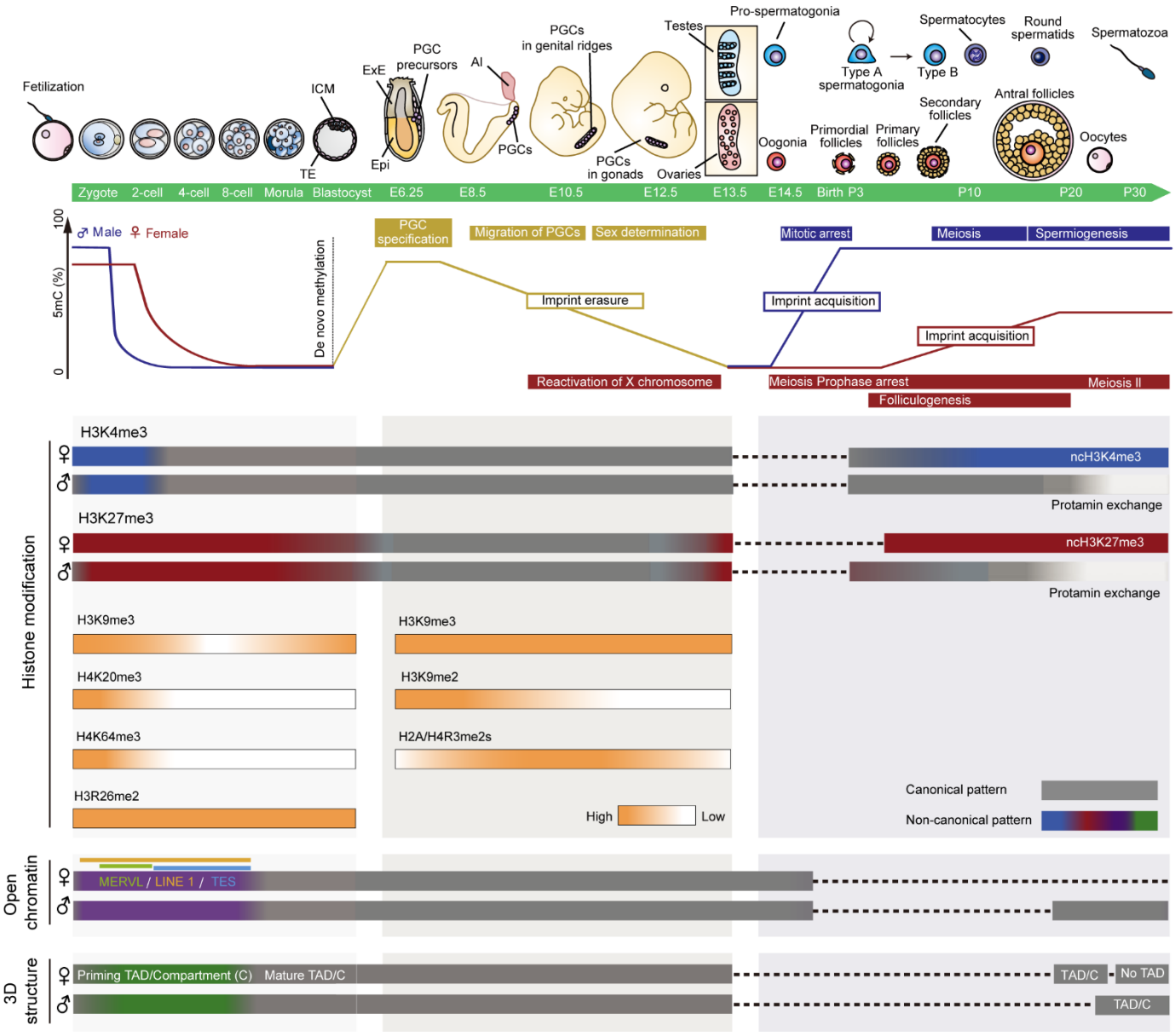


Figure 1-10: Chromatin dynamics during gametogenesis and embryogenesis. Adapted from (Saitou et al., 2012a; Burton et al., 2014; Saitou et al., 2016; Tang et al., 2016; Hanna et al., 2018a; Xu et al., 2018).

1.3.2.2 Chromatin dynamics during spermatogenesis

During spermatogenesis, DNA remethylation starts after E13.5 in gonocytes, by E16.5, global methylation levels already increase to ~50%, and the methylation patterns are fully established at birth (Seisenberger et al., 2012; Stewart et al., 2016).

Following meiotic divisions, spermatids undergo extensive nuclear and morphological changes including the histone-to-protamine transition, during which the majority of core histones are replaced sequentially, first by histone variants and transition proteins and subsequently by protamines (Bao et al., 2016). As a result, only 1% residual histones is retained in mature sperm.

Residual nucleosomes are largely composed of the H3.3 histone variant and H3K4me3 (Erkek et al., 2013), and they (or at least a subset of them) are enriched at CpG-rich sequences with low DNA methylation, although it has been reported that the majority of residual nucleosomes locate at gene-poor regions (Carone et al., 2014). Residual histones in sperm provide other ways for epigenetic inheritance through the male germline without changes of DNA methylation (Gill et al., 2012; Siklenka et al., 2015).

1.3.2.3 Chromatin dynamics during oogenesis

During oogenesis, *de novo* methylation only occurs after birth during oocyte growth, in a transcription-dependent manner, and is largely completed in GV oocyte at ~P21 (Hiura et al., 2006; Smallwood et al., 2011; Stewart et al., 2016). Oocyte acquires lower global methylation than sperm, but with non-canonical genome-wide distribution over transcribed gene bodies (Kobayashi et al., 2013; Veselovska et al., 2015). However, loss of DNA methylation through deletion of *Dnmt3a* or *Dnmt3L* has no effect on oogenesis (Bourc'his et al., 2001; Kaneda et al., 2004).

Intriguingly, a non-canonical pattern of H3K4me3 (ncH3K4me3) with broad peaks is formed during oocyte growth, and overlaps almost exclusively with partially methylated DNA domains (Zhang et al., 2016; Dahl et al., 2016; Liu et al., 2016; Hanna et al., 2018b). Formation of ncH3K4me3 is through the recruitment of MLL2 to unmethylated CpG-rich regions in a transcription-independent manner (Hanna et al., 2018b).

The non-canonical form of H3K27me3 that is gained in PGCs, with weak promoter enrichment and relatively high enrichment at non-promoter /distal regions, is present broadly at unmethylated genomic regions throughout oogenesis (Zheng et al., 2016).

Growing oocytes have relatively high levels of histone acetylation, and undergo abrupt de-acetylation during meiotic resumption (Kim et al., 2003). Double deletion of *Hdac1* and *Hdac2* impairs transcription and oocyte growth, leading to female sterility as follicle

development arrests at the secondary follicle stage (Ma et al., 2012; Ma et al., 2016).

Moreover, histone variant H3.3 is continuously deposited on the chromatin of growing oocyte by HIRA, which is essential for transcription regulation and *de novo* DNA methylation during oogenesis (Nashun et al., 2015). Besides, growing oocyte chromatin also contains histone variant macroH2A, which remains associated with maternal chromatin following fertilization (Chang et al., 2005).

During oocyte maturation, chromatin undergoes dramatic conformational changes in GV oocytes, from non-surrounded nucleolus (NSN) state to partially surrounded nucleolus (PSN) state, and eventually surrounded nucleolus (SN) state (Mattson et al., 1990; Zuccotti et al., 1995; Bogolyubov, 2018). Those conformational changes correlate with, but do not determine oocyte transcriptional activity: NSN oocyte shows transcriptional activity while SN oocyte is transcriptionally silenced (De La Fuente et al., 2004).

Hi-C studies has shown that chromosome interactions, such as TADs and chromosome loops, exist in GV oocytes and start to decrease with the NSN to SN transition (Flyamer et al., 2017). With resumption of meiosis, oocytes lose typical higher-order chromatin structures, including TADs and chromatin compartments. Instead, MII oocytes show a uniform interaction pattern along the entire chromosomes that appears to be locus-independent (Ke et al., 2017; Du et al., 2017).

Given the fact that H3K4me3 interacts with transcription regulators such as TAF3 (Vermeulen et al., 2007; van Ingen et al., 2008; Lauberth et al., 2013), it was proposed that ncH3K4me3 at distal sites could function as ‘sponges’ that absorb and sequester transcription factors (TFs) and regulators, therefore diluting transcription resources away from promoters to modulate transcription (Zhang et al., 2016). Moreover, the unusual epigenetic patterns in oocytes is not limited to H3K4me3.

In this scenario, the distinct epigenetic patterns in oocytes may confer oocytes different mechanisms of transcription regulation, even on the level of basal transcription machinery, together with specific transcription factors, facilitating the acquisition of competencies required for fertilization and embryogenesis.

1.3.2.4 Chromatin dynamics during early embryo development

Following fertilization, the paternal DNA methylation is rapidly lost mainly through a combination of TET3-mediated active demethylation and passive dilution, whereas maternal DNA methylation is mainly lost passively over cell divisions (Wu et al., 2017). However, the main decrease of 5mC in paternal genome occurs before early pronuclear stage (PN) 3, while 5hmC starts to accumulate only after main drop of 5mC has occurred (Santos et al., 2013; Amouroux et al., 2016), and ablation of maternal TET3 prevents accumulation of 5hmC but

does not influence early loss of paternal 5mC in zygote (Amouroux et al., 2016). Together, these findings suggest the existences of *de novo* methylation and of an alternative unknown mechanism for DNA demethylation in early mouse zygote (Amouroux et al., 2016). As a result, global DNA methylome is largely erased by blastocyst stage except in selected regions, including but not limited to, imprinted regions and some classes of repetitive elements (Smallwood et al., 2011; Lee et al., 2014). Subsequently, DNA methylation is re-established in canonical patterns in the post-implantation embryo during lineage specification (Hanna et al., 2018a).

Paternal protamines are replaced by maternal histones immediately after fertilization, and the paternal genome acquires weak ncH3K4me3 and ncH3K27me3 domains (Zhang et al., 2016; Zheng et al., 2016). Both paternal and maternal ncH3K4me3 are maintained until early 2-cell stage and removed after zygotic genome activation (ZGA) at the late 2-cell stage. Meanwhile, canonical H3K4me3 forms at promoters after ZGA (Zhang et al., 2016; Dahl et al., 2016; Liu et al., 2016). Concomitant with the erasure of ncH3K4me3 after ZGA, H3K27ac appears at promoters and putative enhancers near ZGA genes (Dahl et al., 2016; Wu et al., 2016).

Maternal ncH3K27me3 is erased specifically from the promoters of developmental genes after fertilization, while those at distal regions are retained. Weak canonical H3K27me3 starts to form at promoters of polycomb target genes in the blastocyst, and both paternal and maternal distal ncH3K27me3 persist to blastocyst stage before being converted to canonical pattern in post-implantation embryo (Zheng et al., 2016). Notably, maternal H3K27me3 could control DNA methylation-independent imprinting (Inoue et al., 2017a; Inoue et al., 2017b), and in Drosophila, maternally inherited H3K27me3 regulates the activation of enhancers in the early embryos (Zenk et al., 2017).

The zygote genome also undergoes large-scale H3K9me3 re-establishment. H3K9me3 is highly enriched on LTRs in early embryos and is involved in transcriptional repression of LTRs, activation of which is triggered by DNA demethylation (Wang et al., 2018).

Moreover, TADs and compartments exist in a priming state after fertilization, and become more mature from the 8-cell stage. Consolidation of TADs is independent of ZGA and proceeds during cell cycles. Overall, chromatin of zygote exists in a relatively relaxed state with weak TADs and depleted distal chromatin interactions, and is gradually resolved to the canonical state through preimplantation development (Du et al., 2017; Ke et al., 2017; Xu et al., 2018).

2. Transcription by RNA polymerase II

2.1 The basal transcription machinery

The finding that crude HeLa cell extract directs selective and accurate transcription initiation by purified Pol II at the adenovirus major late promoter (Weil et al., 1979) provided direct biochemical evidences that accessory factors are necessary for site-specific initiation by Pol II.

Further fractionation of this HeLa extract yielded four enzymatically active fractions (A, B, C and D) (***Figure 2-1 A***), and the components within fractions A, C, and D were necessary for accurate transcription initiation by Pol II (Matsui et al., 1980). The complex in fraction A and D were named TFIIA and TFIID, respectively, while fraction C was subsequently fractionated into different complexes named TFIIB, TFIIE, TFIIF and TFIIH (Matsui et al., 1980; Sawadogo et al., 1985; Reinberg et al., 1987; Flores et al., 1989; Gerard et al., 1991; Flores et al., 1992).

Accessory factors necessary for site-specific transcription initiation by Pol II were similarly purified from different species like, rat liver (Conaway et al., 1990), *Drosophila* (Parker et al., 1984; Heberlein et al., 1985; Price et al., 1987) and *S. cerevisiae* (Lue et al., 1987; Sayre et al., 1992), and amazingly, organisms as diverse as human, rat, *Drosophila* and yeast use the same set of conserved complexes to initiate Pol II transcription.

These essential accessory factors, TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH were collectively defined as general transcription factors (GTFs) (Orphanides et al., 1996), which together with Pol II are known as the basal / general transcription machinery (***Figure 2-1 B***) (Smale et al., 2003; Thomas et al., 2006).

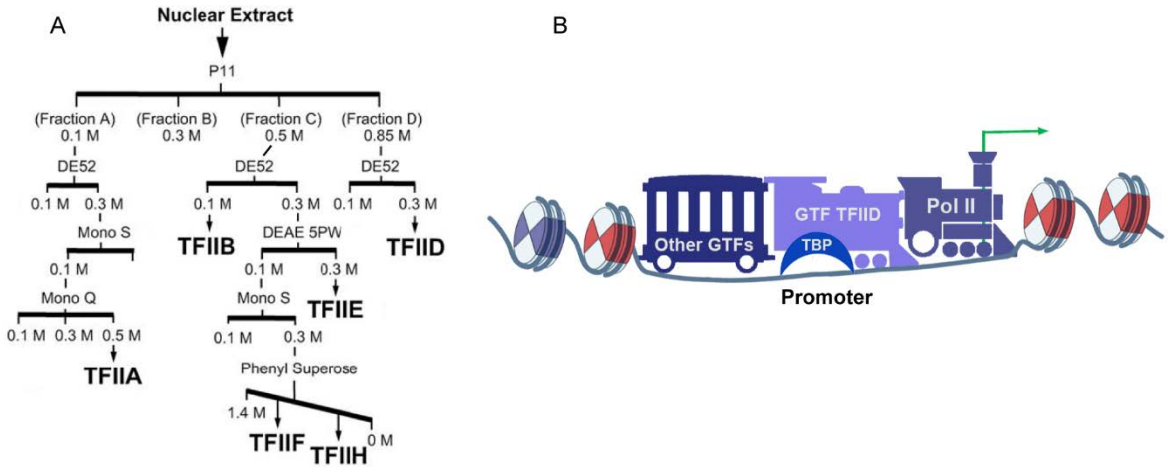


Figure 2-1: Pol II basal transcription machinery. (A) Purification of GTFs by fractionation of HeLa nuclear extract, from (Thomas et al., 2006). (B) Basal transcription machinery on the promoter.

2.1.1 RNA Polymerase II

Pol II is a multi-subunit complex responsible for the transcription of all protein-coding genes, long noncoding RNA, microRNA genes and most snRNA genes, which is composed of 12 highly conserved subunits (RPB1 to RPB12) (***Table 2-1***) (Young, 1991; Sainsbury et al., 2015).

However, not all the subunits are exclusive to Pol II: 5 sub subunits (RPB5, RPB6, RPB8, RPB10 and RPB12) are also present in Pol I and Pol III, and 4 subunits (RPB1, RPB2, RPB3 and RPB11) have homologous counterparts in Pol I and Pol III. Only RPB4, RPB7, RPB9 and the C-terminal domain (CTD) of RPB1 are unique to Pol II (Thomas et al., 2006).

Table 2-1: Subunits of Pol II. Adapted from (Sainsbury et al., 2015)

Subunit	Gene name		Mass (kDa)		Copies
	Yeast	Human	Yeast	Human	
RPB1	*RPO21*	*POLR2A*	191.6	217.2	1
RPB2	*RPB2*	*POLR2B*	138.8	133.9	1
RPB3	*RPB3*	*POLR2C*	35.3	31.4	1
RPB4	*RPB4*	*POLR2D*	25.4	16.3	1
RPB5※	*RPB5*	*POLR2E*	25.1	24.6	1
RPB6※	*RPO26*	*POLR2F*	17.9	14.5	1
RPB7	*RPB7*	*POLR2G*	19.1	19.3	1
RPB8※	*RPB8*	*POLR2H*	16.5	17.1	1
RPB9	*RPB9*	*POLR2I*	14.3	14.5	1
RPB10※	*RPB10*	*POLR2L*	8.3	7.6	1
RPB11	*RPB11*	*POLR2J*	13.6	13.3	1
RPB12※	*RPB12*	*POLR2K*	7.7	7.0	1
Total 12 subunits			513.6	516.7	

※Subunit shared among Pol I, Pol II and Pol III.

Pol II contains a ten-subunit catalytic core, while other two subunits RPB4 and RPB7 form the polymerase stalk (Vannini et al., 2012). The polymerase core can be divided into four distinct mobile modules based on the structure(***Figure 2-1 A***): a core module containing the regions of RPB1 and RPB2 that form the active center and other subunits (RPB3, RPB10, RPB11 and RPB12), a jaw-lobe module comprising of RPB2 and regions of RPB1 and RPB9, a clamp module composed of domains of RPB1 and RPB2, and a shelf module made up of RPB5, RPB6 and regions of RPB1 (Cramer et al., 2001). Interestingly, the 10-subunit

polymerase core represents the Pol II during elongation, and is not able to initiated transcription (Cramer et al., 2001). The RPB4/RPB7 heterodimeric subcomplex binds to a pocket formed by RPB1, RPB2 and RPB6 at the base of clamp module (Bushnell et al., 2003), forming the polymerase stalk (***Figure 2-1 B***). This binding induces a conformational change and locks the clamp in the closed conformation, which suggest that single-stranded DNA enters into the cleft of the polymerase core before binding of RPB4/RPB7 (Bushnell et al., 2003). The RPB4/RPB7 heterodimer is located in the vicinity of the CTD, however, unstructured CTD forms a flexible, tail-like extension from the catalytic core of Pol II, and is not visible due to high mobility in the crystal structures (Cramer et al., 2001; Meinhart et al., 2005).

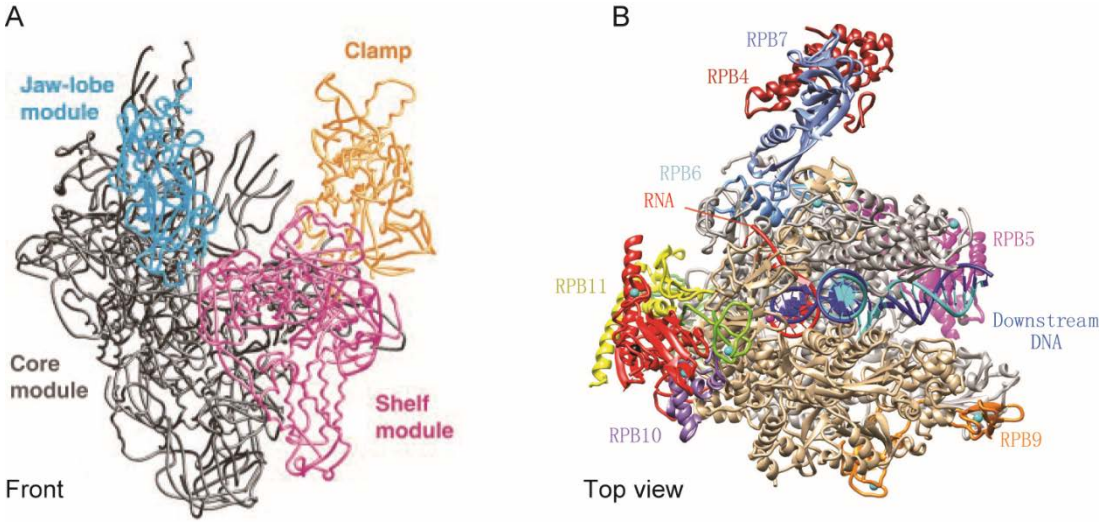


Figure 2-2: Structure of Pol II. (A) Structure of ten-subunit Pol II core with four modules, from (Cramer et al., 2001). (B) Ribbon model of mammalian (human) Pol II, from (Bernecky et al., 2016).

The CTD of RPB1 consists of tandem heptapeptide repeats with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$ (Schuller et al., 2016; Harlen et al., 2017) (***Figure 2-3***). The repeats number varies between species from 26 in *S. cerevisiae* to 52 in vertebrates (Young, 1991; Hsin et al., 2012). However, in vertebrate, out of the 52, only 21 repeats match the consensus perfectly, and are mainly located in the N-terminal half of the CTD, while the remaining 31 heptads have one or more substitutions (Hsin et al., 2012). A 10-residue sequence is present at the C-terminal end of CTD and is important for its stability (Chapman et al., 2004).

Importantly, the CTD interacts with a wide range of regulatory factors via its dynamic binding surfaces generated by post-translational modifications, notably phosphorylation (Buratowski, 2009), thus playing important roles at different steps of transcription, including transcription initiation (see section ***2.2.3***), pause-release (section ***2.2.6***), elongation (section ***2.2.6***) and termination (section ***2.2.7***). Besides those known canonical roles, Pol II CTD also undergoes cooperative liquid phase separation, helping Pol II to form clusters/hubs at active genes (Kwon et al., 2013; Burke et al., 2015; Harlen et al., 2017; Boehning et al., 2018).

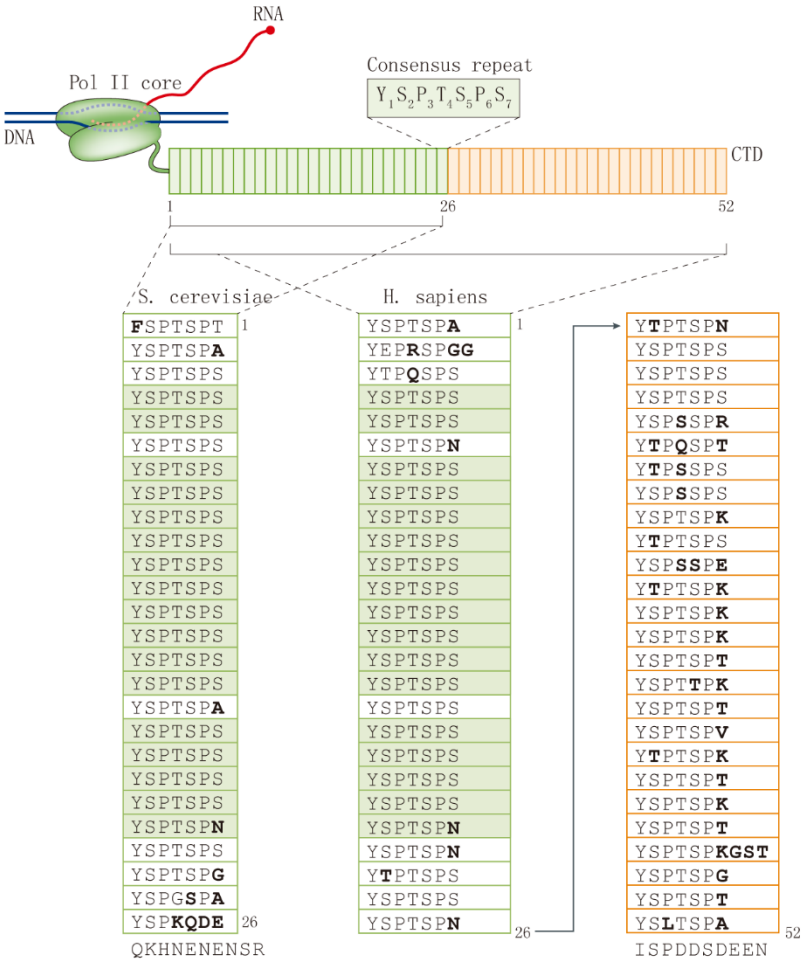


Figure 2-3: The composition and conservation of Pol II CTD. From (Harlen et al., 2017)

In the following part, GTFs are described in the order that they join the PIC, with the exception of TFIIA , which can join the PIC at any step after TFIID binding (Orphanides et al., 1996).

2.1.2 TFIID

TFIID is the first GTF that recognizes and binds to the core promoter, nucleating the assembly of the pre-initiation complex (PIC) (for PIC assembly, see section ***2.2.2***) (Buratowski et al., 1989). It is a multi-subunit complex comprising of the TATA box binding protein (TBP) and 13 TBP-associated factors (TAFs, TAF1 to TAF13 in metazoan or 14 TAFs in yeast) with total molecular weight of 1.2 MDa. The subunits of TFIID are generally conserved from yeast to human, but also with the existence of TAF paralogs, TAF-like proteins and TBP-related proteins in different metazoans (***Figure 2-5 a***) (Tora, 2002; Muller et al., 2010).

2.1.2.1 TBP

TBP plays a crucial role in transcription initiation of all three RNA polymerases in

eukaryotes (Hernandez, 1993). *Tbp* knockout in mice leads to growth arrest and apoptosis at the blastocyst stage (Martianov et al., 2002b), however, while Pol I and Pol III transcription is blocked, Pol II transcription is still active.

In Pol II transcription, TBP is the central DNA-binding subunit of TFIID (Tora et al., 2010), and it binds upstream of the transcription start site (TSS) of all promoters (Rhee et al., 2012). Interestingly, most metazoans have multiple TBP paralogs, the TBP-related factors (TRFs), including insect specific TRF1, metazoan-specific TRF2/TBPL1/TLF/TLP/TRP and vertebrate-specific TBP2/TRF3/TBPL2. (Torres-Padilla et al., 2007; Muller et al., 2010; Goodrich et al., 2010). Function of these TBP-related factors will be discussed in ***section 3.2***.

TBP has a highly conserved C-terminal half, which consists of two symmetric pseudo-repeats folding into a bipartite saddle-like structure (***Figure 2-4 A***). TBP initially binds the TATA box to form an unstable complex containing unbent DNA, subsequently forming a stable complex and inducing a 90° bend of the DNA, thus, providing an asymmetric platform for PIC assembly (Kim et al., 1993a; Kim et al., 1993b; Zhao et al., 2002) (***Figure 2-4 B***). However, it is unclear whether TATA-less promoters display any bending upon TBP or TFIID binding as the structure of a TFIID-containing PIC on TATA-less DNA has not been modeled (Tora et al., 2010; Kamenova et al., 2014). Although it has been shown that single-site variants in the TATA box reduce TBP-induced bending in solution (Wu et al., 2001), however, systematic X-ray crystallographic study shows that co-crystal structures of *A. thaliana* TBP with 10 different TATA box variants are all very similar, indicating that structure of the TBP–DNA complex is independent of TATA element sequence (Patikoglou et al., 1999). It is possible that TATA-less promoters are also bent during PIC formation, by TAFs, TFIIB and / or some other factors (Kamenova et al., 2014).

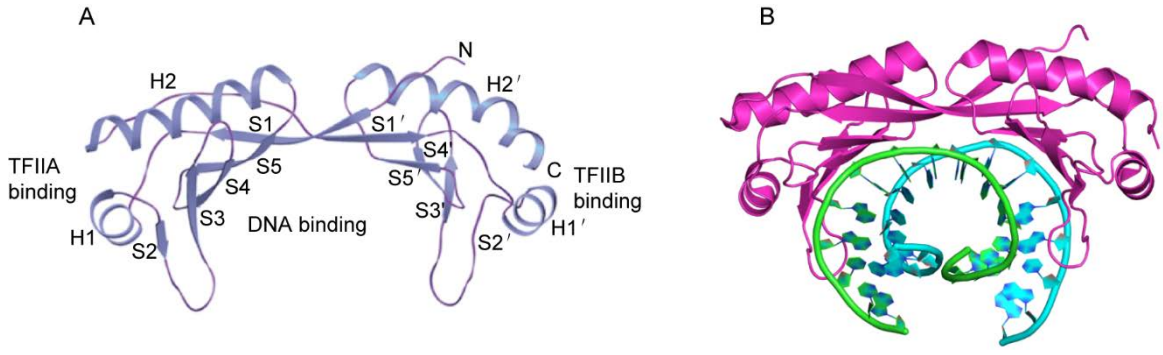


Figure 2-4: Structure of TBP core domain with / without TATA box. (A) Structure of TBP core domain, from (Davidson, 2003). (B) Structure of Human TBP core domain with TATA box, from Protein Data Bank in Europe (PDB EMDB id: 1cdw), (Nikolov et al., 1996)

Although the affinity is ~1000-fold lower compared with TATA sequences, TBP still exhibits binding ability to non-specific DNA, which may lead to the formation of non-productive

PICs (Coleman et al., 1995). Moreover, once TBP binds to DNA, no matter at TATA box or non-specific site, the half-life for TBP dissociation is very long in vitro (20-60 mins or more) (Hoopes et al., 1992; Coleman et al., 1995). Given the central role of TBP in eukaryotic transcription, its activity need to be tightly regulated, including the negative regulation of liberating TBP from stable TBP-promoter binding (especially TBP-TATA complex) and the removal of TBP from non-specific binding sites, which can be achieved by regulators such as ATPase BTAF1 (Mot1 in yeast) and Negative Cofactor 2 (NC2).

BTAF1 was initially identified as part of the B-TFIID complex, which consists of BTAF1 and TBP (Timmers et al., 1992). BTAF1 can bind to the concave surface of TBP to block TBP-promoter binding (Pereira et al., 2001). Moreover, as BTAF1/Mot1 belongs to the SWI2/SNF2-family ATPase, it can bind to the TBP-DNA complex and change TBP DNA binding properties upon ATP-binding and hydrolysis, ultimately leading to the displace and release of TBP from DNA (Pereira et al., 2003; Gumbs et al., 2003), thus facilitating the redistribution of TBP (Klejman et al., 2005).

NC2 is a conserved heterodimeric complex consisting of NC2α and NC2β (Goppelt et al., 1996a). Binding of NC2 to the TBP-promoter complex inhibits the association of TFIIA and TFIIB with TBP, therefore blocking PIC formation (Inostroza et al., 1992; Mermelstein et al., 1996; Goppelt et al., 1996b). In addition, NC2 binding can induce significant conformational changes in the TBP-DNA complex, leading to TBP sliding on DNA (Schluesche et al., 2007).

BTAF1/Mot1 function is intimately linked with NC2. Consistent with that, the binding profiles of BTAF1 and NC2 strongly overlap, furthermore, a stable TBP–NC2–Mot1–DNA complex also exists (van Werven et al., 2008). It has been proposed that bent DNA conformation induced by TBP binding could act as a ‘‘spring” for rapid BTAF1-NC2 mediated TBP release from TATA-containing promoters, allowing TBP redistribution to TATA-less promoters (Tora et al., 2010; Zentner et al., 2013).

TBP recognizes and binds directly to TATA-containing promoters to start nucleation of the PIC. However, only about 24% of human promoters have a TATA-like element, among which only approximately 10% have the canonical TATA box (TATAWAWR) (Yang et al., 2007). Importantly, several TAFs are also able to recognize and bind different elements within the core promoter, or even acetylated or methylated histone tails around.

2.1.2.2 TAFs

As mentioned above, metazoan TFIID contains 13 TAFs (***Figure 2-5 b***). TAFs mediate a broad range of interactions, which allow the regulation of TFIID recruitment and stabilization at both TATA-containing and TATA-less promoter (elements of core promoter, see section ***2.3.1***).

TAFs recognize and bind to promoters (Muller et al., 2007) (***Figure 2-6***), which is of particularly importance since the majority of vertebrate promoters are lacking of TATA-box (Yang et al., 2007). TAF1 and TAF2 can bind to the initiator element (Inr) which overlaps with the transcription start site (TSS) (Chalkley et al., 1999), and possibly bind motif ten element (MTE) (Louder et al., 2016). TAF1 was also shown to bind downstream core element (DCE) (Lee et al., 2005). TAF6/TAF9 heterodimers bind to the downstream promoter element (DPE) (Burke et al., 1997; Shao et al., 2005), which could be the target of TAF1 as well (Louder et al., 2016). TAFs mediated promoter recognition and binding is essential for transcription of the genes with TATA-less promoters (Huisinga et al., 2004).

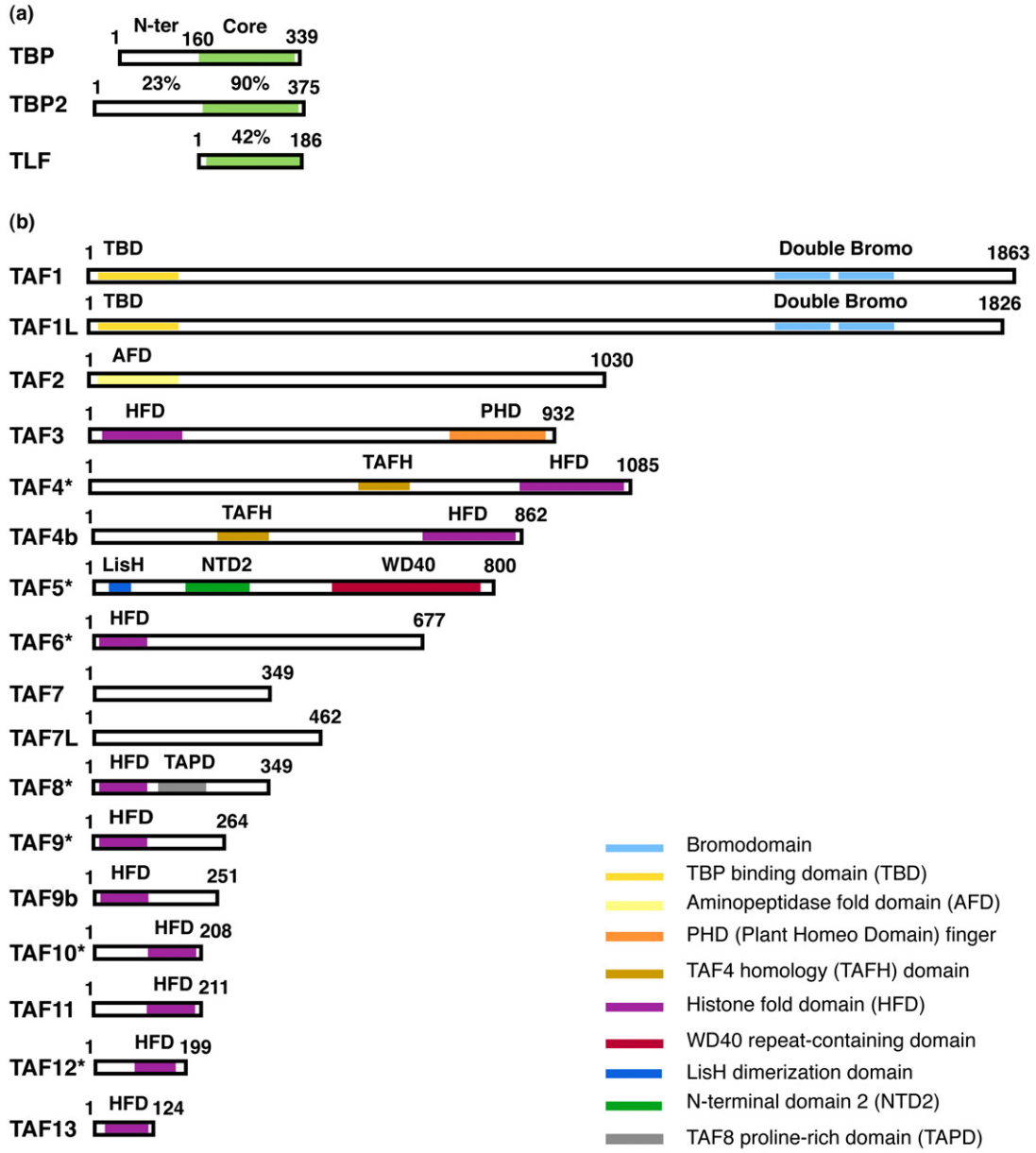


Figure 2-5: Schematic representation of human TFIID subunits. (a) TBP and TBP like factors. (b) Human TAFs, TAF paralogs and TAF-like proteins (except TAF5L and TAF6L, which are TAF-like proteins present in SAGA). TAFs present in the core TAF complex (see

section ***2.1.2.3***) are labelled with a star. From (Muller et al., 2010).

Moreover, TAFs can direct the recruitment of TFIID to promoters through interaction with histone post-translational modifications. Particularly, TAF3 can anchor TFIID to nucleosomes through its plant homeodomain (PHD) finger that selectively binds to the hallmark of active promoters H3K4me3 (Vermeulen et al., 2007; van Ingen et al., 2008). Interestingly, other active promoter marks, H3K9ac and H3K14ac increase TFIID binding to H3K4me3, while H3R2me2a (asymmetric di-methylation) inhibits this interaction (Vermeulen et al., 2007). Further studies show that H3K4me3, through interaction with TAF3, can direct PIC formation either independently or cooperatively with the TATA box (Lauberth et al., 2013). Besides, TAF1 contains two tandem bromodomain modules that bind selectively to multiply acetylated histone H4 peptides (Jacobson et al., 2000) [and it is important to mention that TAF1 has histone acetyltransferase (HAT) activity *in vitro* (Mizzen et al., 1996)].

Interactions between TAFs and histone PTMs provide novel insights for promoter recognition and TFIID recruitment, complementary to the mechanisms involving interactions of TBP and/or TAFs with core promoter elements. These new mechanisms are important specifically for TATA-less promoters (Vermeulen et al., 2007).

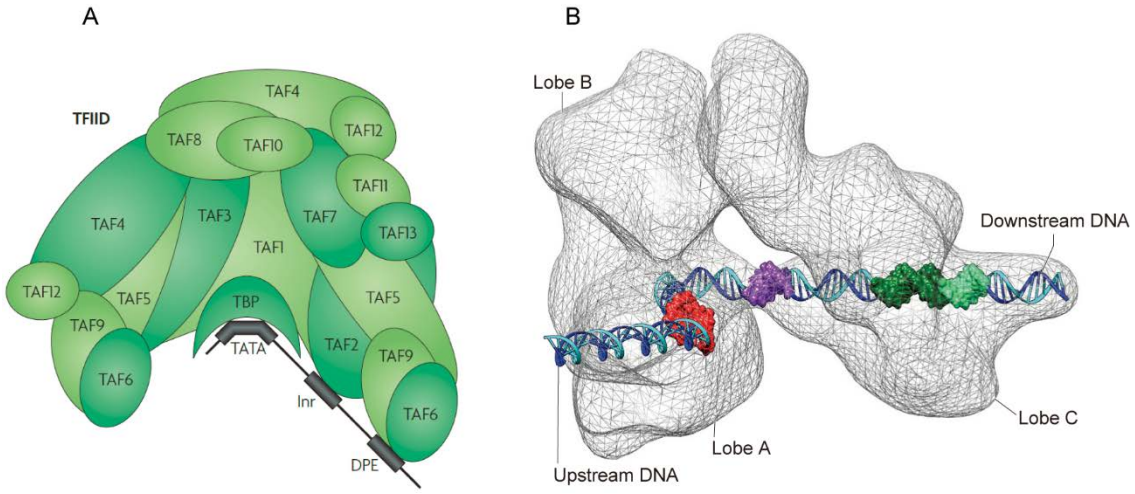


Figure 2-6: Core promoter recognition by TFIID. (A) TFIID binds core promoter elements, from (Goodrich et al., 2010). (B) Low-resolution cryo-EM structure of human TFIID bound to promoter DNA with TATA box (red), Initiator (Inr, purple), motif ten element (MTE, dark green) and downstream promoter element (DPE, light green), from (Cianfrocco et al., 2013; Sainsbury et al., 2015)

In addition, TAFs can also modulate TBP/TFIID activity, for instance, TAF1 inhibits TBP/TFIID binding to TATA-containing core promoter DNA with its N-terminal domains TAND1 and TAND2, which bind respectively to the DNA-binding surface of TBP though mimicking TATA-box and the convex surface of TBP. (Kokubo et al., 1994; Sainsbury et al., 2015).

Binding of TFIIA offsets the inhibition by TAF1 and stabilize the TBP-promoter interaction (Ozer et al., 1998). Alternatively, TAF11/TAF13 also interacts with the DNA binding surface of TBP, thereby blocking TBP from binding TATA-containing promoters, and this inhibition is required for normal TFIID function (Gupta et al., 2017). TAF7 binds to TAF1 and inhibits its HAT activity (Gegonne et al., 2001). Besides, TAFs can also interact with transcription activators (Liu et al., 2009) and GTFs, such as TFIIA (Yokomori et al., 1993), TFIIB (Goodrich et al., 1993) and TFIIH (Ruppert et al., 1995). Notably, TAF7 interacts with TFIIH and p-TEFb, and regulates their kinase activities (Gegonne et al., 2008), thus TAF7 may function as a checkpoint regulator for transcription initiation (Gegonne et al., 2006).

Overall, TAFs modulate TFIID recruitment to promoters containing different combinations of binding elements (including histone PTMs), and also finetune TFIID stability and activity.

2.1.2.3 Structure of the core-TFIID and holo-TFIID assembly

TAFs contain several conserved structural domains (***Figure 2-5 b***), notably, histone fold domains (HFDs), which mediate heterodimerization of TAF3-TAF10, TAF8-10 (Gangloff et al., 2001), TAF4-12 (Werten et al., 2002), TAF6-9 (Xie et al., 1996) and TAF11-13 (Birck et al., 1998). Stoichiometry analysis of TFIID in yeast revealed that six TAFs (TAF4, TAF5, TAF6, TAF9, TAF10 and TAF12) are present in two copies in TFIID, while TBP and other TAFs are present in a single copy (Sanders et al., 2002).

Among the six two-copies TAFs, five (TAF4, TAF5, TAF6, TAF9 and TAF12) form a core-TFIID subcomplex (Wright et al., 2006), which has a symmetric structure, comprising two copies of HFD-containing heterodimer TAF4-12, TAF6-9 and WD40 repeat domain-containing protein TAF5 (Bieniossek et al., 2013) (***Figure 2-7 A***). Interestingly, the symmetry of the core-TFIID is broken upon addition of a TAF8–TAF10 building block, resulting in a 7TAF-TFIID complex with an asymmetric structure, which was proposed to serve as a functional scaffold, enabling the remaining TAFs to assemble along the periphery (Bieniossek et al., 2013) (***Figure 2-7 B***).

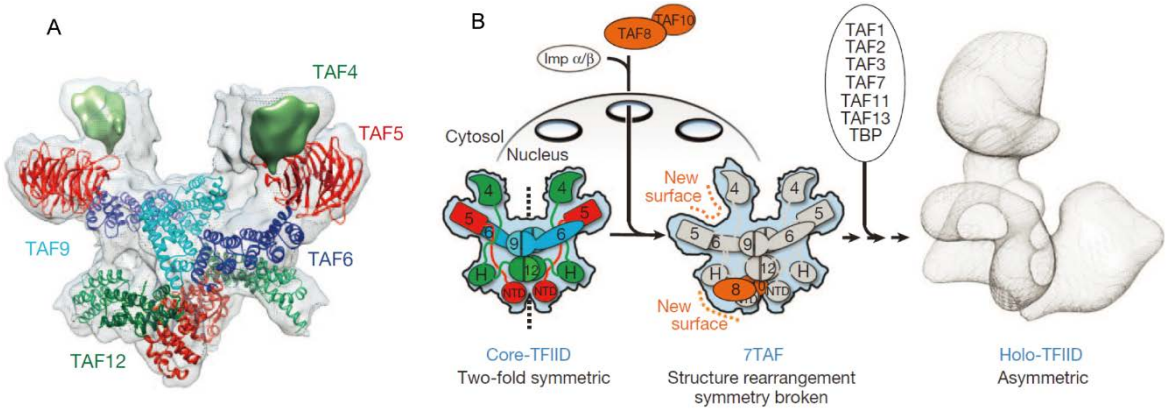


Figure 2-7: Structure of the human core-TFIID and model for TFIID assembly. (A) cryo-EM structure of the human TFIID core complex. (B) Model for holo-TFIID assembly. Adapted from (Bieniossek et al., 2013).

A recent study has shown that, in the cytoplasm, TAF2, TAF8 and TAF10 form a ternary subcomplex, which subsequently translocates into nucleus and incorporates into the core-TFIID complex, forming a 8TAF-TFIID intermediate (Trowitzsch et al., 2015). Moreover, it has been shown that TFIID is assembled co-translationally (termed co-translational assembly) (Kamenova et al., 2018).

However, high resolution structure of holo-TFIID is still not solved (Sainsbury et al., 2015), and also further studies are needed to fully understand the stepwise assembly mechanism of holo-TFIID.

2.1.3 TFIIA, a controversial GTF

Unlike the yeast TFIIA that contains 2 subunits, TFIIA in metazoan is composed of three subunits (TFIIAα, TFIIAβ and TFIIAγ), which are encoded by two genes, *GTF2A1* and *GTF2A2* (Thomas et al., 2006). *GTF2A1* encodes the TFIIAαβ precursor, which is cleaved into TFIIAα and TFIIAβ post-translationally by Taspase1 at the conserved cleavage site "QVD$^{\downarrow}$G", while *GTF2A2* encodes the smallest subunit TFIIAγ (Zhou et al., 2006; Hoiby et al., 2007). In addition, TFIIAαβ has a cell type-specific paralogue, called TFIIA-like factor (ALF) (Upadhyaya et al., 1999; Ozer et al., 2000) (see section ***3.3.2***).

TFIIA was initially classified as a general transcription factor as early studies showed TFIIA was necessary to reconstitute basal transcription *in vitro* (Reinberg et al., 1987). However, later studies showed the requirement for TFIIA in reconstituted transcription varies depending on different reconstituted transcription systems (Orphanides et al., 1996; Thomas et al., 2006). Indeed, TFIIA stimulates both basal and activated transcription *in vitro* when TFIID is used for promoter binding (Thomas et al., 2006; Hoiby et al., 2007).

TFIIA contributes to transcription initiation through different mechanisms. First, TFIIA can bind and stabilize the TBP-promoter complex through direct interaction with TBP and contact with DNA upstream of the TATA box, thus enhancing PIC assembly (Orphanides et al., 1996). Second, TFIIA can counteract the inhibitory effects on TBP binding to DNA that are caused by NC2, BTAF1 or TAF1 (Thomas et al., 2006). In addition, TFIIA can also function as a coactivator to facilitate PIC assembly by direct contacts with several activators and GTFs (Thomas et al., 2006; Hoiby et al., 2007).

2.1.4 TFIIB

TFIIB is the only GTF that consist of a single polypeptide. Human TFIIB contains 316 amino acids (345 aa in yeast), which are organized into 5 functional domains (Sainsbury et al., 2015) (***Figure 2-8 A&B***).

TFIIB is essential for Pol II transcription initiation. First of all, TFIIB can bind to the TBP/TFIID-DNA complex, resulting in the formation of a more stable TFIIB-TFIID-promoter or TFIIB-TFIIA-TFIID-promoter complex, and it can recognize and bind to promoters that contain TFIIB recognition element (BRE) as well (Thomas et al., 2006; Deng et al., 2007).

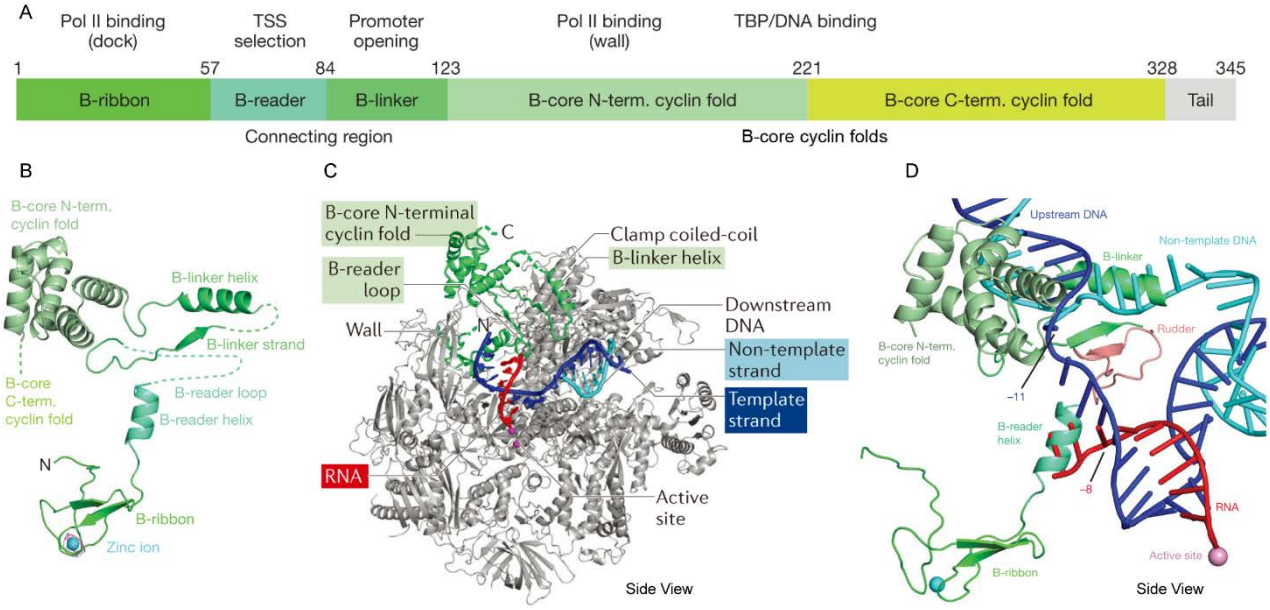


Figure 2-8: Structure of the yeast Pol II-TFIIB complex. (A) TFIIB domain organization. (B) Ribbon model of TFIIB. (C) Ribbon model of Pol II-TFIIB complex. (D) Clash of RNA stand with the B-reader. Adapted from (Kostrewa et al., 2009; Sainsbury et al., 2015)

Secondly, its N−terminal B-ribbon that contacts the dock domain of Pol II (RPB1 subunit) is involved in Pol II recruitment (Buratowski et al., 1993; Bushnell et al., 2004) (***Figure 2-8 A&C***). Thirdly, its C-terminal domain (B-core cyclin folds) interacts with both Pol II and the TBP-promoter complex to orients the DNA. More specifically, the B-core N-terminal cyclin fold binds to the wall of Pol II to position DNA over the Pol II active center cleft (Bushnell et al., 2004; Kostrewa et al., 2009; Liu et al., 2010) (***Figure 2-8 A&C***). Fourthly, its B-linker domain binds to Pol II rudder and clamp coiled-coil domains, helping DNA opening and/or maintenance of the transcription bubble. Its B−reader domain contacts the DNA template strand, assisting in TSS selection and DNA positioning for the initiation of RNA synthesis (Kostrewa et al., 2009; Sainsbury et al., 2013; Sainsbury et al., 2015) (***Figure 2-8 A&D***).

Furthermore, TFIIB stimulates initial RNA synthesis (Sainsbury et al., 2013), and stabilizes the early initiation complex containing a short transcript (Bushnell et al., 2004). B-reader loop

of TFIIB blocks the path of the transcript longer than 6 nucleotides and directs it to its exit tunnel, and this blocking by the B-reader loop may play a role in DNA-RNA strand separation (Sainsbury et al., 2013) (***Figure 2-8 D***). Lastly, TFIIB is released from Pol II when transcript grows to 12-13 nucleotides, as the RNA starts to clash with B-ribbon, triggering TFIIB displacement (Cabart et al., 2011; Sainsbury et al., 2013).

2.1.5 TFIIF

In mammals, TFIIF is a heterodimer comprising of the subunits TFIIFα and TFIIFβ (also known as RAP74 and RAP30, respectively) (Burton et al., 1988; Flores et al., 1988; Flores et al., 1990). However, yeast TFIIF has 3 subunits (Henry et al., 1992; Henry et al., 1994): Tfg1 and Tfg2 are essential and correspond to human TFIIFα and TFIIFβ, respectively, while the third subunit Tfg3 is non-essential for transcription, and it is also present in yeast TFIID (as TAF14) and SWI/SNF chromatin remodeling complexes (as ANC1) (Cairns et al., 1996).

Interesting, transcription can be initiated to some extent *in vitro* without TFIIE and TFIIH, but TFIIF is absolutely needed, indicating the critical role of TFIIF for transcription initiation (Pan et al., 1994). Indeed, TFIIF plays multiple roles during PIC formation. First, TFIIF tightly associates with Pol II and enhances the affinity of Pol II for TFIIB-TFIID-promoter complex (Robert et al., 1998), preventing non-specific interaction of Pol II with DNA (Conaway et al., 1991), facilitating Pol II recruitment (Flores et al., 1991) and stabilizing the PIC (Tan et al., 1994), in particular stabilizing TFIIB within the PIC (Cabart et al., 2011; Fishburn et al., 2012). Second, TFIIF is required for subsequent recruitment of TFIIE and TFIIH through direct interactions with TFIIE (Maxon et al., 1994; Orphanides et al., 1996). Third, TFIIF influences TSS selection (Ghazy et al., 2004).

After PIC assembly, TFIIF stimulates early RNA synthesis and is required for efficient Pol II promoter escape (Yan et al., 1999). TFIIF also enhances the efficiency of Pol II elongation and suppresses transient Pol II pausing (Zhang et al., 2004; Thomas et al., 2006). Moreover, TFIIF assists in the stabilization of the transcription bubble (Pan et al., 1994).

2.1.6 TFIIE

Like TFIIF, human TFIIE is also a heterodimer consisting of the two subunits, TFIIEα and TFIIEβ (Tfa1 and Tfa2 in yeast) (Peterson et al., 1991; Ohkuma et al., 1991; Itoh et al., 2005; Jawhari et al., 2006). TFIIEα contains an N-terminal extended winged WH domain (eWH), a central zinc ribbon domain (E-ribbon) connected by E-linker and a C-terminal acidic domain. TFIIEβ has two WH domains and an E-tether (Sainsbury et al., 2015; Plaschka et al., 2016) (***Figure 2-9 A***). E-tether binds the E-linker and is essential for TFIIE subunit dimerization (Plaschka et al., 2016) (***Figure 2-9 B***).

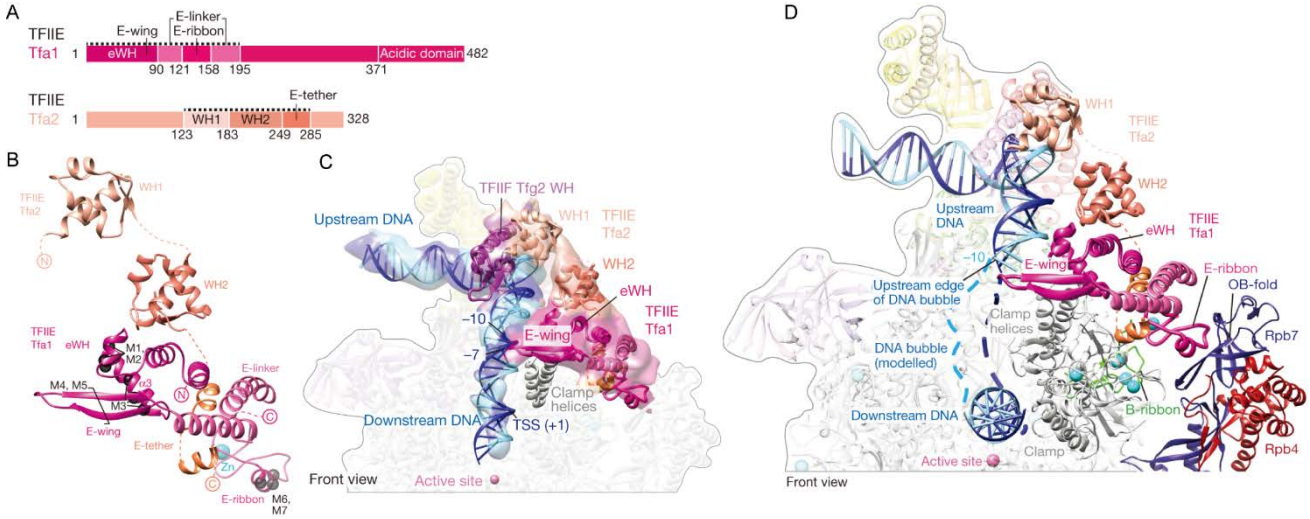


Figure 2-9: TFIIE architecture and interactions. (A) Domain organization of yeast TFIIE. (B) TFIIE domain architecture. (C) TFIIE interactions within the closed complex. (D) TFIIE interactions within the open complex. (Closed complex and open complex see section ***2.2.2***). Adapted from (Plaschka et al., 2016).

Once recruited, it interacts directly with TFIIF, TFIIB, Pol II and promoter DNA (Thomas et al., 2006). TFIIE is located between the clamp and the RPB4–RPB7 stalk of Pol II. The eWH domain of TFIIEα anchors TFIIE to Pol II clamp, and it contacts DNA backbone at positions −13/−14 upstream of TSS. Moreover, TFIIEα eWH contacts the TFIIFβ WH domain above upstream DNA, together with other two WH domains from TFIIEβ, encircling and retaining promoter DNA (Grunberg et al., 2012; Plaschka et al., 2016) (***Figure 2-9 C&D***). The E-ribbon of TFIIEα interacts with the Pol II clamp, as well as stalk subunit RPB7 of Pol II and B-ribbon of TFIIB (Plaschka et al., 2016) (***Figure 2-9 D***).

Furthermore, TFIIE assists in the recruitment of TFIIH (Maxon et al., 1994; Holstege et al., 1996), and it also stimulates ATPase and kinase activity of TFIIH (Ohkuma et al., 1994), thus facilitating the formation of an initiation-competent Pol II complex. In addition, TFIIE, together with TFIIH, is essential for promoter melting and the transition from initiation to elongation (Holstege et al., 1996).

2.1.7 TFIIH

TFIIH is a complex with 10 subunits, organized into a seven-subunit core and a three-subunit kinase module (Compe et al., 2012; Schilbach et al., 2017; Greber et al., 2017). Human TFIIH core is composed of two ATPases (XPB and XPD) and 5 subunits (p62, p52, p44, p34 and p8), which are respectively known as Ssl2, Rad3, Tfb1, Tfb2, Tfb4, Tfb5 and Ssl1 in yeast (Sainsbury et al., 2015). The TFIIH kinase module comprises of CDK7, cyclin H and MAT1 in human, which correspond respectively to Kin28, Ccl1 and Tfb3 in yeast (Sainsbury et al., 2015) (***Table 2-2***). Over all, TFIIH accommodates three different enzymatic activities: ATPase activity and helicase activity from XPB and XPD, and kinase activity from CDK7 (Compe et al., 2012).

Table 2-2: Subunits of TFIIH. Adapted from (Sainsbury et al., 2015).

TFIIH	Subunit	Gene name		Mass (kDa)		Copies
		Yeast	Human	Yeast	Human	
TFIIH *(core)*	Subunit 1 (p62)	*TFB1*	*GTF2H1*	72.9	62	1
	Subunit 2 (p44)	*SSL1*	*GTF2H2*	52.3	44	1
	Subunit 3 (p34)	*TFB4*	*GTF2H3*	37.5	34.4	1
	Subunit 4 (p52)	*TFB2*	*GTF2H4*	58.5	52.2	1
	Subunit 5 (p8)	*TFB5*	*GTF2H5*	8.2	8.1	1
	XPD subunit	*RAD3*	*ERCC2*	89.8	86.9	1
	XPB subunit	*SSL2*	*ERCC3*	95.3	89.3	1
	7 subunits			414.5	377.3	
TFIIH *(kinase module)*	Cyclin H	*CCL1*	*CCNH*	45.2	37.6	1
	CDK7	*KIN28*	*CDK7*	35.2	39.0	1
	MAT1	*TFB3*	*MNAT1*	38.1	35.8	1
	3 subunits			118.5	112.4	1

After its recruitment by TFIIE, TFIIH binds to Pol II and functions in promoter opening and escape (Goodrich et al., 1994; Holstege et al., 1996; Moreland et al., 1999). In general, without TFIIH, Pol II tends to stall on the promoter-proximal region, resulting in abortive transcription (Thomas et al., 2006). It has been shown that promoter opening by TFIIH is dependent on the ATPase activity of XPB but not XPD (Tirode et al., 1999; Coin et al., 1999). As XPB engages with promoter DNA around 25-30 bp downstream of TSS (Schilbach et al., 2017) (***Figure 2-10 A***) and does not bind the transcription bubble, it apparently does not function as a classical helicase, which would bind to the unwound region. Consistently, mutational analysis of XPB

revealed that DNA opening does not rely on its helicase activity (Lin et al., 2005). Current evidences support the translocase model for ATP-dependent DNA opening, according to which, XPB uses its ATP hydrolysis activity to track along the DNA template strand in the 3'-5' direction and translocates DNA away from Pol II, resulting in insertion and rotation of promoter DNA into the Pol II cleft, and leading to DNA unwinding (Grunberg et al., 2012; Fishburn et al., 2015; Schilbach et al., 2017) (***Figure 2-10 B***).

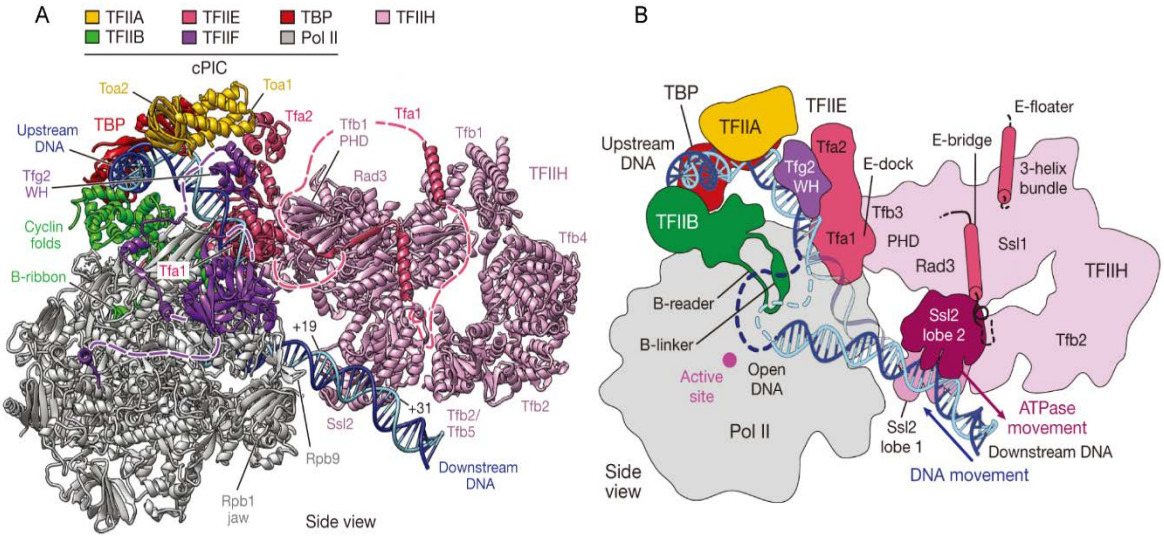


Figure 2-10: Structure of yeast TFIIH within PIC and DNA opening. (A) cryo-EM structure of yeast PIC (core PIC + TFIIH). (B) Schematic cross-section of the PIC with open and closed DNA, during DNA opening, Ssl2/XPB ATPase translocates to the right and DNA moves to the left. Adapted from (Schilbach et al., 2017).

The TFIIH kinase module, specifically CDK7, is responsible for the phosphorylation of Pol II CTD at serine 5 residue (Serizawa et al., 1995), which aids in promoter escape by Pol II (Harlen et al., 2017) and enhances the association of the Pol II CTD with the 7-methylguanosine (m7G) RNA capping machinery (Cho et al., 1997; Komarnitsky et al., 2000). In addition, CDK7 also phosphorylates Ser7 of CTD, which seems to be important for snRNA gene expression (Egloff et al., 2007; Compe et al., 2012).

Besides its roles in Pol II transcription, TFIIH also participates in Pol I transcription, probably Pol III transcription as well, and it is also known be to essential for DNA repair (Compe et al., 2012).

2.2 The cycle of Pol II transcription

The cycle of Pol II-mediated transcription contains at least eight major steps (***Figure 2-11***) (Fuda et al., 2009). It begins with chromatin opening (**Step 1**), which facilitates GTFs and Pol II gaining access to the promoter. GTFs and Pol II bind the core promoter and form a preinitiation complex (PIC) (**Step 2**). DNA is then opened (forming the 'transcription bubble') and RNA synthesis commences, termed initiation (**Step3**). Early elongating Pol II synthetizes only a short stretch of nascent RNA (~30–50 nucleotides) and then undergoes promoter-proximal pausing (**Step 4**). Positive transcription elongation factor b (P-TEFb) phosphorylates Ser2 of Pol II CTD, DRB sensitivity- inducing factor (DSIF) and negative elongation factor (NELF), leading to the dissociation of NELF. Pol II escapes from the pausing (**Step 5**), and enters either premature termination or productive elongation. If not termination, Pol II then productively elongates through the gene body (**Step 6**). After transcribing the gene, Pol II undergoes termination (**Step 7**), and the released free Pol II can reinitiate to start a new round of transcription (**Step 8**). Apparently, all these steps are the targets for transcription regulation.

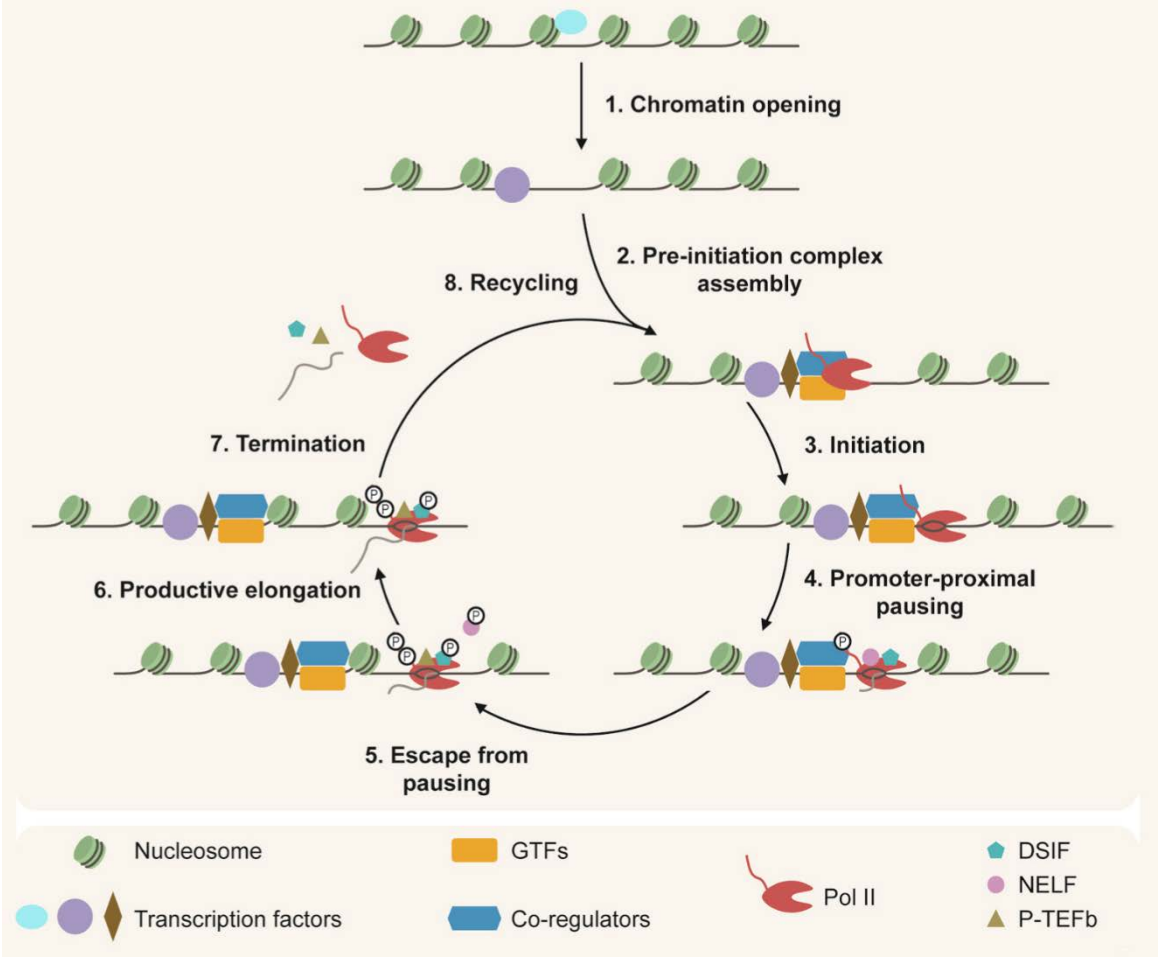


Figure 2-11: The transcription cycle and its potentially regulated steps. Adapted from the cover of cold spring harbor transcription meeting 2017 "Mechanisms of eukaryotic transcription" abstract book, which is adapted from (Fuda et al., 2009).

2.2.1 Chromatin opening

2.2.1.1 Binding of activators

Gene-specificity of transcription is achieved and greatly stimulated by activators or conversely repressors (also referred as gene-specific transcription factors), which mark or pre-mark promoter-proximal regions and/or distant enhancers for activity through DNA sequence-specific binding (Koster et al., 2015) (***Figure 2-14***).

Among the activators, some can be referred as pioneer factors (***Table 2-2***), and have the unique ability to engage their nucleosomal target sites in condensed or inaccessible chromatin (Vernimmen et al., 2015; Iwafuchi-Doi et al., 2016).

Table 2-3: Pioneer factors. Adapted from (Vernimmen et al., 2015)

Pioneer factors	DNA binding domain	Refs
AP-1	Basic leucine zipper	(Biddie et al., 2011)
AP-2γ (TFAP2C)	Basic helix–span–helix	(Tan et al., 2011)
FOXA1 (HNF-3α)	Forkhead	(Cirillo et al., 1998; Serandour et al., 2011)
FOXA2 (HNF-3β)	Forkhead	(Cirillo et al., 2002; Donaghey et al., 2018)
FOXE1	Forkhead	(Cuesta et al., 2007)
FOXD3	Forkhead	(Xu et al., 2009)
GATA2	2X GATA-type zinc fingers	(Wu et al., 2014)
GATA3	2X GATA-type zinc fingers	(Shoemaker et al., 2006)
GATA4	2X GATA-type zinc fingers	(Cirillo et al., 2002; Donaghey et al., 2018)
KLF4	3X C2H2-type zinc fingers	(Soufi et al., 2012; Soufi et al., 2015)
NF-Y (CBF)	NF-YA/HAP2	(Oldfield et al., 2014)
OCT4	POU-specific + POU-Homeodomain	(Buecker et al., 2014; Donaghey et al., 2018)
OTX2	Homeodomain	(Buecker et al., 2014)
PAX7	Paired + Homeodomain	(Budry et al., 2012; Mayran et al., 2018)
PBX1	Homeodomain	(Berkes et al., 2004)
PU.1	Ets	(Barozzi et al., 2014)
SOX2	HMG box	(Soufi et al., 2015)
SOX9	HMG box	(Adam et al., 2015)
TP53	p53	(Sammons et al., 2015)
P63	p53	(Sammons et al., 2015)
RFX	Rfx-type winged helix	(Masternak et al., 2003)

2.2.1.2 Recruitment of coactivators

Following the binding of activators, coactivators are recruited to the regulatory elements by activators, chromatin modifications, DNA, and/or regulatory RNAs (***Figure 2-14***).

Usually, coactivators are multi-subunit complexes that enhance transcription, through reorganizing nucleosomes, like chromatin remodelers (see ***section 1.2.3***); or through modifying histone covalently, like histone-modifying complexes, especially the histone acetyltransferases; or through interacting directly with GTFs and Pol II, like the Mediator complex (see ***2.2.1.3***). Moreover, some coactivators harbor more than one enzymatic activities, therefore they can target chromatin through chromatin remodeling and/or histone modifications. Altogether, coactivators can alter chromatin structure, open the promoter and facilitate the binding of GTFs and Pol II.

In the following sections, I will mainly focus on the Mediator complex, the SAGA complex and NuA4/TIP60 complex.

2.2.1.3 The Mediator complex

The Mediator is an evolutionarily conserved multi-subunit complex that is generally required for Pol II transcription (Allen et al., 2015). It comprises of 25 subunits in yeast and up to 30 subunits in humans, which are organized into four distinct modules: a head module, a middle module, a tail module and a transiently associated CDK8 kinase module (Soutourina, 2018) (***Figure 2-12***).

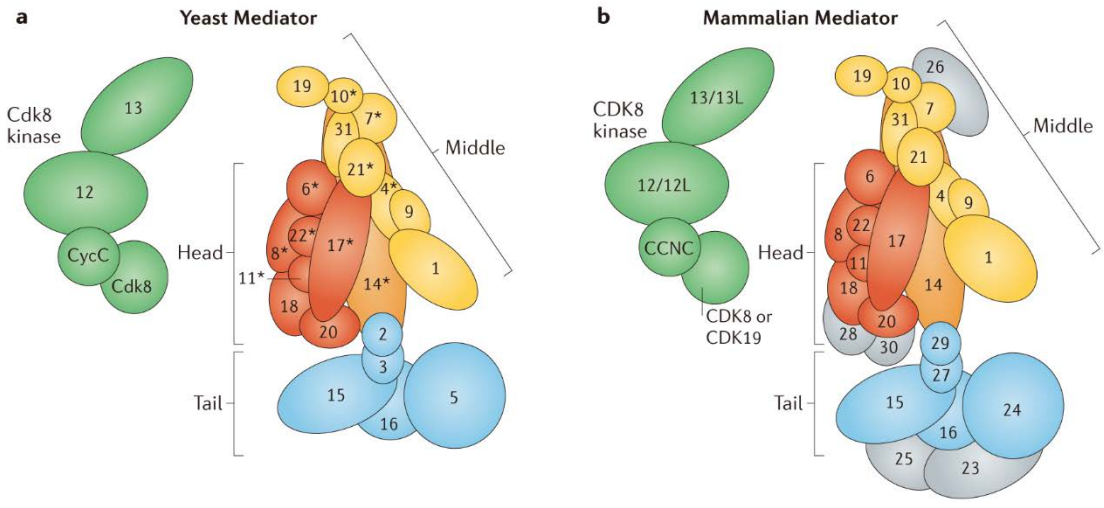


Figure 2-12: Subunit composition of the Mediator complex. (a) Yeast Mediator complex, asterisks indicate the 10 subunits that are essential for yeast viability. (b) Mammalian Mediator complex. From (Soutourina, 2018).

As mentioned above, the Mediator is generally required for transcription. After its

recruitment at enhancers (or UASs in yeast) by activator, the Mediator serves as a 'molecular bridge' between activators and the basal transcriptional machinery, integrating and transmitting regulatory signals from activators directly to Pol II and GTFs, and promoting PIC assembly at core promoters (Kornberg, 2005; Soutourina et al., 2011; Soutourina, 2018).

After PIC formation, the Mediator stimulates phosphorylation of Pol II CTD through regulating the enzymatic activity of TFIIH CDK7 (Kim et al., 1994; Nair et al., 2005; Boeing et al., 2010), thus triggering Pol II release from promoters. Moreover, the Mediator also appears to play a role in promoter-proximal pausing and / or pause release (Wang et al., 2005a; Kremer et al., 2012; Allen et al., 2015), and even transcription re-initiation (Yudkovsky et al., 2000). In addition to its canonical roles in transcription, Mediator also interacts with the TREX2 complex to couple transcription with mRNA export (Schneider et al., 2015).

2.2.1.4 The SAGA general coactivator

The Spt-Ada-Gcn5 acetyltransferase (SAGA) complex is an evolutionarily conserved, multi-functional and multi-subunit coactivator (Spedale et al., 2012). It contains 18-20 subunits, which are organized into 4 or 5 separate modules with distinct activities: an activator-binding module, a histone acetyltransferase (HAT) module, a histone deubiquitinase (DUB) module, a core structural module and a metazoan-specific splicing module (Helmlinger et al., 2017). Interestingly, most subunits within different modules are not exclusive to SAGA, instead, they are shared with other regulatory complexes (Helmlinger et al., 2017) (***Figure 2-13***).

As indicated in the modular names, SAGA plays multiple roles in regulating transcription, including but not limited to activator interaction, histone acetylation, histone deubiquitination and regulation of the basal transcription machinery (Spedale et al., 2012).

The activator-binding module contains TRRAP (Tra1 in yeast) is the largest component of SAGA (~420 kDa), and is also present in the NuA4/TIP60 complex (see ***2.2.1.5***). TRRAP interacts with different transcription factors and may serve as a major target of promoter-bound activators, thus plays a crucial role in SAGA recruitment at gene-specific promoters (McMahon et al., 1998; Brown et al., 2001; Bhaumik et al., 2004; Helmlinger et al., 2011; Weake et al., 2012). It is important to mention that, in addition to TRRAP, SAGA can also be recruited potentially by interactions with chromatin marks, TAF12 interactions with activators and/or binding of Spt3 and Spt8 to TBP in yeast (Weake et al., 2012).

After SAGA recruitment, the HAT module preferentially acetylates histone H3 (Helmlinger et al., 2017), and the DUB module deubiquitinates both histone (H2Bub and H2Aub) and non-histone substrates (Weake et al., 2012).

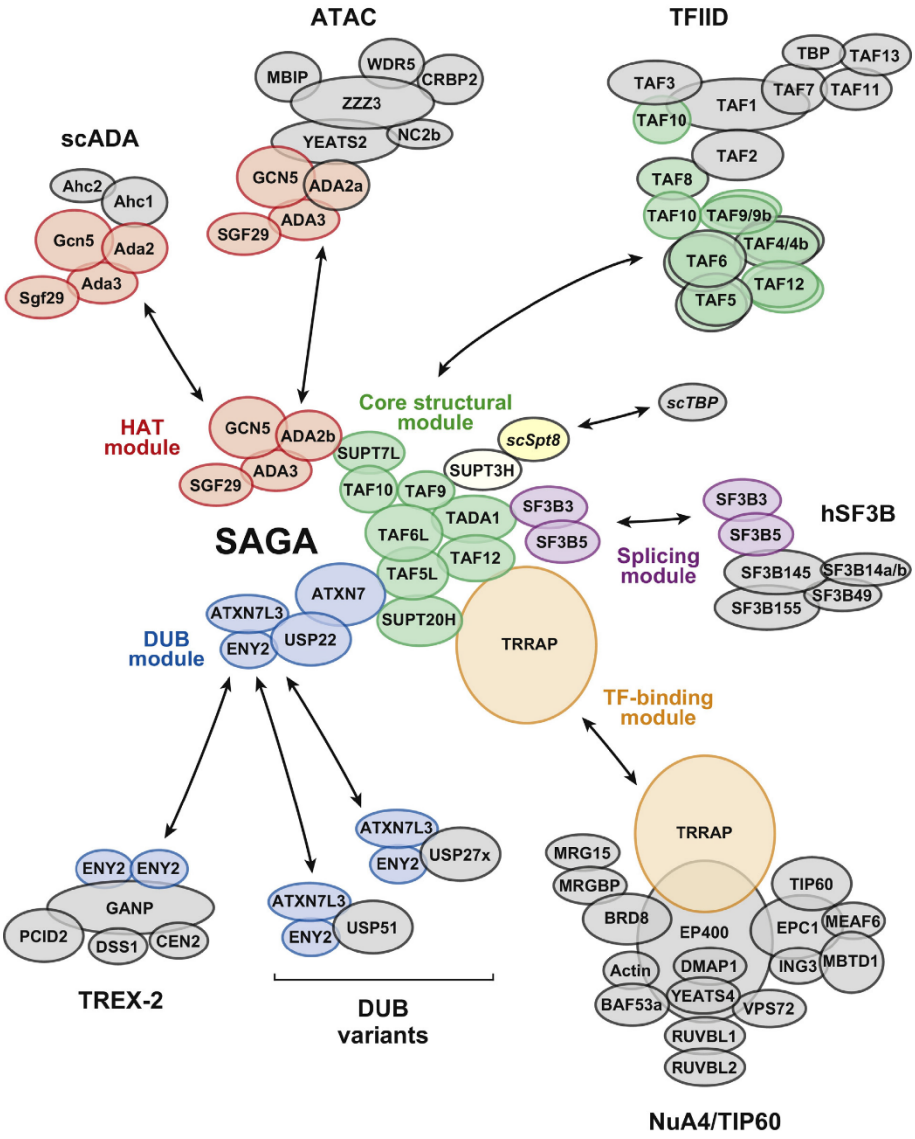


Figure 2-13: SAGA modular organization and the subunits shared with other complexes. From (Helmlinger et al., 2017).

Moreover, recent work showed that SAGA is required for essentially all RNA polymerase II transcription, acting as a general cofactor (Baptista et al., 2017).

2.2.1.5 NuA4/TIP60 complex

The mammalian NuA4/TIP60 complex is a conserved transcriptional coactivator with multiple subunits (Jha et al., 2009), among which, TIP60 is a histone acetyltransferase (HAT) belonging to the MYST family of HATs (Sapountzi et al., 2011), and p400 is a ATPase belonging to the SWI2/SNF2 class of chromatin remodelers (Fuchs et al., 2001) and a activator-binding subunit TRRAP shared with SAGA (***Figure 2-13***). Thus the NuA4/TIP60 complex accommodates three distinct enzymatic activities: a histone H2A/H4 acetyltransferase activity, an ATP-dependent histone (H2A/H2A.Z) exchange activity and a

helicase activity (Ikura et al., 2000; Gevry et al., 2007; Auger et al., 2008). After its recruitment by activators, NuA4/TIP60 complex acetylates histone H4, H2A and H2A.Z (Allard et al., 1999; Keogh et al., 2006; Babiarz et al., 2006) and facilitates histone (H2A/H2A.Z) exchange, therefore allowing chromatin opening and playing important roles in transcription regulation.

In addition, TIP60 not only acetylates histones, but also acetylates other cellular proteins involved in transcription, for instance, the Androgen Receptor (Gaughan et al., 2002) and p53 (Sykes et al., 2006; Tang et al., 2006). Besides, the NuA4/TIP60 complex also plays a key role in DNA repair (Rossetto et al., 2010; Jacquet et al., 2016).

2.2.2 Pre-initiation complex assembly

After chromatin opening, GTFs and Pol II start to form PIC. According to the canonical sequential assembly model, PIC assembly starts with core promoter binding by TFIID, followed by the recruitment of TFIIA and TFIIB, which stabilize the TFIID-promoter complex. As mentioned above (***2.1.3.1***), TFIIA is not absolutely required for basal transcription, but it can stabilize the TFIID–DNA complex, and it can join the PIC at any step after TFIID binding. TFIIB subsequently anchors Pol II to the promoter with the associated TFIIF. TFIIF stabilizes the PIC, in particular TFIIB within the PIC, forming a stable TFIID-TFIIA-TFIIB-Pol II-TFIIF-promoter complex (core PIC). Recruitment of TFIIE and entry of TFIIH complete the PIC assembly, leading to the formation of a closed PIC (Buratowski et al., 1989; Orphanides et al., 1996; Cheung et al., 2012; Sainsbury et al., 2015; Plaschka et al., 2016; Schilbach et al., 2017) (***Figure 2-14*** & ***Figure 2-15***).

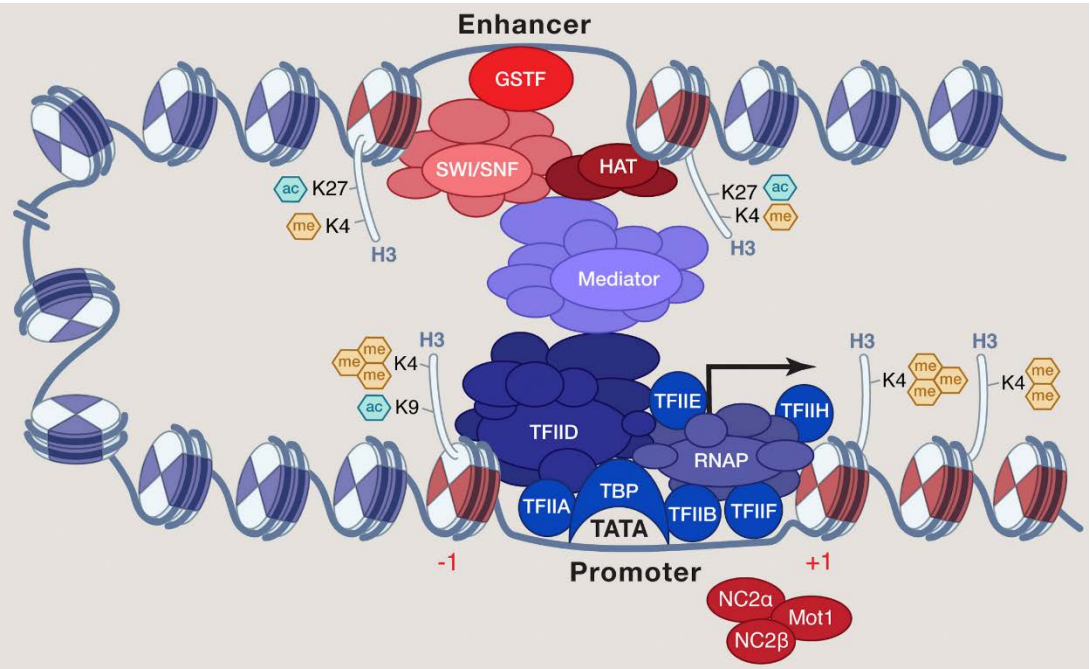


Figure 2-14: Chromatin opening and PIC assembly. From (Koster et al., 2015).

It is important to mention that, besides the sequential assembly model, another model, the RNA Pol II holoenzyme pathway also has been proposed. This model came up after the observation that Pol II could be purified as a preassembled holoenzyme containing several GTFs (but no TFIID), Mediator and chromatin remodelers/modifiers. According to this model, TFIID binds to the core promoter and is stabilized by TFIIA, facilitating the recruitment of preassembled Pol II holoenzyme (Koleske et al., 1994; Kim et al., 1994; Ossipow et al., 1995; Thomas et al., 2006).

Both models are supported by *in vitro* studies (Thomas et al., 2006), however, there is no conclusive evidence which one is used *in vivo*. Given the fact that both pathways are not mutually exclusive, they could co-exist or occur in an integrated middle way.

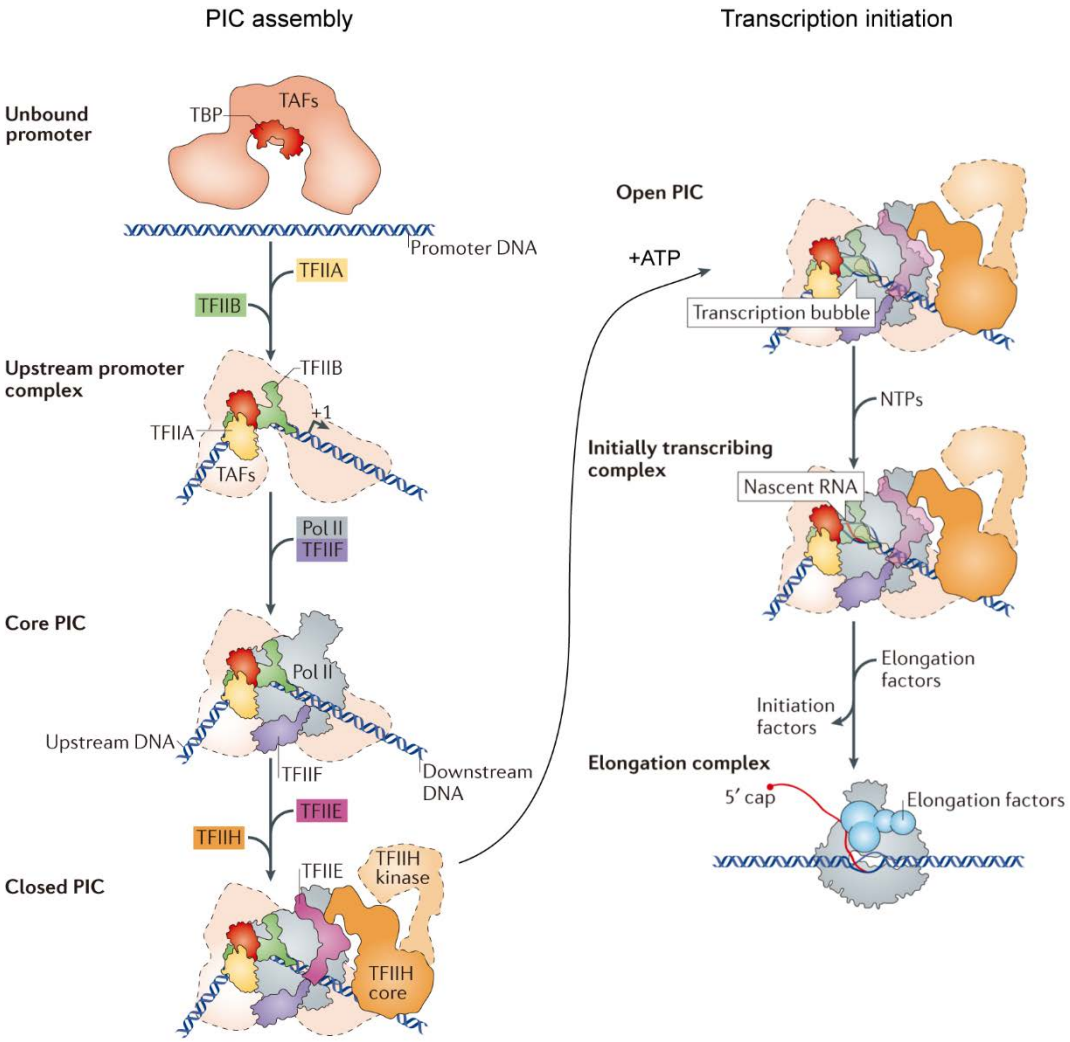


Figure 2-15: Schematic representation of PIC assembly and transcription initiation. Adapted from (Sainsbury et al., 2015).

2.2.3 Transcription initiation and Pol II promoter escape

In the presence of ATP, the promoter DNA (around TSS) inside the closed PIC is unwound by the PIC, especially TFIIH with the help of TFIIB and TFIIE (model for DNA opening see ***2.1.7***), forming the 'transcription bubble', and leading to the formation of open PIC (Sainsbury et al., 2015) (***Figure 2-15***). TFIIB and TFIIF contribute to the stabilization/maintenance of the transcription bubble (Sainsbury et al., 2015). In presence of NTPs, Pol II initiates transcription at TSS and RNA synthesis commences (Sainsbury et al., 2015) (***Figure 2-15***).

To continue transcribing, Pol II need to dissociate from the promoter-binding GTFs (promoter escape), which is mediated by the phosphorylation of Pol II CTD at Ser5 and Ser7 through TFIIH with its CDK7 kinase (Haberle et al., 2018) (***Figure 2-16***). The activity of CDK7 is regulated by TFIIE (Ohkuma et al., 1994) and Mediator (Allen et al., 2015).

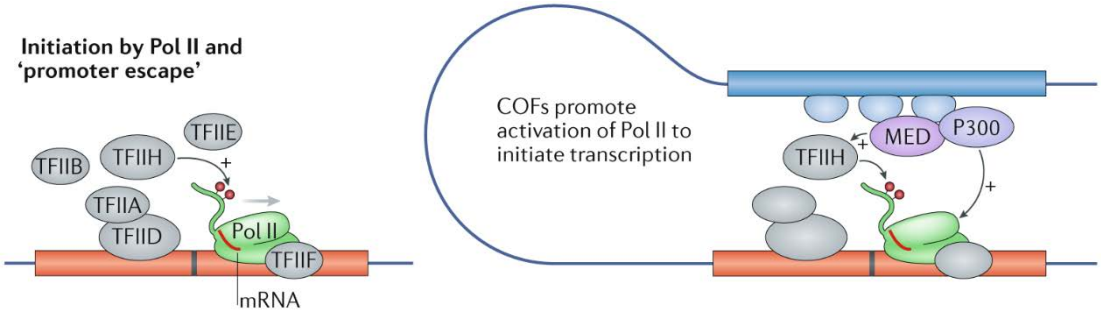


Figure 2-16: Transcription initiation and Pol II promoter escape. TFIIH mediated Pol II CTD Ser5 and Ser7 phosphorylation is accompanied by transcription initiation and Pol II promoter escape. From (Haberle et al., 2018).

Transcription initiation, Pol II promoter escape and promoter-proximal pausing are coupled continuous processes.

2.2.4 Promoter-proximal pausing

In metazoans, at many genes, after promoter escape, Pol II transcribes only a short stretch of nascent RNA (~30-50 nucleotides) and then pauses at downstream of the TSS, undergoing promoter-proximal pausing (Zeitlinger et al., 2007; Muse et al., 2007; Guenther et al., 2007; Core et al., 2008; Adelman et al., 2012; Haberle et al., 2018) (***Figure 2-17***).

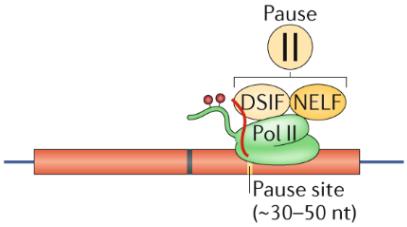


Figure 2-17: Pol II promoter-proximal pausing. From (Haberle et al., 2018).

Pol II promoter-proximal pausing is mediated by the binding of the negative elongation factor (NELF) and the DRB sensitivity-inducing factor (DSIF, a heterodimer of STP4 and SPT5) (Adelman et al., 2012). It has been suggested that the core promoter features and barrier effect of +1 nucleosome play important roles in triggering Pol II pausing (Kwak et al., 2013a; Li et al., 2013a; Weber et al., 2014; Jonkers et al., 2015).

However, interestingly, NELF and DSIF require a nascent transcript longer than 18 nucleotides to stably associate with the Pol II elongation complex (Missra et al., 2010; Yamaguchi et al., 2013), and both NELF and DSIF can contact nascent RNA exiting from Pol II (Bernecky et al., 2017; Qiu et al., 2017; Ehara et al., 2017; Vos et al., 2018a; Vos et al., 2018b), together suggesting that NELF and DSIF interaction with nascent RNA plays a role in establishing Pol II pausing, and that Pol II pausing might be triggered independently of promoter sequence and chromatin properties (Haberle et al., 2018).

Different hypotheses for the functions of Pol II pausing have been proposed, including establishing permissive chromatin, rapid / synchronous gene activation, integrating multiple regulatory signals and acting as a checkpoint for coupling elongation and RNA processing (Adelman et al., 2012).

2.2.5 Escape from the pausing (pause-release)

Release of paused Pol II is mediated by the positive transcription elongation factor b (P-TEFb), which is composed cyclin-dependent kinase 9 (CDK9) and cyclin T (Peterlin et al., 2006; Zhou et al., 2012). P-TEFb can be recruited directly by activators like NF-κB and MYC, or indirectly via coactivator such as BRD4 (Jonkers et al., 2015). After its recruitment, P-TEFb phosphorylates NELF, DSIF and Ser2 of the Pol II CTD (Kwak et al., 2013b), leading to the dissociation of NELF and conversion of DSIF into a positive transcription elongation factor, thereby triggering paused Pol II release (Jonkers et al., 2015) (***Figure 2-18***). A detailed list of pausing- related factors see ***Table 2-4***.

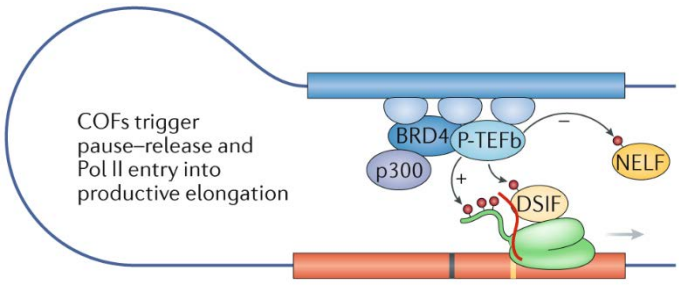


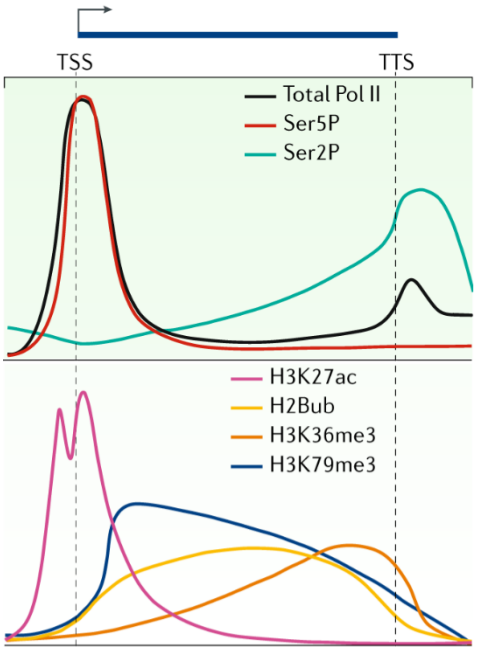
Figure 2-18: Paused Pol II release. From (Haberle et al., 2018)

Table 2-4: Pausing- related factors. From (Chen et al., 2018a).

Factors	Occupancy	Function in pausing
NELF	Promoter (P)	Stabilizes paused Pol II by preventing premature promoter proximal termination
DSIF	P, Gene body (G)	Promotes the recruitment of NELF and capping factors
PAF1C	Enhancer (E), P, G	Modulates enhancer activity and maintains paused Pol II by hindering its release into productive elongation
Gdown1	P	Blocks TFIIF recruitment and prevents early termination of promoter- proximal Pol II
PARP1	E, P	ADP-ribosylates NELF and inhibits its function in pausing
P-TEFb	E, P, G	Phosphorylates the Pol II CTD, NELF and the SPT5 CTR to promote release from pausing
SEC	E, P, G	Most active P-TEFb-containing complex; promotes rapid release of paused Pol II into productive elongation
BRD4	E, P, G	Stimulates P-TEFb activity and promotes pause release
7SK	P	Sequesters P-TEFb and prevents pause release

2.2.6 Productive elongation

After escaping from the promoter-proximal pausing, Pol II enters into productive elongation. During this process, the dynamic nucleosome turnover that allows Pol II to move forward while prevents cryptic intragenic transcription, is mainly driven by chromatin remodelers and histone chaperones (Venkatesh et al., 2015; Talbert et al., 2017; Lai et al., 2017). For instance, FACT travels with Pol II during elongation and mediates eviction of H2A-H2B dimers, facilitating the passage of Pol II (Chen et al., 2018a).



During elongation, besides the rapid disassembly and reassembly of nucleosomes, histone PTMs also significantly change, H2Bub, H3K36me3, H3K79me2 and H3K79me3 are enriched on the gene body (Vakoc et al., 2006; Fuchs et al., 2014; Chen et al., 2018a) (***Figure 2-19***). Notably, in mammals, H3K36me3 is deposited by Pol II CTD associated SETD2, and DNMT3B can recognize this mark and methylate the associated DNA on gene body to help prevent cryptic transcription (Baubec et al., 2015; Neri et al., 2017).

Figure 2-19: PTMs of Pol II and histones at gene bodies. Adapted from (Chen et al., 2018a)

Moreover, transition from Pol II pausing to productive elongation is not an ON-OFF switch. The Pol II elongation rate (kb/min) varies between genes, and also varies between different parts of the gene. In mammals, productive elongation is not very efficient within the first kilobase, and it increases from ~0.5 kb/min within the first few kilobases to 2-5 kb/min after ~15 kb (Jonkers et al., 2015). In addition, Pol II can be slowdown by mRNA cleavage and presence of exons, as well as polyadenylation sites (Jonkers et al., 2015).

2.2.7 Transcription termination

Transcription termination is necessary to partition the genome by defining the boundaries of transcription units (Porrua et al., 2015).

In metazoans, transition from productive elongation to termination is triggered by cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CSTF), cleavage factor I (CFI) and CFII, which are all bound to Pol II. Notably, CSTF and CFI-CFII bind to phosphorylated Ser2 of CTD (Kuehner et al., 2011; Porrua et al., 2015). CPSF and CSTF recognize the 3' UTR of the nascent RNA, cleavage of which is mediated by CPSF at 18-30 nucleotides downstream of the polyadenylation signal 'AAUAAA'. Released nascent RNA is polyadenylated at the 3'end and subsequently exported to the cytoplasm (Porrua et al., 2015) (***Figure 2-20 A&B***).

Two alternative models have been proposed for Pol II expulsion after transcript cleavage, namely, the torpedo model (West et al., 2004) and the allosteric model (Zhang et al., 2006). According to the torpedo model, cleavage of the nascent RNA allows the entry of the 5'-3' exoribonuclease XRN2, which degrades the transcript downstream of the poly(A), subsequently leading to transcription termination. On the other hand, the allosteric model posits that poly(A) signal-dependent loss of elongation factors and/ or conformational changes of Pol II destabilizes the elongation complex, triggering termination (Porrua et al., 2015; Chen et al., 2018a) (***Figure 2-20 C***).

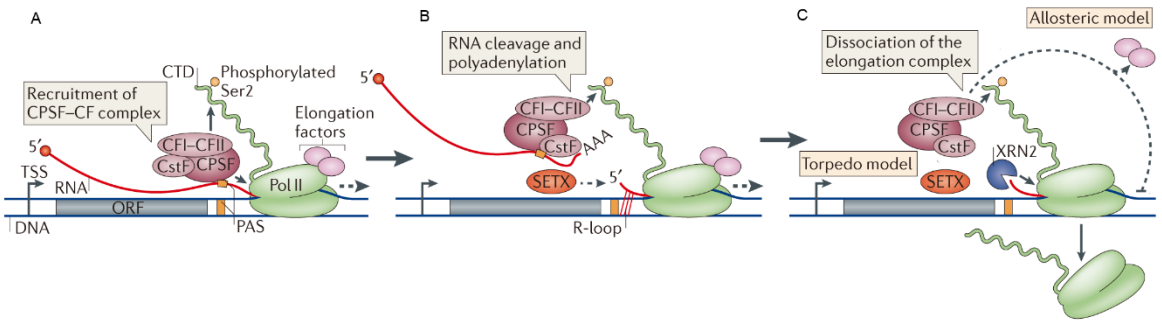


Figure 2-20: Transcription termination. Adapted from (Porrua et al., 2015).

Transcription termination occurs when the when Pol II is released from chromatin. Free

Pol II can be recycled and reinitiate to start a new round of transcription.

2.2.8 Transcription re-initiation

It has been reported that after transcription initiation, a subset of the transcription machinery remains at the promoter, forming a platform for the assembly of a second PIC (Roberts et al., 1995; Zawel et al., 1995; Sandaltzopoulos et al., 1998; Hahn, 1998).

Notably, after Pol II promoter escape, some GTFs (including TFIID, TFIIA, TFIIH and TFIIE) and Mediator can remain associated with the promoter as a re-initiation intermediate, which can be stabilized by activator, act as a scaffold for re-initiation complex formation and allow Pol II to initiate efficiently in successive rounds of transcription (Yudkovsky et al., 2000).

Moreover, a recent study showed that transcription induced downstream promoter binding of TAFs promotes subsequent activator-independent transcription re-initiation, suggesting that TAFs could function as re-initiation factors (Joo et al., 2017). It was proposed that after the pioneer round of PIC assembly and transcription initiation, TAFs stably interact with downstream promoter DNA and the acetylated +1 nucleosome, acting as a re-initiation intermediate, which could further stimulate re-initiation, or facilitate the incorporation of H2A and may act as a memory of recent transcription to allow rapid re-initiation, or return to the inactive state (Joo et al., 2017).

2.3 *Cis*-acting elements, code behind the cycle

The regulatory DNA sequences plus the associated epigenetic modifications of a specific gene provide the code that dictates when, where and at which level the gene should be transcribed. Beside epigenetic modifications, the regulatory sequences code usually derives from three parts (Fuda et al., 2009) (***Figure 2-21***): the core promoter (Juven-Gershon et al., 2010; Muller et al., 2014), the transcription factor binding sites (TFBSs) proximal to the core promoter (Ong et al., 2011; Lenhard et al., 2012), and the more distant *cis*-regulatory elements (≥ 1 kb from TSS, including enhancers, silencers and insulators (Ong et al., 2011). The promoter-proximal TFBSs can also exist in clusters, forming *cis*-regulatory modules (CRMs) (Lenhard et al., 2012).

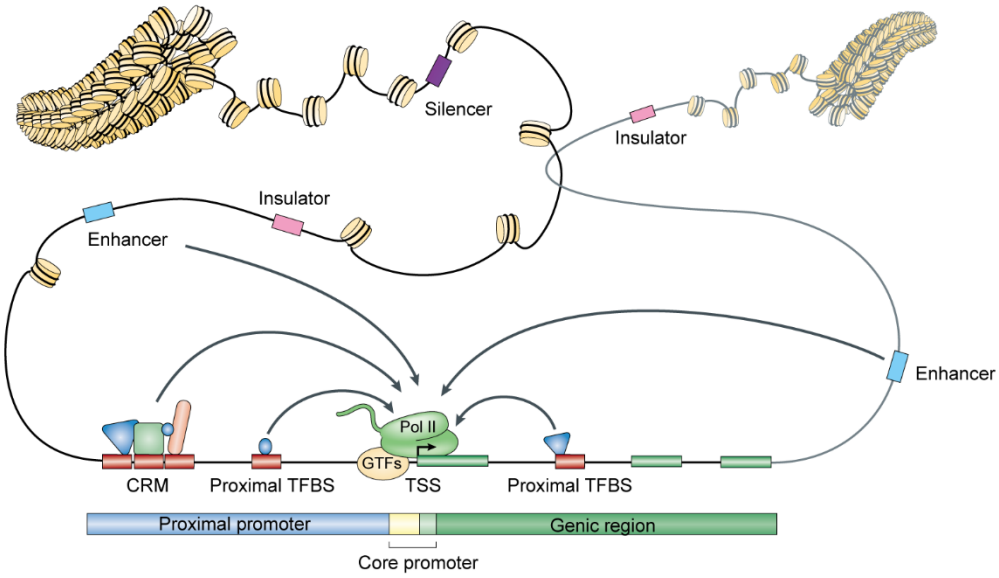


Figure 2-21: Transcription regulatory interactions with *Cis*-acting elements. Adapted from (Lenhard et al., 2012).

Promoter-proximal TFBSs and enhancers direct the binding of activators/repressors, which further mediate the recruitment of coactivators/repressors, followed by PIC assembly on the core promoter. In the following part, I will mainly focus on the core promoter and enhancer.

2.3.1 Core promoter

Core promoter, the gateway to transcription (Vo Ngoc et al., 2017b), is defined as a short sequence flanking the TSS that is sufficient to assemble the RNA polymerase II transcription machinery and to initiate transcription. It encompasses ~50 bp upstream and ~50 bp downstream of the TSS (Haberle et al., 2018).

Extensive works from many groups have characterized different features of core promoters, including the sequence motifs, TSS patterns and chromatin properties [reviewed in (Muller et al., 2007; Juven-Gershon et al., 2010; Lenhard et al., 2012; Muller et al., 2014; Roy et al., 2015; Vo Ngoc et al., 2017b; Haberle et al., 2018)].

2.3.1.1 Core promoter elements

The activity of the core promoter largely depend on its sequence motifs, which are diverse in terms of sequence differences and also their positions relative to TSS (Vo Ngoc et al., 2017b; Haberle et al., 2018) (***Figure 2-22***) (***Table 2-***).

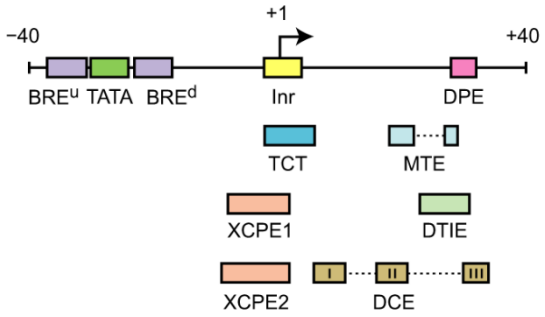


Figure 2-22: Pol II core promoter motifs. Typically, a core promoter might contain zero to three of the shown motifs, locations of which are roughly scaled. From (Vo Ngoc et al., 2017b).

The TATA box was the first identified core promoter motif in eukaryotes (Goldberg, 1979), and is located ~30 bp upstream of TSS and serves as the conserved binding site for the TBP (Patikoglou et al., 1999). However, only a minority of core promoters contain a TATA box, for instance, ~5% of core promoters in flies and ~10% in humans (Ohler et al., 2002; FitzGerald et al., 2006; Yang et al., 2007). Among the TATA-less promoters, many contain the initiator (Inr) motif, although it is also present often in TATA-containing promoters (Roy et al., 2015).

The Inr motif spans the TSS and can be recognized by TAF1 and TAF2 (Chalkley et al., 1999; Louder et al., 2016). It is more abundant than the TATA box but still not universal. In TATA-less promoters, the Inr motif is often accompanied by the downstream promoter element (DPE).

DPE was initially discovered in *Drosophila*, it is located downstream of TSS and functions

cooperatively with the Inr for TFIID binding and transcriptional activity (Burke et al., 1996). DPE commonly exists in flies (~30% of core promoters) but appears to be rare in humans (Burke et al., 1997; Kutach et al., 2000). Moreover, DPE rarely co-exists with TATA box in flies, therefore they were suggested to be associated with genes of different functional categories (Haberle et al., 2018).

The motif ten element (MTE) was identified as an overrepresented sequence in *Drosophila* core promoters (Ohler et al., 2002), and it was found to be a TFIID binding motif (Lim et al., 2004; Theisen et al., 2010). The downstream core element (DCE) comprising three sub-elements, is another core promoter motif that can be recognized by TFIID (Lewis et al., 2000; Lee et al., 2005).

The TFIIB recognition element (BRE) was initially identified as a TFIIB binding site that is located immediately upstream of TATA box (Lagrange et al., 1998). Later on, a second BRE located downstream of TATA box was found (Deng et al., 2005), thus they were renamed as BRE^{u} and BRE^{d}, respectively. During PIC assembly, TFIIB binds to TFIID-DNA complex forming a ternary complex, in which TFIIB interacts with TBP as well as DNA in both flanks of TATA box (Sainsbury et al., 2015), namely, BRE^{u} and BRE^{d}.

The TCT motif is also known as the polypyrimidine initiator, which is a rare motif that was estimated can be found in only ~1% of human core promoters (Vo Ngoc et al., 2017a). However, interestingly, the TCT motif is present in the core promoters of nearly all ribosomal protein genes, as well as some translation initiation and elongation factor coding genes in both flies and humans (Hariharan et al., 1990; Perry, 2005; Roepcke et al., 2006; Parry et al., 2010). Moreover, *in vitro* foot-printing experiments revealed that TCT motif is not recognized by the canonical TFIID (Parry et al., 2010). It seems that TCT motif-containing promoters preferentially employ a specialized PIC without canonical TFIID to transcribe translation related genes (Muller et al., 2014). Specifically, in *Drosophila*, TCT-dependent transcription involves the use of TRF2/TBPL1 instead of the canonical TBP (Wang et al., 2014) (discussed in more detail in section ***3.2.2.4***).

In addition to the core promoter motifs discussed above, some other motifs also have been discovered (see ***Table 2-5***). These motifs, together with other promoter features, can be recognized and bound by GTFs, thus potentially mediating PIC recruitment and assembly.

It is important to mention that, the sequence motifs described above are mostly studied in 'focused' promoters (see section ***2.3.1.2***) with single dominant TSS, and there are no universal core promoter motifs. Indeed, although various core promoter motifs have been identified, many core promoters do not contain any of these known motifs, probably these promoters possess other features or there are other core promoter motifs remain to be discovered.

Table 2-5: Known core promoter motifs. From (Haberle et al., 2018).

Motifs	Sequence logo	Consensus sequence	Position relative to TSS	Bound by	Fly	Human
TATA box		TATAWAWR	-31 to -24	TBP	+	+
Inr (fly)		TCAGTY	-5 to -2	TAF1, TAF2	+	−
Inr (human)		YR	-1 to +1	NA	−	+
		BBCABW	-3 to +3			
DPE		RGWCGTG	+28 to +34	TAF6,TAF9 possibly TAF1	+	Possibly Rarely
		RGWYVT	+28 to +33			
		GCGWKCGGTTS	+24 to +32		+	−
MTE		CSARCSSAACGS	+18 to +29	Possibly TAF1 and TAF2	+	−
Ohler 1		YGGTCACACTR	-60 to -1	M1BP	+	−
Ohler 6		KTYRGTATWTTT	-100 to -1	NA	+	−
Ohler 7		KNNCAKCNCTRNY	-60 to +20	NA	+	−
DRE		WATCGATW	-100 to -1	Dref	+	+
TCT		YYCTTTYY	-2 to +6	NA	+	+
BRE^{u}		SSRCGCC	-38 to -32	TFIIB	+	+
BRE^{d}		RTDKKKK	-23 to -17	TFIIB	+	+
DCEI – DCEIII	NA	CTTC	+6 to +11	TAF1	−	+
		CTGT	+16 to +21			
		AGC	+30 to +34			
XCPE1	NA	DSGYGGRASM	-8 to +2	NA	?	+
XCPE2	NA	VCYCRTTRCMY	-9 to +2	NA	?	+
Pause button		KCGRWCG	+25 to +35	NA	+	?

2.3.1.2 TSS patterns

Transcription typically is initiated at a defined position, called transcription start site (TSS), at the 5' end of a gene. However, transcription initiation sites mapping at single nucleotide resolution revealed that core promoters have distinct TSS patterns (Carninci et al., 2006), according to which core promoters have been classified into three types: 'focused', 'dispersed' and 'mixed' core promoters (Carninci et al., 2006; Juven-Gershon et al., 2010; Vo Ngoc et al., 2017b). Specifically, 'focused' core promoters have a single well-defined TSS or a narrow cluster of TSSs probably derived from a single PIC, 'dispersed' core promoters have multiple weak TSSs with similar use frequency over an ~50 to ~100 bp region, and a variety of 'mixed' TSS patterns commonly exist (Vo Ngoc et al., 2017b) (***Figure 2-23***).

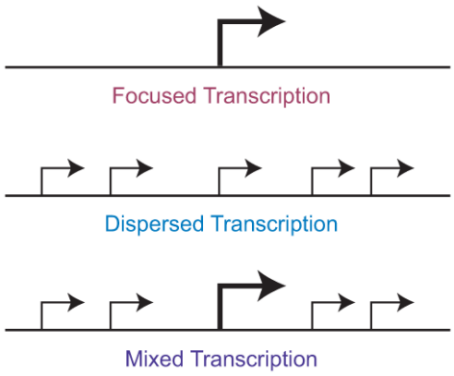


Figure 2-23: Core promoter TSS patterns. From (Vo Ngoc et al., 2017b).

Indeed, ~70% of vertebrate genes have 'dispersed' (or mixed) promoters (Juven-Gershon et al., 2010), which add another layer of control for transcription through alternative TSSs selection/usage. Interestingly, it has been reported that during zebrafish early embryonic development, there is a widespread switch in TSS usage throughout maternal to zygotic transition (Haberle et al., 2014).

2.3.1.3 Chromatin signals at the core promoter

Transcription occurs in the context of chromatin, therefore, besides the presence of sequence motifs, chromatin structure and composition can also influence promoter activity (Muller et al., 2014; Vo Ngoc et al., 2017b) (***Figure 2-24***).

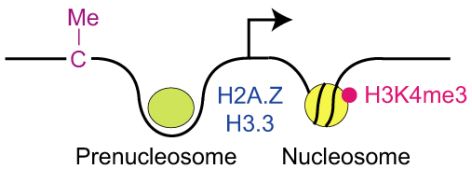


Figure 2-24: Chromatin signals at the core promoter. From (Vo Ngoc et al., 2017b).

First of all, a prominent feature of active gene promoter associated chromatin is the enrichment of specific histone PTMs, such as H3K4me3, H4ac and H3K27ac (Barski et al.,

2007) (H3K27ac pattern see ***Figure 2-19***) (***Figure 2-25***). It has been shown that TAF3 binds to H3K4me3 and facilitates PIC assembly (Vermeulen et al., 2007; van Ingen et al., 2008; Lauberth et al., 2013). Although it has been reported that, in yeast, H3K4me3 deposition is mediated by SET1 that recruited by Pol II and thus occurs downstream of transcription (Ng et al., 2003). However, in mammals, it is known that zinc finger CXXC proteins such as CFP1, MLL and KDM proteins specifically recognize non-methylated CpG islands (CGIs) and regulate lysine methylation on histone tails (Long et al., 2013). CFP1 and MLL proteins can deposit H3K4me3, and as a result, CGIs at promoters tend to be marked by H3K4me3 independently of gene activity (Mikkelsen et al., 2007; Thomson et al., 2010; Muller et al., 2014). Besides, TAF1 binds to multiply acetylated histone H4 peptides (Jacobson et al., 2000), thus H4ac on promoters may serve as a target for TFIID recruitment as well. The role of promoter associated H3K27ac is not clear. Strikingly, in *Drosophila*, it has been reported that H3K9me3 could direct the transcription of PIWI-interacting RNA (piRNA) clusters in heterochromatin (Andersen et al., 2017) (discussed in more detail in section ***3.2.2.4***). Moreover, in plants, Pol IV (a variant of Pol II) is recruited to promoters containing methylated H3K9 (Law et al., 2013; Zhang et al., 2013a).

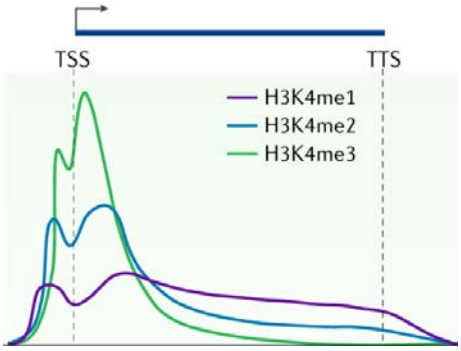


Figure 2-25: H3K4me enrichment at active genes. Adapted from (Chen et al., 2018a).

Secondly, in vertebrates, many core promoters overlap with CGIs, but the mechanisms by which CGIs confer core promoter function are still unknown (Haberle et al., 2018). Although non-methylated CGIs can nucleate the deposition of H3K4me3, there is no evidence that they can nucleate the recruitment of transcriptional machinery to a promoter, instead, 5mC is a stable repressive regulator of promoter activity. However in plants, it has been shown that Pol V (a variant of Pol II) can be recruited to promoters containing methylated DNA via factors that bind to both polymerase and methylated DNA (Johnson et al., 2014; Liu et al., 2014). It is possible that similar mechanisms also exist in vertebrates. Indeed, *in vitro* studies have uncovered that many TFs bind the methylated motifs (Hu et al., 2013b; Zhu et al., 2016). Moreover, many TFs, including some developmentally important proteins, even have higher affinity for CpG-methylated sequences (Yin et al., 2017). Further studies in physiological

contexts are needed. In addition, DNA hydroxy-methylation and 6mA are also the potential targets of the TFs (Vo Ngoc et al., 2017b).

Thirdly, another hallmark of active core promoters is the presence of nucleosome depleted regions (NDRs) (Yuan et al., 2005; Mavrich et al., 2008; Jiang et al., 2009). Interestingly, it has been shown that those NDRs can be occupied by prenucleosome (nucleosome isomer associated with ~80bp of DNA) (Fei et al., 2015; Khuong et al., 2015), prenucleosome-like particles (histone-containing particles associated with ~61 to 100bp of DNA) (Rhee et al., 2014; Ishii et al., 2015), or H3.3- and H2A.Z- containing nucleosomes (Jin et al., 2009). These features are proposed to ensure chromatin accessibility to facilitate transcription, however, it is not clear whether they participate directly in the recruitment of the transcriptional machinery, for instance, by being the TF targets. Given that H3.3 constitutes ~10% of total histone H3 and H2A.Z constitutes ~1%-3% of total H2A, it has been suggested that presence of H3.3 and H2A.Z at the promoters could provide some specificity to core promoter function (Dang et al., 2016; Vo Ngoc et al., 2017b).

2.3.1.4 Core promoter types

Base on their distinct features (including sequence motifs, TSS patterns and chromatin properties that were discussed above) and the function of their dominant genes, major core promoters have been grouped into three general functional classes, referred as types I, II and III (Lenhard et al., 2012; Haberle et al., 2018) (***Figure 2-26***).

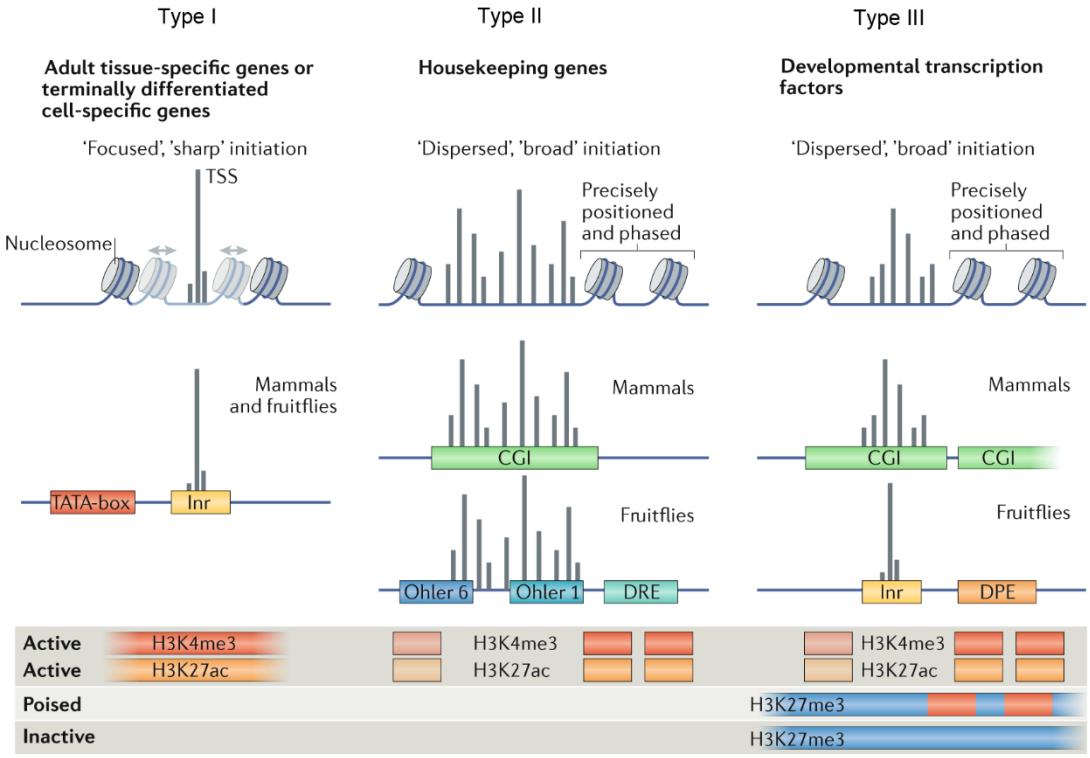


Figure 2-26: Three types of core promoters. Adapted from (Haberle et al., 2018).

Type I consists of adult tissue-specific promoters, which have TATA-box, Inr and 'sharp' TSS pattern, but lack CGIs. Type II promoters are associated with 'housekeeping' genes, and are TATA-less promoters with 'dispersed' TSS patter. In mammals, they overlap individual CGIs. Finally, Type III promoters are preferentially associated with developmental transcription factors. In mammals, they are associated with large CGIs or multiple CGIs (Lenhard et al., 2012; Haberle et al., 2018). TCT promoters are minor promoters not included in this classification. They have 'sharp' TSS pattern, but other features differ substantially from type I promoters (Lenhard et al., 2012).

2.3.1.5 Divergent transcription with unidirectional core promoters

Mammalian promoter regions frequently exhibit divergent transcription, a phenomenon in which transcription of downstream protein-coding genes is coupled with the transcription of short (50-2000 nucleotides) and unstable upstream antisense RNAs (uaRNAs) in the reverse direction (Core et al., 2008; Preker et al., 2008; Seila et al., 2008; Wu et al., 2013; Core et al., 2014; Andersson et al., 2015) (***Figure 2-27***).

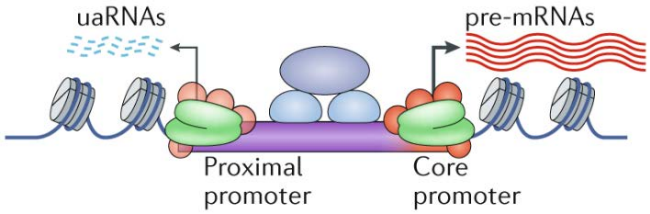


Figure 2-27: Divergent transcription with unidirectional core promoters. From (Haberle et al., 2018)

Further analysis revealed that bidirectional transcription arises from two distinct and inherently unidirectional transcription complexes, and the uaRNAs are transcribed by separate Pol II from their own reverse-directed core promoters (Duttke et al., 2015; Scruggs et al., 2015) .

2.3.2 Enhancer

Enhancers are distal regulatory DNA sequences that contain activator binding sites and can increase the transcription level of target genes independently of distance and orientation (Benoist et al., 1981; Banerji et al., 1981; Haberle et al., 2018), thus playing critical roles in spatial and temporal control of gene expression. Prior to activation, enhancers can exist in a primed state, which are associated with H3.3/H2A.Z incorporation, pioneer TFs binding and presence of H3K4me1. Upon activation, enhancers initiate bidirectional eRNAs transcription, and active enhancers are typically marked by H3K4me1 and H3K27ac (***Figure 2-28***), and are

also associated with the incorporation of H3.3/H2A.Z histone variants, relatively high DNA accessibility, TFs and coactivators binding, and enrichment of Pol II (Ser5P but not Ser2P). Based on these features, large number of putative enhancers (>400,000 to ~1 million) have been annotated in the human genome. Active enhancer can turn to poised state, which is characterized by the presence of both H3K4me1 and H3K27me3. In addition, latent enhancers also exist: they are located in closed chromatin and are not labelled by any enhancer-associated marks, but they acquire H3K4me1 and H3K27ac marks upon stimulation (Spitz et al., 2012; Calo et al., 2013; Rivera et al., 2013; Shlyueva et al., 2014; Vernimmen et al., 2015; Li et al., 2016).

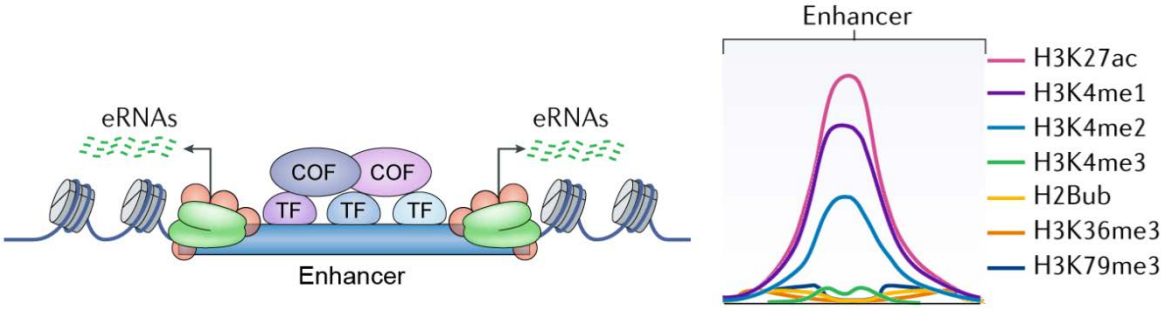


Figure 2-28: Features of active enhancer. Left: Transcription of enhancer RNAs from active enhancer. Right: Histone PTMs at active enhancer. Adapted from (Chen et al., 2018a; Haberle et al., 2018)

2.3.3 Enhancer-promoter communication

As mentioned above (section ***2.3.2***), enhancers increase the transcription level of target genes independently of distance. The distance between enhancers and promoters can range from a few kilobases to over 1000 kb in metazoans (Noonan et al., 2010). How can enhancer convey regulatory information across such distance? Several models for enhancer-promoter communication have been proposed based on studies of specific loci and genome-wide analysis (Vernimmen et al., 2015; Furlong et al., 2018) (***Figure 2-29***).

The tracking model (***Figure 2-29 A***) proposes that enhancer-bound Pol II / transcription complexes move progressively along the DNA, dragging the enhancer towards the target promoter, thus resulting in the formation of a loop, and size of the loop increases progressively until Pol II reaches the promoter (Kong et al., 1997; Blackwood et al., 1998; Hatzis et al., 2002; Wang et al., 2005b; Vernimmen et al., 2015). The linking model (***Figure 2-29 B***) suggests that an activator binds to enhancer first and facilitates the recruitment of other TFs, forming protein-protein oligomers to bridge the enhancer and the target promoter (Bulger et al., 1999; Dorsett, 1999; Vernimmen et al., 2015; Furlong et al., 2018). The looping model (***Figure 2-29***

C) implies that enhancer-bound TFs and promoter-bound TFs physically interact with each other by looping out the intervening DNA. The first evidence for looping came from *E. coli* (Lee et al., 1989), and after, many proteins have been proposed to bridge chromatin looping, including Mediator, CTCF, TAF3, BRG1 [reviewed in (Vernimmen et al., 2015)]. The looping-tracking/linking model (***Figure 2-29 D***) proposes that long-range loops can bring enhancers close to (but not directly contacted with) the target promoter, tracking or linking would bridge the distance left (Furlong et al., 2018).

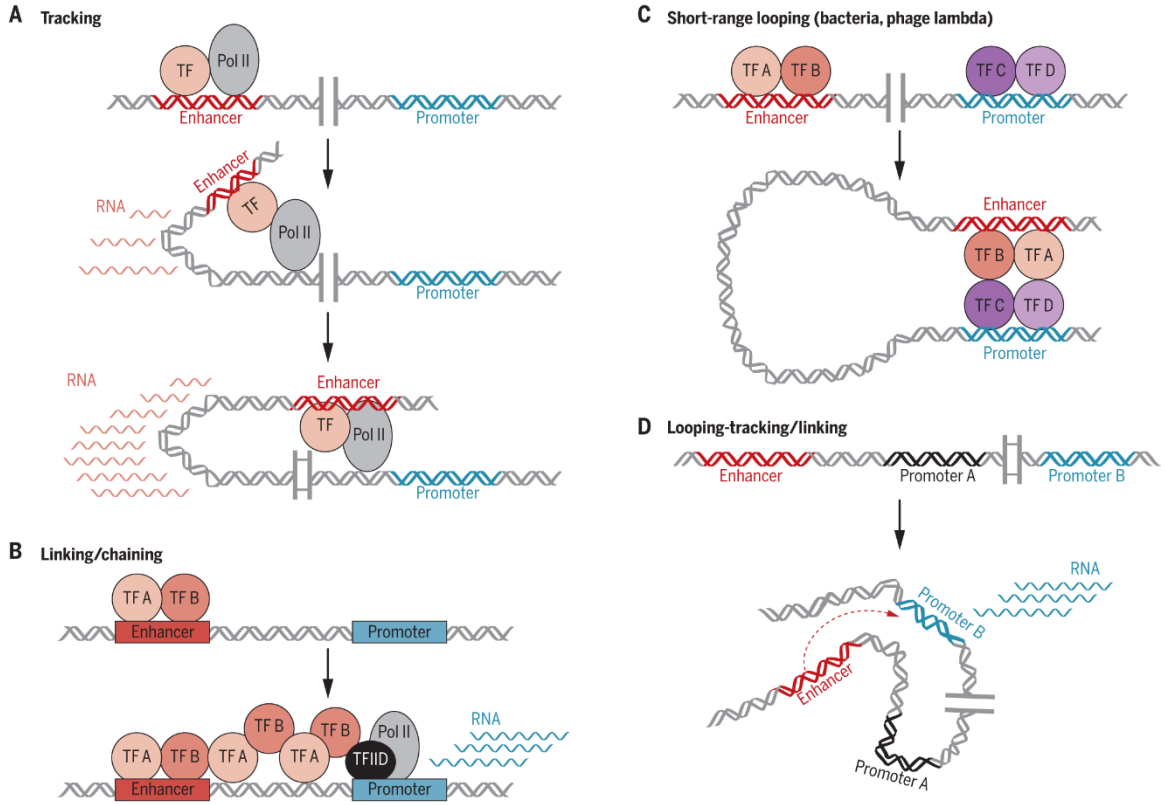


Figure 2-29: Models of enhancer-promoter communication. (A) Tracking model. (B) Linking model. (C) Short range looping. (D) Long-range looping coupled with tracking or linking. From (Furlong et al., 2018).

3. Diversity of basal transcription machinery

TBP and TAFs are important for embryogenesis. For instance, in mice, knockout of *Tbp* (Martianov et al., 2002b)*,Taf7* (Gegonne et al., 2012), *Taf8* (Voss et al., 2000) and *Taf10* (Mohan et al., 2003) lead to embryonic lethality between E3.5 and E6.5. *Taf4* knockout embryos survive until E9.5 because of the compensation of the paralog TAF4b, and from E9.5 TAF4b cannot fully compensate the loss of TAF4 anymore, especially for the transcription of critical differentiation genes (Langer et al., 2016). No knockout of other prototypical TAFs has been reported. However, it has been shown that TAFs are differentially required during development (Metzger et al., 1999; Indra et al., 2005; Fadloun et al., 2008; Liu et al., 2011; Bardot et al., 2017; El-Saafin et al., 2018), which provide evidences for the heterogeneity of TFIID. Moreover, emerging evidences show that, apart from epigenetic regulations and gene-specific transcription factors, variation in the composition of basal transcription machinery, especially the heterogeneity and specialized functions of TFIID with different TAFs and TAF paralogs, as well as specialized role of TBP-related factors, plays vital roles in driving cell-type-specific and gene-specific transcriptions (D'Alessio et al., 2009; Muller et al., 2010; Goodrich et al., 2010; Levine et al., 2014). It is important to mention that, many evidences for diversity of the basal transcription machinery come from germ cells. In this chapter, I will mainly focus on the studies from mice, and if necessary, in comparation with studies in *Xenopus*, zebrafish, *Drosophila* and *C. elegans*.

3.1 Heterogeneity and specialized functions of TFIID

3.1.1 Differential requirement of several TAFs during development

3.1.1.1 Differential requirement of TAF10

TAF10 is a ubiquitously expressed TFIID subunit, and knockout of *Taf10* leads to embryonic lethality between E3.5 and E6.5 (Mohan et al., 2003). Many evidences show requirement of TAF10 varies in different contexts.

First of all, TAF10 disruption leads to early embryonic lethality shortly after implantation, while inner cell mass cells die by apoptosis, trophoblastic cells survive. Although with reduced DNA replication and transcription, these trophoblastic cells can survive in culture for at least 12 days (Mohan et al., 2003). Secondly, TAF10 is essential for the survival and proliferation of F9 carcinoma cells, however, interestingly, it is dispensable for retinoic acid (RA)-induced primitive endodermal differentiation (Metzger et al., 1999). Thirdly, *Taf10* conditional deletion in keratinocytes has shown that ablation of *Taf10* in foetal keratinocytes impairs their terminal

differentiation and alters skin barrier function. However, TAF10 is dispensable for epidermal keratinocyte proliferation and differentiation in adult mice (Indra et al., 2005). Fourthly, conditional knockout of *Taf10* in foetal and adult hepatocytes also has shown different requirement of TAF10 for transcription in embryo and adult (Tatarakis et al., 2008). Moreover, recently, it has been shown that absence of TAF10 does not affect global steady-state mRNA and the dynamic transcription initiation of cyclic genes in presomitic mesoderm (PSM) at ~E9.5 (Bardot et al., 2017).

In addition, in chicken DT40 cells, loss of TAF10 eventually leads to apoptotic cell death, however, activation of *c-fos* transcription can still occur efficiently in absence of TAF10 (Chen et al., 2000). In *C. elegans*, TAF10 is needed for a significant fraction of transcription, but apparently is not required for the expression of many metazoan-specific genes (Walker et al., 2001). Altogether, these studies indicate that TAF10 is differentially required depending on the cellular and developmental context.

3.1.1.2 TAF8 mutation does not impair Pol II transcription

Loss of TAF8 in mouse embryonic stem cells (ESCs) leads to cell death and a global decrease in Pol II transcription, however, in human patient cells with homozygous *Taf8* frameshift mutation, global Pol II transcription is not impaired, although canonical TFIID assembly is altered as unstable TAF8 mutant protein is undetectable (El-Saafin et al., 2018). In addition, it also has been reported that, TAF8 is not detected in preadipocytes, but is dramatically upregulated during adipogenesis (Guermah et al., 2003). These observations suggest that partial TFIID without TAF8 or altered TFIID with mutated TAF8 are sufficient for Pol II transcription, at least in certain context.

3.1.1.3 TAF7 is not essential for mature T cell survival or differentiation

Homozygous *Taf7* deletion results in early embryonic lethality between E3.5 and E5.5, and ablation of *Taf7* in embryonic fibroblasts ceases global transcription. However, functional analysis of TAF7 in thymocytes shows that, while TAF7 is required during the early steps in thymic development [double-negative ($CD4^-CD8^-$) to double-positive ($CD4^+CD8^+$) transition] for the differentiation and proliferation of immature thymocytes, it is not essential for the survival or differentiation of mature T cells (Gegonne et al., 2012).

3.1.1.4 TAF4 is differentially required in T-RA-induced gene activation

Characterization of the role of TAF4 in the activation of cellular genes by all-*trans* retinoic acid (T-RA) revealed that T-RA regulates ~1000 genes in $Taf4^{lox/-}$ MEFs and less than 300 genes in $Taf4^{-/-}$ MEFs, indicating the existence of TAF4 dependent and independent

mechanisms for T-RA induced transcription activation (Fadloun et al., 2008). The TAF4 independent activation could due to compensation of TAF4 by its paralog TAF4b or TAF4-independent transcription initiation. Moreover, it has been shown that *Taf4*$^{-/-}$ ESCs are viable due to the compensation by TAF4b, and the observation that *Taf4* is dispensable for primordial germ cell generation may also due to the same mechanism (Langer et al., 2016).

It also has been shown that TAF4 (also TBP and some other TAFs) protein level is dramatically decreased in adult hepatocytes compared with E13.5 hepatoblasts, suggesting the a change in core promoter recognition complex in hepatocytes (D'Alessio et al., 2011). Later studies confirmed the reduced expression of TAF4 in adult hepatocytes, but not TBP, and showed that TAF4 is required for post-natal hepatocyte differentiation (Alpern et al., 2014).

3.1.1.5 TAF3 is differentially required during ESCs lineage commitment

In mice, TAF3 is highly expressed in embryonic stem cells (ESCs), interestingly, TAF3 expression (protein level) is reduced progressively during the formation of embryoid bodies (EBs) while expression of other TFIID subunits (TAF4 and TBP) remain mostly unchanged (Liu et al., 2011). Functional analysis showed that TAF3 is required for endoderm lineage differentiation, interacting with CTCF and mediating DNA looping between distal enhancer sites and core promoters of endoderm specification genes, while it prevents premature specification of neuroectoderm and mesoderm by repressing neuroectodermal genes (Liu et al., 2011).

Altogether, these observations show that the composition of TFIID can be variable depending on the cellular and developmental context.

3.1.2 Role of TAF paralogs during development

Besides differential requirement of TAFs, the heterogeneity of TFIID also arises from the TAF paralogs. In vertebrates, apart from the 13 prototypical TAFs, several TAF paralogs have been identified, including TAF1L, TAF4b, TAF5L, TAF6L, TAF7L and TAF9b [reviewed in (Kolthur-Seetharam et al., 2008; Muller et al., 2010; Goodrich et al., 2010)].

TAF1L, a retroposed copy of TAF1, has been identified as a human specific TAF1 paralog (also present in some but not all primates) and is specifically expressed in human male germ cells (Wang et al., 2002). TAF5L and TAF6L are SAGA specific TAF paralogs (Helmlinger et al., 2017). Thus in the following, I mainly focus on TAF4b, TAF7L and TAF9b.

3.1.2.1 Specialized roles of TAF4b in germ cell differentiation

TAF4b was initially identified as a cell-type-specific TAF in human differentiated B-cells (Dikstein et al., 1996). However, it was later found to be expressed in many tissues in mice and

highly enriched in the testis and the ovary (***Figure 3-2 a & b)*** (Freiman et al., 2001).

In mouse testis, the TAF4b protein is present in post-natal gonocytes, spermatogonia and spermatids (Falender et al., 2005a). *Taf4b*-null males are initially fertile, but progressively become infertile (by around 3 months), as testes degenerate with age resulting in seminiferous tubules devoid of germ cells (Falender et al., 2005a). Furthermore, in *Taf4b*-deficient males, gonocyte proliferation is impaired from post-natal day 2 onward, and expression of spermatogonial stem cell markers (such as *c-Ret, Plzf* and *Stra8*) is diminished (***Figure 3-1***).

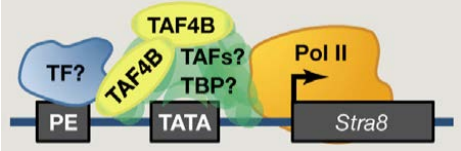


Figure 3-1: TAF4b-containing transcription machinery for male germ cell maintenance in spermatogonia. From (Levine et al., 2014).

In embryonic and early neonatal mouse ovary [from 14.5 d.p.c (day post coitum) up to postnatal day (P) 2], TAF4b protein selectively expressed in the oocytes but not somatic cells (Falender et al., 2005b). In adult mouse ovary, *Taf4b* mRNA is restricted to the granulosa cells (Freiman et al., 2001), and required for granulosa cell survival and proliferation (Voronina et al., 2007). Female mice lacking *Taf4b* are viable but infertile, developmental defects was observed as early as P3 when the number of oocytes is significantly reduced (Falender et al., 2005b). Females mice are with smaller ovaries that undergo progressive follicle loss and lack mature follicles, and many ovarian-specific genes are downregulated (Freiman et al., 2001; Lovasco et al., 2010). It was found later that *Taf4b* null females are able to ovulate, although many of the ovulated oocytes display defects in spindle formation and/or extrusion of polar body, some oocytes achieve metaphase II of meiosis (Falender et al., 2005b). However, few oocytes that are fertilized and embryos cannot proceed beyond the 1-cell stage (Falender et al., 2005b). Moreover, TAF4b controls granulosa-cell-specific expression of *c-Jun* (Geles et al., 2006), and it has been suggested that TAF4b works with c-Jun to regulate the transcription of genes involved in follicle growth (Goodrich et al., 2010).

The fact that *Taf4b*$^{-/-}$ mice are viable and only show defects in germ cells, suggests that TAF4b can be fully replace by TAF4 during embryogenesis, while TAF4b can only partially compensate for TAF4 functions as TAF4-deficient embryo die at ~E9.5 (Langer et al., 2016).

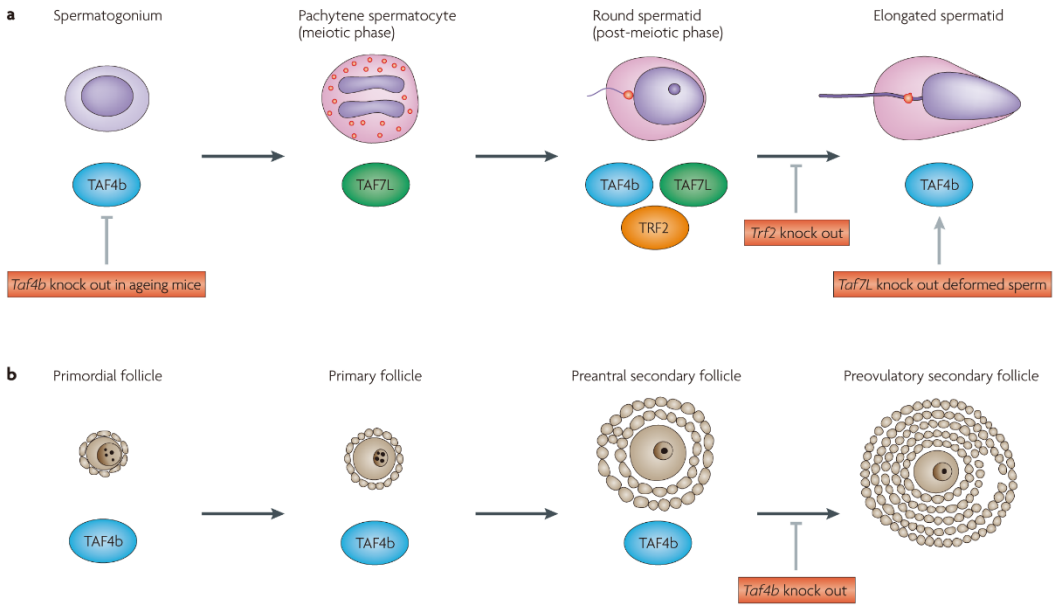


Figure 3-2: Roles of TRF2, TAF4b and TAF7L in germ cell differentiation. Adapted from (Goodrich et al., 2010)

3.1.2.2 TAF7L in germ cell and somatic cell differentiation

TAF7L, the paralogue of TAF7, was initially discovered as a spermatogonial expressed gene located on the X chromosome (Wang et al., 2001b; Pointud et al., 2003; Cheng et al., 2007), but later also was found in different mouse cells and tissues, and is highly enriched in testis and adipocytes (Zhou et al., 2013b; Zhou et al., 2014).

In mouse testis, TAF7L protein is localized in the cytoplasm of spermatogonia and early primary spermatocytes, and is imported into the nucleus from mid-pachytene stage onwards, eventually accumulates strongly in the nucleus of post-meiotic round spermatids (Pointud et al., 2003) (***Figure 3-2 a***). Interestingly, import of TAF7L into the nucleus correlates with increased TBP expression and decreased TAF7 expression, and TAF7L is associated with TBP in pachytene spermatocytes and round spermatids, indicating it can replace TAF7 as a TFIID subunit in these stages (Pointud et al., 2003). Knockout of the *Taf7l* gene leads to the development of deformed sperm (Cheng et al., 2007). The male mice are initially fertile with reduced litter size, however, after two to four additional rounds of back-crossings, the TAF7-deficient male mice become essentially sterile (Cheng et al., 2007; Zhou et al., 2013a). Importantly, TAF7 binds to the promoters of target genes in the testis and *Taf7l* ablation impairs the expression of many post-meiotic spermiogenic-specific as well as metabolic genes (Zhou et al., 2013a). Moreover, TAF7L forms a complex with TRF2 (for TRF2, see section ***3.2.2***), and they coregulate the expression of post-meiotic genes (Zhou et al., 2013a) (***Figure 3-3 A***).

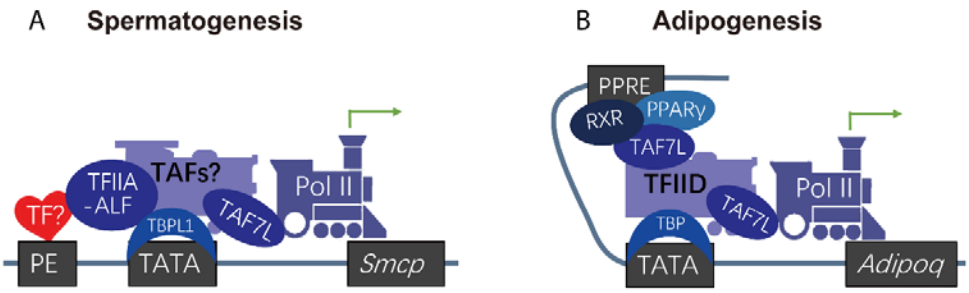


Figure 3-3: TAF7L regulates spermatogenesis and adipogenesis.

In adipocytes, TAF7L is required for adipocyte-specific gene expression (Zhou et al., 2013b). Depletion of TAF7L compromises adipocyte differentiation and white adipose tissue (WAT) development, while ectopic expression of TAF7L transdifferentiate C2C12 myoblast into adipocytes upon adipogenic induction (Zhou et al., 2013b). During adipogenesis, TAF7L binds to the promoters and/or promoter proximal regions of most adipocyte-specific genes, and interestingly, it co-localizes with PPARγ (a key adipogenic transcriptional activator) genome-widely (Zhou et al., 2013b). Moreover, TAF7L physically associates with PPARγ and TBP/TFIID, respectively (Zhou et al., 2013b). Altogether, these data suggest that TAF7L plays important roles during adipogenesis by serving as a cofactor for PPARγ and also a component of the core transcriptional machinery (Zhou et al., 2013b) (***Figure 3-3 B***). Furthermore, TAF7L also modulates brown adipose tissue formation (Zhou et al., 2014).

3.1.2.3 TAF9b

TAF9b, the paralog of TAF9, is present widely in many tested cell lines (Frontini et al., 2005). Human TAF9b can fully restore the function of chicken TAF9 in TAF9-deficient DT40 cells, suggesting partially redundant functions for these two proteins (Chen et al., 2003b). Like TAF9, TAF9b is also present in both TFIID and SAGA (Frontini et al., 2005). TAF9 and TAF9b are both essential for HeLa cell viability, however, transcriptome analysis revealed that they have minimal overlap of regulated genes, indicating they have distinct roles in gene transcription regulation (Frontini et al., 2005). Moreover, recent studies have shown that TAF9b is dispensable for global gene expression in mouse ES cells, and that TAF9b is up-regulated upon neuronal differentiation (Herrera et al., 2014). TAF9b controls neuronal gene expression, thus it is required for efficient differentiation of ES cells to motor neurons, where it is preferentially associated with SAGA rather than the canonical TFIID complex (Herrera et al., 2014).

In addition, *Drosophila* homologues of TAF4 (*No hitter*), TAF5 (*Cannonball*), TAF6 (*Meiosis I arrest*), TAF8 (*Spermatocyte arrest*) and TAF12 (*Ryan*) are selectively expressed in

primary spermatocytes and required for the regulation of testis-specific gene expression program during spermatogenesis (Hiller et al., 2001; Hiller et al., 2004).

Collectively, the differential requirement for TAFs and specialized functions of TAF paralogs suggest that, heterogeneous 'TFIIDs' with different combinations of subunits can participate in the gene-specific and tissue-specific transcriptional regulatory functions.

3.2 Specialized functions of TBP-related factors

Initially, TBP was thought to be a universal transcription factor, but later on, genomic and cDNA sequencing revealed that three TBP paralogs exist: the insect-specific TRF1 (also known as TRF), the metazoan-specific TRF2 (also known as TBPL1, TLP, TRP and TLF) and the vertebrate-specific TBP2 (also known as TBPL2 and TRF3) (Crowley et al., 1993; Berk, 2000; Davidson, 2003; Persengiev et al., 2003; Muller et al., 2010; Akhtar et al., 2011; Vo Ngoc et al., 2017b).

3.2.1 TRF1, an insect-specific TBP-related factor

TRF1 was the first identified TBP-related factor (Crowley et al., 1993), and has been found only in insects. In *Drosophila*, TRF1 associates with BRF1 (an RNA Pol III transcription factor) to form the TFIIIB complex and mediates Pol III-dependent tRNA gene transcription instead of TBP (Takada et al., 2000; Isogai et al., 2007b; Verma et al., 2013). Although TRF1 predominantly regulates Pol III transcription, it also has been reported that TRF1 can also bind to the TATA box and form a stable TRF1/TFIIA/TFIIB complex, substituting for TBP in Pol II transcription *in vitro* (Hansen et al., 1997; Holmes et al., 2000).

3.2.2 TRF2, a metazoan-specific TBP-related factor

TRF2 is the second identified TBP-related factor, and it has been given various names, such as TBP-like protein (TLP) (Ohbayashi et al., 1999a), TBP related factor 2 (TRF2) (Rabenstein et al., 1999; Teichmann et al., 1999), TBP-like factor (TLF) (Dantonel et al., 1999), TBP-related protein (TRP) (Moore et al., 1999) and also TBPL1. As TRF2 is more commonly used in the literature, I will also use the term TRF2 (but not to be confused with telomere-repeat binding factors, which are also termed TRF1 and TRF2).

TRF2 is a metazoan-specific TBP-related factor that shares ~40% identity with the TBP core domain and that interacts with TFIIA and TFIIB (Rabenstein et al., 1999; Teichmann et al., 1999). Unlike TBP and the other TRFs, TRF2 alone does not bind to the TATA box and does not appear to have any sequence-specific DNA-binding activity (Dantonel et al., 1999; Rabenstein et al., 1999; Wang et al., 2014; Vo Ngoc et al., 2017b). However, TRF2 has many functions in different organisms.

3.2.2.1 TRF2 regulates male germ cell differentiation in mouse

In mammals, TRF2 is broadly expressed at low levels in different tissues but is highly enriched in testes (Ohbayashi et al., 1999b; Rabenstein et al., 1999; Teichmann et al., 1999; Moore et al., 1999; Martianov et al., 2001).

In mice testis, dynamic expression of TRF2 is tightly controlled during spermatogenesis, TRF2 is highly expressed only in late-pachytene spermatocytes, late round spermatids and elongating spermatids, but not in other stages (Martianov et al., 2001; Zhang et al., 2001a; Martianov et al., 2002a) (***Figure 3-4***).

Trf2$^{-/-}$ mice are viable, but mutant male mice are sterile due to a severe defect in spermiogenesis (Martianov et al., 2001; Zhang et al., 2001b). In the TRF2-deficient males, spermatogonia and spermatocytes develop normally, however, acrosome formation is impaired in early stage round spermatids, and spermiogenesis is arrested at the transition of round spermatids to elongating spermatids, and most round spermatids undergo apoptosis during the arrest (Martianov et al., 2001; Zhang et al., 2001b; Martianov et al., 2002a). Importantly, transcription of multiple post-meiotic testes-specific genes is severely decreased, including the transition protein and protamine genes (Martianov et al., 2001; Zhang et al., 2001b). Further investigation showed that *Trf2*$^{-/-}$ mutant round spermatids display a fragmentation of the chromocenter (Martianov et al., 2002a), a nuclear structure containing centromeric heterochromatin from each chromosome. However, it is not yet clear whether the defect of spermatogenesis in TRF2-deficient mice is mainly due to the impaired gene expression or due to the defect of chromocenter formation.

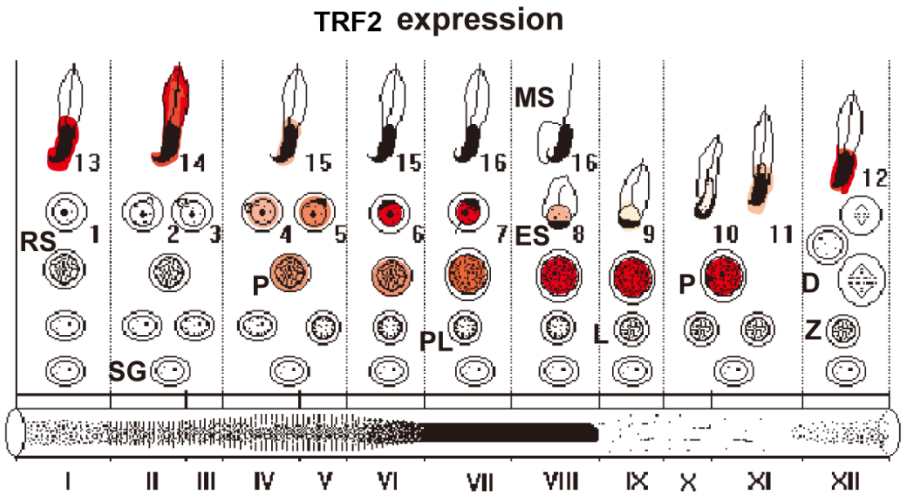


Figure 3-4: Dynamic expression of TRF2 during spermatogenesis. SG, spermatogonia; PL, preleptotene; L, leptotene; Z, zygotene; P, pachytene; D, diakinetic; RS, round spermatids; ES, elongating spermatids; MS, mature spermatozoa. Adapted from (Martianov et al., 2002a).

Later on, it has been reported that TRF2 can activate a number of genes, and together with TFIIA, it even binds to the promoter of neurofibromatosis type 1 gene (Chong et al., 2005). Moreover, TRF2 interacts with TAF7L in testis and they coregulated a subset of post-meiotic genes required for spermiogenesis (Zhou et al., 2013a). Furthermore, TRF2 associates with TFIIA and/or ALF, forming a stable complex chaperoned by heat shock proteins in the cytoplasm (Catena et al., 2005; Martianov et al., 2016), and it is recruited to active spermatid gene promoters together with TBP, TAF7L and Pol II (Martianov et al., 2016). Based on these observations, it has been proposed that TRF2 is recruited to the preinitiation complex as a testis-specific subunit of TFIIA/ALF that cooperates with TBP and TAF7L to regulate spermatid gene expression (Martianov et al., 2016) (***Figure 3-5***). It is also possible that haploid gene expression is driven by two distinct set of PICs, one containing TRF2-TFIIA/ALF while the other containing TBP-TAF7L (Martianov et al., 2016) (***Figure 3-5***).

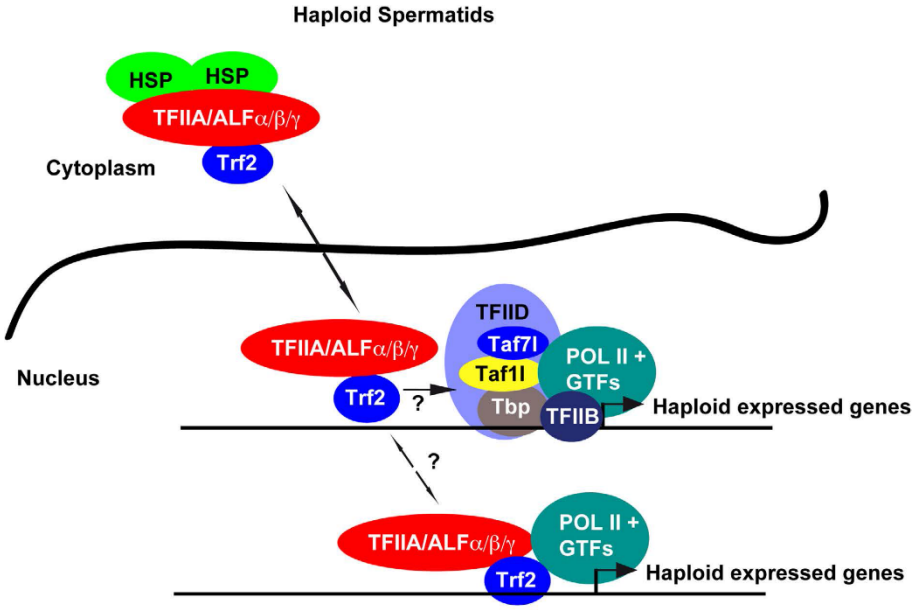


Figure 3-5: TRF2-containing PIC in spermatids. In spermatid, TRF2 is proposed to act as a testis-specific subunit of TFIIA/ALF, cooperating with TBP and TAF7L to regulate haploidcell gene expression. From (Martianov et al., 2016).

3.2.2.2 TRF2 functions in *Xenopus*

In *Xenopus*, TRF2 is highly expressed in ovary, testis and embryos, and knock down of *Trf2* blocks embryos development past the mid-blastula stage, showing that TRF2 is essential for early embryogenesis and contributes to transcription *in vivo* (Veenstra et al., 2000; Xiao et al., 2006). *Trf2* knockdown combined with transcriptome profiling in blastula stage embryos showed that a large number of transcripts require TRF2, and among these genes, a significant proportion is preferentially expressed in embryos (Jacobi et al., 2007). Moreover, TRF2 shows

a functional specialization in catabolism (Jacobi et al., 2007).

3.2.2.3 TRF2 functions in zebrafish

In zebrafish, function of TRF2 has been investigated by inhibiting TRF2 through the expression of a dominant negative *Trf2* mutant (Muller et al., 2001). TRF2-blocked embryos develop normally until the mid-blastula stage, however, they fail to initiate epiboly or arrested before dome stage and eventually fail to gastrulate (Muller et al., 2001). Interestingly, this phenotype is similar to inhibition of Pol II transcription by injecting α-amanitin, indeed, block of TRF2 abolishes the expression of many zygotic regulatory genes (Muller et al., 2001).

3.2.2.4 Versatile TRF2 in *Drosophila*

In *Drosophila*, two TRF2 protein isoforms of 632aa and 1715aa exist, and both isoforms are expressed in embryos and different tissues of adult flies (Kopytova et al., 2006). Both isoforms can interact with ISWI (Hochheimer et al., 2002; Kopytova et al., 2006). Study of *Trf2* mutation showed that TRF2 has essential functions during embryonic Drosophila development, and that the *Trf2* function is essential for differentiation of both male and female germ cells (Kopytova et al., 2006). Intensive studies has shown TRF2 is involved in many different transcriptional programs.

First of all, TRF2 associates with DNA replication-related element (DRE)-binding factor (DREF) and activates the transcription of target genes involved in DNA replication and cell proliferation through the binding of DREF to DRE motifs in promoters.

Secondly, TRF2 is required for transcription from TCT-dependent core promoters (Wang et al., 2014; Isogai et al., 2007a). Earlier ChIP-on-chip and RT-PCR analysis showed that TRF2 is required for expression of a cluster of ribosomal protein genes (Isogai et al., 2007a). As already described before in section ***2.3.1.1***, the TCT core promoter motif is present in most ribosomal protein (RP) genes in *Drosophila* and human. Later, i*n vitro* experiments showed that purified TRF2 activates TCT-containing promoters and that in *Drosophila* S2 cells, TCT-dependent transcription shows increase or decrease upon overexpression or depletion of TRF2, respectively (Wang et al., 2014). Consistently, ChIP-seq experiments showed that TRF2 is enriched at TCT-dependent promoters *in vivo* (Wang et al., 2014). Altogether, these data indicates a specialized TRF2-based Pol II transcription machinery driving the expression of RP genes.

Thirdly, TRF2 is required for transcription from DPE-dependent core promoters (Hsu et al., 2008; Kedmi et al., 2014). It has been shown that RNAi depletion of TBP decreases TATA-dependent but not DPE-dependent transcription, and that TBP overexpression increases TATA-dependent transcription but decreases DPE-dependent transcription (Hsu et

al., 2008). Upon ecdysone induction, TBP occupancy does not increase at DPE-containing promoter while Pol II occupancy does. These observation suggest that DPE-dependent transcription may occur in a TBP-independent manner (Hsu et al., 2008). Later on, it was found that TRF2 is enriched in protein fractions supporting DPE transcription, and that short TRF2 isoform preferentially activates DPE-containing promoters (Kedmi et al., 2014). Moreover, RNAi depletion of TRF2 reduces DPE-dependent but not TATA-dependent transcription, and genes induced by TRF2 overexpression are enriched for Inr and DPE motifs (Kedmi et al., 2014). *In vitro* affinity analysis also showed the enrichment of Inr and DPE motifs in DNA oligos bound by TRF2-containing complexes (Kedmi et al., 2014). Taken together, these findings suggest that TRF2 is functionally important for DPE-dependent transcription.

Fourthly, TRF2 selectively regulates the TATA-less Histone H1 gene promoter (Isogai et al., 2007a), apparently, by a DRE, TCT and DPE motifs independent mechanism (Vo Ngoc et al., 2017b).

Lastly, TRF2 is involved in piRNA cluster transcription (Andersen et al., 2017) (***Figure 3-6***). In *Drosophila*, TFIIA-L (homolog of human TFIIAαβ precursor) has a ovary-specific paralog Moonshiner, which lacks the TBP interaction domain. Moonshiner interacts with TRF2 to form an alternative TFIIA-TRF2 complex at bidirectional piRNA clusters through its interaction with Deadlock, a binding partner of HP1 variant Rhino, thus facilitating the transcription of piRNA (Andersen et al., 2017).

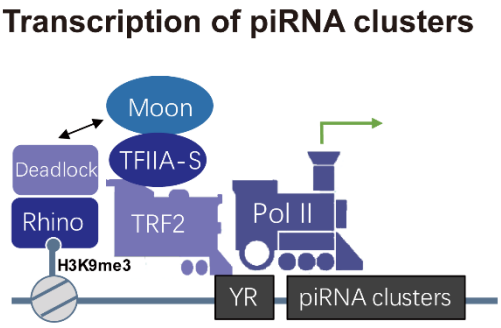


Figure 3-6: TRF2-dependent piRNA cluster transcription.

3.2.2.5 TRF2 is in *C. elegans*

In *C. elegans*, TRF2 is required for zygotic transcription during embryogenesis, and RNAi depletion of TRF2 results in embryonic lethality (Dantonel et al., 2000; Kaltenbach et al., 2000). TRF2-deficient embryo arrest as clusters of 80-350 undifferentiated cells before gastrulation. Analysis of gene expression revealed that TRF2 is required to for the expression of differentiation makers and also the establishment of bulk transcription during early embryogenesis (Dantonel et al., 2000; Kaltenbach et al., 2000).

3.2.3 TBP2, the vertebrate-specific TBP-related factor

TBP2 (also known as TBPL2 and TRF3) is the last identified TBP-related factor. It shares more than 90% identity with the TBP core domain and is able to bind the TATA-box, TFIIA and TFIIB (Persengiev et al., 2003; Bartfai et al., 2004; Jallow et al., 2004). Although early studies suggested a widespread expression of *Tbp2* (Persengiev et al., 2003; Yang et al., 2006), it has become apparent that in mice, TBP2 is exclusively expressed in the oocytes (Bartfai et al., 2004; Xiao et al., 2006; Gazdag et al., 2007). In *Xenopus* and *z*ebrafish, expression of *Tbp2* is broader, but is also highly enriched in the ovaries (Bartfai et al., 2004; Jallow et al., 2004; Xiao et al., 2006; Akhtar et al., 2009), suggesting a fundamental role for TBP2 in the vertebrate ovary.

In this section, I will mainly describe the function of TBP2 in *Xenopus* and *z*ebrafish, and the role of TBP2 during mouse oogenesis will be described in chapter ***4***.

3.2.3.1 TBP2 function in *Xenopus*

Knockdown studies in *Xenopus* embryos showed that TBP2 is required for embryonic transcription and gastrulation (Jallow et al., 2004). In embryos, TBP2 overexpression can partially rescue TBP knockdown and restores the transcription of many TBP-dependent genes, suggesting that TBP2 may function as a substitute for TBP (Jallow et al., 2004). Transcriptome analysis showed that TBP2 is linked to vertebrate-specific embryonic genes and involved in ventral specification (Jacobi et al., 2007).

In *Xenopus* oocytes, TBP2 but not TBP is present at the protein level. After meiotic maturation, TBP2 is actively degraded following global repression of transcription, and only residual levels of TBP2 remain in the eggs and in the early embryos. After fertilization, maternal TBP mRNA is translated and TBP starts to accumulate during cleavage stages of development. Both TBP and residual TBP2 contribute to zygotic transcription (Akhtar et al., 2009; Muller et al., 2009; Akhtar et al., 2011) (***Figure 3-7***). Interestingly, TBP2 is recruited to transcriptionally active loops of ‘lampbrush’ chromosomes in oocytes (Akhtar et al., 2009). TBP2 can promote transcription from TATA-containing promoters, and this function can be replaced by TBP when ectopically expressed in oocytes. Moreover, analysis of TBP2 occupancy in oocytes revealed that TBP2 is associated with active Pol II promoters. However, TBP can also promote transcription for these promoters when ectopically expressed in oocytes. Furthermore, TBP2 is also recruited to 5S rRNA Pol III promoter, suggesting that TBP2 probably can also participate in Pol III transcription in *Xenopus* (Akhtar et al., 2009).

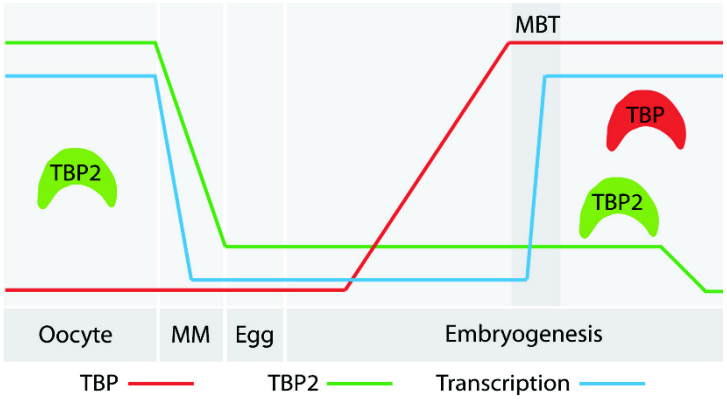


Figure 3-7: Switching of TATA-binding proteins between oocytes and embryos. Adapted from (Akhtar et al., 2011)

Interestingly, it has been reported that after nuclear transfer in *Xenopus* oocytes, somatic nuclei progressively loses somatic TBP and accumulates TBP2, indicating there is an exchange in basal transcription machinery during nuclear reprograming (Jullien et al., 2014). It has been proposed that this basal transcription machinery exchange mediates reprogramming by the *Xenopus* oocyte (Jullien et al., 2014).

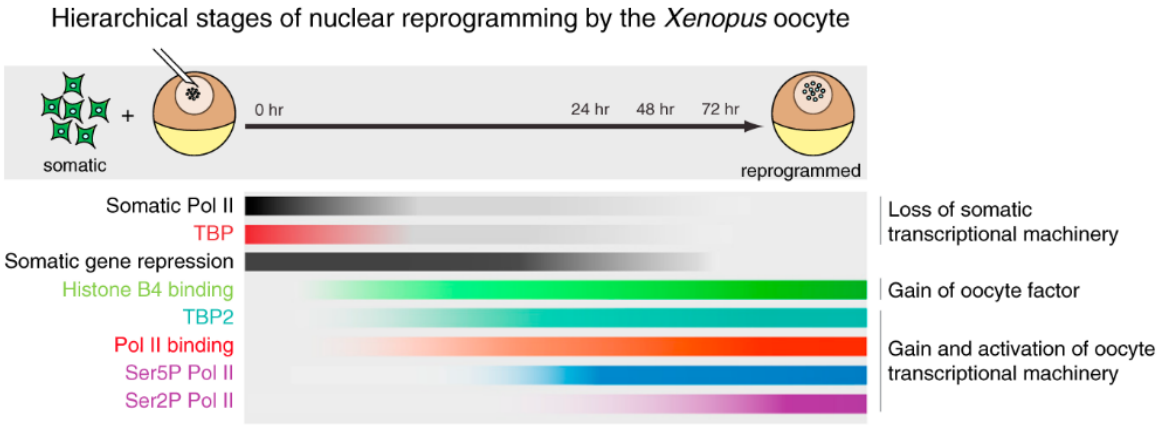


Figure 3-8: Basal transcription machinery exchange following nuclear transfer to *Xenopus* oocyte. From (Jullien et al., 2014).

Altogether, these findings indicates that TBP2 is more than a substitute for TBP. In *Xenopus*, the TBP2-containing basal transcription machinery has specialized functions in oocytes, early embryogenesis and even nuclear reprograming by *Xenopus* oocytes.

3.2.3.2 TBP2 function in zebrafish

In zebrafish, knockdown studies showed TBP2 is required for embryonic development and expression of some differentiation maker genes (Bartfai et al., 2004). A later study reported that TBP2-depleted zebrafish embryos exhibit multiple developmental defects, in particular, fail to undergo haematopoiesis (Hart et al., 2007). Expression profiling for

TBP2-dependent genes revealed that *mespa* [murine orthologues (*Mesp1* and *Mesp2*) of *mespa* are required for mesoderm specification] is the target gene required for embryonic development (Hart et al., 2007). *mespa* further targets *cdx4*, a caudal-related gene required for haematopoiesis. Thus it was proposed that TBP2 initiated the transcription factor pathway for commitment of mesoderm to haematopoietic lineage (Hart et al., 2007).

Study of TSS usage during zebrafish early embryonic development by cap analysis of gene expression (CAGE) revealed that there is a widespread switch in TSS usage throughout maternal to zygotic transition (Haberle et al., 2014). Moreover, maternal TSS selection is associated with the presence of an A/T-rich motif (referred as W-box). The similarity of the W-box with the TATA-box suggests that transcription initiation in the oocyte may be mediated by the oocyte-enriched TBP-related factor TBP2 (Haberle et al., 2014).

3.2.3.3 Controversial function of TBP2 in myogenesis

It has been reported that, during skeletal muscle differentiation, canonical TFIID is replaced by a TBP2-TAF3 complex for the activation of muscle genes (Deato et al., 2007; Deato et al., 2008). However, further studies revealed that *Tbp2* is not expressed in muscles, and *Tbp2* null mice do not display any skeletal muscle phenotype (Gazdag et al., 2009). Moreover, it has been shown that TBP, but not TBP2, is the essential component of the PIC that promotes muscle gene expression in differentiated skeletal muscles (Malecova et al., 2016).

3.3 TFIIA and its paralog ALF

As mentioned in section ***2.1.3***, metazoan TFIIA contains three subunits, TFIIAα, TFIIAβ and TFIIAγ. TFIIAα and TFIIAβ are produced by Taspase1 cleavage of the TFIIAαβ precursor, and TFIIAαβ has a cell type-specific paralogue, called TFIIA-like factor (ALF).

3.3.1 TFIIAαβ, cleave or not make a difference

Initially, it was thought that TFIIA(α+β), the cleaved form, was the functional form. Later, it was reported that TFIIAαβ, the uncleaved form, together with TFIIAγ, interact stably with TBP and form a functional TBP-TFIIA(αβ+γ)-containing complex (TAC) in P19 embryonal carcinoma cells (Mitsiou et al., 2000, 2003). *Taspase1*-/- MEF cells with only uncleaved TFIIA are viable, although with cell cycle defects, indicating that the uncleaved TFIIA is functional (Zhou et al., 2006). In addition, uncleavable TFIIA mutant can rescue TFIIA knock-down in Xenopus (Zhou et al., 2006). Uncleavable TFIIA is more stable, as processed TFIIA can be more efficiently degraded by the ubiquitin-proteasome pathway (Hoiby et al., 2004).

Moreover, *Taspase1* knockout mice are viable, although the majority die after birth and the survivors are smaller (Takeda et al., 2006). Surprisingly, analysis of the surviving *Taspase1* null mice shows that males are infertile, while *Taspase1-/-* females are fertile. Further analysis with noncleavable *TFIIAαβ* mutant mice revealed that both *Taspase1-/-* and noncleavable *TFIIAαβ* testes release immature germ cells with impaired transcription of *Transition proteins* (*Tnp*) and *Protamines* (*Prm*), exhibiting chromatin compaction defects, and recapitulating the observations in *Trf2-/-* testes (Oyama et al., 2013). Interestingly, although the noncleavable TFIIA(αβ+γ) can still interact with TRF2, they are not able to target and activate *Tnp1* and *Prm1* promoters, indicating Taspase1-mediated cleavage of TFIIAαβ is essential for TFR2-dependent testis-specific transcription (Oyama et al., 2013) (***Figure 3-9***).

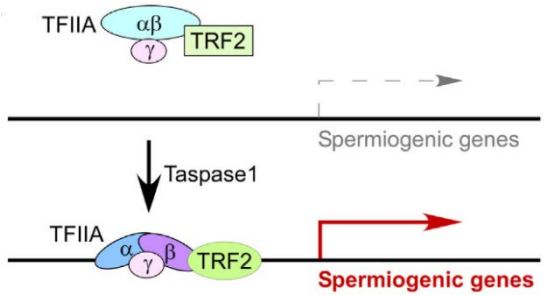


Figure 3-9: Taspase1-TFIIA-TRF2 axis regulates spermatogenesis. From (Oyama et al., 2013).

3.3.2 ALF

ALF is a paralogue of TFIIAαβ precursor (Upadhyaya et al., 1999; Ozer et al., 2000). Like TFIIA, the majority of ALF is cleaved into α and β subunits, which can form TFIIA with TFIIAγ. ALF is only found in the gonads (Xiao et al., 2006). In testis, it is expressed in late pachytene spermatocytes and round spermatids (Catena et al., 2005), and in ovary, it is expressed in the oocytes (Xiao et al., 2006). Immunoprecipitation with testis extracts revealed that full combination of hybrid TFIIA complexes exist (Catena et al., 2005) (***Figure 3-10***). It is known ALF interacts with TRF2 (Catena et al., 2005; Martianov et al., 2016), however, its function is not clear yet.

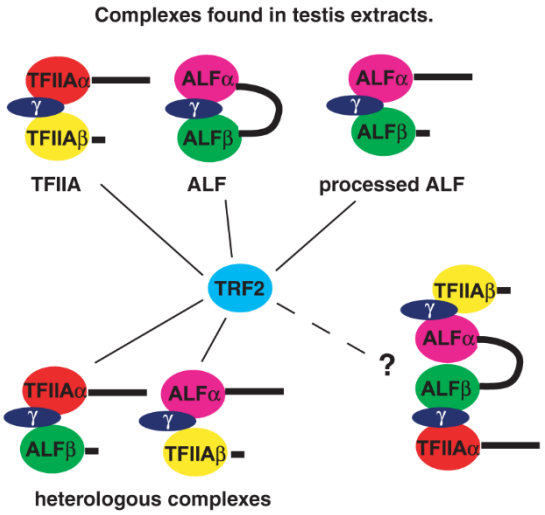


Figure 3-10: Summary of TFIIA complexes exist in testis extracts. From (Catena et al., 2005).

Altogether, these findings show that, diversified sets of basal transcription machinery orchestrate cell specific transcription programs through the variation of different TAFs, TAF paralogs, TBP-related factors and TFIIA/ALF.

4. TBP2 and transcription regulation during oocyte growth

Oogenesis is a complex process with dynamic changes in gene expression that are regulated by a vast number of well-coordinated transcription factors (reviewed in (Jagarlamudi et al., 2012; Sanchez et al., 2012)) (***Figure 4-1***). Following their colonization by PGCs, the female gonad continues its differentiation mostly under the influence of somatic cell expressed transcription factors LHX9, FOXL2, WT1, SF1 and GATA4 (reviewed in (Sanchez et al., 2012)). In the embryonic ovary, the oogonia form into germ cell cysts, which later break apart as individual oocyte surrounded by several pre-granulosa cells, forming primordial follicles after birth (Pepling, 2006). Transcription factors such as FIGLA (Soyal et al., 2000), NOBOX (Rajkovic et al., 2004), FOXO3 (Castrillon et al., 2003; Hosaka et al., 2004), SOHLH1 (Pangas et al., 2006), SOHLH2 (Choi et al., 2008b) and LHX8 (Pangas et al., 2006; Choi et al., 2008a) that are expressed in the oocytes, as well as transcription factors such as ZGLP1 (Li et al., 2007; Strauss et al., 2011) and FOXL2 (Schmidt et al., 2004; Uda et al., 2004), that are expressed in the soma, have been revealed as crucial transcriptional regulators for follicle formation and early folliculogenesis. Further progression of follicle development requires YY1 (Griffith et al., 2011), TAF4b (Freiman et al., 2001) and TBP2 (Gazdag et al., 2009). Here I mainly focus on the role of TBP2 in transcription regulation during mouse oocyte growth.

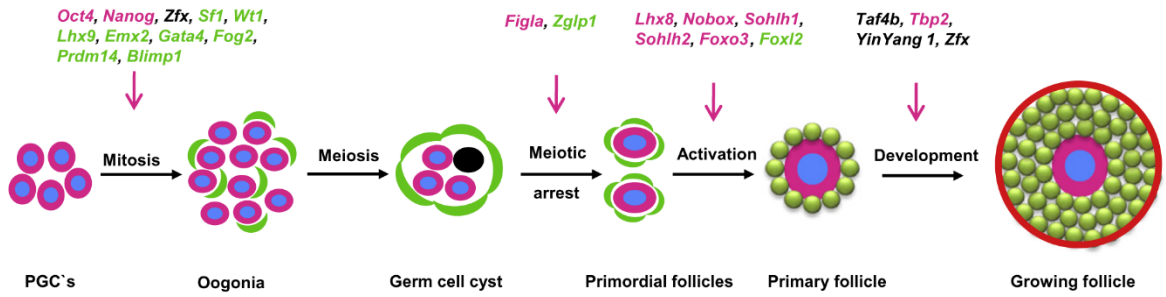


Figure 4-1: Crucial transcription factors during mouse oogenesis. Transcription factors act at different stages of folliculogenesis as revealed by mouse knockouts. TFs expressed in germ cells, granulosa cells, or both germ cell and granulosa cells are marked in pink, green and black, respectively. From (Jagarlamudi et al., 2012).

In the mouse, TBP2 is exclusively expressed in the oocytes (Xiao et al., 2006; Gazdag et al., 2007). During folliculogenesis, expression of TBP protein is decreasing, becoming undetectable during oocyte growth and reappearing only after fertilization. To the contrary, TBP2 is highly expressed in growing oocytes, declining to very low levels by the preovulatory follicle stage, only some traces of TBP2 persist up to the 2 cell-stage (Gazdag et al., 2007; Muller et al., 2009) (***Figure 4-2***). Interestingly, the global transcriptional activity in oocyte is

increased during oocyte growth, reaches its peak at early antral follicles, subsequently declines and becomes quiescent by preovulatory follicle stage (De La Fuente, 2006) (***Figure 4-2***).

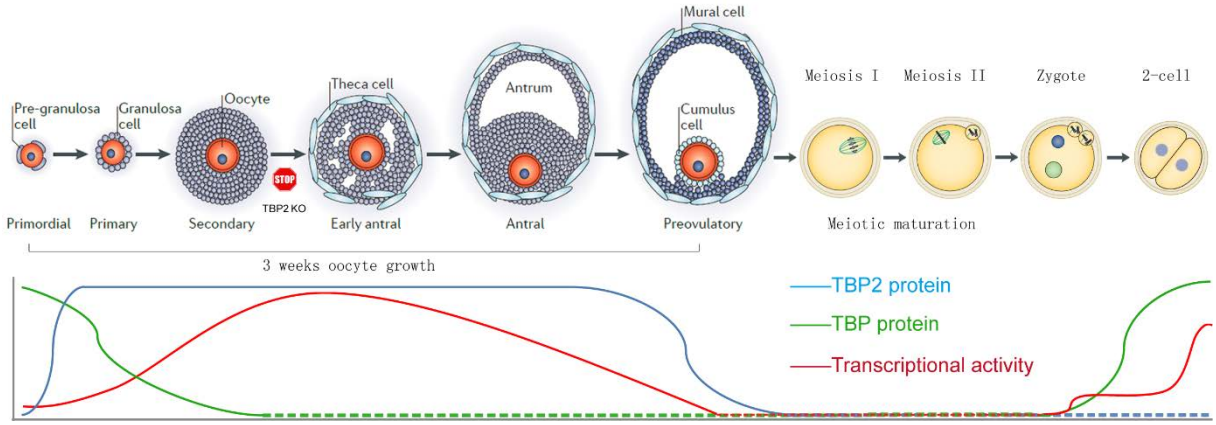


Figure 4-2: Scheme of TBP-TBP2 switch and global transcription activity during oogenesis and early embryogenesis. Adapted from (De La Fuente, 2006; Gazdag et al., 2007; Li et al., 2013b; Clift et al., 2013).

Moreover, it has been show that *Tbp2*-defecient female mice are infertile due to a defect in folliculogenesis: ovaries from *Tbp2*$^{-/-}$ mice have an increased number of primordial and primary follicles, but diminished number of secondary follicles compared to wild-type (Gazdag et al., 2009). TBP2 depletion impairs Pol II activity and chromatin structure in oocytes, and results in the downregulation of many oocyte-specific genes are downregulated, including *Gdf9*, *Bmp15* and *Zp3* (Gazdag et al., 2009). TBP2 binds to the promoters of several active genes in oocytes that have been tested (Gazdag et al., 2009). In addition, TBP2 is not able to fully compensate the depletion of TBP in MEFs (Gazdag, 2008), and TBP2 overexpression in zygotes altered pre-implantation embryo development (Gazdag et al., 2009).

Altogether, these observations indicate that there is a switch between TBP and TBP2-mediated transcription during oocyte growth and that TBP2 plays important role in oocyte transcription regulation.

Aims of the project

I discussed in the introduction that diversified basal transcription machineries can direct cell type-specific transcription. In agreement with this concept, it has been shown that TBP is replaced by TBP2 during oocyte growth. Moreover, *Tbp2*$^{-/-}$ females are sterile due to defective folliculogenesis, altered chromatin organization and misregulation of key oocyte-specific genes. These data indicate that TBP2 is essential for oocyte development, and that it plays a critical and specialized role in the regulation of oocyte-specific gene expression program.

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

(a) To determine the role of TBP2 is controlling transcription initiation during oogenesis;
(b) To map TBP2 occupancy in oocytes and to decipher the exact role of TBP2 in regulating oocyte gene expression program;
(c) To characterize the oocyte specific basal transcription machinery, in particular, the TBP2-containing complex;
(d) To understand what is the biological significance of the TBP replacement by TBP2 during oocyte growth

Material & Methods

1. Mouse lines

Tbp2 knockout mouse line (Gazdag et al., 2009), *Zp3-Cre*$^{Tg/Tg}$ mouse line (Lewandoski et al., 1997), *Taf*10$^{f/f}$ mouse line (Mohan et al., 2003), Taf7$^{f/f}$ mouse line (Gegonne et al., 2012), *Uprt* transgenic mouse line (Gay et al., 2013), *Rosa26*^{Cre} and *Rosa26*^{Flp} (Birling et al., 2012) have already been described. *Tbp-Tbp2* swap knock-in mouse line was generated at the Mouse Clinical Institute/Institut Clinique de la Souris.

Oocyte-specific *Taf7* and *Taf10* depletion was achieved by crossing Taf7$^{f/f}$ and *Taf*10$^{f/f}$ mice with *Zp3-Cre* mice, respectively. Oocyte-specific *Uprt* mice were obtained by crossing *Uprt* transgenic mice with *Zp3-Cre* mice.

Superovulation: 5U PMS was injected to 4 week old female mice between 2-4 pm. After 44-46 hours, GV oocytes were obtained from the ovaries by puncturing with needles.

4TU delivery for oocyte-specific *Uprt* mice (Gay et al., 2014): 4TU (Aldrich, #440736) was dissolved in DMSO at concentration of 210mg/ml and stored at −80°C. Before injection, stock 4TU was diluted to 30mg/ml with corn oil, and 4TU (430mg/kg body weight) was delivered to P14 pups by intraperitoneal injection. 6 hours after 4TU injection, oocytes were collected (see ***8.1***) for SLAM-seq (Herzog et al., 2017).

Animal experimentations were carried out according to animal welfare regulations and guidelines of the French Ministry of Agriculture and procedures were approved by the French Ministry for Higher Education and Research ethical committee C2EA-17 (project n° 2018031209153651)

2. Cell culture

TBP2 stable overexpression cell line (3T3-II10) and mock cell line (3T3-K2) has already been described (Gazdag et al., 2007). NIH 3T3 cells were routinely cultured in DMEM (4.5g/l glucose) with 10% newborn calf serum (P122207N) at 37 °C and 5 % CO2.

3. Whole cell extracts

3.1 Whole cell extracts of II10 and K2 cells

Cells cultured in 15cm dish were washed twice with 1xPBS, subsequently harvested by scrapping on ice. Harvested cells were centrifuged 2000rpm at 4°C for 5min and then

resuspended in 1 packed cell volume of whole cell extraction buffer consisting of 20mM TrisCl pH7.5, 2mM DTT, 20% Glycerol, 400mM KCl and 1x Protease inhibitor cocktails (PIC, Roche). Cell lysates were froze in liquid nitrogen and thawed on ice for 3 times, followed by centrifuge at 13000rpm, 4°C for 15 min. The supernatant was collected and protein concentration was measured by Bradford protein assay (see ***3.3***). The cell extract was used directly for immunoprecipitations (see ***5.1***), or western blot (see ***4***), or stored at −80°C.

3.2 Whole cell extracts from ovaries

P13.5 ovaries collected from CD1 mice were homogenized in whole cell extraction buffer consisting of 20mM TrisCl pH7.5, 2mM DTT, 20% Glycerol, 400mM KCl and 5x Protease inhibitor cocktails. Cell lysates were froze in liquid nitrogen and thawed on ice for 3 times, followed by centrifuge at 14000rpm, 4°C for 15 min. The supernatant extracts were used directly for immunoprecipitations (see ***5.2***).

3.3 Bradford protein assay

1ul of a whole cell extract was mixed with the 1ml 1/5th diluted dye reagent (Bio-rad protein assay, Bio-Rad), and the absorbance at 595 nm was measured. The concentration of whole cell extracts was calculated from a standard curve determined by six standard BSA samples with different concentration.

4. Western blot

Protein samples (15ug-25ug cell extracts or 15ul of IP elution) were mixed with 1/4th volume of loading blue (100 mM Tris pH 6.8, 30% glycerol, 4% SDS, 0.2% bromophenol blue and fresh added 100 mM DTT) and boiled for 10 min. Samples were then resolved on 10 % SDS-PAGE and transferred to nitrocellulose membrane (Protran, Amersham). Membranes were blocked in 3% non-fat milk in 1xPBS at room temperature (RT) for 30 min, and subsequently incubated with the primary antibody (see ***6.3***) overnight at 4°C. Membranes were washed three times (10 min each) with 1xPBS plus 0.05% Tween20. Membranes were then incubated with HRP-coupled secondary antibodies for 1h at RT, followed by ECL detection (Thermo Fisher). The signal was acquired with the Chemidoc imaging system (Bio-Rad).

5. Immunoprecipitations

5.1 IP with sepharose-beads

a) IP buffers

IP100mM KCl buffer	100ml	stock
25 mM Tris Cl pH 7,9	2.5ml	1M
0.1% NP40	1ml	10%
5mM $MgCl_2$	0.5ml	1M
~10% Glycerol	10ml	87%
100mM KCl	3.33ml	3M
2mM DTT	200ul	1M
1X PIC	2 tablets	
Water	82.7ml	

IP500mM KCl buffer	100ml	Stock
25 mM Tris Cl pH 7.9	2.5ml	1M
0.1% NP40	1ml	10%
5mM $MgCl_2$	0.5ml	1M
~10% Glycerol	10ml	87%
500 mM KCl	16.66ml	3M
2mM DTT	200ul	1M
1X PIC	2 tablets	
Water	69.4ml	

b) Washing/Preparation of beads

Resuspend sepharose-beads by vorexing for 10-15 seconds, transfer 150ul 50% sepharose beads-Pro A/G slurryto a 1.5 EP tube (For each IP from 4mg WCE, 100ul 50% beads slurry + 10ul antibody was used, also take 1/10 volume of beads to pre-clean WCE). 3 times quick wash with 1ml 1×PBS and 2 times quick wash with 1ml IP 100mM KCL buffer (centrifuge 2min at 600g/2500rpm after each wash). Remove the SN and add IP 100mM KCL buffer to the beads to make it as 50% slurry, put on ice.

c) Antibody coupling to Dynabeads-Pro A/G

Aliquot 100ul IP 100mM KCL buffer (1 beads volume) into 1.5ml tubes, add 100ul pre-washed beads-Pro G. Add 10ul/20-100ug antibodies to each tube, place at 10-30 rpm on rotator for at least 2h at 4℃ (approximately 1mg antibody per ml of protein G sepharose beads). 3 times wash with 500ul IP 500mM KCL buffer and 2 times wash with 500ul IP 100mM KCL buffer (5 min wash for each). Remove the supernatant, keep the beads and put it on ice.

d) Pre-clean of INPUT

4mg WCE + 1/10 volume sepharose beads, pre-clean at 4℃ 2h, then centrifuge 5min at 600g/2500rpm at 4℃, keep the SN as INPUT sample.

e) Protein fixation/Binding IP

Add the INPUT sample to the beads-antibody complex, incubate overnight at 4℃. Centrifuge 3min at 600g/2500rpm at 4℃, keep the supernatant as SN for WB. Wash the beads

with 1ml IP 500mM buffer three times on rotator, 5min each. Wash the beads with 1ml IP 100mM buffer twice on rotator, 5min each. Keep 10ul of Beads as PBE (protein before elution, optional), and before add Glycine, remove all the IP buffer.

f) Elution

Add 100ul IP Glycine 0.1M pH=2.8, incubate 5 min at RT with gentle agitation, then centrifuge 3min at 600g/2500rpm at 4℃, take the SN to a new tube with 6.7ul Tris (Tris-base, 1.5M, pH8.8, adding before) to neutralize it quickly. Wash the beads with 20ul IP 100mM buffer, centrifuge 3min at 600g/2500rpm at 4℃, take the SN and mix with elution, add 5ul 100×PIC. Centrifuge the elution at 13000prm for 10min at 4℃, then pipet the SN to a new tubes as ELUTION, and keep the beads at 4℃ for until finished the IP-WB. Do the western immediately or freeze it in liquid N2 and stored in -80℃.

5.2 Ovaries WCE IP with Dynabeads

a) Washing/Preparation of Dynabeads

Resuspend Dynabeads Protein A (100.01D) by vorexing for 10-15 seconds, and transfer 200ul beads to a 1.5 EP tube (For ovaries WCE IP, using 150ul beads for IP and 50ul beads for pre-clean, and make sure the stock bead suspension is homogenous before pipetting). Put on the magnet for 20-30s and remove the supernatant, then remove the tube from the magnet and quick wash the beads three times with 1ml IP 100mM KCL buffer, and remove the supernatant. Add 200ul IP 100mM KCL buffer to the beads, resuspend, put on ice.

b) Antibody coupling to Dynabeads-Pro A

Place 200ul PCR tubes on ice and add 150ul pre-washed beads-Pro A, and add 10ul TBP2 (3024) antibody, place at RT on shaker for 30min, then 10-20 rpm on rotator for 2h at 4℃. Place the tubes on magnet and remove the supernatant, add 150ul IP 500mM KCL buffer, transfer to a 500ul tube, wash the 150 tubes twice with 150ul IP 500mM KCL buffer make sure all the beads are transferred, followed by three times wash with 500ul IP 500mM KCL buffer (5min wash on rotator), remove the supernatant, and wash twice with 500ul IP 100mM KCL buffer (5min wash on rotator). Remove the supernatant, and leave beads on ice.

c) Pre-clean of INPUT

Ovary WCE sample was incubated with 50ul dynabeads @ cold room for 1h, then centrifuge at 14000rpm for 5min at 4℃, keep the supernatant as INPUT sample.

d) Protein fixation/Binding IP

Add the INPUT sample to the beads-antibody complex, and incubate at 4℃ overnight. Put on the magnet, keep the supernatant as SN for WB. Wash the beads with 500ul IP 500mM

buffer three times on rotator (5min each), followed by two washes with 500ul IP 100mM buffer. Before add Glycine, remove all the IP buffer.

e) Elution

Add 80ul IP Glycine 0.1M pH=2.8, vortex to mix, incubate 5 min at RT with gentle agitation, vortex, incubate 5min again, then put on the magnet, take the SN to a 200ul tube with 5.4ul Tris (Tris-base, 1.5M, pH8.8, adding before) to neutralize it quickly. Add 20ul IP Glycine 0.1M pH=2.8 to the beads, mix, incubate 5 min at RT with gentle agitation, take the supernatant to the tube and add another 1.33ul Tris. Mix the elution, and put it on the magnet, transfer the supernatant to a low bind 1.5 EP tube, add 2ul 100xPIC, as ELUTION, totally around 100ul. Take 15ul to do the western blot immediately, the left 85ul was used for mass-spectrometry analysis, which was described in the manuscript.

6. Antibodies

6.1 Rabbit polyclonal antibody generation from protein

Antibody generation is performed with the help of IGBMC animal facility. Two months old female rabbit (White New-Zealand strain, 2-2,5 kg) were used for immunization. Rabbit is anesthetized by intramuscular injection of Xylazine/Ketamine solution, and blood sample collected to produce the pre-immune serum as negative control. 1 ml of antigen solution (containing ~200 ug of recombinant protein or ~300 µg peptide) is mixed with 1 ml of "Complete Freund Adjuvant" and was injected intradermally to rabbit on 40 to 60 different locations for immunization. One month after, blood sample is collected every week for one month to obtain antibody-containing antiserums. Another injection of antigen (containing ~100ug recombinant protein or ~200 µg peptide) were performed and twelve days later, rabbit is anesthetized and then sacrificed for the boosted antisera. The antiserums were tested and purified to obtain polyclonal antibodies.

6.2 Antibody purification

Affinity purification of rabbit polyclonal antibodies was performed with SulfoLink™ Coupling Resin from Thermo Fisher Scientific according to the user manual.

6.3 List of antibodies

Antibody	Name	Type	Application	Position	Reference
Anti-TBP2	**2B12**	Mouse monoclonal	**WB**, IF, IP, ChIP	N-termial	(Gazdag et al., 2007)
Anti-TBP2	2481	Rabbit polyclonal	IP	84-103 of	New antibody

Anti-TBP2	2482	Rabbit polyclonal	IP	TBP2	New antibody
Anti-TBP2	3017	Rabbit polyclonal	WB	N-termial 17-30 of TBP2	New antibody
Anti-TBP2	3018	Rabbit polyclonal	none		New antibody
Anti-TBP2	3019	Rabbit polyclonal	none		New antibody
Anti-TBP2	3022	Rabbit polyclonal	IP	N-termial 111-129 of TBP2	New antibody
Anti-TBP2	**3023**	Rabbit polyclonal	IP, IF, **ChIP**		New antibody
Anti-TBP2	**3024**	Rabbit polyclonal	**IP**		New antibody
Anti-TBP2	3499	Rabbit polyclonal	WB, IP, ChIP	1-140	New antibody
Anti-TAF5	1TA1C2	Mouse monoclonal	WB	(Jacq et al., 1994)	
Anti-TAF6	2G7	Mouse monoclonal	WB	(Bell et al., 2001)	
Anti-TAF7	3435	Rabbit polyclonal	IP	(Bardot et al., 2017)	
Anti-TAF10	2B11	Mouse monoclonal	WB, IP	(Mohan et al., 2003)	
Anti-TBP	3G3	Mouse monoclonal	WB, IP	(Lescure et al., 1994)	
Anti-TBP		Rabbit polyclonal	ChIP	Abcam, ab28175	
Anti-Pol II	7G5	Mouse monoclonal	ChIP	(Acker et al., 1997)	
anti-Rabbit / anti-Mouse IgG Peroxydase conjugate		Goat polyclonal	WB	Jackson ImmunoResearch 111-035- 144 / 111-036- 071	
Anti-TIP60		Rabbit polyclonal	ChIP	(Frank et al., 2003)	
Anti-H3K4me3		Monoclonal	ChIP	Millipore, Cat. # 04-745	
Anti-Lhx8		Rabbit polyclonal	Staining	Abcam, ab41519	
Anti-SCP3		Mouse monoclonal	Staining	Abcam, ab97672	
Anti-Flag M2		Mouse monoclonal	WB	Sigma, Cat. # F1804	

7. Gel filtration

A Superose 6 (10/300) column was equilibrated with buffer consisting of 25mM Tris Cl pH 7.9, 5mM MgCl2, 150mM KCl, 5% Glycerol, 1mM DTT and 1xPIC. 500 ul of whole cell extracts containing ~5 mg of protein were injected in an Akta Avant chromatography device and run at 0.4 ml per min. Protein detection was performed with absorbance at 280nm and 260nm. 500 ul fractions were collected and analyzed by WB and IP.

8. RNA preparation from P7 and P14 oocytes

8.1 P7 and P14 growing oocytes collection

The ovaries is rapidly dissected and freed from adhering tissues in PBS. For each 6 ovaries, add 500µl digest mix consisting of 429 µl PBS, 33.4 µl Collagenase (30 mg/ml stock in PBS, final concentration 2mg/ml, ref. C2674-100MG), 12.5 µl Trypsin (1% stock in PBS, final concentration 0.025%, ref. 93615-5G) and 25ul hyaluronidase (type IV-S, 10mg/ml stock in H2O, final concentration 0.5mg/ml, ref.H3884). Incubate at 37 °C in thermomixer at 600 rpm for ~20 mins, pipette samples up and down gently every 5mins with a P1000. Once suspension is uniformly digested (~20 mins), stop digestion by adding 1ml of 37°C pre-warmed αMEM (with 5% FBS) to quench the digestion. Transfer the 1.5ml solution to a 35mm dish, quickly pick the oocytes with good size with mouth-pipette.

8.2 RNA preparation and sequencing

Oocytes collected above were washed through several M2 drops, and total RNA was isolated using NucleoSpin RNAXS kit from Macherey-Nagel according to the user manual. RNA quality and quantity were evaluated by Bioanalyzer. RNA-sequencing and data analyses (including RNA-seq analyses, repeat element analyses and core promoter motif analyses) were described in the manuscript.

9. ChIP-seq

9.1 Chromatin immunoprecipitation

a) Crosslinking and nuclei preparation

When cells reache 80-90% confluence, remove the medium, wash the cells once with room temperature PBS, fix cells by adding 20ml 1xPBS with 1% concentration formaldehyde, incubate for 10min at RT while mixing by rotating agitation. Stop crosslink by adding 2.9ml of

1M glycine to a final concentration of 125mM. Incubate 10 min at RT while mixing by rotating agitation, then trash the solution in the formaldehyde waste. Wash the fixed cell twice with cold 1xPBS. Scrape cells using scrapper in 5ml cold 1XPBS per 15cm-dish and collect in 15ml tubes, then spin at 2000 rpm or 600g for 5min at 4˚C to pellet cells. Wash once again with 5ml 1XPBS/PIC and discard SN. Resuspend cells in 2 ml / large dish (about 1ml per 1.0×107cells, or 6 volumes to the packed pellet) of L1 buffer, and incubate for 10min on a rocker at 4°C. Spin down to pellet nuclei for 5 min at 800g or 2500 rpm at 4°C in 1.5ml Eppendorf centrifuge, discard the supernatant. Resuspend nuclei in L2 buffer at 2.0×10^7cells / ml, incubate for 10min on a rocker at 4°C.

b) Sonication and sonication check

Samples (nuclei) were distributed 600ul per tube into tubes for covaris sonication, sonication conditions should be optimized for each cell type with Covaris E210 sonicator. For NIH3T3 cells, sonication was performed with Duty cycle 20%, Intensity 8 and cycle burst 200 for 15min. After sonication transfer the sonicated chromatin into 1.5ml Eppendorf, and centrifuge for 15min at 14000 rpm, take the supernatant and transfer into new Eppendorf tubes. Measure the concentration of chromatin, aliquot 200ug-500ug per tubes to avoid de-freezing. Save 15ug chromatin to reverse crosslink and check sonication efficiency, freeze the chromatin at Liquid Nitrogen and kept in -80°C. 15ug chromatin were add up to ~300ul with L2 buffer. Add 12.5ul 5M NaCl to 0.2M final and 1.5ul of RNase A (10mg/ml) to final 50ug/ml, incubate at 37°C for 1h, then add 20ug Proteinase K (1ul of 20mg/ml Proteinase K) and heat with 400 rpm shaking at 65°C for 5 hours or overnight in a thermomixer to reverse crosslink. DNA extraction by phenol/chloroform, and resuspend in 40ul TE buffer. Check the quality of the chromatin by running 10ul and 20ul on 1.5% agarose gel, the fragment size should be between 250 to 750 bp.

c) Pre-block of protein A/G-Sepharose beads

Take protein A/G-sepharose 50% beads slurry (100ul per ChIP) into a 2ml tube, and wash twice with 1ml TE (Vortex, spin 2 min at 600g/2500rpm). For 1 ml beads (50% slurry), add 100ul (final 1ug/ul) of denatured tRNA (10mg/ml, denature 5min at 95°C) and 50ul (final 1ug/ul) of BSA (20mg/ml) in 1ml beads slurry solution. Incubate for 3 hours with rotation at 4°C, followed by two washes with 1ml TE. Add 500ul TE final and keep as 50% slurry at 4°C (saturated beads are stable for 1 week in the fridge).

d) Immunoprecipitation

Use 50ug chromatin per ChIP (for ChIP-seq, 50ug is enough; for ChIP-seq, I used 200ug chromatin per ChIP), and dilute chromatin ~10 times with chromatin dilution buffer to decrease

SDS concentration down to 0.1% maximum. Pre-clear chromatin with 30ul blocked beads (50% slurry) for 2h at 4°C with overhead shaking, then spin down the beads at 600g/2500 rpm for 2min and use SN for ChIP Input. Save 1% Input fraction and add it to 300ul ChIP lysis buffer L2 (don't forget to save 1% Input fraction as it will be used for qPCR normalization). Add ~5ug antibody per ChIP (antibody amount is highly depends on the antibodies, normally 3-6ug of antibody per 40ug of chromatin). Incubate overnight at 4°C with overhead shaking. Add 50ul pre-blocked protein A/G-Sepharose beads 50% slurry and incubate 2h at 4°C with overhead shaking. Centrifuge at 600g/2500 rpm for 2 min and discard SN. Wash twice with 1ml low salt washing buffer (150mM NaCl) at 4°C with overhead shaking, 10 mins each. Then wash twice with high salt washing buffer (500mM NaCl), 10 mins each. Optionally, wash twice (5 mins each) with LiCl wash buffer if the background is too high. TE Wash 10 min X 2 times.

e) Elution

Elute twice with 160ul freshly prepared elution buffer (10 minutes overhead shaking at RT, centrifuge at 600g/2500 rpm for 2mins). Pool the elution, centrifuge at 8000rpm for 3 min, take ~300ul SN as final elution.

f) Reversal of crosslink and isolation of DNA

For ChIP-qPCR, add 12.5ul 5M NaCl (0.2M final conc.) to 300ul elution as well as input sample, then add 50ug/ml of RNase A (1.5ul 10mg/ml), and incubate for 30min at 37°C, then add 1ul 20mg/ml Proteinase K, incubate at 65°C O/N with 400 rpm shaking in a thermomixer. For ChIP-seq, add 12.5ul 5M NaCl (0.2M final conc.) to 300ul elution, then add 50ug/ml of RNase A (1.5ul 10mg/ml) and incubate at 65°C O/N with 400 rpm shaking in a thermomixer. Next day add 20ug Proteinase K, 20ul of tris pH7.9 (1M), 10 ul of EDTA (0.5M), incubate at 45°C for 1hr. Extract DNA with Phenol/Chlorofom/Isoamylalcohol(25:24:1), and resuspend DNA in 60ul TE buffer, and it ready for qPCR and sequencing.

g) ChIP buffers

L1 lysis buffer	**50ml**	**Stock**
50mM Tris-HCl pH 8.0	2.5ml	1M
2mM EDTA pH 8.0	200ul	0.5M
0.5% NP40	2.5ml	10%
10% Glycerol	5.75ml	87%
1X PIC	1 tablet	
Water	Up to 50ml	

L2 lysis buffer	**50ml**	**Stock**
50mM Tris-HCl pH 8.0	2.5ml	1M
10mM EDTA	1ml	0.5M
1% SDS	2.5ml	20%
1X PIC	1 tablet	
Water	Up to 50ml	

Low salt wash buffer	50ml	Stock
20 mM Tris-HCl pH 8.0	1ml	1M
2mM EDTA pH 8.0	200ul	0.5M
0.5% NP40	2.5ml	10%
0.1% SDS	250ul	20%
150 mM NaCl	1.5ml	5M
1X PIC	1 tablet	
Water	Up to 50ml	

High salt wash buffer	50ml	Stock
20 mM Tris-HCl pH 8.0	1ml	1M
2mM EDTA pH 8.0	200ul	0.5M
0.5% NP40	2.5ml	10%
0.1% SDS	250ul	20%
500 mM NaCl	5ml	5M
1X PIC	1 tablet	
Water	Up to 50ml	

ChIP Dilution buffer	50ml	Stock
16.7mM Tris-HCl pH8.0	840ul	1M
1.2 mM EDTA	120ul	0.5M
0.5% NP40	2.5ml	10%
167mM NaCl	1.67ml	5M
1X PIC	1 tablet	
Water	Up to 50ml	

LiCl wash buffer	50ml	Stock
10 mM Tris-HCl pH 8.0	500ul	1M
1 mM EDTA	100ul	0.5M
0.5% NP40	2.5ml	10%
250mM LiCl	2.5ml	5M
0.5% Sodium deoxycholate	2.5ml	10%
1X PIC	1 tablet	
Water	Up to 50ml	

TE buffer	50ml	Stock
10mM Tris-HCl pH 7.5	500ul	1M
1mM EDTA	100ul	0.5M
Water	Up to 50ml	

Elution buffer	1ml	Stock
0.1M $NaHCO_3$	100ul	1M
1% SDS	50ul	20%
Water	850ul	
Prepare fresh 1M NaHCO3 every time		

9.2 Sequencing and data analysis

ChIP-seq libraries were prepared and sequenced on an Illumina Hi-seq4000 as single-end 50-base reads. After sequencing, peak calling and quantitative comparisons were performed using the MACS, seqMINER and R.

10. DamID

10.1 DamID in cells

Dam-Tbp2 cDNA and *Dam-only* cDNA were cloned into pX vector, under the control of the TRE inducible promoter. *Dam-Tbp2*-NIH3T3 stable cell lines and *Dam-only*-NIH3T3 stable cell lines were obtained after transfection, FCAS sorting and puromycin selection. In the stable cell lines, leakage expression of *Dam-Tbp2* and *Dam* could methylate the 'GATC' site to 'G^{m}ATC'. Thus, genomic DNA was extract from both *Dam-Tbp2*-NIH3T3 and *Dam-only*-NIH3T3 cells, and used for 'G^{m}ATC'-methylation-specific PCR amplification (Vogel et al., 2007; Marshall et al., 2016). PCR products were purified by column (MACHEREY-NAGEL #REF740609.50) and sonicated. Libraries were prepared with the sonicated PCR products and sequenced on an Illumina Hi-seq4000 as single-end 50-base reads. DamID-seq data was analysed using the published pipeline (Marshall et al., 2015).

10.2 Oocyte DamID

Dam-Tbp2 cDNA and *Dam-only* cDNA were cloned into pRN3P vector, and capped mRNA of *Dam-Tbp2* and *Dam-only* were obtained by *in vitro* transcription with mMESSAGE mMACHINE® Kit (AM1344) from Life Technologies. mRNA of *Dam-Tbp2* (50ng/ul) was injected into oocytes (20-50 per group) collected from P7 and P13 mice, while oocytes injected with *Dam-only* mRNA (20ng/ul) are controls. After injection, oocytes were cultured in αMEM (Pfender et al., 2015) with penicillin G, streptomycin, ITS, 5% fetal bovine serum and 0.01 mg/ml oFSH for ~24 hours. Then the positively injected oocytes were collected and subsequently used for 'G^{m}ATC'-methylation-specific PCR amplification. PCR products were purified by column. Libraries were prepared with purified PCR products and sequenced on an Illumina Hi-seq2500 as single-end 50-base reads.

11. CUT&RUN

Oocyte CUT&RUN were performed following the published protocol (Hainer et al., 2018). Briefly, oocyte nuclei prepared from P7 and P14 oocytes (500 oocytes used per experiment) was immobilized on Concanavalin A coated beads. After the incubation of primary antibody (TBP2 and Pol II antibodies were used), protein A-micrococcal nuclease (pA-MN) was added. pA-MN binds specifically to the primary antibody, and upon the addition of Ca^{2+}, pA-MN can cleave the DNA that bound by primary antibody. Cleaved DNA fragments were released from nuclei to the supernatant, and recovered by Phenol/chloroform/isoamyl alcohol extraction. Libraries were prepared from these recovered DNA fragments and sequenced on Illumina Hi-seq4000.

12. List of primers

Gene name	Sequence(5' to 3')	Application
2654	F: CCATAGCTCCATATCCAGGGGG	TBP2 KO genotyping
2655	R: CAATGTCTAGGTCTGTTCTCTACAC	
2659	R:CTTTCCAAACAAGATTCCAAATGATGAAGCAAAG	
Gadd45g	F: ATCGGACTCTGGGAATCTTTACCT	TBP2 ChIP II10 cells
	R: AGAGGACCCTGTAAGACCACTACCA	
Mef2a	F: AAAAACATAGTCCGCCCCTCTTGTC	
	R: CTCGGCTTCCTCTCTTTCTTCTCTC	
Cdca8	F: ATTCACAAGAACGAACTCACCACTC	
	R: GTCCCAAAACACAGTCTGAGGAAC	
mouse GAPDH	F: CTCTGCTCCTCCCTGTTCC	ChIP primers
	R: TCCCTAGACCCGTACAGTGC	
mouse HPRT	F: CCAAGACGACCGCATGAGAG	
	R: CAACGGAGTGATTGCGCATT	
mouse MyoD1	F: GTCTCTCTGCCCTCCTTCCT	
	R: GTGTAGTAGGGCGGAGCTTG	
IR	F: TGATGCAACACATGGACATTTCTG	Intergenic region for ChIP
	R: TTCAGGGGTTGGGACAAAGTG	
UPRT	F: AGT GAC AAC CCC TCT GGA TG	Transgene
	R: CAT CGG ATC TAG CAG CAT CA	
	F: CAA ATG TTG CTT GTC TGG TG	Internal positive control
	R: GTC AGT CGA GTG CAC AGT TT	
Cre	F: TGATGAGGTTCGCAAGAACC	Cre genotyping
	R: CCATGAGTGAACGAACCTGG	
Sr-9264	GCTGAACAGTTGAGACATAGCTGGAGG	TBP-TBP2 swap mice genotyping
Xf-9265	CCTGGCTCATCAACTCCTTCTCTGC	
Er-9266	GTCTGGGTAATAGCCTGCCTCTTGG	
Sf-9267	CCTCCCACAACGAGGACTACACCATC	
AdRt	CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGA	DamID adaptor and
AdRb	TCCTCGGCCG	

AdR_PCR	GGTCGCGGCCGAGGATC	PCR primers
BMP15	F: CATCATCAGCAGCAGCAT	qPCR primers for TBP2 DamID in Oocytes
	R: ATCCAACTTGTAGACCTCTTAG	
Fbxw18	F: AACTGTGGCTGTGTATCC	
	R: GGTAGAACTTATGTCCAGGA	
Gli3	F: CGGTATTCTGGAATGTCTCTT	
	R: GAGTGCCTGTCCTTGGTT	
Oosp3	F: CTCTCCAGGAAGTACACAT	
	R: GAGTCCAGTATTCTTCAGTG	
Scml2	F: GCTTCTGGTGGTTGTACTT	
	R: CCGATGACTGAGCAAGATT	
Dcp1a	F: CAGGAGTCAGGTTCATTCA	
	R: CGTGGTACTTCAGCTAATCT	
Kdm6a	F: TTGTGTCTCCTACGAATCC	
	R: TAGAGGTGAAGGCAGAGG	
Dnmt1	F: TAGGAGGACTGCCACATT	
	R: CAACACTGAGGAGGAGGA	
Btg4	F: TTCCTGCGTGAATCTGAC	
	R: GGAGGTTGTATCAAGAAGAAC	
ZP3	F: GAGGTAGGAGAATTGGAGTT	
	R: TCAGTGGAGTTGCTTGTC	
Gdf9	F: TTGCTATCTTGCCATTCCA	
	R: TGCTCTTCTTAGAACAACCA	
Rbm39	F: CCTTATGGCGACCTTGAC	qPCR primers for TBP2 DamID in 3T3 cells
	R: GGATTCCTTGCTGAGTGG	
Epc1	F: GTCTGCCAAGTTAGGTGAA	
	R: TTGAGTCTGAACGCCTTAC	
Sf3b3	F: AGTCAAGCCAATTCGGTATA	
	R: ACCTGTCAGTTCTACTCAAG	
Cnot11	F: GGAAGACGCTCTTGGTATT	
	R: GGATGCTCTATGACAACTAAT	
Prdm5	F: AGACTTCCTACAGAGCAATC	
	R: GCATTCCATTCAACAAGACT	

Ino80d	F: TTCCAGCGATGAGAGGTA	
	R: TCCTACAGCACCACTACA	
Ythdf3	F: TCCAGCCTCAATTCCTCT	
	R: TCTCTTGTGTTCCTTCTGTT	
Myod1	F: CCGTGGAAGAACAGATATTC	negative controls
	R: GGCAAGAGACAGTGAGAC	
Ins1	F: CACCTGGAGACCTTAATGG	
	R: ATACCTGCTTGCTGATGG	
Intergenic region	F: CTCTGCTTTACCTAATGTCTC	IR DamID
	R: GCCTGTGCCTATTGAGTA	
SG12	F: CTGCCCACTAGCACGGCC	TAF10 genotyping with SE89
SE89	R : CAGTCTAACCTGCTCCGAG	
Taf10_ex2	F: GTAGTGTCCAGCACACCTCT	
TAF7 WT	F: ATGAAAGGCAAGCTCCAAGA	TAF7 genotyping
	R: ATTCCAGCTCTTCCTGCAAA	
TAF7 KO	F : CGAAGAGTTCGTTCACTCCC	
	R: GAAGGCAAGTTCTCAATGAAAGGG	
TAF7 loxp_loxp	F: GTATGAAAACCTGTGTCCTGGTCTG	
	R: GAAGGCAAGTTCTCAATGAAAGGG	
GAPDH	F: TCACCACCATGGAGAAGGC	Housekeeping genes for qPCR
	R: GCTAAGCAGTTGGTGGTGCA	
18rRNA	F: GTAACCCGTTGAACCCCATT	
	R: CCATCCAATCGGTAGTAGCG	
TBP mCherry-F	F: GGATTCAGGAAGACCACAA	TBP2 SWAP mice mRNA qPCR primers
TBP mCherry-R	R: GCATGAACTCCTTGATGATG	
TBP2 venus-F	F: ACAGGTTGTGTTGCTAATCT	
TBP2 venus-R	R: AGGCTGAAGTTGGTTGCT	

Results

TBP2-TFIIA-containing transcription machinery, different from TFIID, drives mRNA and MaLR endogenous retroviral element expression in growing oocytes

Changwei Yu[1,2,3,4], Emese Gazdag[1,2,3,4,5], Matthieu Jung[1,2,3,4], Mathilde Joint[1,2,3,4], Luc Negroni[1,2,3,4], Stéphane D. Vincent[1,2,3,4,*] and László Tora[1,2,3,4,*]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France;

[2]Centre National de la Recherche Scientifique (CNRS), UMR7104, Illkirch, France;

[3]Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Illkirch, France;

[4]Université de Strasbourg, Illkirch, France;

[5]Ichan School of Medicine at Mount Sinai, New York, USA;

[*]Correspondence should be addressed to S.D.V. vincent@igbmc.fr or L.T. laszlo@igbmc.fr

Running title: TBP2-TFIIA regulate oocyte specific transcription

Key words: TFIID, TBP, RNA polymerase II, TFIIA, TBP-associated factors (TAFs), transcription, promoter, MaLR (mammalian apparent LTR retrotransposons), ERV (endogenous retroviral elements), oocytes, primary follicles, secondary follicles, mass spectrometry, RNA-seq,

Abstract

Mammalian oocytes go through consecutive differentiation process, during which the synthesis and accumulation of RNAs and proteins are essential for oocyte growth, maturation, fertilization and early embryogenesis. Little is known about the nature and function of the transcriptional machinery that is involved in RNA polymerase II (Pol II) transcription initiation during oogenesis. In somatic cells, the Pol II general transcription factor (GTF), TFIID, is the first to bind to gene promoters to nucleate pre-initiation complex (PIC) formation together with TFIIA, -IIB, -IIE, -IIF, and -IIH. In metazoans, TFIID is composed of the TATA binding protein (TBP) and 13 TBP-associated factors (TAFs). During oocyte growth TBP is replaced by a vertebrate specific TBP-type protein, TBP2 (also called TRF3 or TBPL2) and *Tbp2*$^{-/-}$ females are sterile. To understand whether and how TBP2 is controlling transcription initiation during oogenesis, we carried out RNA-seq analyses from wild-type and *Tbp2*$^{-/-}$ oocytes from primary and secondary follicles. These analyses show a main decrease in the expression of the most abundant genes as well as specific down-regulation of the expression of the MaLR (mammalian apparent LTR retrotransposons)-type endogenous retroviral elements. To identify the nature of the complex associated with TBP2 in the oocytes, we carried out immunoprecipitation followed by mass spectrometry. We demonstrate that, in the oocytes, TBP2 associates with TFIIA, but does not assemble into a TFIID-type complex. Altogether, our data show that a specific TBP2-TFIIA-containing transcription machinery, different from canonical TFIID, drives transcription in mouse oocytes.

Introduction

Regulation of transcription initiation by RNA Polymerase II (Pol II) is central to any developmental process. Female germ cells develop during oogenesis leading to the formation of a highly differentiated and specialised cell, the oocyte. Oogenesis comprises a well-defined series of events, each of which must be regulated. In females, the oocytes enter meiosis during the embryonic life and quiescent primordial follicles composed of meiotically arrested oocytes at the late diplotene stage surrounded by granulosa cells, are formed perinatally in mice (reviewed in (Choi and Rajkovic, 2006)). Shortly after birth, some primordial follicles enter folliculogenesis and undertake a growth phase during which oocytes increase their size until the pre-antral follicular stage (Pedersen and Peters, 1968). Later, in antral, preovulatory follicles, the fully-grown germinal vesicle (GV) stage oocyte is ready for ovulation. Maturation follows with the resumption of meiosis induced by ovulatory stimulus and proceeds until the metaphase II, at which stage the oocyte awaits for fertilisation (reviewed in (Li and Albertini, 2013)). Genes specifically expressed in the oocyte are necessary for either growth or for communication with follicular cells and a remarkable feature of oocyte is the high expression of the retrotransposons driven by Pol II transcription. These elements are interspersed repetitive elements that can be mobile in the genome. They represent a threat for the integrity and functionality of the genome and suppression mechanisms have been selected throughout evolution in the soma and in germ cells (reviewed in (Crichton et al., 2014)). There are 3 major classes of retrotransposons in mammals: long interspersed nuclear elements (LINE), short interspersed nuclear elements (SINE) and long terminal repeat (LTR) retrotransposons derived for retroviruses, also known as endogenous retroviruses (ERVs). There are 3 main sub classes of LTRs: ERV1, ERVK and endogenous retrovirus like (ERVL)-MaLR (mammalian apparent LTR retrotrosposons) defined by their phylogenetic relationships. Remarkably, some retrotransposons, LTRs in particular, are actively expressed in female germ cells. The class III

MaLR family is highly expressed in oocytes and can regulate oocyte gene expression by acting for example as alternative promoters or first exons for somatic genes (Peaston et al., 2004). Although transcription factors regulating oocyte-specific gene expression have been characterized (reviewed in (Jagarlamudi and Rajkovic, 2012)), little is known about the role of general transcription factors (GTFs) in regulating transcription initiation in female germ cells.

Pol II transcription requires the stepwise assembly of protein complexes on core promoters of genes forming the preinitiation complex (PIC) (reviewed in (Goodrich and Tjian, 2010)). A functional PIC consists of the Pol II and several multi-protein complexes called general transcription factors (GTFs). The evolutionary conserved TFIID complex plays a major role in transcription initiation as it is the first GTF to initiate the assembly of the PIC by recognizing the promoter. TFIID is a large multi-protein complex composed of the TATA box-binding protein (TBP) and 13 TBP-associated factors (TAFs) in metazoans (Tora, 2002). Among the 13 metazoan TAFs, TAF9, TAF10 and TAF12 are also shared by the Spt Ada Gcn5 acetyl transferase (SAGA) activator complex conserved in metazoans (reviewed in (Spedale et al., 2012)). The traditional text-book model suggesting that transcription is always regulated by the same transcription complexes has been challenged in metazoans in the last decade by the discovery of cell-type specific complexes containing specialized TAF paralogs (reviewed in (Goodrich and Tjian, 2010; Ho and Crabtree, 2010; Müller et al., 2010)) and of the context-dependent requirement of different TAFs for transcription (Bardot et al., 2017; Gegonne et al., 2012; Indra et al., 2005; Mohan et al., 2003; Tatarakis et al., 2008). Moreover, in the absence of TBP, TBP paralogues are able to mediate Pol II transcription of developmentally important genes in vertebrate embryos (Ferg et al., 2007; Jacobi et al., 2007; Martianov et al., 2002b; Müller et al., 2001; Veenstra, 2000). Three metazoan TBP paralogs have been identified. The TBP related factor (TRF) TRF1 is insect specific (Crowley et al., 1993). The TBP like factor (TLF, also known as TRF2 or TBPL1) has been identified in several metazoan species

(Dantonel et al., 2000; Kaltenbach et al., 2000; Ohbayashi et al., 1999; Teichmann et al., 1999) while the TBP-related factor 3 (TRF3, also known as TBPL2, here after called TBP2) has only been described in vertebrates (Bártfai et al., 2004; Persengiev et al., 2003). While TLF and TBP2 share significant identity with the highly conserved saddle-like-C-terminal DNA binding domain (Hernandez, 1993) (78% and 92-93%, respectively (Bártfai et al., 2004; Crowley et al., 1993; Persengiev et al., 2003)), TRF2 is more distant with only 42% identity (Bártfai et al., 2004). A consequence of this difference is that TBP2, but not TRF2, is able to bind the TATA-box (Bártfai et al., 2004; Ohbayashi et al., 1999; Rabenstein et al., 1999). TFIIA and TFIIB are two GTFs that are important for the stabilization of TFIID on the DNA. While TFIIB is composed of only one protein, TFIIA is a complex formed by 3 polypeptides α and β produced by taspase 1 cleavage of the TFIIA-αβ precursor, and TFIIA-γ (reviewed in (Høiby et al., 2007)). Variability is also present in the TFIIA complex, as a TFIIAα-β homolog called ALF (TFIIA-like factor or TFIIAτ), has been identified in male germ cells (Upadhyaya et al., 1999). Similarly to TBP, all three TBP related factors are able to interact with TFIIA and TFIIB to mediate Pol II transcription initiation *in vitro* (Bártfai et al., 2004; Hansen et al., 1997; Jallow et al., 2004; Rabenstein et al., 1999). In mice, *Tbp2* is exclusively expressed in the oocytes (Bártfai et al., 2004; Xiao et al., 2006). In zebrafish and xenopus, expression of *Tbp2* is broader, but is enriched in the ovary, suggesting a fundamental role for TBP2 in the vertebrate ovary (Bártfai et al., 2004; Jallow et al., 2004; Xiao et al., 2006).

We have previously shown in the mouse ovary that TBP is expressed in the oocytes in primordial follicles and becomes undetectable during oocyte growth whereas TBP2 is highly expressed in the growing oocytes, strongly suggesting that TBP2 is replacing TBP during folliculogenesis (Gazdag et al., 2007). The crucial role of TBP2 for oogenesis was demonstrated by the absence of phenotype of *Tbp2*$^{-/-}$ mice, except female sterility due to defect in secondary follicles production (Gazdag et al., 2009). TBP2 binds to actively transcribed genes promoter

and in its absence, transcription is perturbed in the oocyte, primarily at the primary follicle stage, and leads to altered transcriptional profile of oocyte-specific genes in a transcriptome analysis of 2 weeks old ovaries (Gazdag et al., 2009). Altogether, these data strongly suggest that TBP2 is playing a specialized role during oocyte development.

To understand whether and how TBP2 is controlling transcription initiation during oogenesis, we carried out RNA-seq analyses from wild-type and *Tbp2*$^{-/-}$ oocytes from primary and secondary follicles. These analyses show an important impairment of gene expression, with a strong tendency for down regulation of the most abundantly expressed genes, as well as specific down-regulation of the expression of the MaLR (mammalian apparent LTR retrotransposons)-type endogenous retroviral elements. To identify the nature of the complex associated with TBP2 in the oocytes, we carried out immunoprecipitation followed by mass spectrometry. We demonstrate that, contrary to TBP, TBP2 associates in the oocytes with TFIIA, but does not assemble into a TFIID-type complex. Altogether, our data show that a specific TBP2-TFIIA containing transcription machinery, different from canonical TFIID, drives transcription in mouse oocytes.

Results

RNA polymerase II transcribed genes are affected in *Tbp2*$^{-/-}$ mouse primary and secondary follicle oocytes

To analyse whether oocyte specific transcription is affected by the ablation of the mouse *Tbp2* gene (official symbol *Tbpl2*), we have isolated primary and secondary follicles [post-natal days (P) 7 and P14], isolated polyA+ RNA and carried out RNA-seq analyses of wild-type and *Tbp2*$^{-/-}$ oocytes. Each RNA sample was derived from three independent biological replicates and sequenced at 50 nucleotides single ends. Specific sequence-reads were mapped to the mouse genome, and unique reads were considered for further analyses. Next, we verified the

expression profile of the *Tbp2* gene on the IGV genome browser in wild type (WT) and knock-out P7 and P14 oocytes with the data coverage normalization parameter (Fig. 1A). As expected the expression of the *Tbp2* gene was hardly detectable, with no signal in the deleted exon 4, while *Tbp2* expression was readily detected in the WT controls, validating the RNA-seq experiment. Besides *Tbp2*, we observed the down regulation of a number of oocyte specific genes, such as for example *Ooep* (Fig. 1B).

For the comparative analyses, all datasets were normalized across samples with the median-of-ratio method (Anders et al., 2013) and differential expression was accessed using DESeq2 (Love et al., 2014). To analyse the RNA-seq data more precisely a principal component analysis (PCA) and a hierarchical clustering were carried out (Fig. 2A and 2B, Supplementary Fig. 1A). Both analyses showed that the four distinct RNA samples separated in individual groups, but the triplicate samples belonging to the same group clustered together, indicating that the main explanation for the variance is the genotype, and not the stage. These analyses further validated the high throughput sequencing and the knock-out strategy. Next, we analysed how *Tbp2* loss of function influences mRNA expression in both P7 and P14 oocytes. At both of these oocyte development stages we found that in WT P7 and in WT P14 about 10^4 of genes were expressed, (10622 and 10697 reads divided by the median length of transcripts in kb, respectively) although many of these transcripts maybe maternally deposited in the oocytes. Out of these transcripts using an absolute log2 fold change cut off of 1, slightly more genes were downregulated (1720 in P7 and 1794 in P14) than up regulated (1577 in P7 and 1358 in P14) in the *Tbp2*$^{-/-}$ oocytes (Fig. 2C and 2D, Supplementary Fig. 1B-C). Interestingly, the most highly expressed genes were down regulated in the two oocyte developmental stages (Fig. 2C and 2D, Supplementary Fig. 1B-C). When either the down regulated or the upregulated genes were compared between the two developmental stages in the *Tbp2*$^{-/-}$ cells, we found that in each category about 50% or more genes of the TBP2 regulated genes were also influenced the same

way in the two oocyte developmental stages (Fig. 2E). Next, we carried out a direct fold-change comparison between P7 and P14. This analysis indicated that loss of TBP2 has a major effect on the expression of the most abundantly expressed genes in oocytes at P7 and P14 (Fig. 2F). These results together suggest that TBP2 has an important role in gene expression regulation in the growing oocytes. However, as TBP2 is suggested to play a role in Pol II transcription it is surprising that large proportion of genes are upregulated following *Tbp2* ablation. As in the above analyses we measured steady state mRNAs changes, it is possible that the direct transcript synthesis effects have been masked by transcript buffering mechanisms (reviewed in (Timmers and Tora, 2018)).

To analyse and find potential differential gene function categories in transcripts down regulated in P7 and P14 oocytes, we carried out Gene Ontology (GO) analyses on the identified common down regulated categories of genes (Supplementary Table 1, carried out with DAVID (Huang et al., 2009)). Interestingly, many down regulated genes were classified in GO categories linked to chromatin binding, transcription factor activity, RNA Pol II activity, Pol II binding, DNA binding, core promoter specific binding (Fig. 3A), which are in good agreement with the hypothesis that TBP2 functions as a Pol II specific promoter binding factor in growing oocytes. One of the GO categories that came out with the highest significance was "polyA specific ribonuclease activity" containing many genes coding for two exonuclease complexes contributing to the majority of the deadenylation activity in eukaryotes: CCR4-NOT and PAN2-PAN3. This finding suggests that the activities of the CCR4-NOT and PAN2-PAN3 complexes are impaired, and thus mRNA decay is down regulated in oocytes lacking TBP2. Importantly, mRNAs stored in the oocytes undergo general decay during maternal-zygotic transition and their stability is tightly regulated (reviewed in (Walser and Lipshitz, 2011)). In addition, it was shown that an ERK1/2 triggered *Btg4*-mRNA translation is a key step in oocyte maturation and revealed that BTG4-CNOT7 and BTG4-CNOT8 mediated mRNA decay is required for the

successful development of the oocytes (Yu et al., 2016). Thus, next we analysed the expression of factors regulating mRNA decay in the oocytes. Interestingly, many genes coding for factors participating in the oocyte specific mRNA decay were significantly down regulated in the *Tbp2*$^{-/-}$ mutant oocytes (Figure 3B, Supplementary Table 2), including *Btg4* and *Cnot8*. Thus, it is possible that many transcripts, which were found upregulated in *Tbp2*$^{-/-}$ mutant oocytes are indirectly regulated by the depletion of TBP2, due to the lack of the decay of maternal mRNAs.

Several core promoter motifs are enriched in TBP2 regulated genes

Several elements such as the TATA box, the initiator (Inr), the Motif Ten element (MTE) and the distal promoter element (DPE) are bound by TFIID (reviewed in (Danino et al., 2015)). However, there are no individual core promoter sequence elements, or sequence element combinations that can be used to define TSSs, and the structure-function relationship of core promoters remains poorly understood. Nevertheless, as the C-terminal DNA binding domains of TBP2 and TBP are highly similar, TBP2 is able to bind the TATA box (Bártfai et al., 2004). We hypothesized that the down regulated genes may be direct targets of TBP2-mediated transcription and thus, we carried out analyses to uncover the potential enrichment of different individual core promoter elements of the down regulated genes within a -50/+50 genomic sequence using regulatory sequence analysis tools (RSAT) (Turatsinze et al., 2008). We tested the presence of TATA box, Inr, MTE, DPE and two core promoter motifs localized around the TSS initially identified in the promoter of the hepatitis B virus X gene (Danino et al., 2015). We analyzed the enrichment of predicted motifs between the common down regulated genes at P7 and P14 (down) compared to an array of random sequences (random), all the annotated genes (all) or to the protein coding genes of the analysis (CDS) (Fig. 4A-F). While our pipeline of analysis did not detect significant enrichments using the Inr element, we observed a small enrichment of all the other core promoter elements that was statistically significant for the

TATA box, MTE and XCPE1 elements (Fig. 4A, 4C and 4E). Thus, it seems that the TBP2-regulated genes do not dramatically differ at the sequence level from canonical TFIID-regulated genes, but are more enriched in TATA boxes, MTE and XCPE1 elements.

Expression of mouse class III MaLR endogenous retroviral (ERV) elements are down regulated in *Tbp2*$^{-/-}$ primary and secondary follicle oocytes

During the course of the above mRNA analyses we realized that expression from many mouse repetitive elements are significantly down regulated (Fig. 5A). Mammalian repetitive elements belong to the following major classes of repeat elements: long terminal repeats (LTRs), long interspersed nucleotide elements (LINEs), short interspersed nuclear elements (SINEs) class I retrotransposons and class II DNA transposons and are transcribed by Pol II (reviewed in (Smit, 1999)). When the relative abundance of transcripts initiated from these repetitive elements was analysed in wild-type P7 and P14 oocytes, the expression of the LTR retrotransposon class [also called endogenous retroviruses (ERVs)], was found to be the most abundant at both P7 and P14 (Fig. 5B). As the mouse LTRs (ERVs) are further divided in ERV classes I, II and III as well as into solo LTRs and gypsy LTRs (Hubley et al., 2016), we analysed which of these classes is affected in the P7 and the P14 *Tbp2*$^{-/-}$ mutant oocytes (Fig. 5C and Supplementary Fig. 2A). Our analyses indicated that expression from the class III ERVs was the most down regulated (Fig. 5D and Supplementary Fig. 2B). ERVs belonging to the mouse class III family contain ERVL and MaLR apparent LTR retrotransposons (reviewed in (Thompson et al., 2016)). In our WT oocytes RNA-seq analysis MaLR family of non-autonomous retrotransposons were found to be highly expressed, in agreement with previous observation (Peaston et al., 2004) and importantly the most down regulated in both P7 and P14 *Tbp2*$^{-/-}$ mutant oocytes (Fig. 5D and Supplementary Fig. 2B). There is more than 100 distinct MaLR LTRs (Hubley et al., 2016) and among them, three members MT-Int, MTA-Mm, and

MTA-Mm-Int are highly expressed and highly affected in P7 and P14 *Tbp2*$^{-/-}$ mutant oocytes (Fig. 5E and 5F, and Supplementary Fig. 2C). In order to get a better mapping precision, we resequenced one sample of each triplicates in 100 nucleotides paired-ends and similar results were obtained (Supplementary Fig. 3A-H). MaLRs encode no known proteins, but it has been shown that MaLR-dependent transcription is key in initiating synchronous developmentally regulated transcription to reprogram the oocyte genome(Peaston et al., 2004). As TBP2 depletion is reducing MaLR transcription by 4-fold in both P7 and P14 oocytes, it could thus seriously deregulate oocyte specific transcription of neighbouring genes and consequently stop oocyte growth.

TBP2 is interacting with TFIIA in TBP2 in mouse oocytes

To characterize TBP2-containing transcription complexes, first we have carried out immunoprecipitations (IPs) from NIH3T3 cells artificially overexpressing TBP2 (NIH3T3-II10 cells) (Gazdag et al., 2007). To be able to use antibodies for anti-TBP2 IPs that recognize mouse (m) TBP2 with high affinity we have generated novel anti-mTBP2 rabbit polyclonal antibodies (pAb) and affinity purified them (Materials and Methods, Supplementary Fig. 4). Using the two of the best generated anti-TBP2 pAbs and the previously published anti TBP2 monoclonal antibody (mAb) (Gazdag et al., 2007), TBP2-containing complexes were IP-ed from NIH3T3 whole cell extracts, IP-ed complexes were eluted by pH 2.5 buffer and proteins were identified by mass spectrometry. In parallel, as a control we have also carried out an anti-TBP IP and the eluted TBP-containing complexes were also analysed. To be able to compare the composition of the IP-ed complexes, normalized spectral abundance factor (NSAF) values were calculated (Zybailov et al., 2006). All 3 tested anti-TBP2 antibodies gave very similar results (Supplementary Fig. 5A). All the known TBP-containing complex subunits [Pol II TAFs (1-13), Pol I TAF1s (A-D) and Pol III subunit BRF1/TF3B) subunits were identified by mass

spectrometry analysis (Fig. 6A and Supplementary Fig. 5B). In contrast, the anti-TBP2 immunoprecipitations showed that the artificially expressed TBP2 can incorporate in TFIID-like complexes, as TFIID TAFs were co-IP-ed with TBP2, however with lower stoichiometry (NSAF values) than that of TBP (Fig. 6A, Supplementary Fig. 5A). Interestingly, in spite of the high similarity between the core domains of TBP2 and TBP, we have not identified any of the Pol I or Pol III TBP-associated factors (Fig. 6A and Supplementary Fig. 5A). In addition, in the anti-TBP2 IPs, while TFIIB was not detected, we have identified TFIIA-αβ and TFIIA-γ as TBP2 partners (Fig. 6A and Supplementary Fig. 4A). To analyse whether the TBP2 associates with TFIID TAFs and TFIIA in same complex, NIH3T3-II10 whole cell extract was purified on a Superose 6 gel filtration column that separates proteins and complexes with molecular weights between 5 MDa and 5000 Da. This separation indicated that the most of TBP2 and TFIIA isolated from NIH3T3-II10 cells could be found in the same fractions (22-26) eluting around 150 kDa, while TBP2 protein was below the detection threshold of the western blot assay in the TFIID-containing fractions 9-15 (Fig. 6B). To verify whether in the fractions 22-26 TBP2 and TFIIA would form a complex, we have IP-ed TBP2 from fractions 23-25 pooled and subjected to mass spectrometric analysis. This anti-TBP2 IP confirmed that in the fractions eluting from the Superose 6 around 150 kDa TBP2 and TFIIA form a complex, with a calculated molecular weight of about 110-120 kDa (Fig 6C and Supplementary Fig. 5C).

To investigate whether TBP2 would form an endogenous complex with TFIID TAFs and/or TFIIA, we prepared mouse ovary whole cell extracts (WCE) from 120 P14 stage ovaries, where TBP2 expression is the highest in the growing oocytes. Three independent anti-TBP2 IPs identified TFIIA-α or unprocessed TFIIA-αβ and TFIIFA-γ subunits, as well as in one of the triplicate experiments we also detected one peptide corresponding to TAF5 and TAF10 (Fig. 7A). Nevertheless, in a control anti-TBP IP we have detected all TFIID subunits with high NSAF values, as well as background levels of TFIIFA-γ and TFIIB (Fig. 7B). As TBP2 is only

expressed in the growing oocytes present in the ovaries, we wanted to further verify whether TBP2 could associate with TFIID TAFs. To this end we have designed a triple IP strategy (Fig. 7C). First, we have carried out an anti-TAF7 IP, to deplete all TAF7-containing TFIID, from the ovary WCE and analysed whether in the IP-ed complexes we would detect TBP2. All the TFIID subunits, except TAF7L, were identified in the 3 technical replicates with high confidence (Fig. 7D and Supplementary Fig. 6A). In contrast, TBP2 was detected with only one TBP2-specific peptide with a medium score in only 1 of the 3 technical replicates (data not shown) indicating that this detection is not significant. Remarkably, no TFIIA was detected in the TAF7-associated complexes (Fig. 7D and Supplementary Fig. 6A). As in the crude anti-TBP2 IPs (Fig. 7A) we detected background levels (one peptide) of TAF10 (a subunit of both TFIID and SAGA complexes), the anti-TAF7 IP flow through was re-immunoprecipitated with an anti-TAF10 antibody. This second IP demonstrated the anti-TAF7 depletion of TFIID was very efficient, as in the second anti-TAF10 IP only the shared SAGA-TFIID TAFs were detected (Supplementary Fig. 6B). Again, in the anti-TAF10 IP we did not detect any TBP2 nor TFIIA, but SAGA subunits (Fig. 7D and Supplementary Fig. 6B). Next from these TAF7 and TAF10 double depleted IP flow through extracts we have carried out an anti-TBP2 IP. The analysis of this third consecutive IP indicated that TBP2 forms a complex with TFIIA-αβ and TFIIFA-γ subunits (Fig. 7E and 7F and Supplementary Fig. 6C). Interestingly, we did not detect the germ cell specific TFIIA-αβ paralog ALF (Upadhyaya et al., 1999) despite its high level of expression in oocytes (Xiao et al., 2006). Thus, in conclusion we demonstrate that in mouse oocytes TBP2 forms a TFIID TAFs-free stable complex with TFIIA.

TAF7 is not required for oocyte growth and maturation

In order to functionally validate that TFIID TAFs, TAF7 in particular, are not required for transcription during oocyte growth, we carried out a conditional deletion of *Taf7* during

oocyte growth using the *Zp3-Cre* transgenic line (Lewandoski et al., 1997). We chose TAF7 since it is a TFIID specific subunit. We obtained Tg(*Zp3-Cre*/+);*Taf7*$^{flox/\Delta}$ females (Fig. 8A) and performed superovulation and histological analyses. As shown in Table 1, the oocyte specific deletion of *Taf7* did not affect the numbers of collected mature oocytes after superovulation. Histological analyses of 6 weeks-old female ovaries confirmed the presence of antral follicles (Fig. 8C compared to 8B, magnification in Fig. 8D and 8E). Altogether, these data further suggest that expression of *Taf7* in the oocyte is not necessary for oocyte growth and is in good agreement with our finding that the oocyte specific functional TBP2-containing transcription complex does not contain TFIID TAFs.

Discussion

Three genes encoding different TBP-type factors, TBP, TLF/TRF2/TBPL1 and TBP2/TRF3/TBPL2 exist in vertebrate genomes (reviewed in (Müller et al., 2010)). TBP-type factors are bipartite proteins with variable N-terminal domains and a relatively well conserved shared C-terminal domain forming a saddle-like structure with a concave surface that is known to bind to DNA. TBP2 proteins from different vertebrates show a high degree of similarity in their C-terminal core domains amongst themselves, but also with TBP. On the other hand, the N-terminal domains of the TBP2s are highly variable amongst the different vertebrate homologues of TBP2s and also very different from TBPs. While TBP has a rather ubiquitous expression pattern in mammalian cells and tissues, the expression of TBP2 and TLF/TRF2 are highly specialized. TBP2 is expressed in growing oocytes and thus, essential for female germ cell differentiation in mice (Gazdag et al., 2007; Xiao et al., 2006). In a mirroring situation, TRF2 expression is enriched during spermatogenesis and male germ cells lacking TRF2 are blocked between the transition from late round spermatids to early elongating spermatids {Martianov:2001kk}. However, while TBP2 and TBP show contrasting expression patterns in

the oocytes (Gazdag et al., 2007), TRF2 and TBP are co-expressed in spermatids (Martianov et al., 2002a; Schmidt and Schibler, 1997). Nevertheless, it seems that during gonad development specific transcription programs driven by either TBP2 during oocyte growth, or TRF2 during spermatogenesis, are required, and that these specified gonad specific transcription programs cannot be carried out by TBP.

In this study, we show that TBP2 forms a stable complex with TFIIA in mouse oocytes, but does not associate with either the oocyte expressed TFIIA-αβ paralogue, ALF protein (Upadhyaya et al., 1999; Xiao et al., 2006). While we detected the presence of TFIIA in all our TBP2-IPs from ovaries, the detection of TAF5 and TAF10 was observed only in 1 out of 3 biological replicates. In the same line, the detection of TBP2 in the anti-TAF7 IP has been observed in only 1 out 3 technical replicates with only a medium score. Altogether, these data strongly support the idea that the main TBP2-associated complex in the growing oocytes contains TFIIA but not TFIID TAFs. This conclusion is supported by our IPs experiments in the NIH3T3-II10 cell line that overexpresses TBP2. While TFIID is present in these cells and TBP2 able to interact with TFIID TAFs, our gel filtration experiment indicates that the vast majority of the TBP2-associated complex does not elute with TFIID and mass spectrometry experiment confirmed the absence of any TFIID TAFs in the 120 kDa fractions. Our oocyte specific *Taf7* KO experiments, showing no phenotypic and functional problems following TAF7 depletion (Fig. 8), is in accordance with the lack of functional TFIID in mouse oocytes. A possible explanation of the absence of significant interactions between TFIID TAFs and TBP2 in the oocyte could be the absence of TFIID TAFs proteins. Interestingly, we found that all TFIID *Tafs*, except *Taf7l*, are expressed at the mRNA level in oocytes (see Supplementary Table 4), however, whether they are also expressed in oocytes at the protein level is not known, except for TAF4B that has been detected in female neonate' oocytes (Falender et al., 2005). We did not detect any TAFs proteins in crude oocyte extracts by proteomic analyses (data not

shown) but this could be due to their potential relative low abundance. It could be conceivable that, similarly to *Tbp* mRNA that is transcribed but not translated in oocytes (Akhtar and Veenstra, 2009), TFIID *Taf* mRNA translations are also inhibited and as a result the canonical TFIID is not present in oocytes. Another reason why TBP2 does not interact with TAFs nor ALF, but rather interacts with TFIIA could be its N-terminal domain that is very different from TBP.

TFIIA was initially classified as a GTF, when it was first purified (DeJong et al., 1995; Ma et al., 1993; Ozer et al., 1994) but since it is dispensable in transcription (Van Dyke et al., 1988; Wu and Chiang, 1998), it can be considered more as a general cofactor. TFIIA is the proteolytical cleavage of TFIIA-αβ into a TFIIA-α and TFIIA-β moiety (Høiby et al., 2007). Recent studies have shown that TFIIA is cleaved by Taspase1 (Zhou et al., 2006). In our proteomic experiments, we cannot determine whether the anti-TBP2 IPs bring down uncleaved TFIIA-αβ or processed TFIIA-α and TFIIA-β. Preliminary *in vitro* interaction experiments suggest that TBP2 interacts preferentially with uncleaved TFIIA (I. Berger personal communication). However, cleavage of TFIIA does not appear to serve as a step required for its activation in oocyte as Taspase1 knock-out female mice are fertile (Oyama et al., 2013). It is interesting to note that while TRF2 can interact with unprocessed TFIIA, this complex is not functional (Oyama et al., 2013). The crystal structure between the human and yeast TFIIA/TBP/DNA complex revealed conserved architecture (reviewed in (Hantsche and Cramer, 2017)) and indicated that the β-sandwich domain of TFIIA interacts with the N-terminal stirrup region of TBP and with the backbone of the TATA box, whereas the four-helix bundle of TFIIA projects away from the TFIIA/TBP/DNA complex and is available for additional interactions with transcription factors. Further structural studies will determine whether TBP2 interacts with uncleaved or processed TFIIA, how TBP2-TFIIA interact with DNA and whether TFIIA/TBP2/DNA complex is different from that of its counterpart with TBP or TRF2.

An interesting parallel between TBP2 and TRF2 is that both TBP-type factors form endogenous stable complexes with TFIIA. The beginning of TBP2 accumulation in the oocyte nuclei, or TRF2 accumulation in male germ cell nuclei coincides with the phase of meiosis I (Gazdag et al., 2007; Martianov et al., 2001). It is thus conceivable that TBP2-TFIIA in oocytes, or TRF2-TFIIA during spermatogenesis are involved in the control of gene expression in a meiotic context and that probably both transcription complexes function in a more compacted DNA environment in which TBP/TFIID cannot. Moreover, in *Drosophila* TLF/TRF2 has been suggested to bind to promoters harbouring a non-canonical, TCT-containing initiator (Wang et al., 2014). In the future genome-wide TBP2 binding analyses with sensitive low-cell techniques, such as TBP2-Dam-ID or CUT&RUN (Marshall et al., 2016; Skene et al., 2018), will be needed to map more precisely the sequence motifs to which TBP2 is binding in the oocytes.

Oocytes display remarkable post-transcriptional regulatory mechanisms that control mRNA stability and translation. During oogenesis, the oocyte genome is transcriptionally active, and the newly synthesized maternal mRNAs are either translated or stored in a dormant form (reviewed in (Bettegowda, 2007)). The newly synthesized transcripts receive a long poly(A) tail and subsequently undergo poly(A) shortening in the cytoplasm, preventing translation. Until resumption of meiosis, mRNAs with a short poly(A) tail are stored in the cytoplasm in a dormant form. Thus, poly(A) tail deadenylation, amongst other activities, coordinates post-transcriptional regulation of the oocyte mRNA pool and is critical for normal the progression of early embryonic development. In this respect, it is interesting to emphasize that in *Tbp2*$^{-/-}$ mutant oocytes the activity of the two major deadenylation complexes, CCR4-NOT and PAN2-PAN3, seems to be impaired. This in turn would result in the increase of transcripts with long(er) poly(A) tails and their unwanted translation of dormant maternal mRNAs in the oocytes. In addition, TBP2 could also control other oocyte–specific negative post-transcriptional regulatory mechanisms, which target mRNAs for degradation. In addition

to mRNA deadenylation, such negative regulatory mechanisms could be mediated by the interaction of transcripts with RNA-binding proteins in a nonspecific or sequence-specific fashion, and/or potentially via actions of microRNA and repeat- associated small interfering RNA, which degrade maternal RNA transcripts. It is this possible that in the absence of TBP2 most of the upregulated transcripts are those that escape such negative regulation.

LTR retrotransposons, also known as ERVs, constitute ~10% of the mouse genome (reviewed in (Crichton et al., 2014)). While their expression is generally suppressed by DNA methylation and/or repressive histone modifications, a subset of ERV subfamilies retain transcriptional activity in specific cell types ((Faulkner et al., 2009). ERVs are especially active in germ cells and early embryos (reviewed in (Thompson et al., 2016)). Indeed, many genome-wide many transcripts are initiated in LTRs, such as for example of MaLRs in mouse oocytes, which constitute ~5% of the genome (Consortium, 2002; McCarthy and McDonald, 2004). Members of the MT subfamily of MaLRs are particularly active in oocytes and hundreds of MT LTRs have been co-opted as oocyte-specific gene promoters (Franke et al., 2017; Peaston et al., 2004). The fact that in the TBP2 ablated oocytes MaLR MTA_Mm, MTA_Mm-int and MT-int transcripts are four-fold down regulated in general strongly suggests that TBP2 can binding to many of these MaLR LTR-initiated transcription units to nucleate the imitation of Pol II transcription. However, further experimental and bioinformatics investigations would be required to understand the mechanisms by which TBP2 is able to regulate these LTR-initiated transcription units. As LTR-initiated transcription units shape also the oocyte methylome, it will be important to analyse also how TBP2 influences DNA methylation in oocytes.

References

Akhtar, W., and Veenstra, G.J.C. (2009). TBP2 is a substitute for TBP in Xenopus oocyte transcription. BMC Biol. *7*, 45.

Anders, S., McCarthy, D.J., Chen, Y., Okoniewski, M., Smyth, G.K., Huber, W., and Robinson, M.D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc *8*, 1765–1786.

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

Bardot, P., Vincent, S.D., Fournier, M., Hubaud, A., Joint, M., Tora, L., and Pourquié, O. (2017). The TAF10-containing TFIID and SAGA transcriptional complexes are dispensable for early somitogenesis in the mouse embryo. Development *144*, 3808–3818.

Bártfai, R., Balduf, C., Hilton, T., Rathmann, Y., Hadzhiev, Y., Tora, L., Orbán, L., and Müller, F. (2004). TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. Curr Biol *14*, 593–598.

Bettegowda, A. (2007). Mechanisms of maternal mRNA regulation: implications for mammalian early embryonic development. Frontiers in Bioscience *12*, 3713.

Brou, C., Chaudhary, S., Davidson, I., Lutz, Y., Wu, J., Egly, J.M., Tora, L., and Chambon, P. (1993). Distinct TFIID complexes mediate the effect of different transcriptional activators. Embo J *12*, 489–499.

Choi, Y., and Rajkovic, A. (2006). Genetics of early mammalian folliculogenesis. Cell Mol Life Sci *63*, 579–590.

Consortium, M.G.S. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Crichton, J.H., Dunican, D.S., Maclennan, M., Meehan, R.R., and Adams, I.R. (2014). Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cell Mol Life Sci *71*, 1581–1605.

Crowley, T.E., Hoey, T., Liu, J.K., Jan, Y.N., Jan, L.Y., and Tjian, R. (1993). A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila. Nature *361*, 557–561.

Danino, Y.M., Even, D., Ideses, D., and Juven-Gershon, T. (2015). The core promoter: At the heart of gene expression. Biochim Biophys Acta *1849*, 1116–1131.

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

DeJong, J., Bernstein, R., and Roeder, R.G. (1995). Human general transcription factor TFIIA: characterization of a cDNA encoding the small subunit and requirement for basal and activated transcription. Proceedings of the National Academy of Sciences *92*, 3313–3317.

Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., and Torres-Padilla, M.-E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nat Struct Mol Biol *20*, 332–338.

Falender, A.E., Shimada, M., Lo, Y.K., and Richards, J.S. (2005). TAF4b, a TBP associated factor, is required for oocyte development and function. Developmental Biology *288*, 405–419.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. Nat Genet *41*, 563–571.

Ferg, M., Sanges, R., Gehrig, J., Kiss, J., Bauer, M., Lovas, A., Szabo, M., Yang, L., Straehle, U., Pankratz, M.J., et al. (2007). The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. Embo J *26*, 3945–3956.

Franke, V., Ganesh, S., Karlic, R., Malik, R., Pasulka, J., Horvat, F., Kuzman, M., Fulka, H., Cernohorska, M., Urbanova, J., et al. (2017). Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome Res. *27*, 1384–1394.

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

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

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

Goodrich, J.A., and Tjian, R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet *11*, 549–558.

Hansen, S.K., Takada, S., Jacobson, R.H., Lis, J.T., and Tjian, R. (1997). Transcription properties of a cell type-specific TATA-binding protein, TRF. Cell *91*, 71–83.

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

Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? Genes Dev *7*, 1291–1308.

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

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44–57.

Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F.A., and Wheeler, T.J. (2016). The Dfam database of repetitive DNA families. Nucleic Acids Res *44*, D81–D89.

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

Indra, A.K., Mohan, W.S., Frontini, M., Scheer, E., Messaddeq, N., Metzger, D., and Tora, L. (2005). TAF10 is required for the establishment of skin barrier function in foetal, but not in adult mouse epidermis. Developmental Biology *285*, 28–37.

Jacobi, U.G., Akkers, R.C., Pierson, E.S., Weeks, D.L., Dagle, J.M., and Veenstra, G.J.C. (2007). TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation. Embo J *26*, 3900–3909.

Jagarlamudi, K., and Rajkovic, A. (2012). Oogenesis: Transcriptional regulators and mouse models. Mol Cell Endocrinol *356*, 31–39.

Jallow, Z., Jacobi, U.G., Weeks, D.L., Dawid, I.B., and Veenstra, G.J.C. (2004). Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in Xenopus. Proc Natl Acad Sci USA *101*, 13525–13530.

Kaltenbach, L., Horner, M.A., Rothman, J.H., and Mango, S.E. (2000). The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during C. elegans embryogenesis. Mol Cell *6*, 705–713.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

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

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lewandoski, M., Wassarman, K.M., and Martin, G.R. (1997). Zp3–cre, a transgenic mouse line for the activation or inactivation of loxP-flanked target genes specifically in the female germ line. Current Biology *7*, 148–151.

Li, R., and Albertini, D.F. (2013). The road to maturation: somatic cell interaction and self-organization of the mammalian oocyte. Nat Rev Mol Cell Biol *14*, 141–152.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 31.

Ma, D., Watanabe, H., Mermelstein, F., Admon, A., Oguri, K., Sun, X., Wada, T., Imai, T., Shiroya, T., and Reinberg, D. (1993). Isolation of a cDNA encoding the largest subunit of TFIIA reveals functions important for activated transcription. Genes Dev *7*, 2246–2257.

Marshall, O.J., Southall, T.D., Cheetham, S.W., and Brand, A.H. (2016). Cell-type-specific profiling of protein-DNA interactions without cell isolation using targeted DamID with next-generation sequencing. Nat Protoc *11*, 1586–1598.

Martianov, I., Fimia, G.M., Dierich, A., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2001). Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. Mol Cell *7*, 509–515.

Martianov, I., Brancorsini, S., Gansmuller, A., Parvinen, M., Davidson, I., and Sassone-Corsi, P. (2002a). Distinct functions of TBP and TLF/TRF2 during spermatogenesis: requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. Development *129*, 945–955.

Martianov, I., Viville, S., and Davidson, I. (2002b). RNA polymerase II transcription in murine cells lacking the TATA binding protein. Science *298*, 1036–1039.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

McCarthy, E.M., and McDonald, J.F. (2004). Long terminal repeat retrotransposons of Mus musculus. Genome Biol. *5*, R14.

Mohan, W.S., Scheer, E., Wendling, O., Metzger, D., and Tora, L. (2003). TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice. Mol Cell Biol *23*, 4307–4318.

Müller, F., Lakatos, L., Dantonel, J., Strähle, U., and Tora, L. (2001). TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. Curr Biol *11*, 282–287.

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

Nagy, Z., Riss, A., 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. Cell Mol Life Sci *67*, 611–628.

Ohbayashi, T., Makino, Y., and Tamura, T.A. (1999). Identification of a mouse TBP-like protein (TLP) distantly related to the drosophila TBP-related factor. Nucleic Acids Res *27*, 750–755.

Oyama, T., Sasagawa, S., Takeda, S., Hess, R.A., Lieberman, P.M., Cheng, E.H., and Hsieh, J.J. (2013). Cleavage of TFIIA by Taspase1 activates TRF2-specified mammalian male germ cell programs. Dev Cell *27*, 188–200.

Ozer, J., Moore, P.A., Bolden, A.H., Lee, A., Rosen, C.A., and Lieberman, P.M. (1994). Molecular cloning of the small (gamma) subunit of human TFIIA reveals functions critical for activated transcription. Genes Dev *8*, 2324–2335.

Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell *7*, 597–606.

Pedersen, T., and Peters, H. (1968). Proposal for a classification of oocytes and follicles in the mouse ovary. J. Reprod. Fertil. *17*, 555–557.

Persengiev, S.P., Zhu, X., Dixit, B.L., Maston, G.A., Kittler, E.L.W., and Green, M.R. (2003). TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. Proc Natl Acad Sci USA *100*, 14887–14891.

Rabenstein, M.D., Zhou, S., Lis, J.T., and Tjian, R. (1999). TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. Proceedings of the National Academy of Sciences *96*, 4791–4796.

Schmidt, E.E., and Schibler, U. (1997). Developmental testis-specific regulation of mRNA levels and mRNA translational efficiencies for TATA-binding protein mRNA isoforms. Developmental Biology *184*, 138–149.

Skene, P.J., Henikoff, J.G., and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat Protoc *13*, 1006–1019.

Smit, A.F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev *9*, 657–663.

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

Tatarakis, A., Margaritis, T., Martinez-Jimenez, C.P., Kouskouti, A., Mohan, W.S., Haroniti, A., Kafetzopoulos, D., Tora, L., and Talianidis, I. (2008). Dominant and redundant functions of TFIID involved in the regulation of hepatic genes. Mol Cell *31*, 531–543.

Teichmann, M., Wang, Z., Martinez, E., Tjernberg, A., Zhang, D., Vollmer, F., Chait, B.T., and Roeder, R.G. (1999). Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. Proceedings of the National Academy of Sciences *96*, 13720–13725.

Thompson, P.J., Macfarlan, T.S., and Lorincz, M.C. (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. Mol Cell *62*, 766–776.

Timmers, H.T.M., and Tora, L. (2018). Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation. Mol Cell *72*, 10–17.

Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. Genes Dev *16*, 673–675.

Turatsinze, J.-V., Thomas-Chollier, M., Defrance, M., and van Helden, J. (2008). Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. Nat Protoc *3*, 1578–1588.

Upadhyaya, A.B., Lee, S.H., and DeJong, J. (1999). Identification of a general transcription factor TFIIAalpha/beta homolog selectively expressed in testis. J Biol Chem *274*, 18040–18048.

Van Dyke, M.W., Roeder, R.G., and Sawadogo, M. (1988). Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. Science *241*, 1335–1338.

Veenstra, G.J.C. (2000). Distinct Roles for TBP and TBP-Like Factor in Early Embryonic Gene Transcription in Xenopus. Science *290*, 2312–2315.

Walser, C.B., and Lipshitz, H.D. (2011). Transcript clearance during the maternal-to-zygotic transition. Curr Opin Genet Dev *21*, 431–443.

Wang, Y.L., Duttke, S.H.C., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J., and Kadonaga, J.T. (2014). TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes Dev *28*, 1550–1555.

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. J Biol Chem *273*, 12492–12498.

Xiao, L., Kim, M., and DeJong, J. (2006). Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice. Gene Expression Patterns *6*, 409–419.

Yu, C., Ji, S.-Y., Sha, Q.-Q., Dang, Y., Zhou, J.-J., Zhang, Y.-L., Liu, Y., Wang, Z.-W., Hu, B., Sun, S.-C., et al. (2016). BTG4 is a meiotic cell cycle–coupled maternal-zygotic-transition licensing factor in oocytes. Nat Struct Mol Biol *23*, 387–394.

Zhou, H., Spicuglia, S., Hsieh, J.J.-D., Mitsiou, D.J., Høiby, T., Veenstra, G.J.C., Korsmeyer, S.J., and Stunnenberg, H.G. (2006). Uncleaved TFIIA is a substrate for taspase 1 and active in transcription. Mol Cell Biol *26*, 2728–2735.

Zybailov, B., Mosley, A.L., Sardiu, M.E., Coleman, M.K., Florens, L., and Washburn, M.P. (2006). Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J. Proteome Res. *5*, 2339–2347.

Material and Methods

Animal experimentation

Animal experimentations were carried out according to animal welfare regulations and guidelines of the French Ministry of Agriculture and procedures were approved by the French Ministry for Higher Education and Research ethical committee C2EA-17 (project n°2018031209153651). The Tg(*Zp3-Cre*), *Taf7*^{flox} and *Tbp2*$^{-}$ lines have already been described (Gazdag et al., 2009; Gegonne et al., 2012; Lewandoski et al., 1997).

Histology analyses of ovaries

Ovaries were collected from 6 weeks-old Tg(*Zp3-Cre*/+);*Taf7*$^{flox/+}$ and Tg(*Zp3-Cre*/+);*Taf7*$^{flox/\Delta}$ oocyte specific mutant females, fixed in 4% paraformaldehyde (Electron Microscopy Sciences) over-night at 4°C, washed 3 times in PBS at room temperature and embedded in paraffin. Five µm-thick sections were stained with hematoxylin and eosin and images were acquired using a slide scanner Nanozoomer 2.0HT (Hamamatsu Photonics).

Supervovulation

5 U of pregnant mare serum (PMS) was injected intraperitoneally in 4-week-old female mice between 2-4 pm. After 44-46 hours, GV oocytes were collected from the ovaries by puncturing with needles.

Oocytes collection

After dissection, ovaries are freed from adhering tissues in &x PBS. A pool od 6 ovaries was digested in 500 µL of 2 mg/mL Collagenase, 0.025% Trypsin and 0.5 mg/mL type IV-S hyaluronidase, on a thermomixer at 600 rpm for 20 minutes. The digestion was then stopped

by the addition of 1 mL of 37°C pre-warmed αMEM - 5% FBS. The oocytes were then size-selected under a binocular.

Cell lines and cell culture

The NIH3T3-II10 line overexpressing TBP2 and the control NIH3T3-K2 have already been described (Gazdag et al., 2007) and were maintained in high glucose DMEM supplemented with 10% of newborn calf serum at 37°C in 5% CO_2.

Whole cell extracts

Cells cultured in 15 cm dish were washed twice with 1x PBS, subsequently harvested by scrapping on ice. Harvested cells were centrifuged 1000 rcf at 4°C for 5min and then resuspended in 1 packed cell volume of whole cell extraction buffer (20 mM Tris HCl pH7.5, 2 mM DTT, 20% Glycerol, 400 mM KCl, 1x Protease inhibitor cocktail (PIC, Roche)). Cell lysates were frozen in liquid nitrogen and thawed on ice for 3 times, followed by centrifugation at 20817 rcf, at 4°C for 15 min. The supernatant was collected and protein concentration was measured by Bradford protein assay. The cell extracts were used directly for IP and western blot, or stored at −80°C.

Ovaries collected from P14 CD1 female mice were homogenized in whole cell extraction buffer (20 mM Tris HCl pH7.5, 2 mM DTT, 20% Glycerol, 400 mM KCl, 5x PIC (Roche)). Cell lysates were frozen in liquid nitrogen and thawed on ice for 3 times, followed by centrifugation at 20817 rc, at 4°C for 15 min. The supernatant extracts were used directly for immunoprecipitations or gel filtration.

Antibodies and antibody purification

The 2B12 anti-TBP2, the 3TF13G3 anti-TBP and the 15TF21D10 anti-GST mouse monoclonal antibodies have already been described (Brou et al., 1993; Gazdag et al., 2007; Nagy et al., 2010). The J7 rabbit polyclonal anti TFIIA was a gift of H.G. Stunnenberg. The IGBMC antibody facility generated several anti-TBP2 polyclonal antibodies with peptides (2481, 2482, 3017, 3018, 3019, 3022, 3023 and 3024) or with purified protein (3498 and 3499) (see Supplementary Table 5). The different peptides were synthetized and coupled to ovalbumin and injected into 2 or 3 rabbits and the first 140 amino-acids of the mouse TBP2 fused to a His tag were produced in BL21DE3 bacteria and purified with Ni-NTA beads (Qiagen) and injected in 2 rabbits. The resulting sera were affinity purified by using the Sulfolink Coupling Gel (Pierce) following the manufacturer's recommendations.

Gel filtration

A Superose 6 (10/300) column was equilibrated with buffer consisting of 25mM Tris HCl pH 7.9, 5 mM $MgCl_2$, 150 mM KCl, 5% Glycerol, 1 mM DTT and 1x PIC (Roche). Five hundred µL of whole cell extracts containing ~5 mg of protein were injected in an Akta Avant chromatography device and ran at 0.4 mL per min. Protein detection was performed by absorbance at 280nm and 260nm. Five hundred µL fractions were collected.

Immunoprecipitation

Immunoprecipiation were carried out as already described in (Bardot et al., 2017).

Western blot

Protein samples (15-25 µg of cell extracts or 15 µL of IP elution) were mixed with 1/4th volume of loading buffer (100 mM TrisHCl pH 6.8, 30% glycerol, 4% SDS, 0.2% bromophenol blue and freshly added 100 mM DTT) and boiled for 10 min. Samples were then resolved on

a10 % SDS-PAGE and transferred to nitrocellulose membrane (Protran, Amersham). Membranes were blocked in 3% non-fat milk in 1x PBS at room temperature (RT) for 30 min, and subsequently incubated with the primary antibody overnight at 4°C. Membranes were washed three times (10 min each) with 1x PBS - 0.05% Tween20. Membranes were then incubated with HRP-coupled secondary antibodies for 1 hour at RT, followed by ECL detection (Thermo Fisher). The signal was acquired with the Chemidoc imaging system (Bio-Rad).

RNA preparation

Oocytes collected were washed through several M2 drops, and total RNA was isolated using NucleoSpin RNAXS kit (Macherey-Nagel) according to the user manual. RNA quality and quantity were evaluated using a Bioanalyzer.

Mass spectrometry analyzes and NSAF calculations

Samples were TCA precipitated, reduced, alkylated and digested with LysC and Trypsin at 37°C overnight. After C18 desalting, samples were analyzed using an Ultimate 3000 nano-RSLC (Thermo Scientific, San Jose, California) coupled in line with a linear trap Quadrupole (LTQ)-Orbitrap ELITE mass spectrometer via a nano-electrospray ionization source (Thermo Scientific). Peptide mixtures were loaded on a C18 Acclaim PepMap100 trap column (75 µm inner diameter × 2 cm, 3 µm, 100 Å; Thermo Fisher Scientific) for 3.5 min at 5 µl/min with 2% acetonitrile (ACN), 0.1% formic acid in H_2O and then separated on a C18 Accucore nano-column (75 µm inner diameter × 50 cm, 2.6 µm, 150 Å; Thermo Fisher Scientific) with a 240 minutes linear gradient from 5% to 50% buffer B (A: 0.1% FA in H_2O / B: 80% ACN, 0.08% FA in H_2O) followed with 10 min at 99% B. The total duration was set to 280 minutes at a flow rate of 200 nL/min.

Proteins were identified by database searching using SequestHT (Thermo Fisher Scientific) with Proteome Discoverer 1.4 software (Thermo Fisher Scientific) a combined Mus musculus database (Swissprot, release 2015_11, 16730 entries) where 5 interesting proteins sequences (TrEMBL entries) were added. Precursor and fragment mass tolerances were set at 7 ppm and 0.5 Da respectively, and up to 2 missed cleavages were allowed. Oxidation (M) was set as variable modification, and Carbamidomethylation © as fixed modification. Peptides were filtered with a false discovery rate (FDR) at 5 %, rank 1 and proteins were identified with 1 unique peptide.

Normalized spectral abundance factor (NSAF) (Zybailov et al., 2006) normalized to the bait ($NSAF_{bait}$) were obtained as followed (PSM*; peptide spectrum match, SAF; spectral abundance factor, x; protein of interest) using a R script available on request (R software, version 3.5.1):

$$SAF_{(x)} = \frac{PSM^{*}_{x(IP)} - PSM^{*}_{x(IPmock)}}{length(x)}$$

$$NSAF_{(x)} = \frac{SAF_{(x)}}{\sum_{i=1}^{n} SAF_{(xi)}} \times 100$$

$$NSAF_{bait(x)} = \frac{NSAF_{(x)}}{NSAF_{(bait)}}$$

All the figures were generated using R software version 3.5.1.

RNA-seq analyses

PolyA+ RNA seq libraries were prepared using the SMART-Seq v4 UltraLow Input RNA kit (Clonetch) followed by the Nextera XT DNA library Prep kit (Illumina) according to the manufacturer recommendations from 3 replicates for each conditions (wild-type P7 (SNVT9, SNVT10, SNVT11), P7 mutant (SNVT13, SNVT14, SNVT15), wild-type P14: SNVT1, SNVT2, SNVT4), P14 mutant (SNVT5, SNVT6, SNVT7)) and sequenced 50 pb single end on

an Illumina HiSeq 4000. One sample per condition (SNVT1, SNVT6, SNVT10 and SNVT14) were also resequenced 100 pb paired end.

Reads were preprocessed in order to remove adapter, polyA and low-quality sequences (Phred quality score below 20). After this preprocessing, reads shorter than 40 bases were discarded for further analysis. These preprocessing steps were performed using cutadapt version 1.10 (Martin, 2011). Reads were mapped to spike sequences using bowtie version 2.2.8 (Langmead and Salzberg, 2012), and reads mapping to spike sequences were removed for further analysis. Reads were then mapped onto the mm10 assembly of *Mus musculus* genome using TopHat2 version 2.0.1.4 (Kim et al., 2013). Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.6.1p1 (Anders et al., 2015), with annotations from Ensembl version 93 and “union" mode. Read counts were normalized across samples with the median-of-ratios method (Anders et al., 2013), to make these counts comparable between samples and differential gene analysis were performed using the DESeq2 version 1.16.1 (Love et al., 2014). All the figures were generated using R software version 3.5.1.

Repeat element analyses

For the repeat element analyses, data were processed as already described (Fadloun et al., 2013) using Bowtie1 (Langmead et al., 2009) instead of Maq. The repeatMasker annotation was used to identified the different types of repeat elements (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013-2015 http://www.repeatmasker.org). All the figures were generated using R software version 3.5.1.

Core promoter motif analyses

Sequences were downloaded from the Ensembl data base using EnsDb.Mmusculus.v79 R package version 2.99.0 (Rainer J (2017) *EnsDb.Mmusculus.v79: Ensembl based annotation*

package. R package version 2.99.0.). All the -50/+50 genomic sequences across the TSS of all transcripts were considered in this analysis and 3 set of sequences were obtained: 1/ the sequences for the common genes that were expressed above 100 reads/kb median length of transcript and were downregulated for a log2 fold change below -1 at P7 and P14 (963 genes present in the annotation), 2/ all the genes present in the annotation (43309 genes) and 3/ all the genes confiding for protein present in the annotation (21989). A random sample of 18000 100 nucleotides sequences was also generated as a control using RSAT (Turatsinze et al., 2008). We used the following consensus sequences: TATA box; TATAWAAR, Inr; YYANWYY, MTE; CSARCSSAAC, DPE; GNNNDSWYVY, XCPE1; DSGYGGRAS and XCPE2; VCYCRTTRCMY (IUPAC nomenclature) (Danino et al., 2015)to design Transfac format matrix for each element. We used the matrix scan module of RSAT to detect the presence of these different element in the 4 different sets of sequences using the following parameters: background model estimation method: Markov order 1, organism specific Mus musculus GRCm38 upstream-noorf, scanning options: single strand. The number of motifs detected per gene was then evaluated by a R script. All the figures were generated using R software version 3.5.1.

Acknowledgements

We thank D. Singer and A. Gegonne for the gift of the *Taf7*^{flox} mouse line and H. Stunnenberg for TFIIA antibodies. We would also like to thank all members of the Tora lab for thoughtful discussions and suggestions throughout the course of the work. We are grateful to M.E. Torres Padilla for advice, preliminary experiments, C. Hérouard from the GenomEAST platform for library preparation, P. Eberling for the generation of the peptides, G. Duval for the generation of the polyclonal sera, V. Hisler for help with analyses, the histology platform, the IGBMC cell culture

facility and S. Falcone, M. Poirot and F. Memedov of the IGBMC animal facility for the animal care taking.

Funding

This work was supported by funds from CNRS, INSERM, and Strasbourg University. This study was also supported by the European Research Council (ERC) Advanced grant (ERC-2013-340551, Birtoaction) and grant ANR-10-LABX-0030-INRT (to LT), a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02.

Authors' contribution

CY, SDV and LT designed the study; SDV and LT supervised the work; CY performed all the molecular lab and mouse experiments, EG generated some of the anti-TBP2 polyclonal antibodies and generated the NIH3T3-II10 and K2 cell lines, MJung and SDV analysed the RNA-seq data. MJoint, LN carried out the proteomic analyses which were analysed by SDV. All authors contributed to the text and figure panels. CY, SDV and LT wrote the manuscript. All authors gave final approval for publication.

Competing interests

The authors declare that they have no competing interests.

References

control oocytes	*Taf7* mutant oocytes
Tg(*Zp3-Cre*/+);*Taf7*$^{flox/+}$)	Tg(*Zp3-Cre*/+);*Taf7*$^{flox/\Delta}$
126 oocytes / 4 females	195 oocytes / 6 females
31.5 oocytes/female	32.5 oocytes/female

Table 1: *Taf7* deletion in oocytes does not impair the production of GV oocytes.

Control (Tg(*Zp3-Cre*/+);*Taf7*$^{flox/+}$) and mutant (Tg(*Zp3-Cre*/+);*Taf7*$^{flox/\Delta}$) females were stimulated for oocyte maturation and GV oocytes were counted. Mann and Whitney statistical analysis did not detect a significant difference between the control and the mutant data.

Figure legends

Figure 1: Protein coding genes expression is affected in *Tbp2*$^{-/-}$ mutant oocytes. A-B. Genomic views of the *Tbp2* (A) and *Ooep* (B) genes whose expression is reduced in *Tbp2*$^{-/-}$ mutant oocytes. As expected no reads map to the exon 4 of *Tbp2* (gray dashed box) which is deleted in *Tbp2*$^{-/-}$ mutant mice (A). Data were derived from three independent biological replicates sequenced in single ends.

Figure 2: TBP2 is responsible for the expression of the majority of the genes expressed in oocytes at post-natal days 7 and 14. A. Principal component analysis (PCA) after regularized log transformation (rlog) of all data. **B.** Hierarchical clustering using Euclidean distance on rlog transformed data indicating that the main explanation for the variance is the genotype, and not the stage. **C-D.** Post-natal day (P) 7 (C) and P14 (D) normalized expression to the media size of the transcript in kb comparison between wild-type (WT) and *Tbp2*$^{-/-}$ mutant oocytes. Gray dots correspond to all genes, blue dots to significant genes for a p value ≤ 0.05 and dark blue dots to significant genes for a p value ≤ 0.05 and an absolute log2 fold change above 1, after Cook's distance Wald test and Benjamini-Hochberg correction (DEseq2). The number of up- or down-regulated genes is indicated on the graphs. **E.** Venn diagrams of the up- (above) and down-(below) regulated genes corresponding to the dark blue dots in (C and D). **F.** P7 versus P14 fold change comparison of the colour-scaled expression normalized to the media size of the transcript in kb indicating that loss of TBP2 has a major effect on the expression of the most abundant genes in oocytes at P7 and P14.

Figure 3: Affected expression of genes related to the RNA decay pathway at post-natal days 7 and 14 in growing *Tbp2*$^{-/-}$ mutant oocytes. A. Gene ontology annotation of molecular function for an enrichment fold above 1.5 and a Benjamini-Hochberg p value ≤ 0.05. **B.** P7

versus P14 fold change comparison of the expression of genes involved in the RNA decay pathway. The colour code indicated the different RNA destabilizing activities. The most down-regulated genes are indicated on the graph.

Figure 4: Analysis of the core promoter elements within the promoter of downregulated genes in the *Tbp2*$^{-/-}$ mutant oocytes. A-F. Comparison of the percentage of predicted TATA boxes (A), Inr element (B), MTE (C), DPE(D), XCPE1 (E) and XCPE2 (F) within the -50/+50 gene sequence using RSAT pattern motif search. The proportion of the different predicted core promoter elements between the coding genes (CDS) and the genes downregulated at P7 and P14 (down) was compared using Pearson's Chi^2 test with Yate's continuity correction. down; common down-regulated genes in *Tbp2*$^{-/-}$ mutant oocytes (adjusted p value < 0.05 and log2 fold change ≤ -1, 963 genes in the annotation), CDS; genes annotated as protein coding genes (21989 genes), all; all annotated genes (43309 genes), random, 18000 100 nucleotides random sequences generated using RSAT., ns; non-significant, *; p value ≤0.05, ***; p value ≤0.001.

Figure 5: High expression of repeat elements in oocytes and specific down regulation of the MaLR subfamily in *Tbp2*$^{-/-}$ mutant oocytes. A. Genomic view of the *Tcstv3* gene whose expression is reduced in *Tbp2*$^{-/-}$ mutant oocytes. Reads accumulation outside of the *Tcstv3* gene corresponds to the MTA_Mm_dup703 repeat element after comparison with the RepMasker database. **B.** Differential expression of the major classes of repeat elements in wild-type (WT) and *Tbp2*$^{-/-}$ mutant oocytes at post-natal day (P) 7 and P14. The LTR retrotransposon class is the most abundant at P7 and P14 and is affected in the *Tbp2*$^{-/-}$ mutant oocytes. **C.** Differential expression of the different LTR retrotransposon superfamilies indicated that the sub class-III is the most severely affected in *Tbp2*$^{-/-}$ mutant oocytes at P7 and P14. **D.** Differential expression of the 2 families constituting the class III LTR: ERVL (endogenous retrovirus type L) and

MaLR (mammalian apparent LTR retrotransposons) showing that while ERVL expression is not affected, expression of the MaLR family is downregulated in *Tbp2*$^{-/-}$ mutant oocytes. **E.** P7 versus P14 foldchange comparison for the different members of the MaLR family members indicating that the expression of the most abundant members is affected in *Tbp2*$^{-/-}$ mutant oocytes. The gray scale corresponds to the P7 wild-type read numbers. **F.** Differential expression of the 3 most affected MaLR repeats. Data were derived from three independent biological replicates sequenced in single ends and error bars indicate standard deviation.

Figure 6: TBP2 is interacting with TFIIA in 3T3 cells overexpressing TBP2. A. Comparison of immunoprecipitated (IP-ed) protein from II10 cells whole cell extracts (WCE) using a TBP antibody and one of our specific mouse anti TBP2 antibody indicating that compared to TBP, TBP2 is strongly interacting with TFIIA but is able to interact with some TAF subunits. **B.** Western blot of a Superose 6 gel filtration analysis of NIH3T3-II10 WCE indicating that TBP2 is associated within a complex with TFIIA-α but does not co-purify with TAF6. **C.** LC/LC-MS analysis of the fraction highlighted in red (B) compared to an anti-TBP2-IP from II10 WCE, confirming that TFIIA, but no TAF proteins, is detected in the fraction. LC/LC-MS data were derived from three technical replicates. NSAF; normalized spectral abundance factor, error bars indicate sem.

Figure 7: TBP2 is interacting with TFIIA in the oocytes. A-B. TBP2 (A) and TBP(B) immunoprecipitation (IP) from whole cell extract (WCE) of post-natal day 14 CD1 mouse ovaries. **C.** Scheme of the sequential IP experiment. **D-E.** Comparison of the first TAF7 IP and second TAF10 IP (D) showing that the TAF7 IP efficiently IP-ed the TFIID complex, whereas the TAF10 IP efficiently IP-ed the SAGA complex. Comparison of the second TAF10 IP and the third TBP2 IP (E magnification in F) indicates that TBP2 strongly interacts with the TFIIA

complex. Data were derived from three technical replicates. NSAF; normalized spectral abundance factor, error bars indicate sem.

Figure 8: Conditional deletion of *Taf7* in growing oocytes does not affect oocyte growth. **A.** Scheme of the breeding and experiments. **B-E.** Hematoxylin and eosin stained ovaries section from wild-type (B,D; Tg(*Zp3-Cre*/+);*Taf7*$^{flox/+}$) and oocyte-specific *Taf7* mutant (C,E; Tg(*Zp3-Cre*/+);*Taf7*$^{flox/\Delta}$) ovaries. (D) and (E) are magnifications of (B) and (C), respectively of antral follicle. Scale bars: 500 µm in B,C; 100 µm in D,E.

Supplementary Figure legends

Supplementary Figure 1: RNA-seq analyses of wild-type versus *Tbp2*$^{-/-}$ mutant oocytes at P7 and P14. A. Pearson correlation and Euclidean distance clustering of the 12 sets of data **B-E.** MA plots (B,C) and Volcano plots (D,E) at post-natal day 7 (P7) (B,D) and P14 (C,E). The color scale indicates the significance of the Wald test after Benjamini-Hochberg correction (DESeq2). NA correspond to the genes with high Cook's distance that were filtered out (DESeq2). Data were derived from three independent biological replicates for each stage; post-natal day(P) 7 and P14, and each genotype; wild-type and *Tbp2*$^{-/-}$ mutant).

Supplementary Figure 2: High expression of repeat elements in oocytes and specific down regulation of the MaLR subfamily in *Tbp2*$^{-/-}$ mutant oocytes. A-C. Differential expression (read number) of the major classes of the different LTR retrotransposon superfamilies (A), of the different class III family members (B) and of the 3 most affected MaLR repeats (C), in wild-type (WT) and *Tbp2*$^{-/-}$ mutant oocytes at post-natal day (P) 7 and P14. Data were derived from three independent biological replicates sequenced in single ends. ERVL; endogenous retrovirus type L, MaLR; mammalian apparent LTR retrotransposons. Error bars indicate standard deviation.

Supplementary Figure 3: High expression of repeat elements in oocytes and specific down regulation of the MaLR subfamily in *Tbp2*$^{-/-}$ mutant oocytes. A. Relative abundance of transcript of the major classes of repeat elements in wild-type (WT) and *Tbp2*$^{-/-}$ mutant oocytes at post-natal day (P) 7 and P14. Data derived from a single biological sample sequenced in paired ends. The LTR retrotransposon class is the most abundant at P7 and P14 and is affected in the *Tbp2*$^{-/-}$ mutant oocytes. **B-C.** Differential expression (number of reads (B) and log2 fold

change (C) of the different LTR retrotransposon superfamilies indicated that the sub class-III is the most severely affected in *Tbp2*$^{-/-}$ mutant oocytes at P7 and P14. **D-E.** Differential expression (number of reads (D) and log2 fold change (E) of the 2 families constituting the class III LTR: ERVL (endogenous retrovirus type L) and MaLR (mammalian apparent LTR retrotransposons) showing that while ERVL expression is not affected, expression of the MaLR family is downregulated in *Tbp2*$^{-/-}$ mutant oocytes. **F.** P7 versus P14 fold-change comparison for the different members of the MaLR family members indicating that the expression of the most abundant members is affected in *Tbp2*$^{-/-}$ mutant oocytes. The gray scale corresponds to the P7 wild-type read numbers. **E.** Differential expression (number of reads (G) and log2 fold change (H)) of the 3 most affected MaLR repeats.

Supplementary Figure 4: Testing of the anti-TBP2 polyclonal antibodies. A. Map of the mouse TBP2 protein with the core DNA domain in blue. The position of the different protein fragments used to generate the monoclonal (in blue) and the polyclonal (in red) antibodies are indicated. **B.** Western blot analyses of the different antibodies on the TBP2 overexpressing NIH3T3-II10 and control NIH3T3-K2 cells. The TBP2 protein is migrating with an apparent molecular weight of 50 kDa.

Supplementary Figure 5: Proteomic analyses of the TBP2 containing complex in a NIH3T3 line overexpressing TBP2. A. LC/LC-MS analysis of TBP2 immunoprecipitation (IP) using 3 different anti TBP2 antibodies (3024, 3499 and 2B12) from NIH3T3-II10 cells whole cell extract (WCE). **B.** LC/LC-MS analysis of TBP-IP from NIH3T3-II10 WCE. **C.** LC/LC-MS analysis of TBP2-IP of the gel filtration 23-25 fraction from NIH3T3-II10 WCE. The colored dots indicate the different complexes/proteins analyzed. Data were derived from three technical replicates. NSAF; normalized spectral abundance factor, error bars indicate sem.

Supplementary Figure 6: Proteomic analyses of the TBP2 containing complex in the ovary. **A-C.** Triple sequential immunoprecipitation (IP) of the TBP2associated complex from ovary whole cell extract (WCE). **A.** LC/LC-MS analysis of the TAF7-IP from P14 ovary whole cell extract (WCE). The red star indicates that this detection is not significant. **B.** LC/LC-MS analysis of the TAF10-IP from the TAF7-IP flow through. **C.** LC/LC-MS analysis of the TBP2-IP from the TAF10-IP flow through. The colored dots indicate the different complexes/proteins analyzed. Data were derived from three technical replicates. NSAF; normalized spectral abundance factor, error bars indicate sem.

Supplementary Table 1: Gene ontology analyses of the common down-regulated genes in *Tbp2*$^{-/-}$ mutant oocytes at P7 and P14.

Supplementary Table 2: Expression of mRNA decay-associated genes in *Tbp2*$^{-/-}$ mutant oocytes at P7 and P14.

Supplementary Table 3: Comparison between the 3 anti-TBP2 antibodies used for proteomic analyses of the TBP2-IP from NIH3T3-II10 cells.

Supplementary Table 4: Oocyte-specific RNA-seq data at P7 and P14.

Supplementary Table 5: List of the antibodies used.

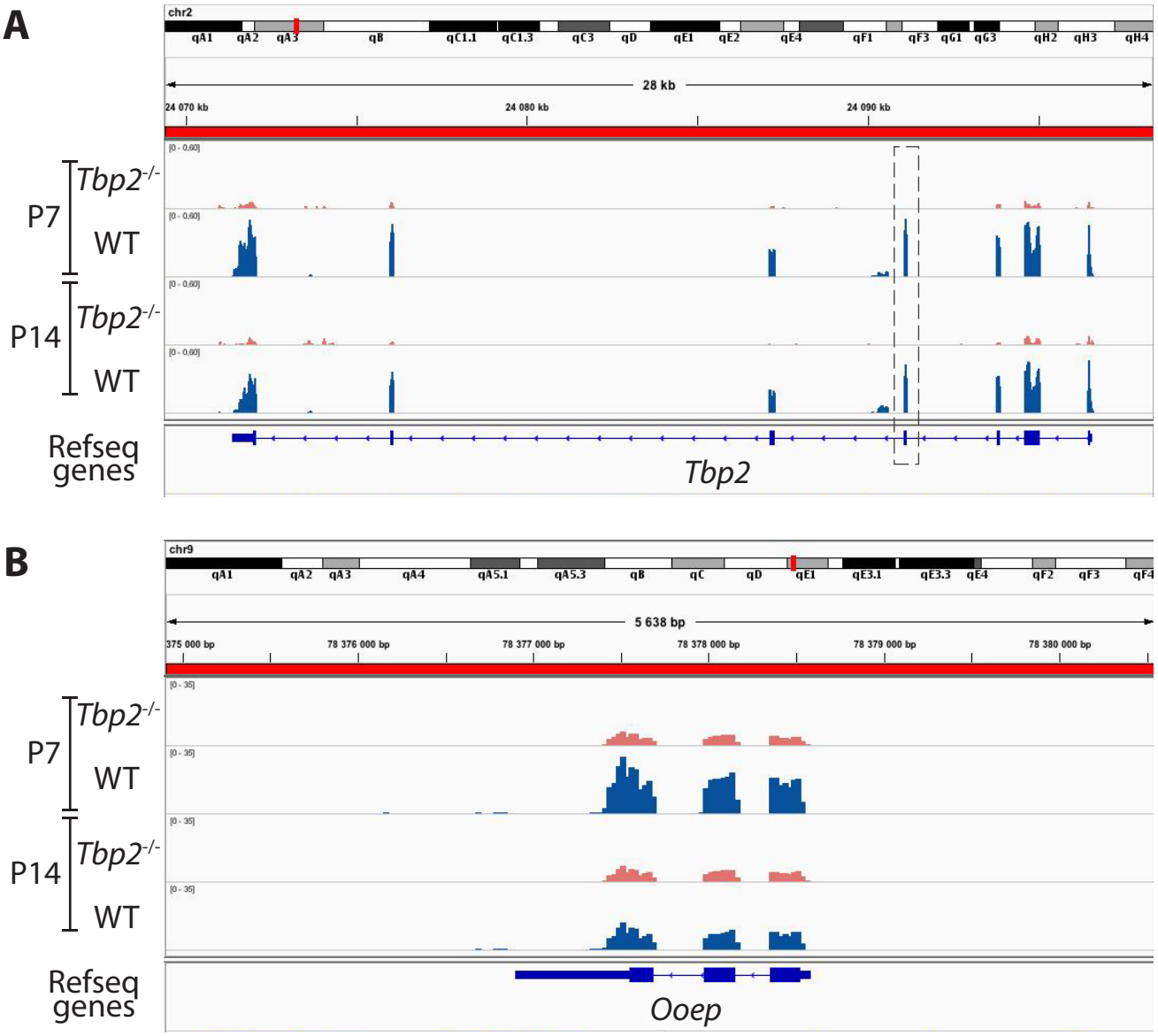

A
chr2
qA1 qA2 qA3 qB qC1.1 qC1.3 qC3 qD qE1 qE2 qE4 qF1 qF3 qG1 qG3 qH2 qH3 qH4
28 kb
24 070 kb
24 080 kb
24 090 kb
P7
Tbp2-/-
WT
P14
Tbp2-/-
WT
Refseq genes
Tbp2
B
chr9
qA1 qA2 qA3 qA4 qA5.1 qA5.3 qB qC qD qE1 qE3.1 qE3.3 qE4 qF2 qF3 qF4
5 638 bp
375 000 bp
78 376 000 bp
78 377 000 bp
78 378 000 bp
78 379 000 bp
78 380 000 bp
P7
Tbp2-/-
WT
P14
Tbp2-/-
WT
Refseq genes
Ooep

Figure 1

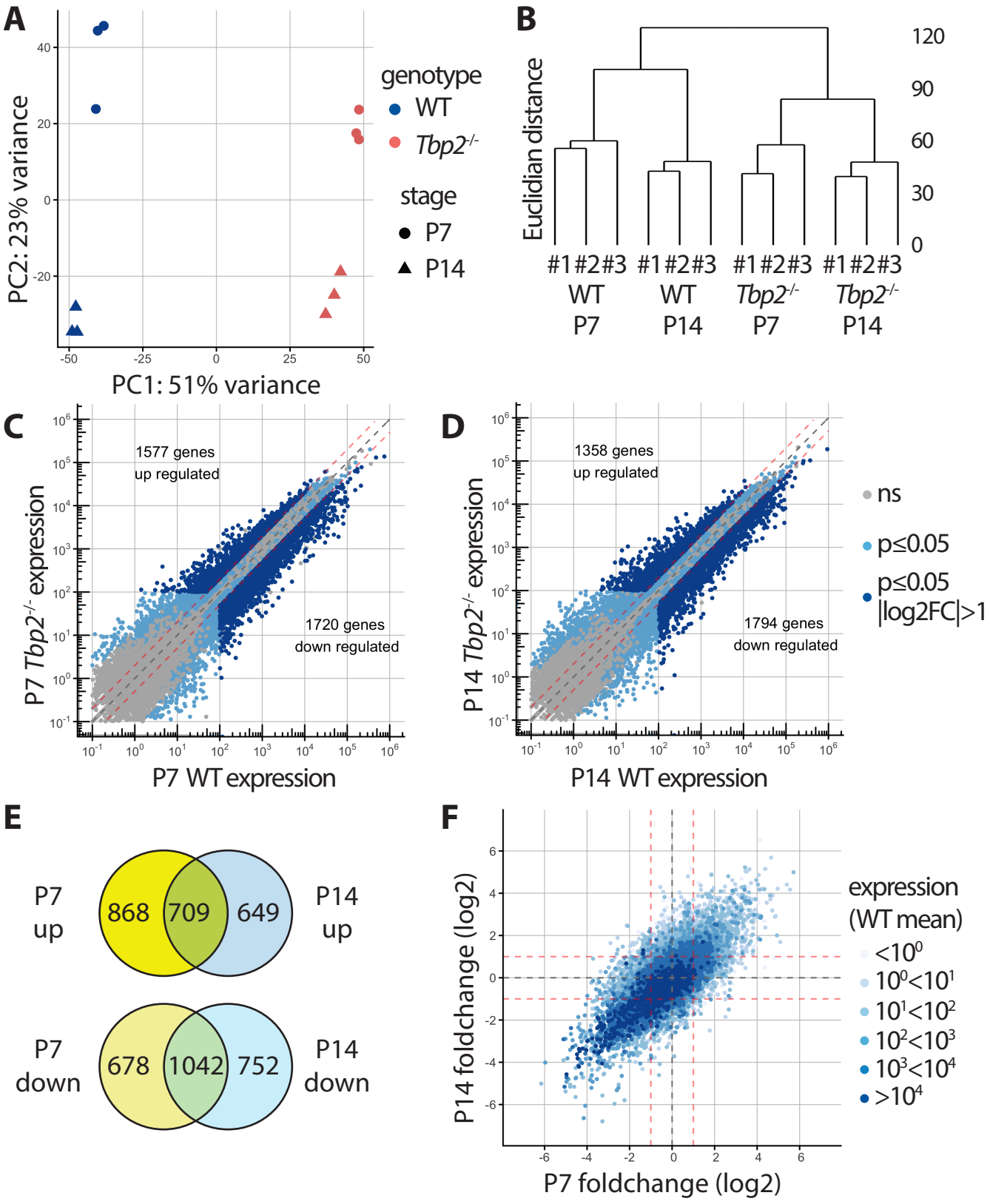
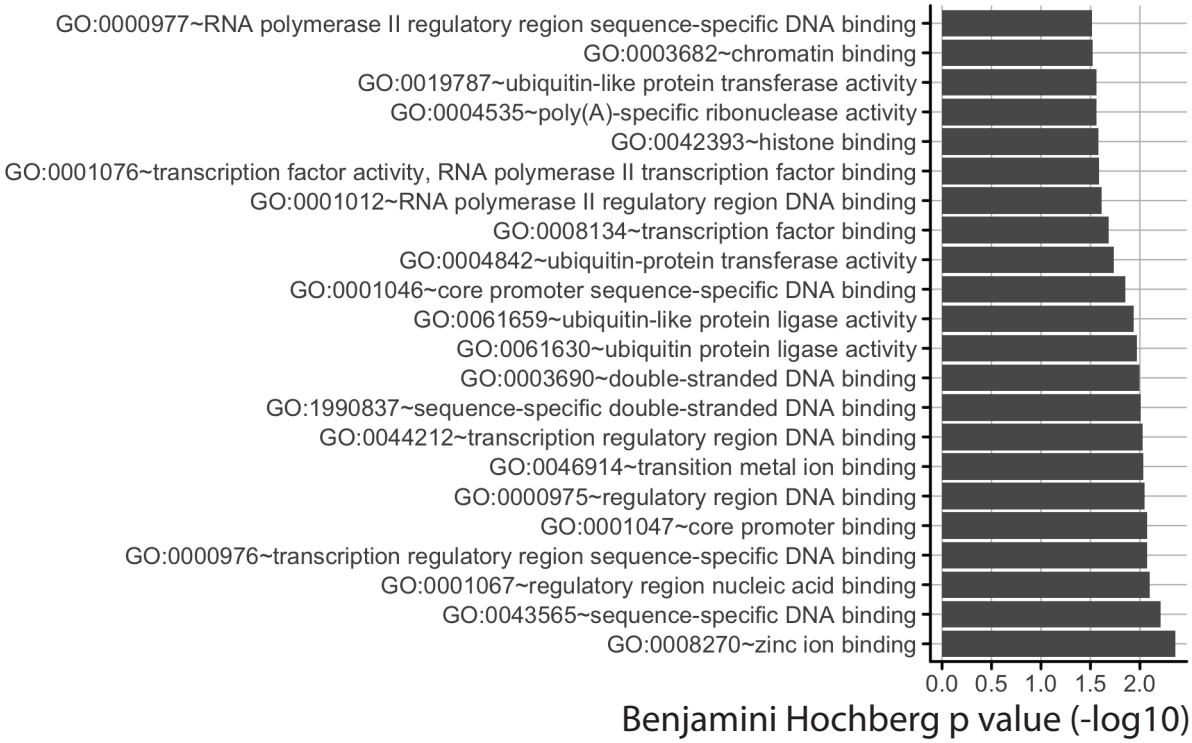

A
PC2: 23% variance
PC1: 51% variance
genotype
WT
Tbp2-/-
stage
P7
P14
B
Euclidian distance
#1#2#3 WT P7
#1#2#3 WT P14
#1#2#3 Tbp2-/- P7
#1#2#3 Tbp2-/- P14
C
1577 genes up regulated
1720 genes down regulated
P7 Tbp2-/- expression
P7 WT expression
D
1358 genes up regulated
1794 genes down regulated
P14 Tbp2-/- expression
P14 WT expression
ns
p≤0.05
p≤0.05 |log2FC|>1
E
P7 up
868
709
649
P14 up
P7 down
678
1042
752
P14 down
F
P14 foldchange (log2)
P7 foldchange (log2)
expression (WT mean)
<10^0
10^0<10^1
10^1<10^2
10^2<10^3
10^3<10^4
>10^4

Figure 2

A


GO:0000977~RNA polymerase II regulatory region sequence-specific DNA binding
GO:0003682~chromatin binding
GO:0019787~ubiquitin-like protein transferase activity
GO:0004535~poly(A)-specific ribonuclease activity
GO:0042393~histone binding
GO:0001076~transcription factor activity, RNA polymerase II transcription factor binding
GO:0001012~RNA polymerase II regulatory region DNA binding
GO:0008134~transcription factor binding
GO:0004842~ubiquitin-protein transferase activity
GO:0001046~core promoter sequence-specific DNA binding
GO:0061659~ubiquitin-like protein ligase activity
GO:0061630~ubiquitin protein ligase activity
GO:0003690~double-stranded DNA binding
GO:1990837~sequence-specific double-stranded DNA binding
GO:0044212~transcription regulatory region DNA binding
GO:0046914~transition metal ion binding
GO:0000975~regulatory region DNA binding
GO:0001047~core promoter binding
GO:0000976~transcription regulatory region sequence-specific DNA binding
GO:0001067~regulatory region nucleic acid binding
GO:0043565~sequence-specific DNA binding
GO:0008270~zinc ion binding
0.0 0.5 1.0 1.5 2.0
Benjamini Hochberg p value (-log10)

B

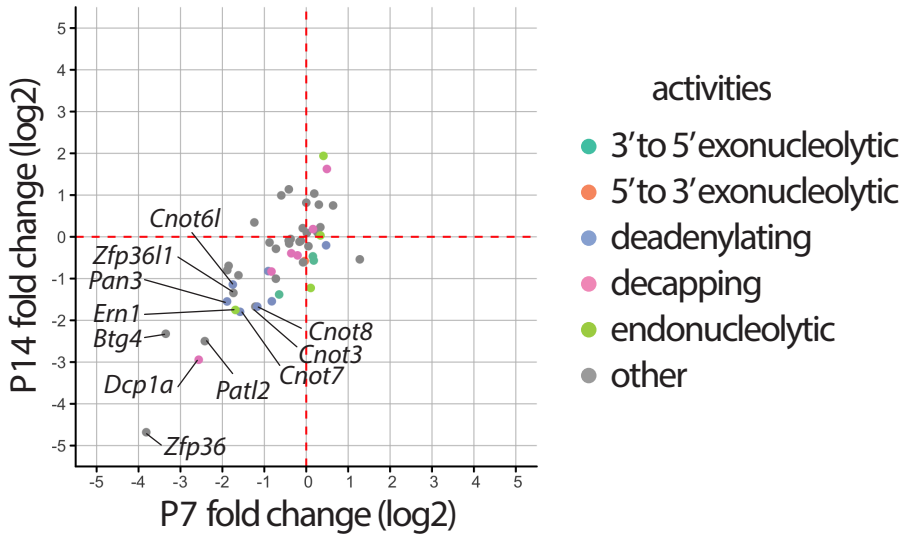

P14 fold change (log2)
P7 fold change (log2)
Cnot6l
Zfp36l1
Pan3
Ern1
Btg4
Dcp1a
Patl2
Cnot7
Cnot3
Cnot8
Zfp36
activities
3'to 5'exonucleolytic
5'to 3'exonucleolytic
deadenylating
decapping
endonucleolytic
other

Figure 3

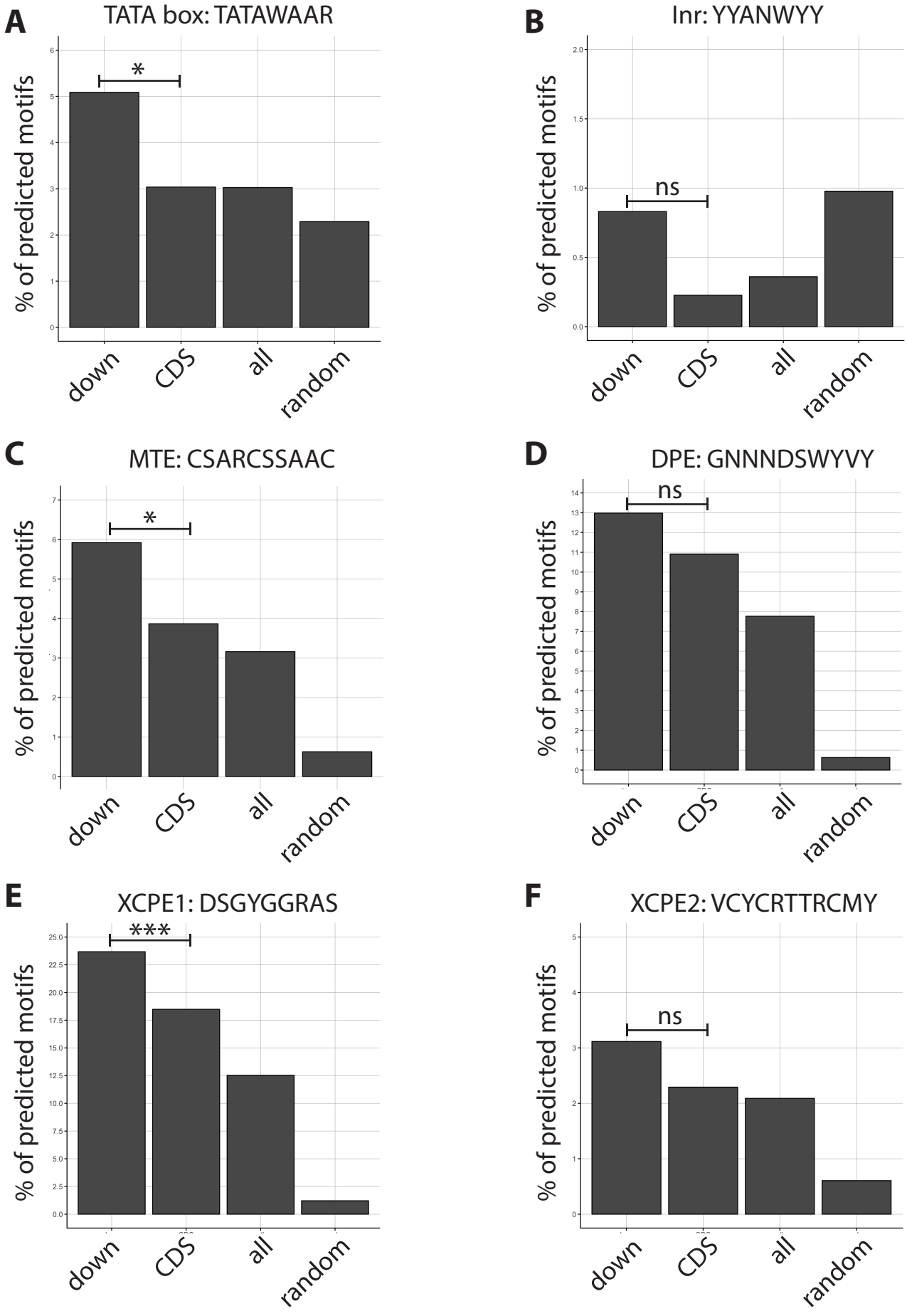

A
TATA box: TATAWAAR
% of predicted motifs
*
down
CDS
all
random
B
Inr: YYANWYY
% of predicted motifs
ns
down
CDS
all
random
C
MTE: CSARCSSAAC
% of predicted motifs
*
down
CDS
all
random
D
DPE: GNNNDSWYVY
% of predicted motifs
ns
down
CDS
all
random
E
XCPE1: DSGYGGRAS
% of predicted motifs
***
down
CDS
all
random
F
XCPE2: VCYCRTTRCMY
% of predicted motifs
ns
down
CDS
all
random

Figure 4

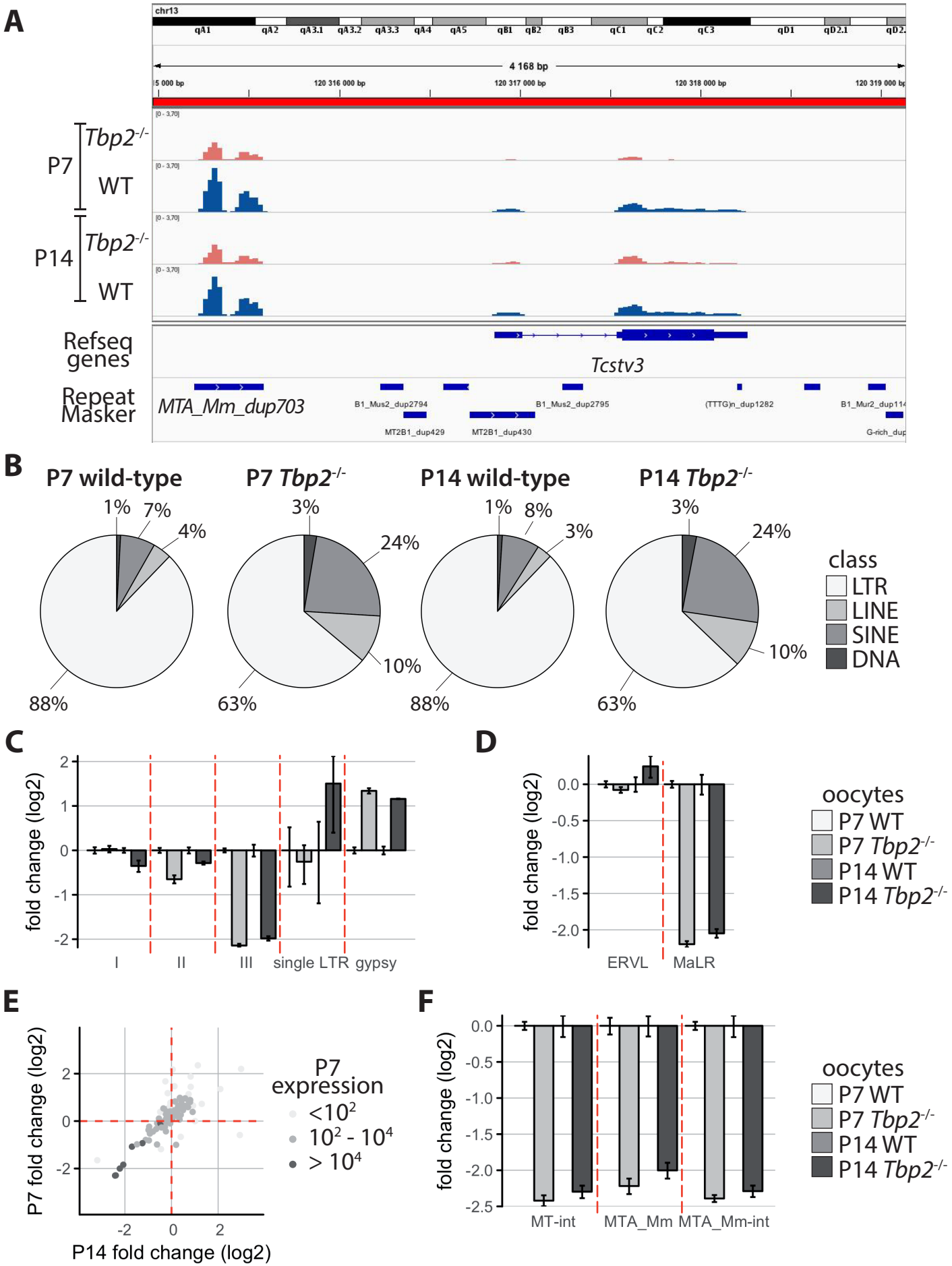

A
chr13
4 168 bp
P7
Tbp2-/-
WT
P14
Tbp2-/-
WT
Refseq genes
Tcstv3
Repeat Masker
MTA_Mm_dup703
B1_Mus2_dup2794
B1_Mus2_dup2795
(TTTG)n_dup1282
MT2B1_dup429
MT2B1_dup430
B
P7 wild-type
P7 Tbp2-/-
P14 wild-type
P14 Tbp2-/-
1%
7%
4%
88%
3%
24%
10%
63%
1%
8%
3%
88%
3%
24%
10%
63%
class
LTR
LINE
SINE
DNA
C
fold change (log2)
I
II
III
single LTR
gypsy
D
fold change (log2)
ERVL
MaLR
oocytes
P7 WT
P7 Tbp2-/-
P14 WT
P14 Tbp2-/-
E
P7 fold change (log2)
P14 fold change (log2)
P7 expression
<10^2
10^2 - 10^4
> 10^4
F
fold change (log2)
MT-int
MTA_Mm
MTA_Mm-int
oocytes
P7 WT
P7 Tbp2-/-
P14 WT
P14 Tbp2-/-

Figure 5

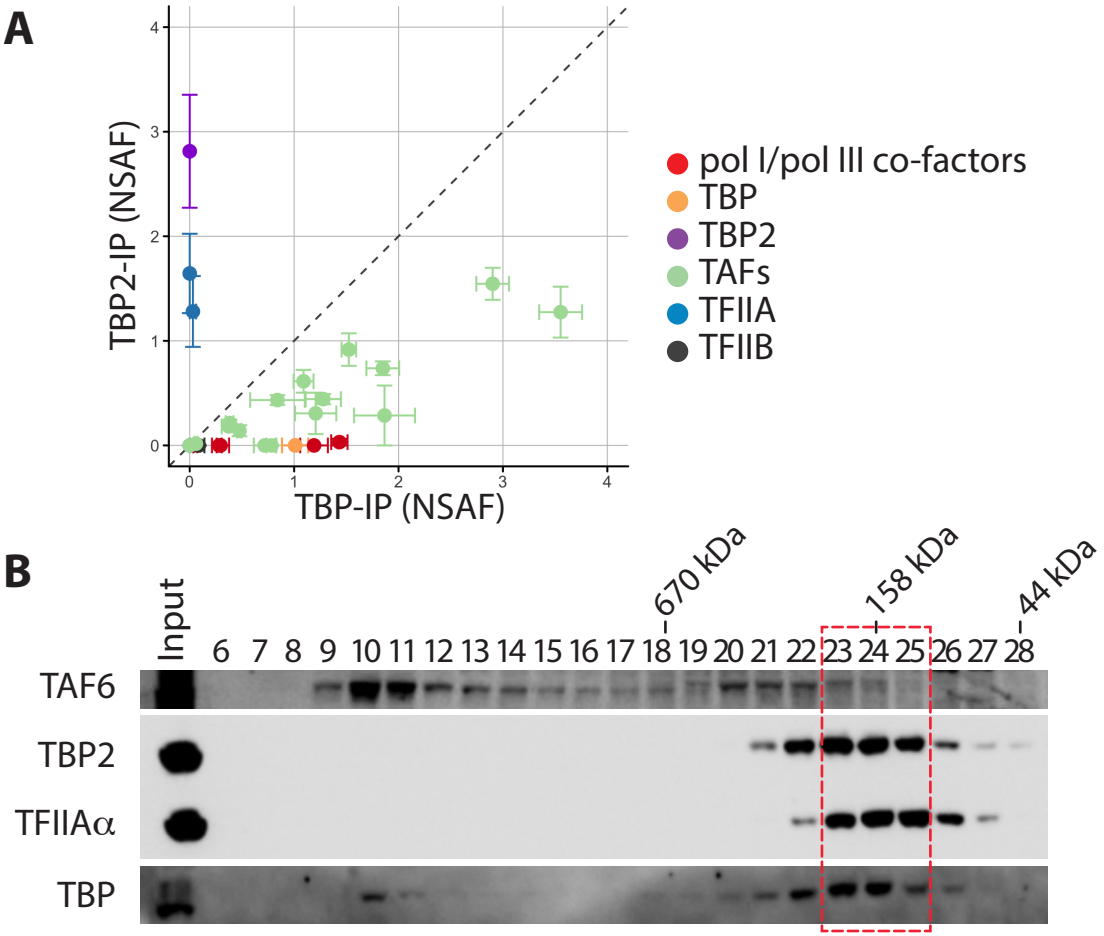

A
TBP2-IP (NSAF)
TBP-IP (NSAF)
pol I/pol III co-factors
TBP
TBP2
TAFs
TFIIA
TFIIB
B
Input
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
670 kDa
158 kDa
44 kDa
TAF6
TBP2
TFIIAα
TBP
C
TBP2-IP (NSAF)
TBP2-IP fraction (NSAF)
pol I/pol III co-factors
TBP
TBP2
TAFs
TFIIA
TFIIB

Figure 6

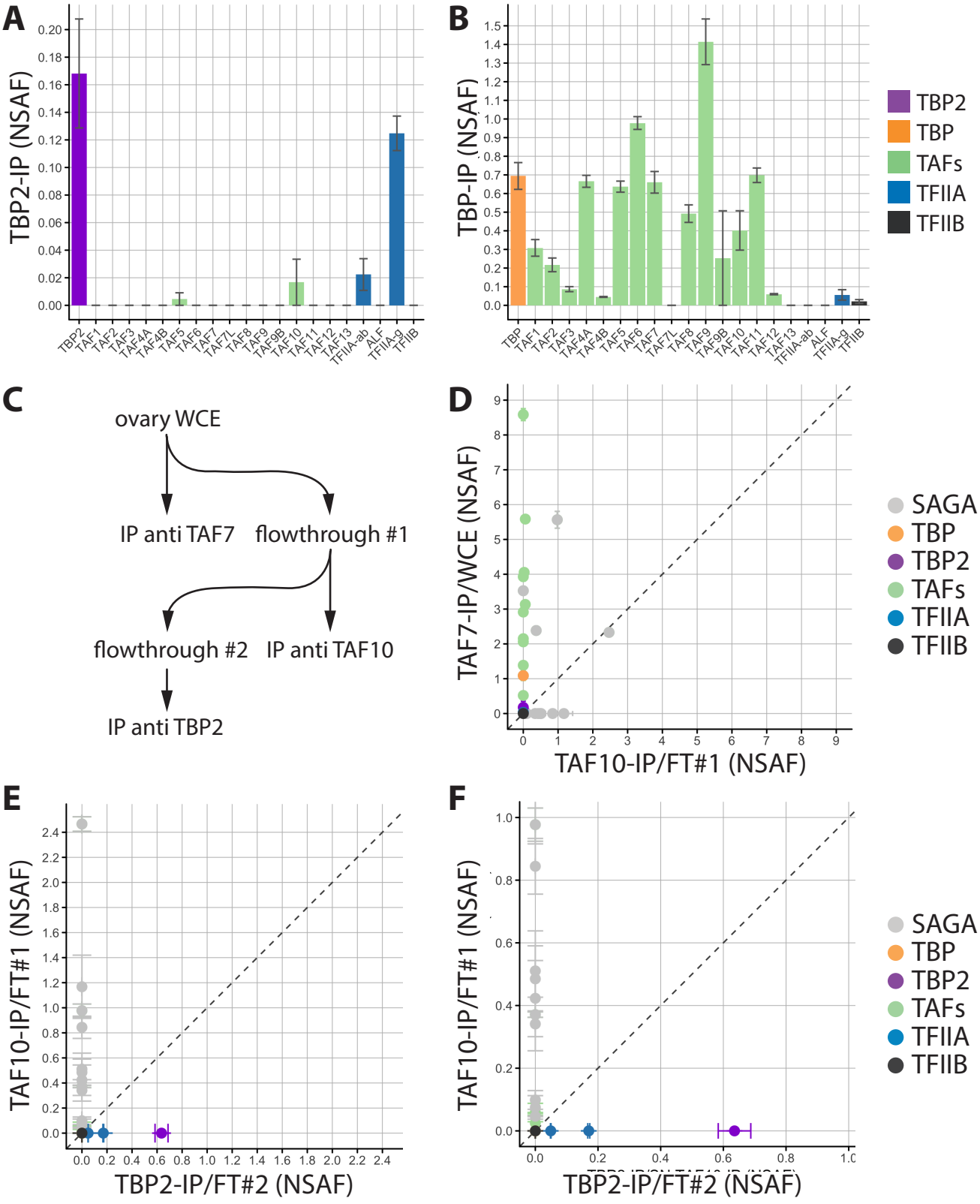

A
TBP2-IP (NSAF)
B
TBP-IP (NSAF)
TBP2
TBP
TAFs
TFIIA
TFIIB
C
ovary WCE
IP anti TAF7
flowthrough #1
flowthrough #2
IP anti TAF10
IP anti TBP2
D
TAF7-IP/WCE (NSAF)
TAF10-IP/FT#1 (NSAF)
SAGA
TBP
TBP2
TAFs
TFIIA
TFIIB
E
TAF10-IP/FT#1 (NSAF)
TBP2-IP/FT#2 (NSAF)
F
TAF10-IP/FT#1 (NSAF)
TBP2-IP/FT#2 (NSAF)

Figure 7

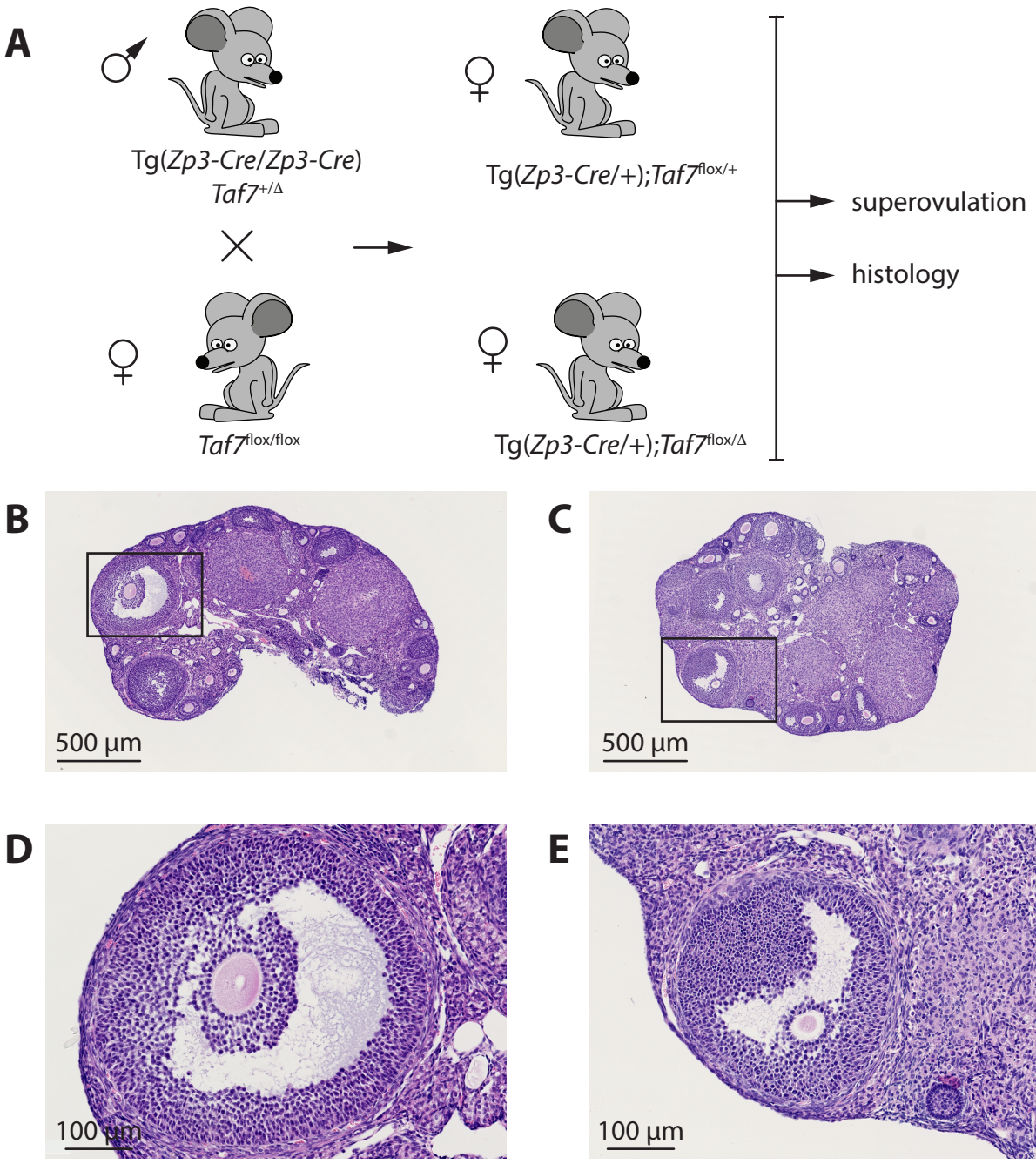

A
Tg(Zp3-Cre/Zp3-Cre)
Taf7+/Δ
Tg(Zp3-Cre/+);Taf7flox/+
Taf7flox/flox
Tg(Zp3-Cre/+);Taf7flox/Δ
superovulation
histology
B
500 µm
C
500 µm
D
100 µm
E
100 µm

Figure 8

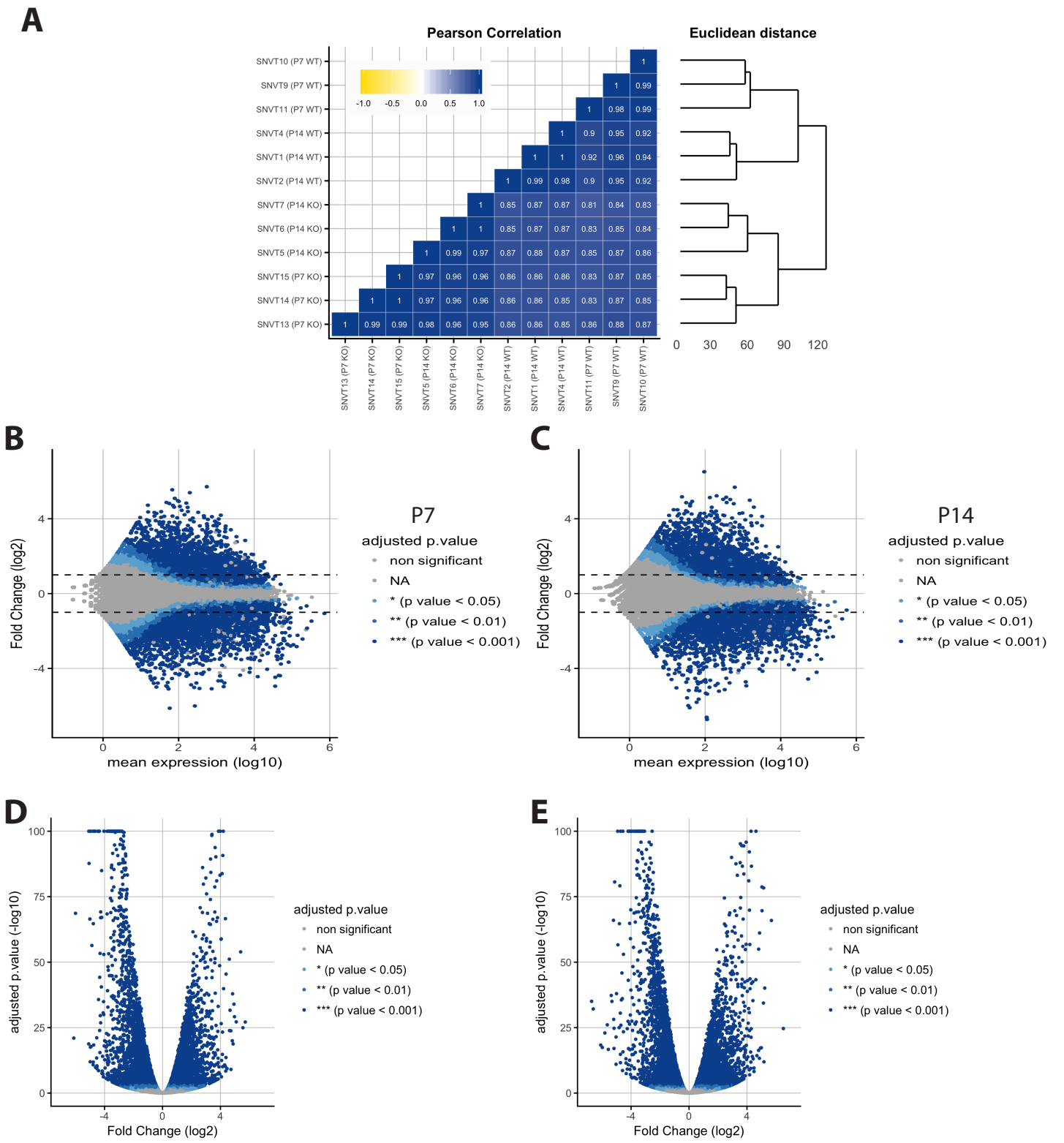

A
Pearson Correlation
Euclidean distance
SNVT10 (P7 WT)
SNVT9 (P7 WT)
SNVT11 (P7 WT)
SNVT4 (P14 WT)
SNVT1 (P14 WT)
SNVT2 (P14 WT)
SNVT7 (P14 KO)
SNVT6 (P14 KO)
SNVT5 (P14 KO)
SNVT15 (P7 KO)
SNVT14 (P7 KO)
SNVT13 (P7 KO)
-1.0 -0.5 0.0 0.5 1.0
0 30 60 90 120
B
C
Fold Change (log2)
mean expression (log10)
P7
P14
adjusted p.value
non significant
NA
* (p value < 0.05)
** (p value < 0.01)
*** (p value < 0.001)
D
E
adjusted p.value (-log10)
Fold Change (log2)

Figure Sup 1

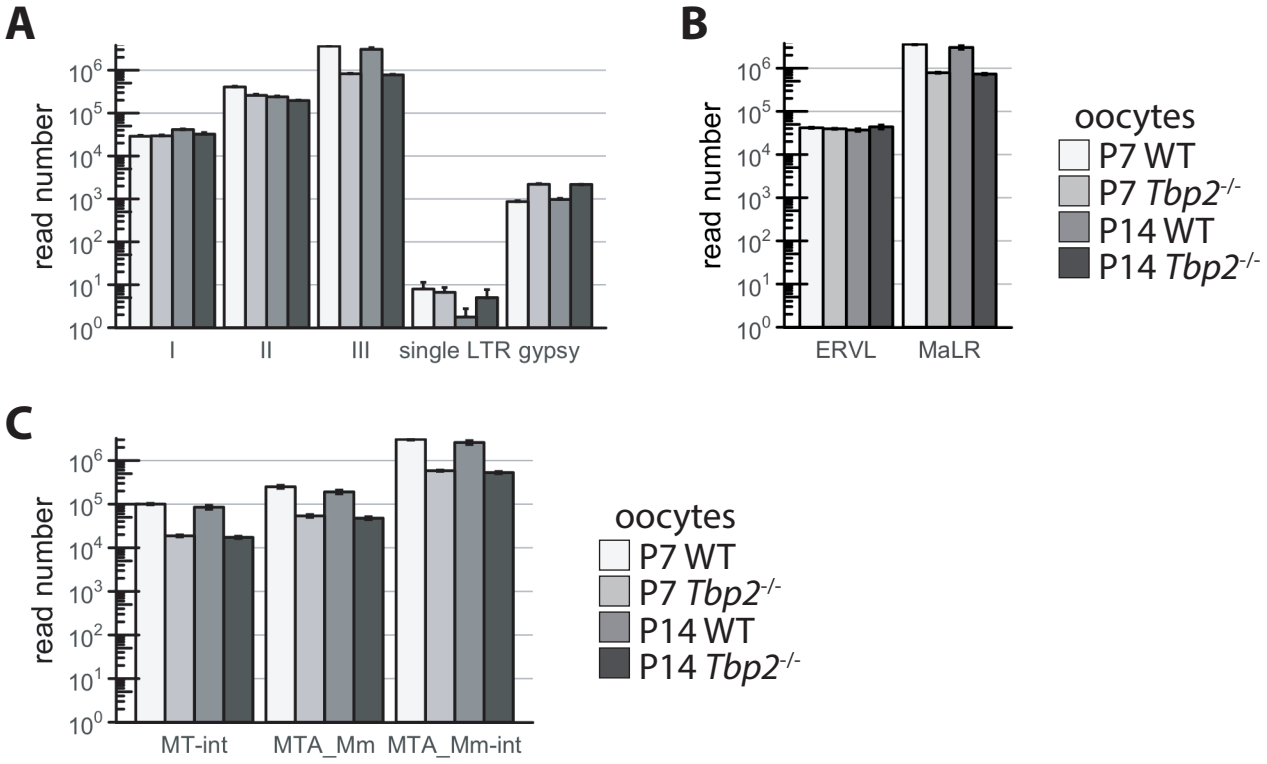

A
read number
10^0
10^1
10^2
10^3
10^4
10^5
10^6
I
II
III
single LTR
gypsy
B
read number
ERVL
MaLR
oocytes
P7 WT
P7 Tbp2-/-
P14 WT
P14 Tbp2-/-
C
read number
MT-int
MTA_Mm
MTA_Mm-int
oocytes
P7 WT
P7 Tbp2-/-
P14 WT
P14 Tbp2-/-

Figure Sup 2

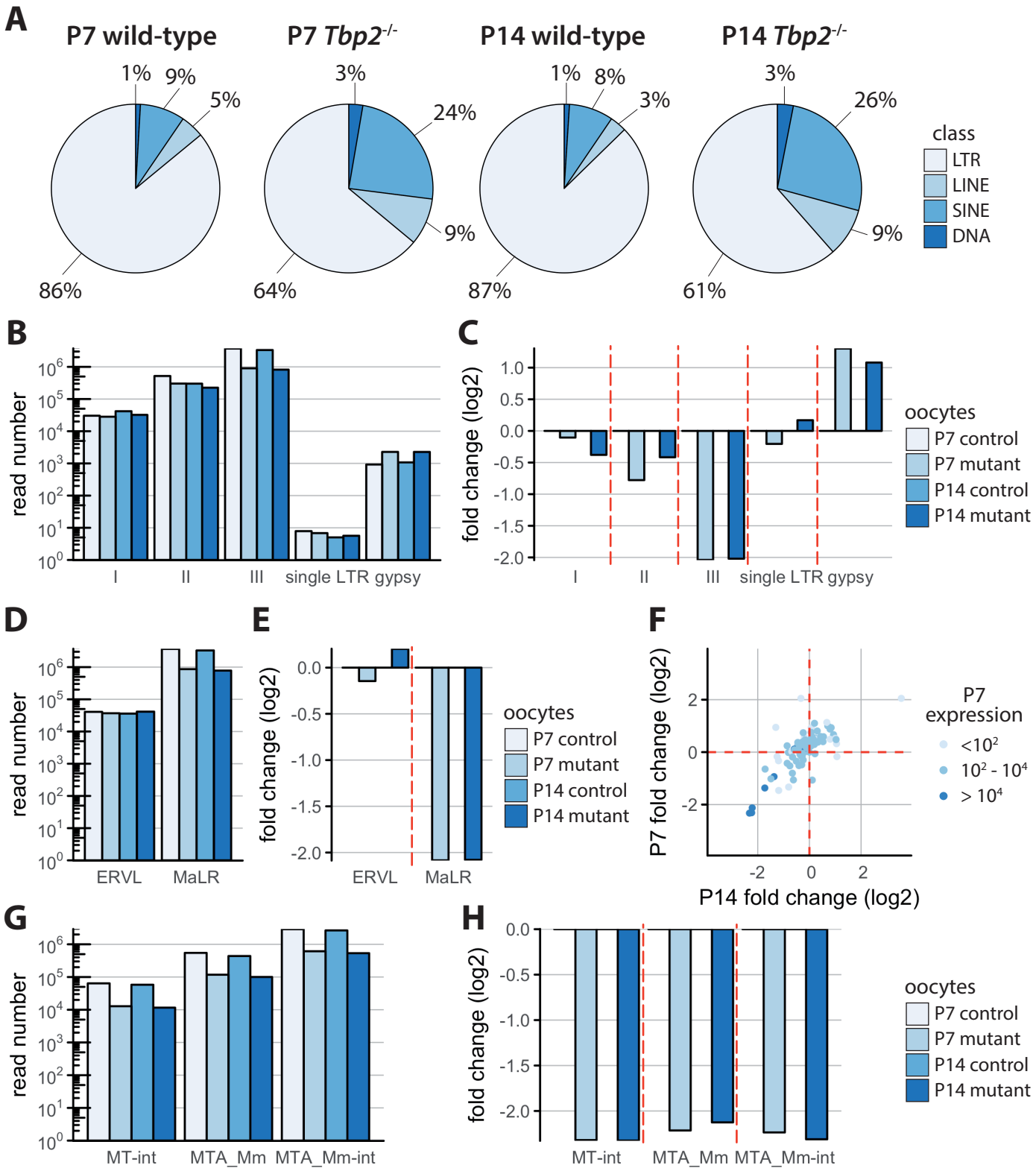

A
P7 wild-type
P7 Tbp2-/-
P14 wild-type
P14 Tbp2-/-
1%
9%
5%
86%
3%
24%
9%
64%
1%
8%
3%
87%
3%
26%
9%
61%
class
LTR
LINE
SINE
DNA
B
read number
I
II
III
single LTR gypsy
C
fold change (log2)
oocytes
P7 control
P7 mutant
P14 control
P14 mutant
D
read number
ERVL
MaLR
E
fold change (log2)
ERVL
MaLR
oocytes
P7 control
P7 mutant
P14 control
P14 mutant
F
P7 fold change (log2)
P14 fold change (log2)
P7 expression
<10²
10² - 10⁴
> 10⁴
G
read number
MT-int
MTA_Mm
MTA_Mm-int
H
fold change (log2)
MT-int
MTA_Mm
MTA_Mm-int
oocytes
P7 control
P7 mutant
P14 control
P14 mutant

Figure Sup 3

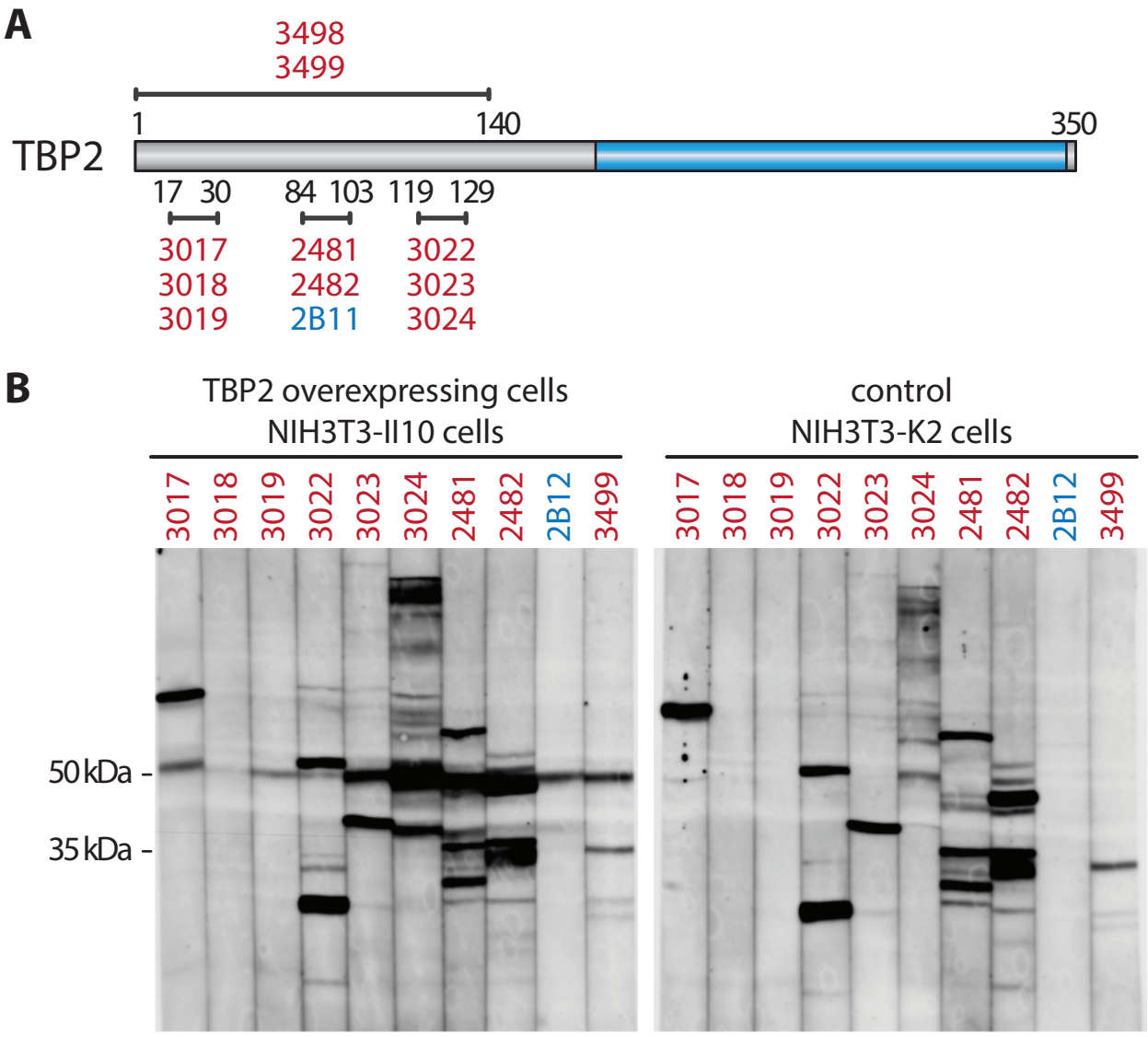

A
3498
3499
1
140
350
TBP2
17 30
84 103
119 129
3017
3018
3019
2481
2482
2B11
3022
3023
3024
B
TBP2 overexpressing cells
NIH3T3-II10 cells
control
NIH3T3-K2 cells
3017
3018
3019
3022
3023
3024
2481
2482
2B12
3499
50 kDa
35 kDa

Figure Sup 4

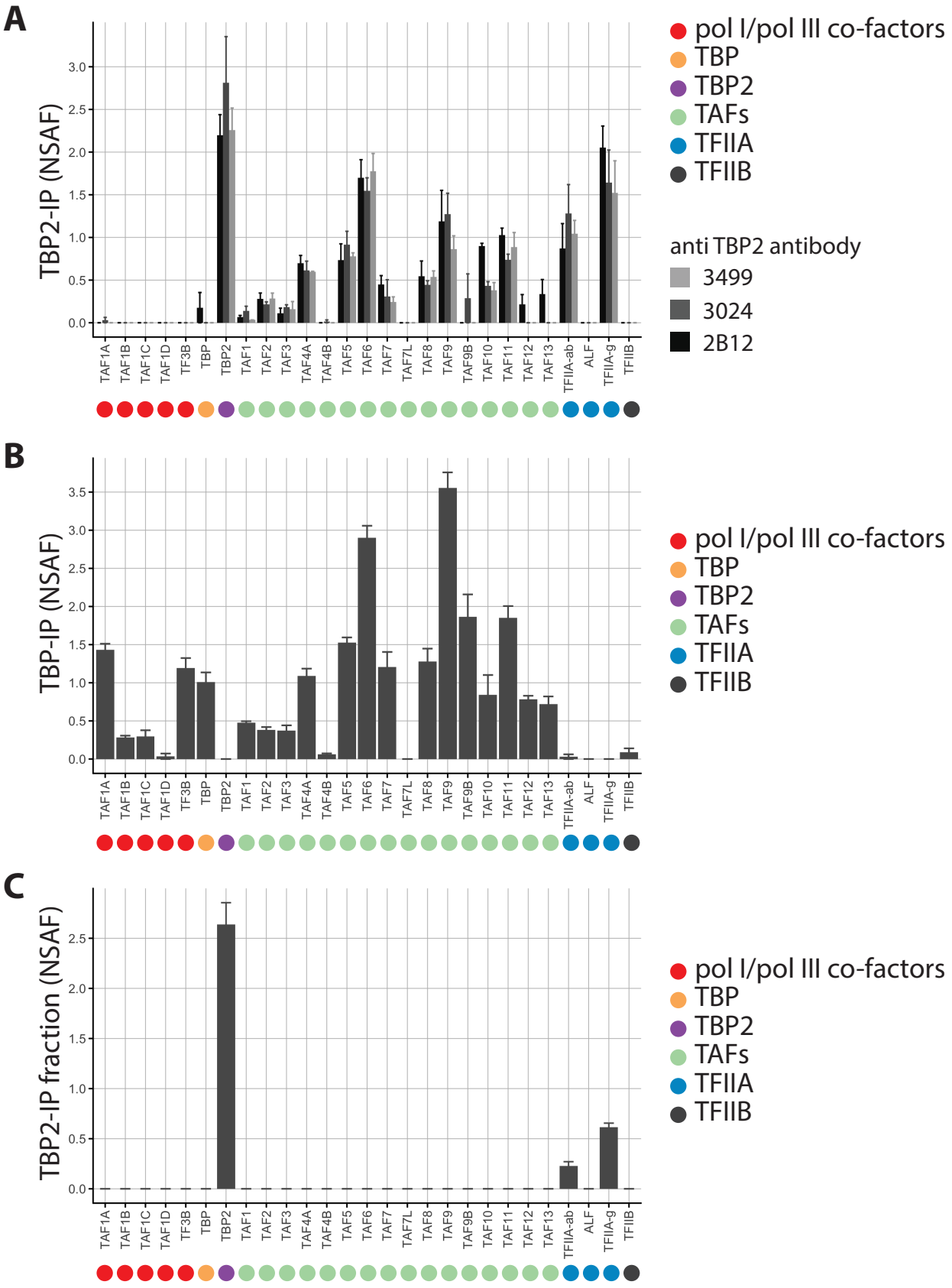

A
TBP2-IP (NSAF)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
TAF1A TAF1B TAF1C TAF1D TF3B TBP TBP2 TAF1 TAF2 TAF3 TAF4A TAF4B TAF5 TAF6 TAF7 TAF7L TAF8 TAF9 TAF9B TAF10 TAF11 TAF12 TAF13 TFIIA-ab ALF TFIIA-g TFIIB
pol I/pol III co-factors
TBP
TBP2
TAFs
TFIIA
TFIIB
anti TBP2 antibody
3499
3024
2B12
B
TBP-IP (NSAF)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
TAF1A TAF1B TAF1C TAF1D TF3B TBP TBP2 TAF1 TAF2 TAF3 TAF4A TAF4B TAF5 TAF6 TAF7 TAF7L TAF8 TAF9 TAF9B TAF10 TAF11 TAF12 TAF13 TFIIA-ab ALF TFIIA-g TFIIB
pol I/pol III co-factors
TBP
TBP2
TAFs
TFIIA
TFIIB
C
TBP2-IP fraction (NSAF)
2.5
2.0
1.5
1.0
0.5
0.0
TAF1A TAF1B TAF1C TAF1D TF3B TBP TBP2 TAF1 TAF2 TAF3 TAF4A TAF4B TAF5 TAF6 TAF7 TAF7L TAF8 TAF9 TAF9B TAF10 TAF11 TAF12 TAF13 TFIIA-ab ALF TFIIA-g TFIIB
pol I/pol III co-factors
TBP
TBP2
TAFs
TFIIA
TFIIB

Figure Sup 5

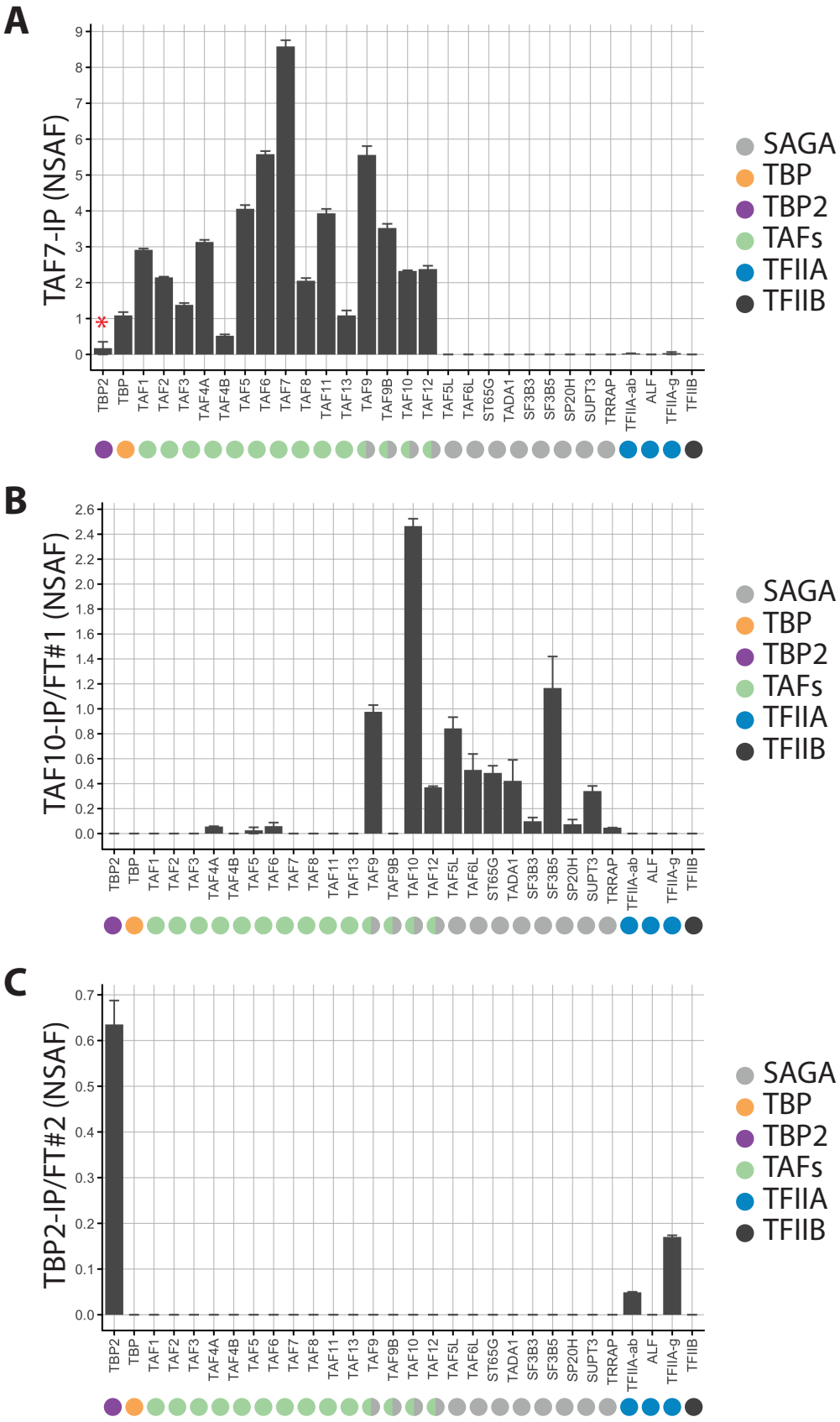

A
TAF7-IP (NSAF)
B
TAF10-IP/FT#1 (NSAF)
C
TBP2-IP/FT#2 (NSAF)
TBP2 TBP TAF1 TAF2 TAF3 TAF4A TAF4B TAF5 TAF6 TAF7 TAF8 TAF11 TAF13 TAF9 TAF9B TAF10 TAF12 TAF5L TAF6L ST65G TADA1 SF3B3 SF3B5 SP20H SUPT3 TRRAP TFIIA-ab ALF TFIIA-g TFIIB
SAGA
TBP
TBP2
TAFs
TFIIA
TFIIB

Figure Sup 6

Unpublished results

1. Technical optimizations

1.1 Antibody generation and validation

As we planned to perform anti-TBP2 IP and ChIP on mouse oocytes that are very limited in numbers, high affinity anti-TBP2 antibodies are required. Besides the published 2B12 mouse monoclonal antibody, there were already 8 rabbit polyclonal anti-TBP2 sera in the lab, all generated with different TBP2 peptides, but they were not purified nor validated. I generated a new anti-TBP2 sera against the N-terminal part of the TBP2 protein, and purified and validated the anti-TBP2 sera.

1.1.1 Antibody generation

As it was thought that immunization of rabbit with protein fragments rather than peptides will have a better chance to obtain ChIP grade antibodies, I generated 2 new anti-TBP2 sera. The N-terminal part of TBP2 (1-140aa) was cloned into the pET-15b vector, and recombinant protein produced in bacteria was used for rabbit immunization to generate anti-TBP2 sera, 3498 and 3499 (***Figure 1-1 B***). In total, besides 2B12, 10 anti-TBP2 sera have been raised (***Figure 1-1 C***).

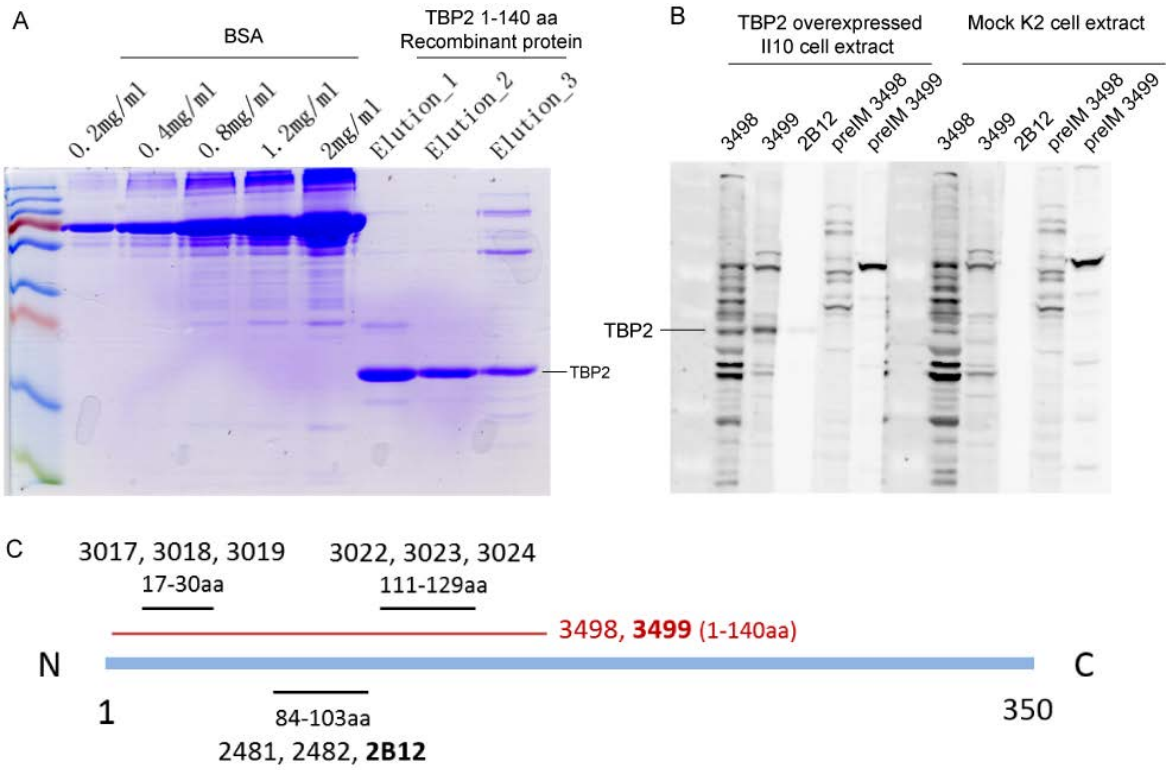


Figure 1-1: Generation of new anti-TBP2 sera. (A) TBP2 (1-140aa) expression and concentration measurement. (B) WB test of the new anti-TBP2 sera before purification. (C) Localization of the peptides or protein fragments used to generate the new sera.

1.1.2 Antibody purification

All the anti-TBP2 sera (10ml boost sera used each) were purified by affinity purification with the corresponding TBP2 peptides or protein and purified antibody fractions were analyzed by 10% SDS polyacrylamide gel electrophoresis, followed by Coomassie-staining (***Figure 1-2***). 3498 was not purified as it was not good (***Figure 1-1 B***).

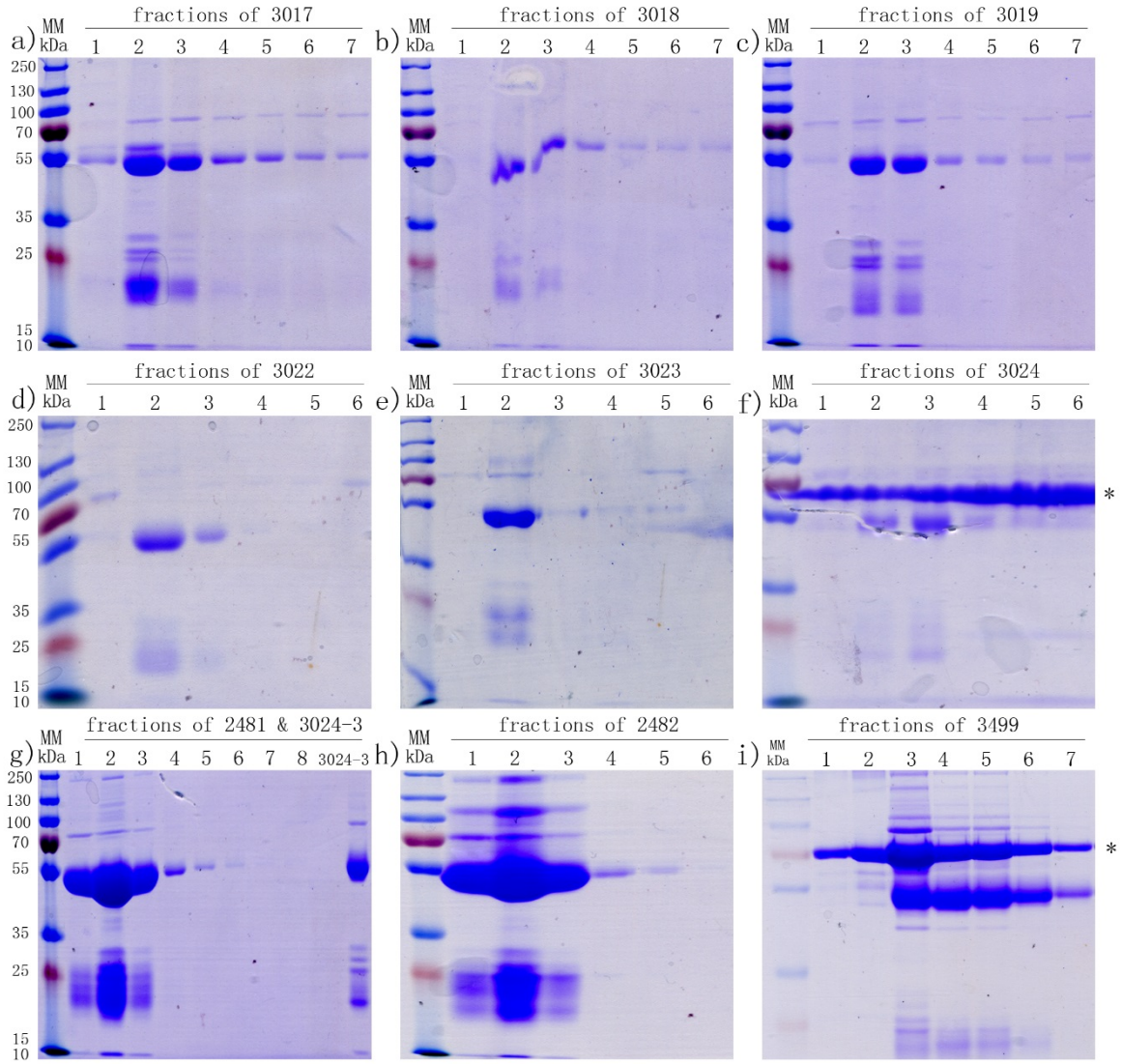


Figure 1-2: Affinity purification of polyclonal antibodies. Affinity-purified antibody fractions were analyzed on 10% SDS-PAGE, followed by Coomassie-staining. Asterisk indicates the albumin contamination.

1.1.3 Antibody validation by WB

All the purified antibodies were then validated by western blots. TBP2 overexpressing NIH3T3-II10 cell extract and TBP2 negative K2 mock cell extract were used for WB. The

monoclonal 2B12 antibody was used as a positive control. The result showed that antibody 3023, 3024 and 3499 were good for WB, although with non-specific signal (***Figure 1-3***).

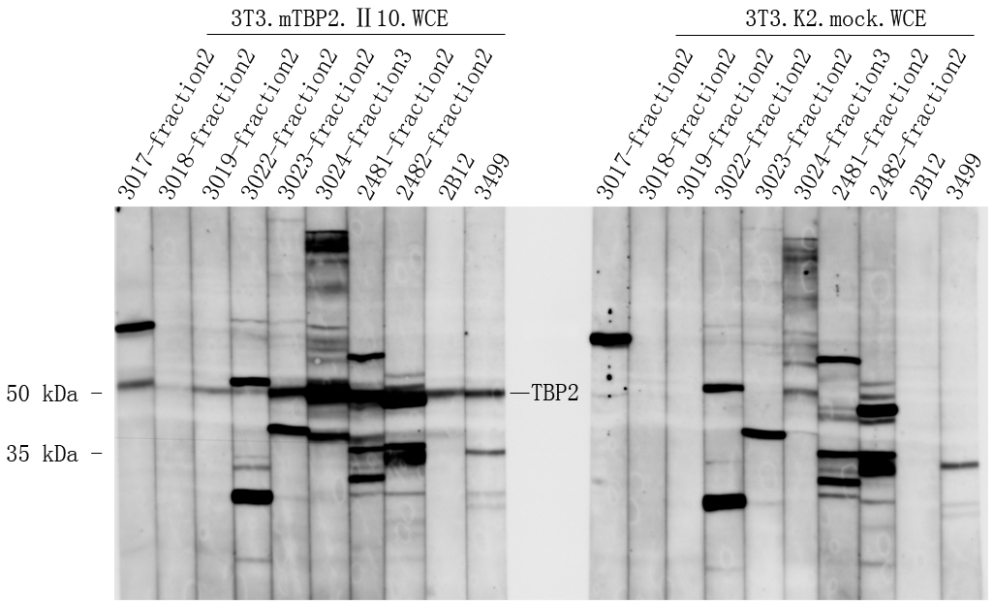


Figure 1-3: Antibody validation by WB with TBP2 overexpressed cell line.

1.1.4 Antibody validation by IP

The anti-TBP2 antibodies were then validated by IP with II10 cell extract, and the IP elutions were analyzed by WB with the 2B12 antibody (***Figure 1-4***), 2B12 being the most specific TBP2 antibody for WB. Results showed that 2481, 2482, 2B12, 3022, 3023 and 3024 were all good for IP, and that 3024 has the highest affinity.

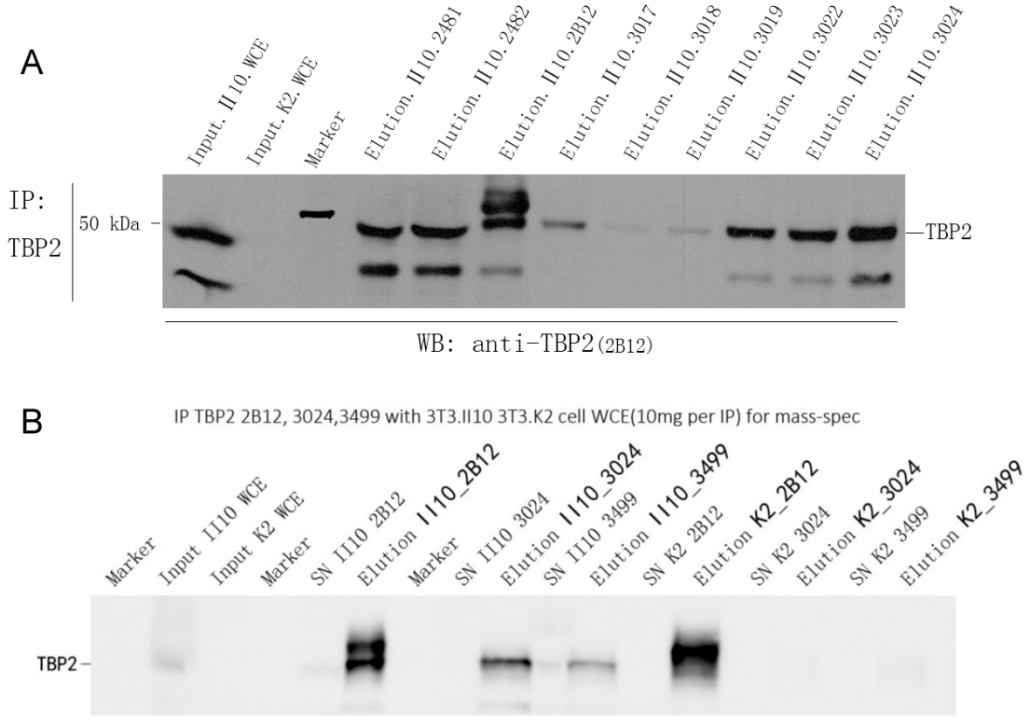


Figure 1-4: Antibody validation by IP.

1.1.5 Antibody validation by ChIP

Microarray analysis showed that, in the II10 cells, some genes such as *Gadd45g* and *Mef2a*, were upregulated and *Cdca8* were downregulated (Gazdag, 2008). TBP2 seemed to bind to the promoters of the upregulated genes (Gazdag, 2008). Thus, I use *Gadd45g* and *Mef2a* as positive targets and *Cdca8* and intergenic region as negative targets to validate the anti-TBP2 antibodies for ChIP using II10 cells.

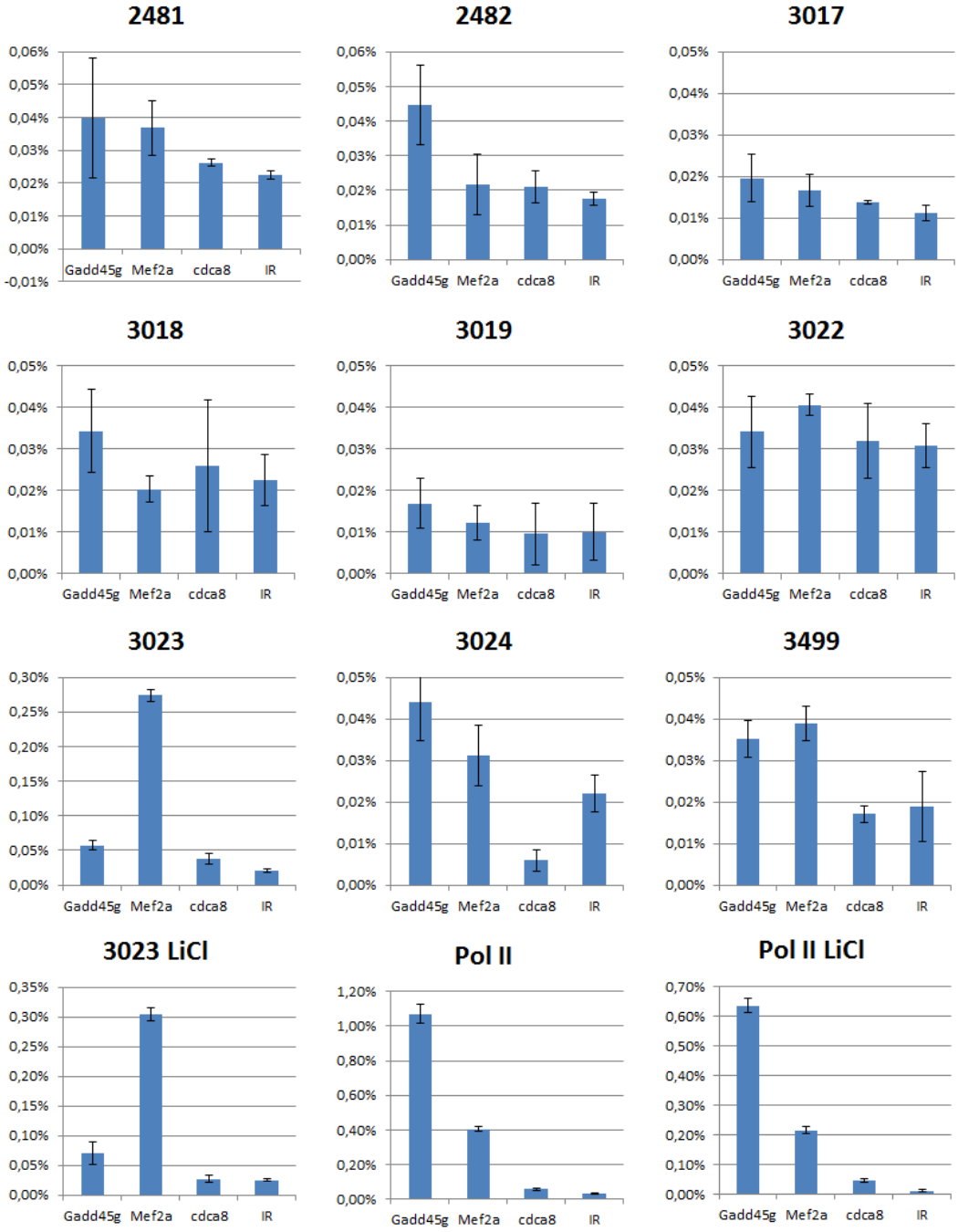


Figure 1-5: TBP2 antibody validation by ChIP. Pol II ChIP was performed as a positive control. '3023 LiCl' means that an additional harsh wash with LiCl buffer was performed during the ChIP.

Out of the 9 anti-TBP2 sera, 3023 and 3499 seemed to work for ChIP, but only 3023 has very high affinity (***Figure 1-5***). 3023 and 3499 antibodies were further validated by IF with the help of Sascha Conic (***Figure 1-6***).

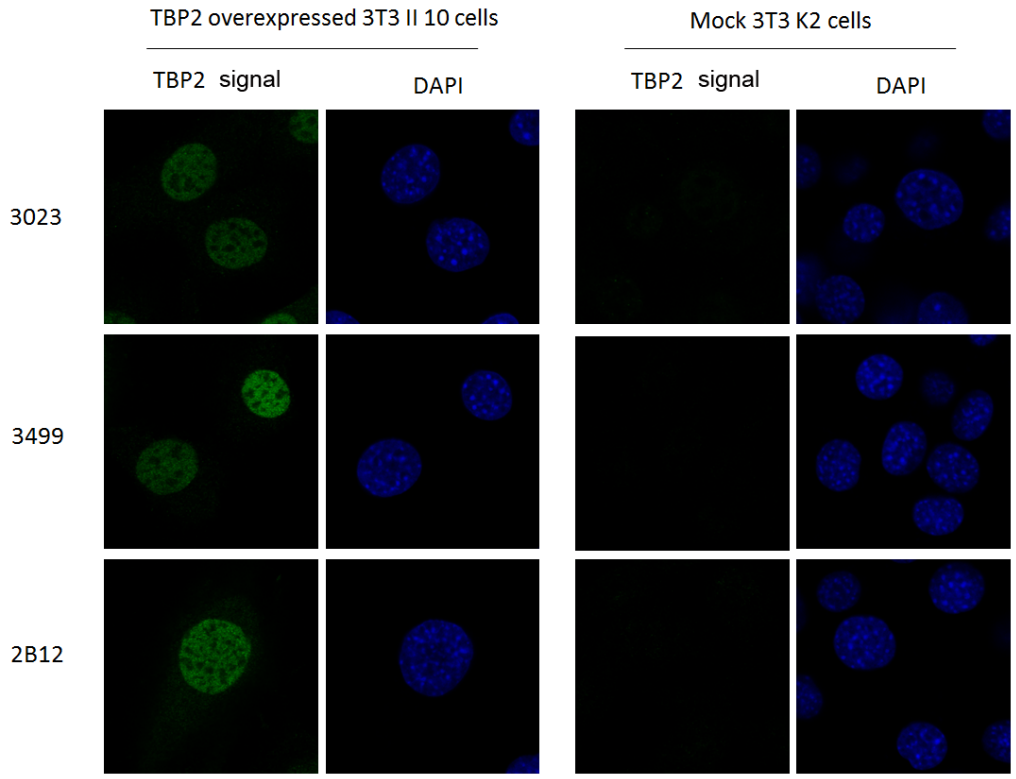


Figure 1-6: Antibody validation for IF.

Altogether, I conclude that we have generated new TBP2 antibodies for different applications. The anti-TBP2 3023 is the best for ChIP and anti-TBP2 3024 is the best for IP.

1.2 Optimization of immunoprecipitation with low input amount

We aimed to identify the TBP2-containing complex/es by performing immunoprecipitation experiments coupled to mass spectrometry (IP mass-spec). TBP2 is only expressed in growing oocytes and a classical protocol for IP mass-spec requires more than 4mg protein extracts, which is not suitable for oocytes. Therefore, optimization of the IP protocol suitable for micrograms of starting material was needed. Based on the micro chromatin immunoprecipitation assay (Dahl et al., 2008), I set up the micro immunoprecipitation (microIP) for low amount of starting material (***Figure 1-7 A***). With this protocol, I succeeded to carry out anti-TBP IP from only micrograms of HeLa nuclei extract input: for instance, in the elution of the anti-TBP microIP from 4ug of HeLa nuclear extract, TBP and TAF6 can be clearly detected by Western blot (***Figure 1-7 B***).

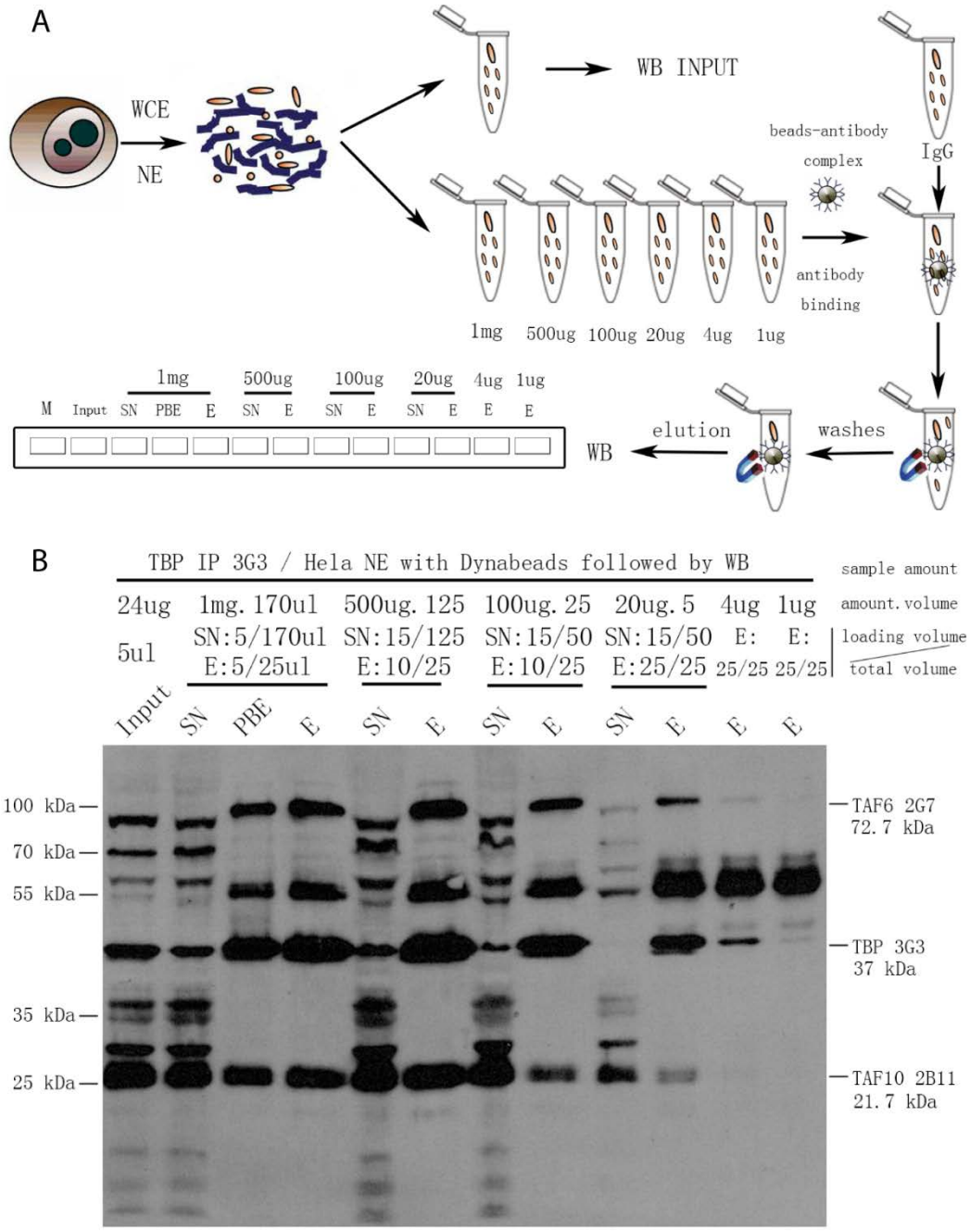


Figure 1-7: Optimization of micro immunoprecipitation. (A) Experimental design. (B) Anti-TBP microIP with different amount of HeLa nuclear extract input. after IP, elution was analyzed by WB with the anti-TBP (3G3) antibody, the anti-TAF6 (2G7) antibody and the anti-TAF10 (2B11) antibody.

I further tested the TBP2 antibodies with different amount of II10 cell extracts (***Figure 1-8***). The results confirmed that 3024 is the best anti-TBP2 antibody for IP and that anti-TBP2 IP can be performed with micrograms of whole cell extracts, which made it possible to carry out anti-TBP2 IP with ovaries.

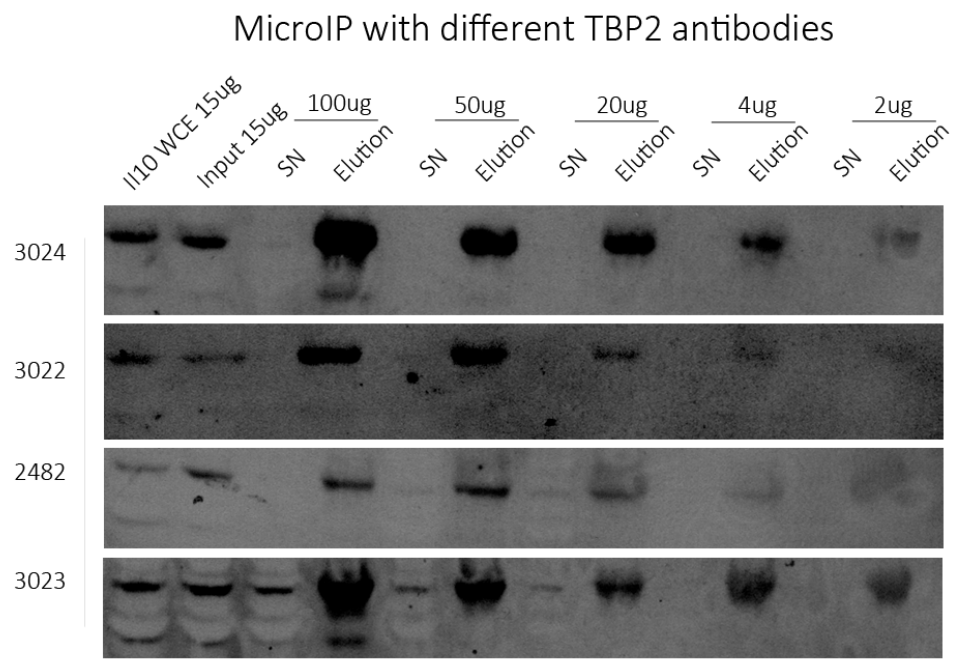


Figure 1-8: Micro immunoprecipitation validation of TBP2 antibodies. Anti-TBP2 microIPs were performed with different amount of II10 WCE input, and the elution was analyzed by WB with the 2B12 anti-TBP2 antibody.

1.3 Optimization of MOWChIP

To decipher the role and function of TBP2 during oocyte growth, initially, we planned to map TBP2 binding profiles in oocyte by carrying out anti-TBP2 ChIP-seq with oocytes.

As the sensitivity of conventional ChIP assays is a major obstacle for the study of low-abundance cells, to perform ChIP-seq with oocytes, we first need to setup a method that can be used for oocytes ChIP-seq.

After comparing the published low input ChIP methods (before 2015), such as uChIP (Dahl et al., 2008), iChIP (Lara-Astiaso et al., 2014), ULI-NChIP (Brind'Amour et al., 2015), ChIPmentation (Schmidl et al., 2015) and MOWChIP (Cao et al., 2015), we decided to set up the MOMChIP, a microfluidic oscillatory washing–based ChIP-seq method that has been reported for epigenomic profiling with as few as 100 cells.

The major steps of MOWChIP contains: (i) formation of a packed bed of IP beads in the microfluidic chamber (~0.71ul volume); (ii) performing ChIP by flowing the chromatin through the packed bed; (iii) oscillatory washing; (iv) removal of the unbound chromatin fragments by flushing the chamber; (v) collection of the IP beads (Cao et al., 2015) (***Figure 1-9 A***).

In collaboration with Igor Kukhtevich (Robert Schneider Lab, Institute of Functional Epigenetics, Munich), we have tested many times with different microfluidic designs for H3K4me3 ChIP with chromatin prepared from 5000 NIH 3T3 cells (***Figure 1-10***). However,

although sometimes it seems promising, most of the tests are not working with huge variation between different experiments. Therefore, I also carried out alternative approaches to study TBP2 occupancy in oocytes (See ***Unpublished data 3.1*** & ***3.2***).

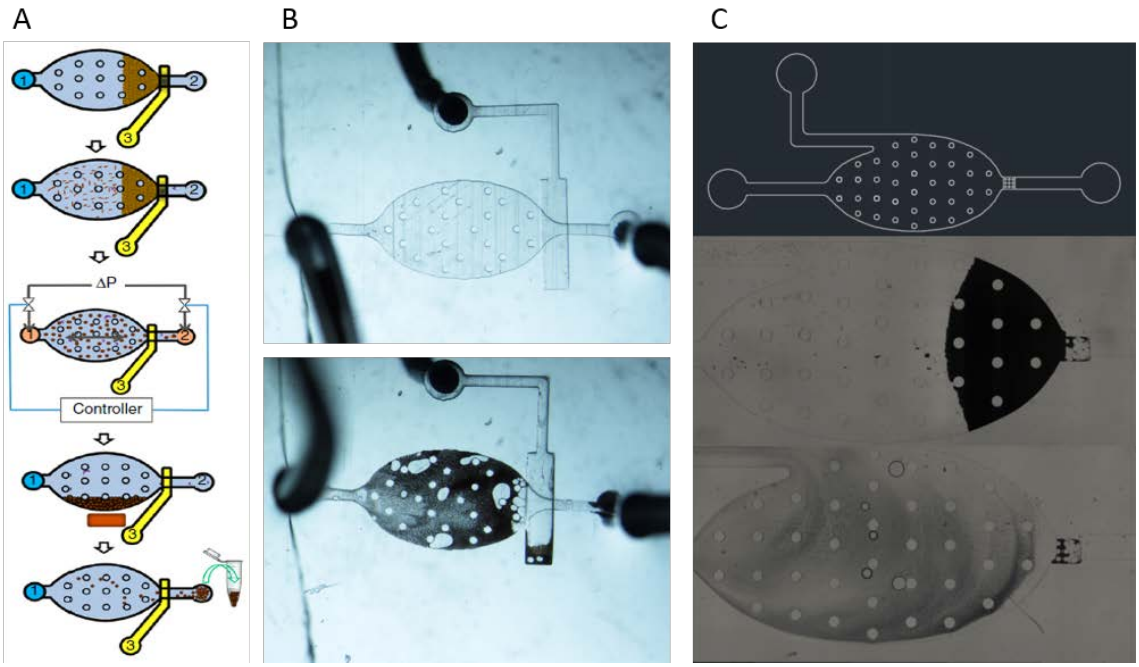


Figure 1-9: Overview of the different microfluidic designs for MOWChIP. (A) Illustration of the major steps of MOWChIP, from (Cao et al., 2015). (B) The first microfluidic design according to the published work (Cao et al., 2015). (C) The second microfluidic design, which has a narrow region with height of ~2 um to help the packing and maintenance of beads (as the diameter of the dynabeads is ~2.8 um).

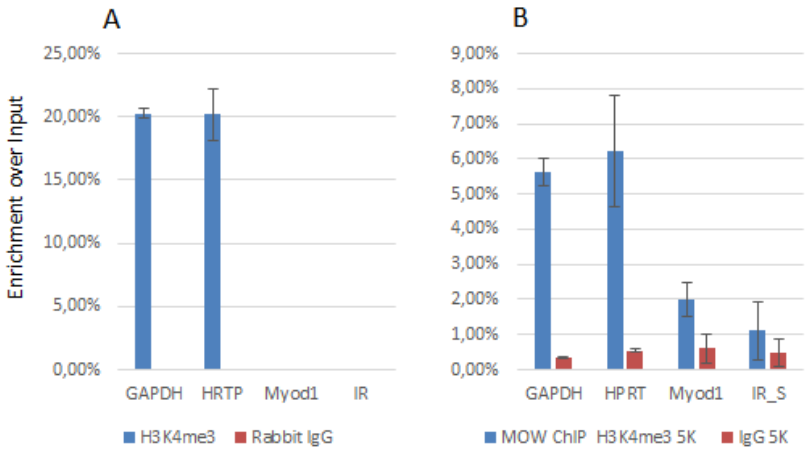


Figure 1-10: Test of the MOWChIP setup. (A) Conventional H3K4me3 ChIP with 50 ug chromatin, as control. (B) Best result obtained by H3K4me3 MOWChIP with chromatin from 5000 NIH 3T3 cells.

2. TBP2 binds genome-widely to gene promoters

Before studying the binding profile of TBP2 in oocytes, we first analyzed its occupancy in TBP2 ectopically overexpressing NIH3T3 cells. We applied both antibody-independent (DamID-seq) and antibody-dependent (ChIP-seq) approaches.

2.1 TBP2 DamID-seq in NIH3T3 cells

2.1.1 Generation of *Dam-TBP2* NIH3T3 stable cell lines

Inducible *Dam-Tbp2*-NIH3T3 and *Dam-only*-NIH3T3 stables cell lines were generated (***Figure 2-1 A&B&C***) as described in the Material and Methods (***10.1)***. Genomic DNA was extracted from both *Dam-Tbp2*-3T3 and *Dam-only*-3T3 cells in the absence of doxycycline and used for 'G^{me}ATC'-methylation-specific PCR amplification (***Figure 2-1 D***).

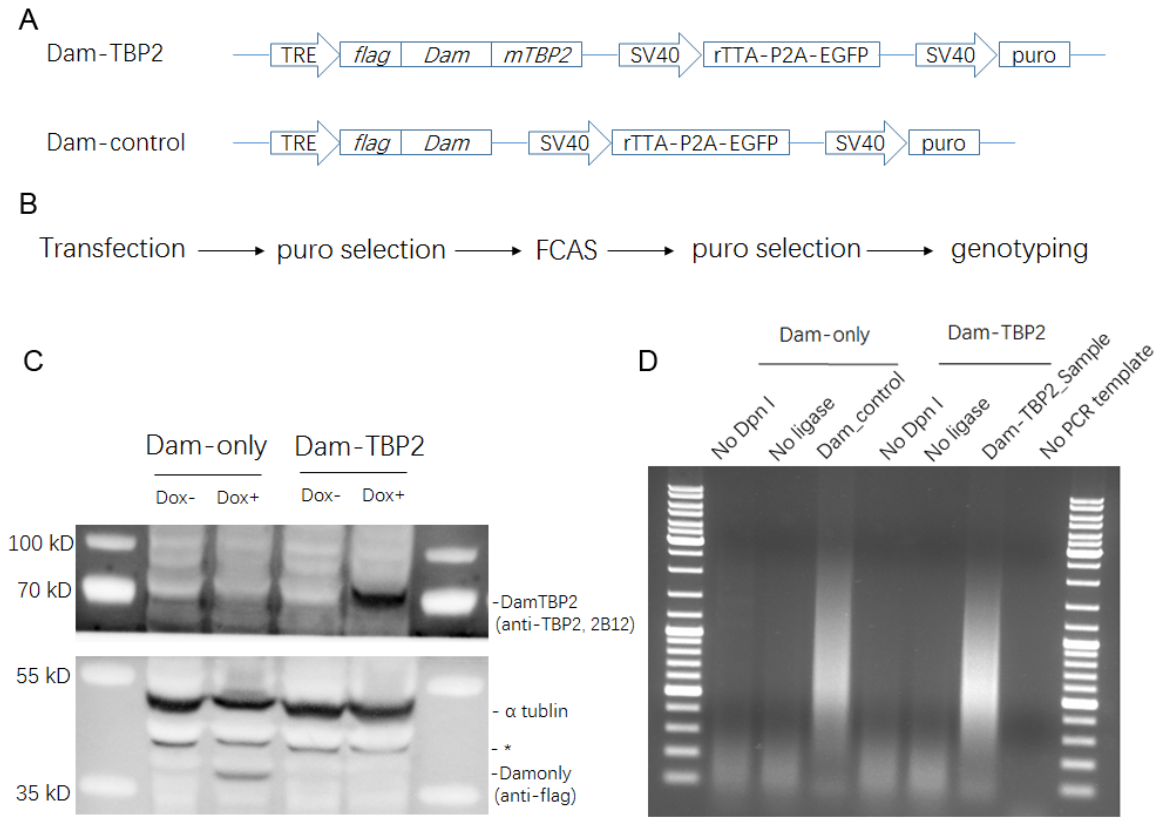
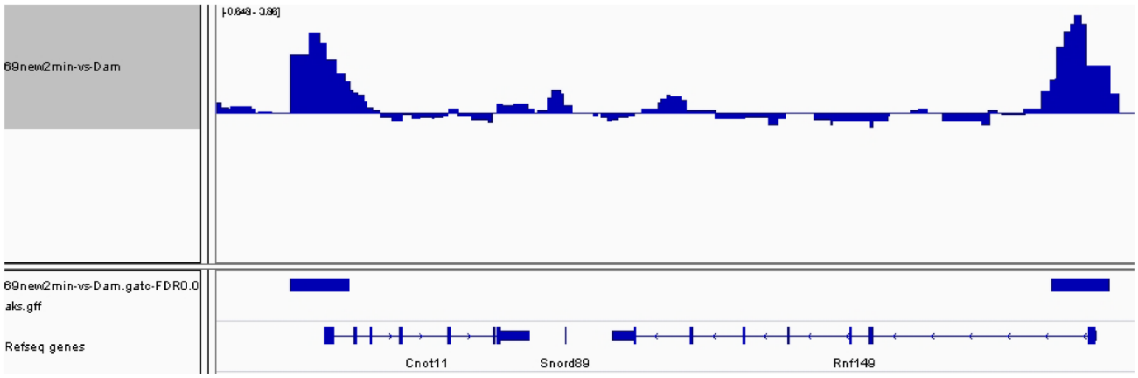


Figure 2-1: TBP2 DamID in NIH3T3 cells. (A) DamID plasmid construction. (B) Generation of *Dam-TBP2* and *Dam-only* NIH3T3 stable cell lines. (C) Validation of the stable cell lines by western blot. (D) Methyl-PCR with processed genomic DNA from *Dam-TBP2* and *Dam-only* NIH3T3 stable cell lines.

2.1.2 TBP2 occupancy revealed by DamID-seq

Methyl-PCR products were purified by column and sonicated for DamID sequencing. DamID-seq revealed that TBP2 binds to more than 13000 genes (Tao Ye, IGBMC) (***Figure 2-2 D***). A screenshot from IGV of some of the TBP2 bound genes is shown below (***Figure 2-2 A***). Some of the binding sites were validated by DamID-qPCR (***Figure 2-2 B***).

A



B

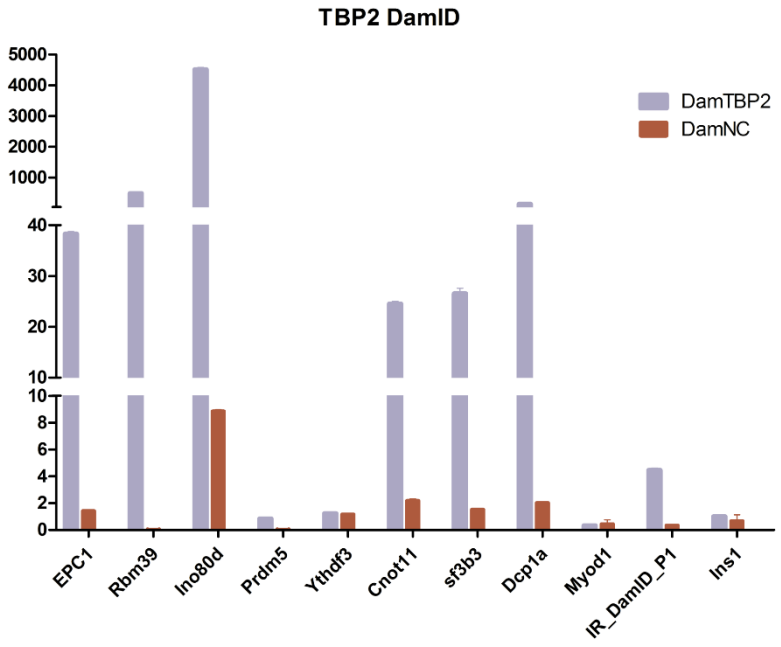


Figure 2-2: TBP2 DamID-seq and qPCR validation. (A) A screenshot (from IGV) of genes bound by TBP2. (B) DamID-qPCR validation of some genes.

2.2 TBP2 ChIP-seq in NIH3T3 cells

TBP2 ChIP-seq analysis was performed with II10 and K2 cells, and in parallel, TBP ChIP-seq was also performed with II10 cells. Astonishingly, TBP2 binds to more than 10000 genes, while TBP binds to less than 4000 genes, and more than 90% of the TBP bound genes are also bound by TBP2 (***Figure 2-2 C***). Consistently, ~80% of TBP2 bound genes are overlapped with the DamID-seq peaks (***Figure 2-2 C***), which shows that TBP2 binds to gene promoters genome-widely.

Moreover, although both TBP2 and TBP bind to gene promoters, the TBP2 average binding profile shifts downstream of the TSS indicating that TBP2 might be involved in a different TSS usage (***Figure 2-2 A&B***).

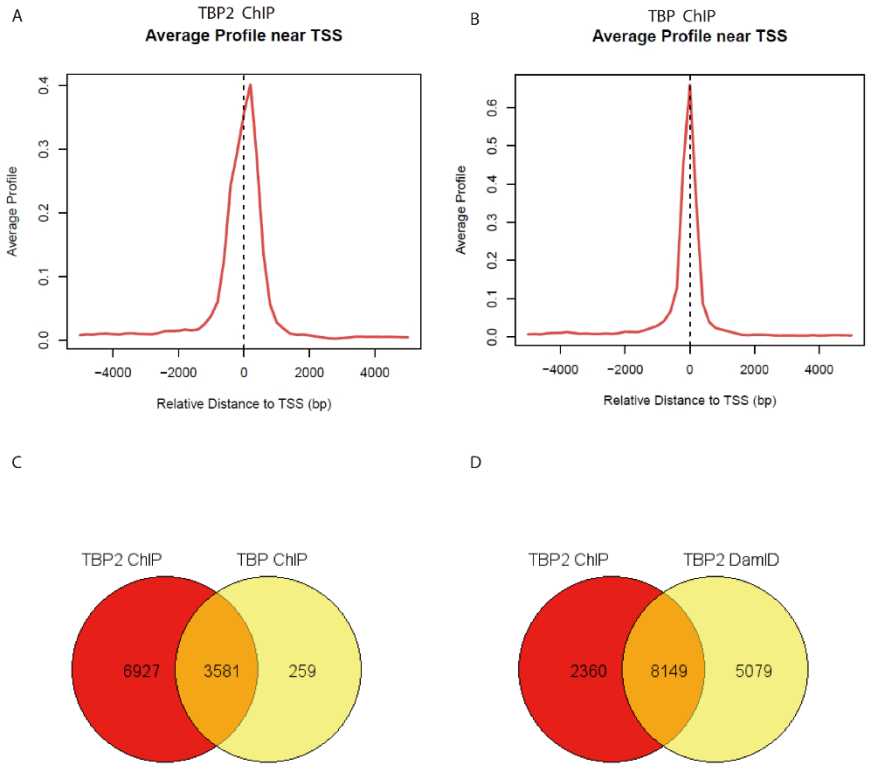


Figure 2-3: TBP2 binds to gene promoters. (A) Plots of average TBP2 binding profiles relative to TSSs of Pol II genes. (B) Plots of TBP profiles relative to TSSs. (C) Overlap of TBP2 and TBP binding genes. (D) Overlap of TBP2 bound genes revealed by ChIP-seq and DamID-seq.

2.3 RNA-seq analysis of TBP2 ectopically overexpressing NIH3T3 cells

To further analyze whether the promoter binding of TBP2 is functionally relevant or not, we performed RNA-seq analysis with II10 and K2 cells. Sequencing has been performed (GenomEAST platform, IGBMC) and data analysis is ongoing.

3. TBP2 binding profiles in oocytes

After analysis of the TBP2 binding profiles in NIH3T3 II10 cells, we performed DamID-seq in growing oocytes, and also tested a new approach named CUT&RUN to map TBP2 occupancy in oocytes.

3.1 TBP2 DamID-seq in growing oocytes

3.1.1 *Dam-Tbp2* and *Dam-only* capped mRNA synthesis

As described in the Material & Methods (***10.2)***, *Dam-Tbp2* cDNA and *Dam-only* cDNA were cloned into the pRN3P vector (***Figure 3-1 A***), and capped *Dam-Tbp2* and *Dam-only* mRNAs were obtained by *in vitro* transcription (***Figure 3-1 B***). Capped mRNAs were validated in GV oocytes (***Figure 3-1 C***): both *Dam-Tbp2* and *Dam-only* mRNAs are translated in GV oocytes and methylate genomic DNA as shown by the signal of a 'G^{m}ATC' tracer (a Dpn I truncation fused to eGFP).

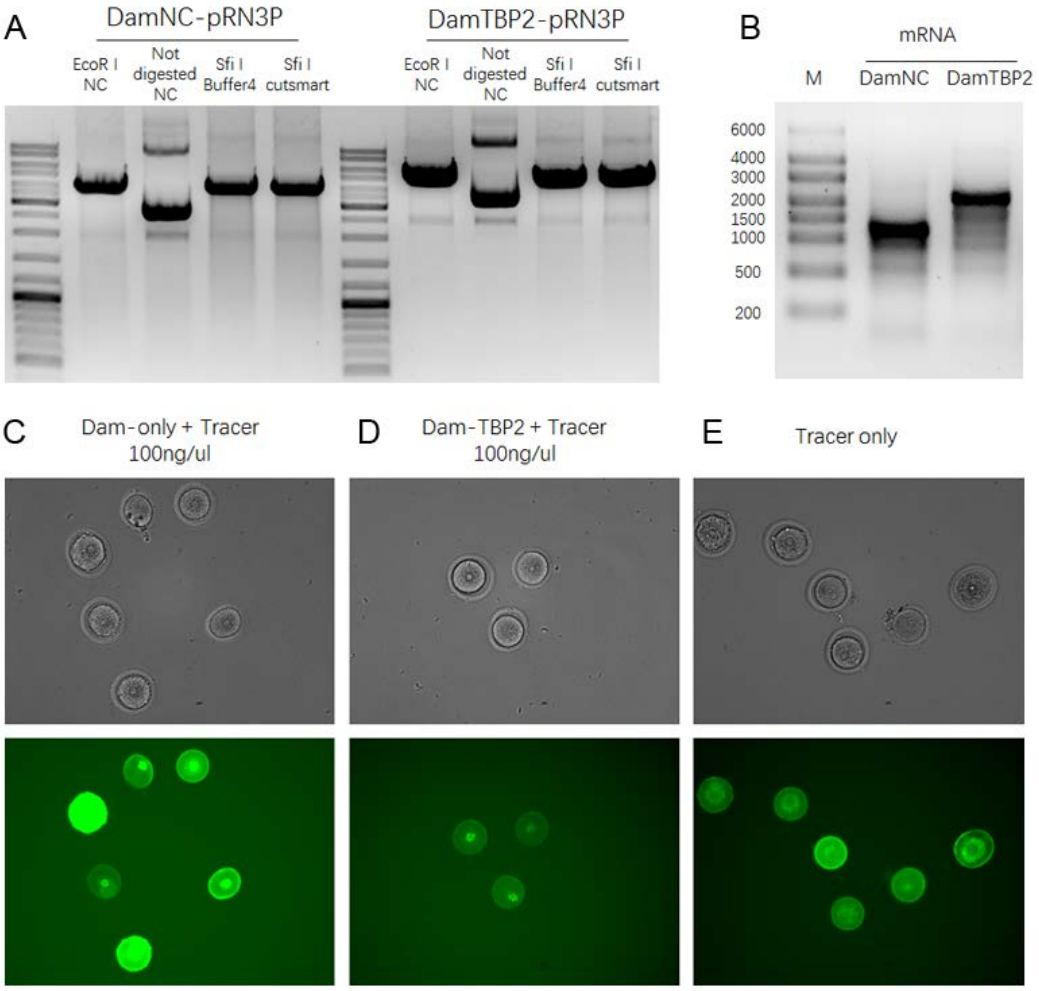


Figure 3-1: *Dam-Tbp2* and *Dam-only* mRNA synthesized by *in vitro* transcription. (A) *Dam-Tbp2* cDNA and *Dam-only* cDNA in pRN3P vector. (B) *Dam-Tbp2* and *Dam-only* capped

mRNAs synthesized by *in vitro* transcription. (C) *Dam-Tbp2* and *Dam-only* mRNAs validation in GV oocytes.

3.1.2 Oocyte injection with *Dam-Tbp2* and *Dam-only* mRNAs

P7 and P13 oocytes were collected as described in the Material & Methods ***8.1***, and cytoplasmically injected with 50ng/µL of *Dam-Tbp2* mRNA (co-injected with *mCherry*) (***Figure 3-2 A***). Control oocytes were injected with 20ng/µL *Dam-only* mRNAs. Oocyte injections were performed by Mate Borsos (Maria Elena Torres-Padilla Lab, Institute of Epigenetics and Stem Cells, Munich). After a ~24 hours culture (***Figure 3-2 B&C***), mCherry positive oocytes were collected, and subsequently used for methyl-PCR amplification.

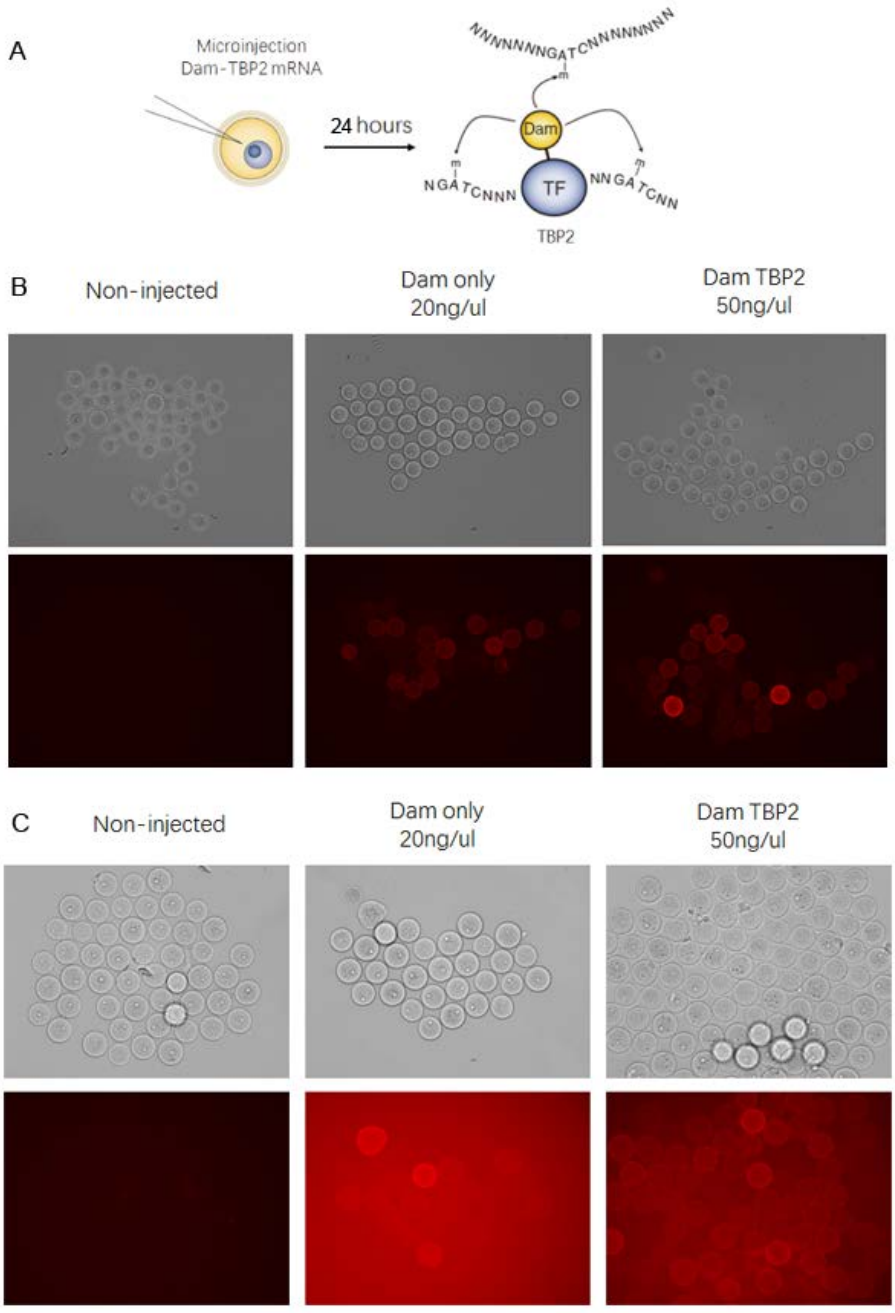


Figure 3-2: TBP2 DamID in oocytes. (A) Oocyte cytoplasmic injection of *Dam-Tbp2* and *Dam-only* mRNAs. (B) P7 oocytes after injection and incubation. (C) P13 oocytes after injection and incubation.

3.1.3 Methyl-PCR with oocyte samples

Genomic DNA of mCherry positive oocytes was digested with Dpn I, followed by DamID adaptor ligation, and the ligation samples were used for methy-PCR amplification (***Figure 3-3***).

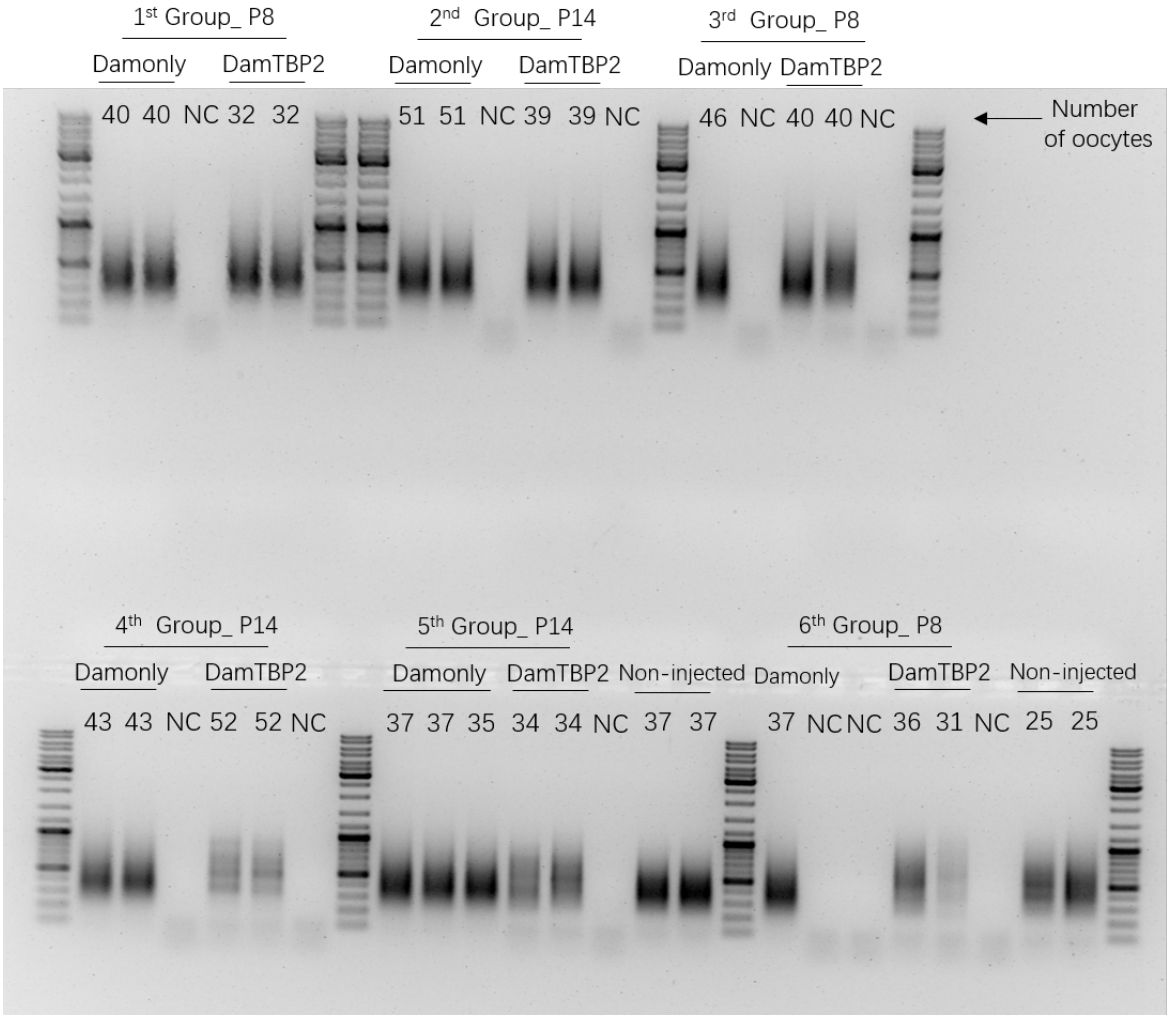


Figure 3-3: Methyl-PCR result of oocytes samples.

PCR products were purified by column. 300ng of purified Methyl-PCR products were used for sequencing libraries preparation (***Figure 3-4***). Sequencing using an Illumina Hi-seq2500 (iGE3 Genomics Platform, Geneva) is ongoing.

Figure 3-4: Size distribution of oocytes TBP2 DamID sequencing libraries. 300ng of purified PCR products were used for libraries preparation, and the size of the libraries mainly distribute between 150bp to 1500bp (GenomEAST platform, IGBMC).

3.2 TBP2 uliCUT&RUN with oocytes

Since the resolution of the TBP2-DamID-seq might not be optimal, I also tested a novel and more sensitive technique named CUT&RUN (Material & Methods ***11***) (***Figure 3-5***). TBP2 CUT&RUN was performed with 500 oocytes, and sequencing libraries were prepared with the recovered DNA fragments, and the libraries are now under sequencing (GenomEAST platform, IGBMC).

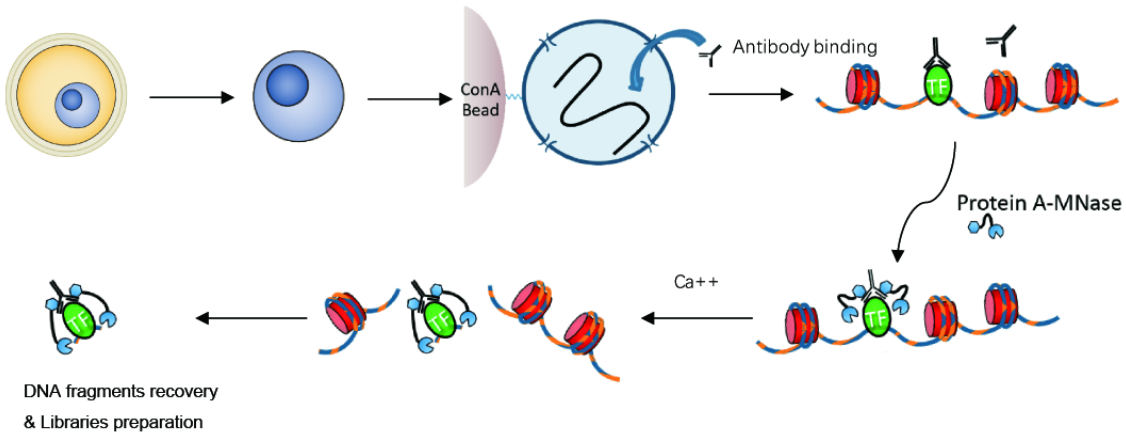


Figure 3-5: Schematic diagram of TBP2 CUT&RUN with oocytes. Adapted from (Skene et al., 2017).

4. Analysis of the direct effect of TBP2 on active transcription

As mentioned in introduction ***2.2.7***, nascent RNAs are released from Pol II during transcription termination. They are rapidly exported to the cytoplasm, where their half-lives are generally determined by the cytoplasmic mRNA decay pathways. Emerging evidences have shown the existence of transcript buffering, a phenomenon that steady-state levels of mRNAs are ‘‘buffered’’ when mRNA synthesis or degradation is impaired by inactivation of either the transcription complexes or the enzymes responsible for cytoplasmic mRNA decay (Shalem et al., 2011; Sun et al., 2012; Haimovich et al., 2013; Sun et al., 2013; Rodriguez-Molina et al., 2016; Baptista et al., 2017; Warfield et al., 2017; Timmers et al., 2018). Moreover, during folliculogenesis, oocyte stores large amount of RNAs (Bettegowda et al., 2007). Therefore, in order to better understand TBP2’s function in oocyte transcription, nascent RNA analysis is required.

Using an oocyte-specific expression of the uracil phosphoribosyl-transferase (UPRT) that allows specific 4-thiouracil (4TU) labeling of nascent RNAs in the oocyte (Lewandoski et al., 1997; Gay et al., 2013), coupled to thiol(S)-linked alkylation for the metabolic sequencing of RNA (SLAM-Seq) (Herzog et al., 2017), I performed nascent RNA analysis in both P14 WT oocytes and P14 *Tbp2*-deficient oocytes. 4TU was intraperitoneally injected in P14 *Uprt*$^{Tg/+}$*;Tg(Zp3-cre/+)* and P14 Tg(*Uprt/+);Tg(Zp3-cre/+)*;*Tbp2*$^{-/-}$ female pups. Six hours after 4TU injection, the oocytes were collected and total RNA were isolated (see ***8.1***) for SLAM-seq (Herzog et al., 2017) (***Figure 3-5***).

Total RNA samples from ~300 oocytes of both WT and *Tbp2*-deficient oocytes were prepared in triplicates. Libraries are prepared using the QuantSeq 3’mRNA-Seq Library Prep Kit, as published (Herzog et al., 2017).

In addition, total RNA from *in vitro* cultured oocytes were also prepared, including one sample from ~600 P14 WT oocytes as a background control, one sample from ~600 P14 WT oocytes cultured with 4sU that can be directly incorporated into the RNA, one sample from ~600 P14 Tg(*Uprt/+);Tg(Zp3-cre/+)* oocytes cultured with 4sU and one last sample from ~600 P14 Tg(*Uprt/+);Tg(Zp3-cre/+)* oocytes cultured with 4TU.

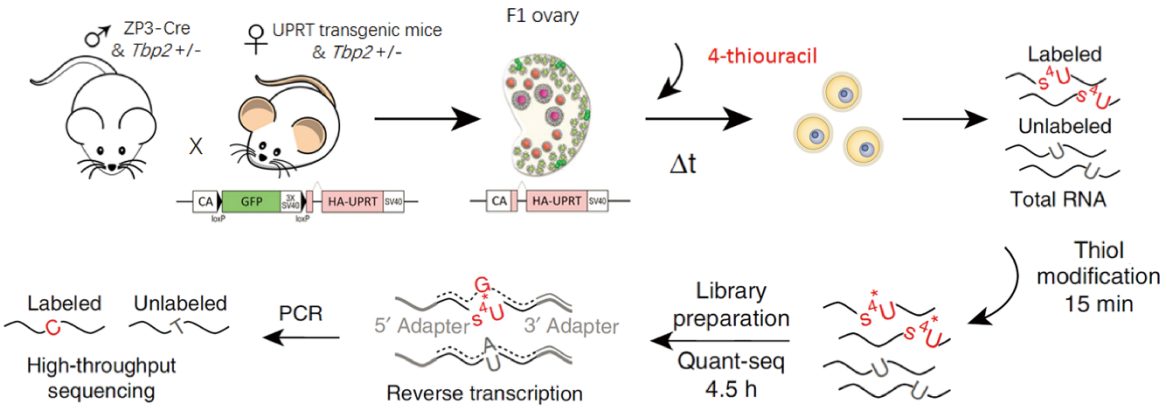


Figure 4-1: Oocyte nascent transcriptome analysis strategy. Nascent RNA study using TU-tagging coupled to SLAM-seq. Adapted from (Cordeiro et al., 2015; Herzog et al., 2017).

5. Potential TBP2-TFIIA interacting factors

In order to identify potential oocyte specific partners of the TBP2-TFIIA complex, we further analyzed the mass-spec results of the anti-TBP2 IP from ovary extracts and we found that among the triplicates, there are 247 common proteins (***Figure 5-1 A***), among which, 161 proteins overlap with proteins pulled down by TBP2 after the triple sequential IP (see manuscript) (***Figure 5-1 B***). We analyzed the 123 common non-ribosomal proteins by GO analysis, many proteins are linked to RNA binding, interestingly, there are also a few proteins are linked to transcription factor and cofactors, such as GATA4, SETDB2, TRIM28 and ZAR1.

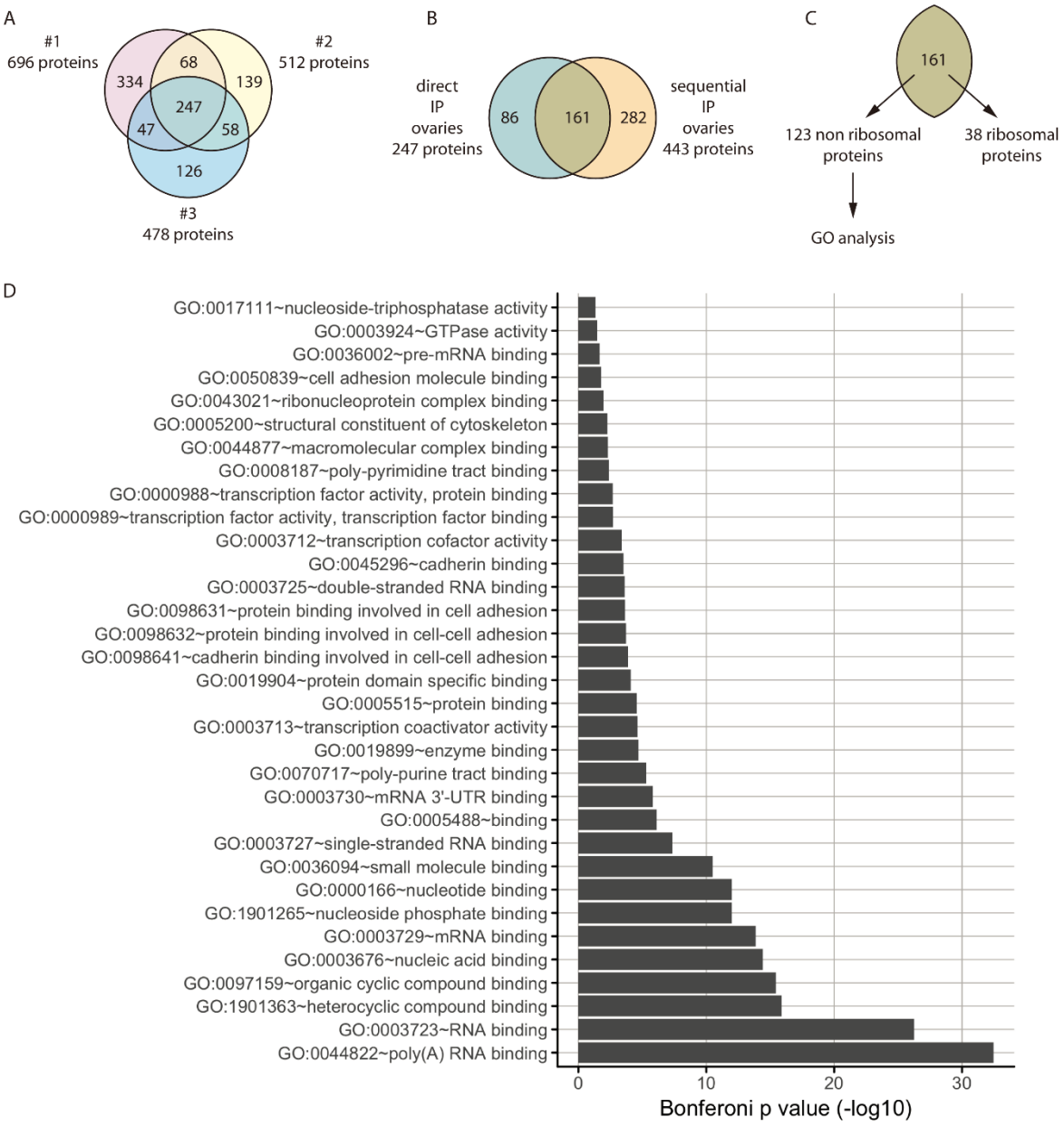


Figure 5-1: Potential TBP2-TFIIA interacting factors. (A) Common proteins in anti-TBP2 IP mass-spec results. (B). (C) Overlap between 'direct TBP2 IP' and 'sequential TBP2 IP'

identified proteins without non-ribosomal proteins. (D) Go analysis of common proteins without non-ribosomal proteins.

6. General coactivator SAGA might be required in oocytes

In parallel with the conditional deletion of *Taf7* during oocyte growth (see in the manuscript) that show that the TFIID-specific TAF7 is not required for transcription in the oocyte, I also carried out a conditional deletion of *Taf10* during oocyte growth using the Tg(*Zp3-Cre)* transgenic line (***Figure 6-1 A***). Interestingly, the ovaries size of 6-week-old Tg(*Zp3-Cre/+)*;*Taf10*$^{flox/\triangle}$ female is much smaller compared to Tg(*Zp3-Cre/+)*;*Taf10*$^{\triangle/+}$ littermates (n≥3) (***Figure 6-1 B***). While control ovaries (Tg(*Zp3-Cre/+)*;*Taf10*$^{\triangle/+}$) show obvious mature follicles, such structures are absent in Tg(*Zp3-Cre/+)*;*Taf10*$^{flox/\triangle}$ ovaries. Moreover, after superovulation, no mature oocytes were obtained (n＞6). As TAF10 is not only part of TFIID but also of the SAGA co-activator complex (Helmlinger et al., 2017), this defect may arise from SAGA specific function, as it has been shown that SAGA assembly is severely affected when *Taf10* is deleted in the embryo (Bardot et al., 2017).

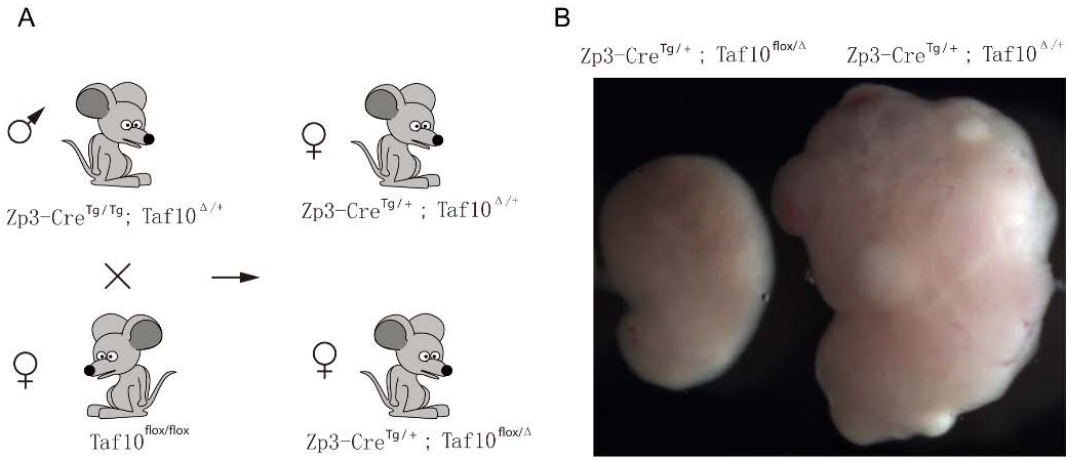


Figure 6-1: Conditional deletion of *Taf10* in growing oocytes impairs oocyte growth. (A) Scheme of the breeding and experiments. (B) Morphology of ovaries from 6-week-old Tg(*Zp3-Cre/+)*;T*af10*$^{\triangle/+}$(control) and Tg(*Zp3-Cre/+)*;*Taf10*$^{flox/\triangle}$ (mutant) females.

7. Importance of the switch between TBP and TBP2-mediated transcription

7.1 Mouse model of *Tbp* cDNA knock-in at *Tbp2* locus

In order to test the importance of the switch between TBP and TBP2-mediated transcription during oocyte growth, we generated a TBP-TBP2 swap mouse model (***Figure 7-1 A***), which was designed to allow us to force the expression of TBP in growing oocytes by knocking in the TBP coding sequence into the *Tbp2* locus (***Figure 7-1 C***).

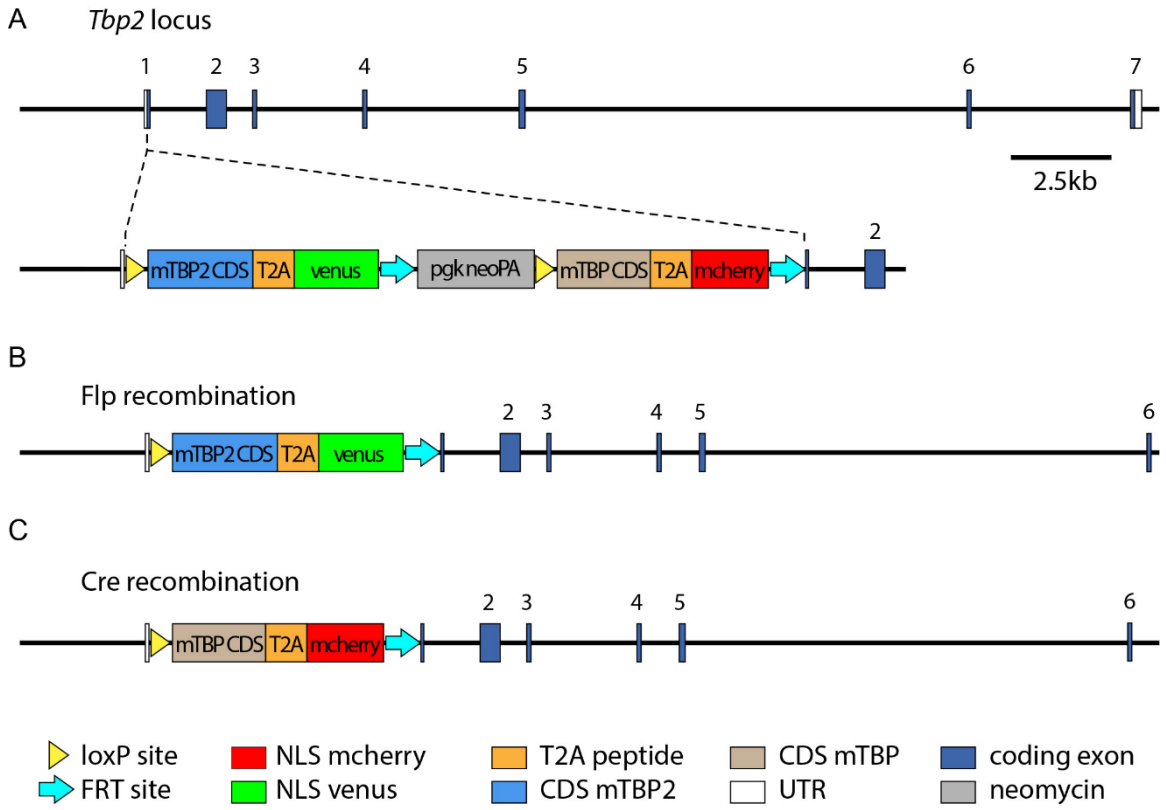


Figure 7-1: Mouse model of *Tbp* and *Tbp2* cDNA knock-in at the *Tbp2* locus. (A) Map of the *Tbp/Tbp2* cDNA knock-in at the *Tbp2* locus. (B) After recombination, the *Tbp2* cDNA is expressed as a control. (C) After recombination, the cDNA of *Tbp* is expressed under the control of the *Tbp2* regulatory elements.

According to our knock-in strategy, after or recombination, either the cDNA of *Tbp2* or the cDNA of *Tbp* will be expressed in the growing oocytes under the control of the *Tbp2* regulatory elements. Moreover, to facilitate the detection of the expression from the diferent modified alleles, *Tbp* and *Tbp2* cDNA were fused to the cDNA of mCherry and Venus, respectively. In order to avoid any dominant negative effects from the fusion, we used a T2A linker to produce

separate proteins. The plasmids were built by Jean Marie Garnier (IGBMC).

However, after recombination, neither Venus signal from *Tbp2*$^{TBP2:Venus}$ mouse nor mCherry signal from *Tbp2*$^{TBP:mCherry}$mouse could be detected. More strangely, control homozygous *Tbp2*$^{TBP2:Venus:TBP2:Venus}$ female are infertile (more than 5 breeding cages).

I analyzed expression of *Tbp*, *Tbp2*, *Tbp-mCherry* and *Tbp2-Venus* by qPCR using cDNA synthesized with mRNA obtained from WT, *Tbp2*$^{TBP:mCherry/+}$, *Tbp2*$^{TBP:mCherry/TBP:mCherry}$, *Tbp2*$^{TBP2:Venus/+}$ and *Tbp2*$^{TBP2:Venus/TBP2:Venus}$ oocytes. While *Tbp* mRNA is present in all the samples, *Tbp-mCherry* and *Tbp2-Venus* are not expressed (***Figure 7-2***). Moreover, the expression of *Tbp2* is also abolished in both *Tbp2*$^{TBP2:Venus/TBP2:Venus}$ and *Tbp2*$^{TBP:mCherry/TBP:mCherry}$ oocytes (***Figure 7-2***), indicating that *Tbp2*$^{TBP:mCherry/TBP:mCherry}$ and *Tbp2*$^{TBP2:Venus/TBP2:Venus}$ mice are *Tbp2* null mice.

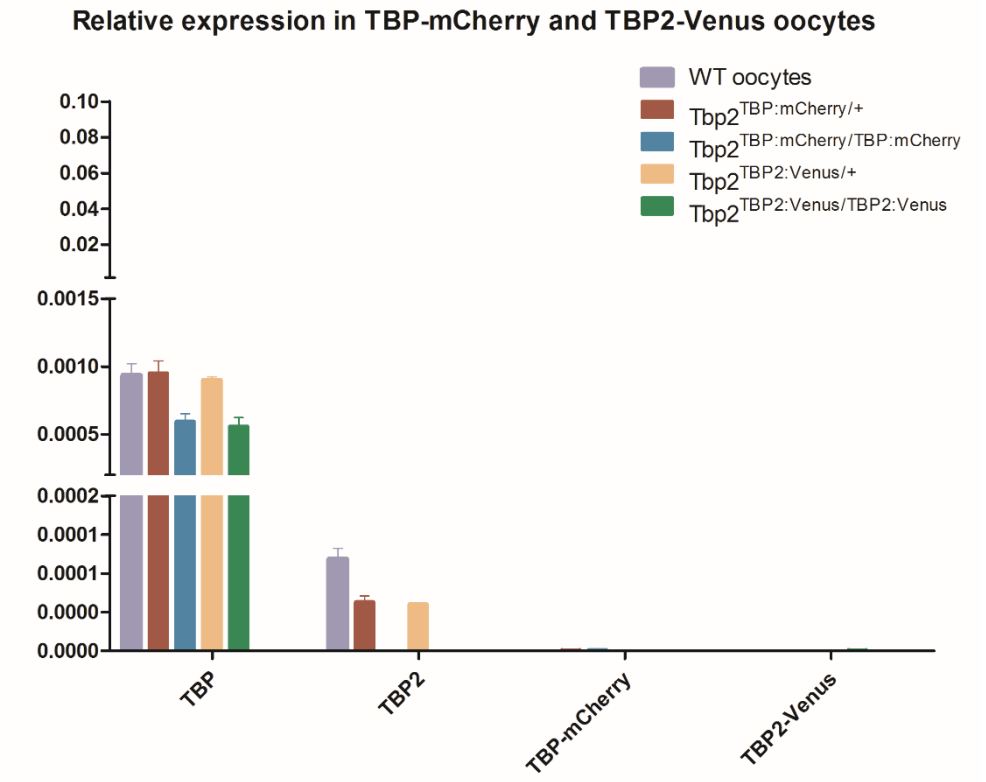


Figure 7-2: Gene expression in *Tbp2*$^{TBP:mCherry}$ and *Tbp2*$^{TBP2:Venus}$ oocytes. Expression of *Tbp*, *Tbp2*, *Tbp-mCherry* and *Tbp2-Venus* were analyzed by RT-qPCR using mRNA obtained from WT, *Tbp2*$^{TBP:mCherry/+}$, *Tbp2*$^{TBP:mCherry/TBP:mCherry}$, *Tbp2*$^{TBP2:Venus/+}$ and *Tbp2*$^{TBP2:Venus/TBP2:Venus}$ oocytes. Expression was normalized to 18s rRNA.

I then asked the question why this knock-in strategy was not working as expected. An hypothesis was that *Tbp2-Venus* mRNA is degraded by RNA non-sense mediated decay (NMD) pathway. Therefore I cultured P14 *Tbp2*$^{TBP2:Venus/+}$ oocytes with or without emetine, a NMD inhibitor. After overnight incubation, the oocytes were collected for mRNA and cDNA

preparation, followed by qPCR analysis (***Figure 7-3***). cDNA samples prepared from both WT oocytes and nontreated *Tbp2*$^{TBP2:Venus/+}$ oocytes were used as control. Expression of *Tbp2-Venus* cannot be detected in any of the samples apart from the genomic DNA (***Figure 7-3***), indicating that the absence of *Tbp2-Venus* mRNA is not due to non-sense mediated decay, suggesting that *Tbp2*$^{TBP2:Venus}$ is not transcribed.

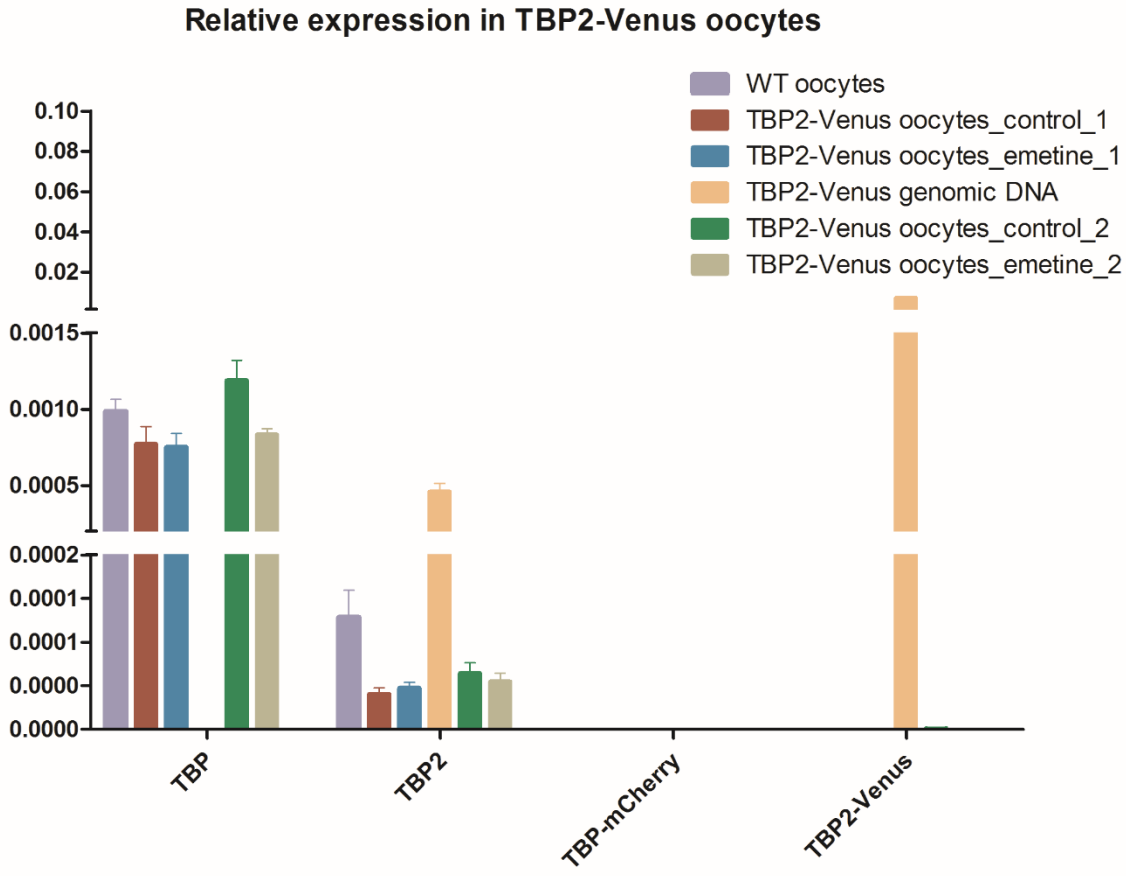


Figure 7-3: Gene expression in *Tbp2*$^{TBP2:Venus/+}$ oocytes after treatment with emetine. Expression of *Tbp*, *Tbp2*, *Tbp-mCherry* and *Tbp2-Venus* were analyzed by RT-qPCR using mRNA obtained from emetine treated *Tbp2*$^{TBP2:Venus/+}$ and control oocytes. Genomic DNA from *Tbp2*$^{TBP2:Venus/+}$ was used as a positive control for the *TBP2-Venus* primer. Expression was normalized to 18s rRNA.

7.2 Proteomics with oocytes

In order to study whether the TAFs are present at protein levels in growing oocytes, we performed proteomic analysis with P14 oocyte whole cell extracts from ~4000 oocytes in duplicates and carried out an anti-TBP2 IP. We then performed proteomic analysis of both the IP elution and the flow through by mass-spec. Unfortunately, as no TBP2 peptide was detected, the experiment was not conclusive.

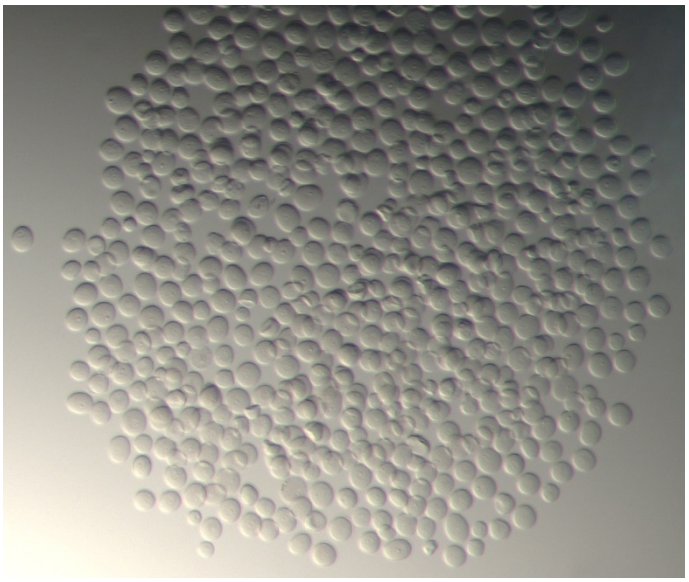


Figure 7-4: Oocytes collection for proteomics.

General discussion and perspectives

1. A TBP2-TFIIA-containing transcription machinery drives transcription in oocytes

The results presented in this thesis demonstrate that a TBP2-TFIIA-containing transcription machinery, different from TFIID, drives transcription in growing oocytes. Importantly, a detailed discussion can be found in the *Manuscript* in *Results* section, where the results are discussed in light of the current knowledge. The following discussion section aims to present the results in a general context and to reveal questions that are still not addressed.

1.1 Which form of TFIIA is associated with TBP2?

As described in introduction ***2.1.3*** and ***3.3.1***, two of the three TFIIA subunits, TFIIAα and TFIIAβ are encoded by a single gene and derived from the proteolytic cleavage of the TFIIAαβ precursor protein by Taspase1 (Zhou et al., 2006; Hoiby et al., 2007). Based on our proteomic results, we cannot determine whether the TBP2 associated TFIIA contains uncleaved TFIIAαβ, or processed TFIIAα and TFIIAβ. However, the cleavage of TFIIA is not necessary for its function in oocytes as *Taspase1*-deficient female mice are fertile (Oyama et al., 2013), indicating indirectly that TBP2 is probably associated with the uncleaved form of TFIIAαβ. Further structural studies are required to determine which form of TFIIA (uncleaved or processed) interacts with TBP2 and how TBP2-TFIIA interacts with DNA.

1.2 Why are there genes upregulated following *Tbp2* ablation?

Our RNA-seq analysis has shown that many genes are downregulated in the *Tbp2*$^{-/-}$ oocytes. However, surprisingly, expression of a large number of genes (1577 at P7 and 1358 at P14) is upregulated following *Tbp2* ablation. As TBP2 is supposed to play a general role in Pol II transcription initiation, then how can this upregulation be explained?

Detailed analyses revealed that most of the upregulated genes are expressed at low levels. Moreover, among the downregulated genes, many are involved in mRNA decay. Thus, it is possible that the upregulated genes are indirectly regulated following *Tbp2* loss of function, and that the upregulation is due to mRNA decay defects in *Tbp2*$^{-/-}$ oocytes.

1.3 Is TBP2 generally required for Pol II transcription in oocytes?

In growing mouse oocytes, all three TBP-type factors are expressed at mRNA levels (Xiao

et al., 2006). However, at the protein level, TBP is undetectable, whereas TBP2 is highly expressed (Gazdag et al., 2007). Although it is not known whether TRF2 protein is present in the oocytes or not, *Trf2*-deficient mice do not show defects in female infertility (Martianov et al., 2001), indicating that TRF2 does not play crucial roles during oocyte growth, or at least its function can be compensated by TBP2. Altogether these data strongly suggest that TBP2 is required for most, if not all, Pol II transcription in oocytes.

In agreement with this hypothesis, we found that, in NIH3T3 cells ectopically overexpressing TBP2 (II10 cells), TBP2 binds to the promoter of more than 10000 genes. It is important to mention that in II10 cells, TBP2 can also be incorporated into a TFIID-like, complex. However, our RNA-seq analyses have shown that, in *Tbp2*$^{-/-}$ oocytes, many genes (1720 at P7 and 1794 at P14) are downregulated, but they do not represent the majority of the genes that are expressed in the oocytes. A possible explanation could be that TRF2 is also involved in RNA pol II transcription, at least as a compensatory mechanism. Another possibility is that, since our RNA-seq analysis only gives a snapshot view of the steady state RNAs, transcript buffering mechanisms (Timmers et al., 2018) could mask more profound effect on transcription initiation. The ongoing experiments of TBP2 chromatin occupancy mapping and nascent transcripts analysis in oocytes will address this question.

1.4 How is TBP2 recruited to promoters without TAFs?

TAFs have been suggested to be important for recognition and binding to specific promoter elements, interaction with GTFs, as well as with activators and recognizing specific chromatin marks through their chromatin reading domains (Introduction ***2.1.2.2*** and ***2.3.1***). However our results show that in oocytes TBP2 is associated with TFIIA, but not with TAFs. TBP2 shares more than 90% identity with the TBP core domain and is able to bind the TATA-box (Persengiev et al., 2003; Bartfai et al., 2004), however, although TATA-box is enriched in TBP2 regulated genes, the majority of which do not contain a TATA-box (Manuscript ***figure 4A***), raising the question how is TBP2 recruited to TATA-less promoters in oocytes?

1.4.1 Possible involvement of histone H3.3 in TBP2 recruitment to promoters

It has been shown that after nuclear transfer in *Xenopus* oocytes, molecular events of nuclear reprogramming take place in a sequential manner (Jullien et al., 2014): firstly, a widespread binding of oocyte linker histone B4 replaces somatic linker histone H1; secondly, B4 around TSS is evicted due to the binding of H3.3 (Braunschweig et al., 2009), finally, it has been suggested that transcription may be driven by TBP2-containing transcriptional machinery

(Jullien et al., 2014) (Introduction ***Figure 3-7***). Interestingly, in mouse, double ablation of *H3f3a* and *H3f3b* in folliculogenesis leads to early primary oocyte death, demonstrating a crucial role for H3.3 in oogenesis (Tang et al., 2015b). Moreover, ablation of H3.3 chaperone *Hira* in developing mouse oocytes results in increased DNA accessibility and aberrant transcription, specifically, decreased expression of highly expressed genes and increased expression of low or non-expressed genes (Nashun et al., 2015).

Taken together, either H3.3 itself or the dynamic DNA accessibility generated by H3.3 deposition plays crucial roles for TBP2-mediated transcription in oocytes. Perhaps in the more compacted 4N oocyte genome, DNA accessibility is more difficult and thus becoming a key determining factor for transcription machinery recruitment. Nevertheless, further investigations will be necessary on genome-wide chromatin accessibility in growing oocytes.

1.4.2 Recruitment through other factors in the TBP2-containing PIC

All the GTFs except TFIID should be present in the TBP2-containing PIC. It is known that TFIIB can bind to BRE (Lagrange et al., 1998; Deng et al., 2005), it is thus possible that TBP2 can be recruited to the promoter though TFIIB. Besides, it is possible that TBP2-TFIIA complex contains other factors (unpublished results ***5***) which may help for the recruitment of TBP2.

1.5 Possible reason for the failure of TBP-TBP2 swap mouse model

In our TBP-TBP2 swap mouse model, the *Tbp* and *Tbp2* cDNA constructions have been knocked into *Tbp2* locus right after the start codon of endogenous *Tbp2*. However, neither *Tbp-mCherry* nor *Tbp2-Venus* is transcribed. Since the knock-in strategy does not change any 5' regulatory sequence of *Tbp2*, a possible explanation would be the existence of promoter-proximal regulatory elements in the introns, especially the first intron (Park et al., 2014). An experiment to test this would be to clone the promoter region and first intron of *Tbp2* with a reporter, followed by injection into growing oocytes.

1.6 Pol I and Pol III transcription in oocytes

TBP plays a crucial role in transcription initiation of all three RNA polymerases (Hernandez, 1993). Therefore, in growing oocytes, replacement of TBP by TBP2 (Gazdag et al., 2007; Muller et al., 2009), raises the question: how is Pol I and Pol III transcription regulated in oocytes?

In *Xenopus*, it has been reported that TBP2 is recruited to 5S rRNA Pol III promoter, suggesting that TBP2 may participate in Pol III transcription (Akhtar et al., 2009). However, in NIH 3T3 cells ectopically overexpressing TBP2, we did not detect any TBP2 association with

Pol I and Pol III co-factors (TAF1A, -1B, -1C and -1D or BRF1, respectively), in contrast to TBP, which interacts with these Pol I and Pol III cofactors. These results together suggest that TBP2 is not involved in Pol I and Pol III transcription in NIH 3T3 cells. Similarly in our anti-TBP2 ovary IPs we did not detect any Pol I and Pol III cofactors, further suggesting that TBP2 may not be involved in Pol I or Pol III transcription. Although in oocytes *Trf2* is expressed at mRNA level (Xiao et al., 2006), it is not known if TRF2 protein is present . Besides, no evidence has been shown that TRF2 is involved in Pol I and Pol III (Vo Ngoc et al., 2017b), and *Trf2*-deficient mice do not display defects in oocytes (Martianov et al., 2001; Zhang et al., 2001b). To explain which TBP-type factor can drive Pol I or Pol III transcription in oocytes, one possibility could be that residual levels of TBP may play a role in Pol I and Pol III transcription, alternatively, Pol I and Pol III transcription may be inactive in oocytes, however, this hypothesis requires further investigations.

2. Why is there a switch between TBP and TBP2-mediated transcription during oocyte growth?

2.1 TAFs proteins seem not be expressed in growing oocytes

Our RNA-seq data have shown that all the *Taf* genes (except *Taf7l*) are expressed in the oocytes at the mRNA levels. However, we were unable to detect TAF proteins in our proteomic analysis carried out on extracts prepared from growing oocytes (***unpublished data 7.2***), raising the question whether these proteins are indeed expressed in oocytes or whether their protein expression is under the detection limit of the experimental setup.

We have shown that in 3T3-II10 cells, besides the TBP2-TFIIA complex, TBP2 can also incorporate into a TFIID-type complex (***Manuscript Figure 6***), indicating that TBP2 can associate with TAFs. This is in good agreement with the high homology between the core domains of TBP and TBP2. However, in oocytes TBP2 associates only with TFIIA, but does not form a TFIID-type complex (***Manuscript Figure 7***). If the TAF proteins were expressed in oocytes, TBP should be able to interact with them, like in the TBP2 ectopically overexpressing cells. Thus, these results further suggest (however indirectly) that TAF proteins may not be expressed in oocytes, and thus this could be one possible reason for the basal transcription machinery switch.

2.2 The non-canonical epigenetic patterns require a different transcription machinery

As discussed in ***Introduction 1.3.2.3***, oocytes possess non-canonical pattern of H3K4me3 (ncH3K4me3) that is enriched at low levels across large genomic regions (Zhang et al., 2016; Dahl et al., 2016; Liu et al., 2016; Hanna et al., 2018b). In non-growing oocytes, where TBP is still present, H3K4me3 is restricted to active promoters (Hanna et al., 2018b). However, at further stages of oogenesis, when TBP is replaced by TBP2 (Gazdag et al., 2007), H3K4me3 accumulates at intergenic regions, putative enhancers and silent promoters through the recruitment of MLL2 to unmethylated CpG-rich regions in a transcription-independent manner (Hanna et al., 2018b), leading to the formation of ncH3K4me3 pattern. ncH3K4me3 pattern is maintained until early 2-cell stage and removed after zygotic genome activation (ZGA) at the late 2-cell stage (Zhang et al., 2016; Dahl et al., 2016; Liu et al., 2016). Interestingly, at the same time TBP protein reappears in zygotes at low level and accumulates by 2-cell stage, allowing Pol II transcription initiation (Gazdag et al., 2007; Muller et al., 2009).

Importantly, it has been shown that TAF3 binds to H3K4me3 and facilitates PIC assembly

(Vermeulen et al., 2007; van Ingen et al., 2008; Lauberth et al., 2013), and that H3K4me3, through interaction with TAF3, can direct PIC formation either independently or cooperatively with the TATA box (Lauberth et al., 2013) (see ***Introduction 2.1.2.2*** and ***2.3.1.3***). Thus, the TAF3-H3K4me3 interaction provides a way to recruit TFIID to the core promoter regions. Moreover, it was proposed that ncH3K4me3 could function as 'sponges' to absorb and sequester transcription factors and regulators, therefore diluting transcription resources away from promoters (Zhang et al., 2016). It is thus conceivable that, with the presence of ncH3K4me3 in oocytes, the canonical TBP-containing TFIID that contains TAF3, would not be able to find its targets efficiently, and that oocytes would require a TAF3-free transcription machinery to drive gene expression. This machinery could then be TBP2-TFIIA.

2.3 Lack of enhancer function in oocytes

It has been reported that mouse oocytes and fertilized eggs are not able to utilize enhancers (Lawinger et al., 1999), and recent studies suggest that that there may be enhancer function at the onset of oogenesis, which may be progressively eroded thereafter (Hanna et al., 2018b). Given the important role of enhancers in the transcription activation in the canonical transcription cycle, it is conceivable that lack of enhancer function in oocytes would require a different, a canonical TFIID-lacking, transcription machinery.

3. Other potential functions of TBP2

3.1 Different TSS usage of in oocytes

Our results also show that after its ectopic expression, TBP2 can bind to gene promoters genome-widely, however, the average binding profile of TBP2 shifts to the downstream of the canonical TBP-dependent TSS (***unpublished data 2.2***), suggesting that TBP2 would use a different TSS than TBP.

Interestingly, it has been shown that in zebrafish there is a switch in TSS usage throughout maternal to zygotic transition (Haberle et al., 2014), and it has been suggested that transcription initiation in the zebrafish oocyte may be mediated by TBP2 (Haberle et al., 2014). With the new published SLIC-CAGE method (Cvetesic et al., 2018), it is possible to analyze TSS usage from as little as 5-10 ng total RNA. Thus, it would be interesting to investigate whether TBP2-mediated transcription uses different TSSs in mouse growing oocytes or not.

3.2 Possible involvement of TBP2 in *de novo* DNA methylation

Several studies have shown that *de novo* methylation in mouse oocytes, at least at imprinted gDMRs, is coordinated with oocyte growth (Obata et al., 2002; Lucifero et al., 2004; Hiura et al., 2006), and acquisition of DNA methylation in oocytes is known to be positively correlated with transcription (Chotalia et al., 2009; Smith et al., 2011; Veselovska et al., 2015). Given the transcription defects in *Tbp2* knockout oocytes, it would be interesting to analyze how TBP2 influences *de novo* DNA methylation in the oocyte.

3.3 Possible function of TBP2 in activation of zygotic genes

It has been shown that, in *Xenopus* and zebrafish embryos, TBP and TBP2 are both indispensable for embryonic development and are both required for activation of zygotic genes (reviewed in (Muller et al., 2009)). In mouse oocyte, TBP2 accumulation gradually decreases after ovulation, is almost undetectable after fertilization by the two-cell stage (Gazdag et al., 2007), however, it is not known whether the residual TBP2 has any function in activation of zygotic genes before it totally disappears.

It has been shown that DUX-family transcription factors regulate ZGA in mammals (De Iaco et al., 2017; Hendrickson et al., 2017; Whiddon et al., 2017), and that *Dppa2* and *Dppa4* directly regulate the Dux-driven zygotic transcription program (Eckersley-Maslin et al., 2018). Interestingly, *Dppa2* is downregulated more than 4-fold in *Tbp2*$^{-/-}$ oocytes at both P7 and P14, and our preliminary data of TBP2 DamID-seq in oocytes showed that TBP2 binds to *Dux* gene. Further study, for instance, by TBP2 Trim-Away in MII oocytes, could address this question.

Conclusions

During my thesis, I have explored the role of TBP2 in controlling transcription initiation during oocyte growth, and have characterized the oocyte-specific TBP2-TFIIA-containing transcription machinery.

We have shown that *Tbp2* loss of function results in the main decrease in the expression of the most abundantly expressed genes as well as a specific down-regulation of the expression of the MaLR retroviral elements by performing RNA-seq analyses of wild-type and *Tbp2*$^{-/-}$ oocytes from both primary and secondary follicles.

Using an optimized immunoprecipitation method coupled to mass spectrometry and mouse molecular genetics, we have shown that in the growing oocytes, TBP2 does not assemble into a TFIID-type complex but associates with TFIIA to form a specific TBP2-TFIIA-containing transcription machinery.

By studying TBP2 occupancy in the II10 cells that ectopically over-express TBP2, we have shown that TBP2 binds to gene promoters genome-widely and the average binding profile shifts a bit downstream of the TSS, suggesting that TBP2 might be involved in a different TSS usage.

Additionally, we have applied TBP2-Dam-ID-seq, TBP2 uliCUT&RUN and carried out nascent transcripts analyses using SLAM-seq coupled with mouse TU-tagging approach, those ongoing studies will give us more insights about the role of TBP2 during oocyte growth.

Altogether, these results show that a specific TBP2-TFIIA-containing transcription machinery, different from canonical TFIID, drives transcription in mouse growing oocytes.

Appendix

Ravens *et al. Epigenetics & Chromatin (2015) 8:45*
DOI 10.1186/s13072-015-0039-z

Epigenetics & Chromatin

RESEARCH

Open Access

Tip60 complex binds to active Pol II promoters and a subset of enhancers and co-regulates the c-Myc network in mouse embryonic stem cells

Sarina Ravens[1], Changwei Yu[1], Tao Ye[2], Matthieu Stierle[1] and Laszlo Tora[1*]

Abstract

Background: Tip60 (KAT5) is the histone acetyltransferase (HAT) of the mammalian Tip60/NuA4 complex. While *Tip60* is important for early mouse development and mouse embryonic stem cell (mESC) pluripotency, the function of Tip60 as reflected in a genome-wide context is not yet well understood.

Results: Gel filtration of nuclear mESCs extracts indicate incorporation of Tip60 into large molecular complexes and exclude the existence of large quantities of "free" Tip60 within the nuclei of ESCs. Thus, monitoring of Tip60 binding to the genome should reflect the behaviour of Tip60-containing complexes. The genome-wide mapping of Tip60 binding in mESCs by chromatin immunoprecipitation (ChIP) coupled with high-throughput sequencing (ChIP-seq) shows that the Tip60 complex is present at promoter regions of predominantly active genes that are bound by RNA polymerase II (Pol II) and contain the H3K4me3 histone mark. The coactivator HAT complexes, Tip60- and Mof (KAT8)-containing (NSL and MSL), show a global overlap at promoters, whereas distinct binding profiles at enhancers suggest different regulatory functions of each essential HAT complex. Interestingly, Tip60 enrichment peaks at about 200 bp downstream of the transcription start sites suggesting a function for the Tip60 complexes in addition to histone acetylation. The comparison of genome-wide binding profiles of Tip60 and c-Myc, a somatic cell reprogramming factor that binds predominantly to active genes in mESCs, demonstrate that Tip60 and c-Myc co-bind at 50–60 % of their binding sites. We also show that the Tip60 complex binds to a subset of bivalent developmental genes and defines a set of mESC-specific enhancer as well as super-enhancer regions.

Conclusions: Our study suggests that the Tip60 complex functions as a global transcriptional co-activator at most active Pol II promoters, co-regulates the ESC-specific c-Myc network, important for ESC self-renewal and cell metabolism and acts at a subset of active distal regulatory elements, or super enhancers, in mESCs.

Keywords: Histone acetyltransferase (HAT), KAT5, H3K27ac, H3K4me3, Enhancers, Super enhancers, Mouse, Pluripotency, Bivalent genes, c-Myc, Mof, NSL, MSL

Background

The histone acetyl trasferase (HAT), Tip60 (Tat interactive protein 60 kDa, also called KAT5) belongs to the MYST family of HATs that play key roles in acetylation of histones and other nuclear factors and thus influence chromatin structure and transcription regulation in the eukaryotic nucleus [1]. The defining feature of the MYST family of HATs is the presence of the highly conserved MYST domain, composed of an acetyl-CoA binding motif and a zinc finger [2]. The majority of cellular Tip60 exists in a stable nuclear multiprotein complex, called the mammalian Tip60 complex, consists of at least

*Correspondence: laszlo@igbmc.fr
[1] Cellular Signalling and Nuclear Dynamics Programme, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U964, Université de Strasbourg (UdS), BP 10142, 1 Rue Laurent Fries, CU de Strasbourg, 67404 Illkirch Cedex, France
Full list of author information is available at the end of the article

18 subunits and performs most transcription- and DNA damage-related Tip60 functions [3, 4]. The yeast (y) homologue of Tip60 is the yEsa1 HAT that is a subunit of the yNuA4 complex [5]. This yNuA4 HAT complex, as well as the human Tip60-containing complex, contains a large number of homologue subunits [6, 7]. In addition, the mammalian Tip60 complex seems to combine the functions of the yNuA4 HAT and the ySWR1 ATP-dependent chromatin remodelling complexes into a single complex [8]. The ATPase p400, belonging to the SWI2/SNF2 class of ATP-dependent chromatin remodelers [9], is an E1A-interacting protein essential for E1A-dependent apoptosis and cellular transformation [10]. The isolated mammalian Tip60 complexes were suggested to be heterogeneous, with a population that would contain p400 and another that would not, suggesting a dynamic assembly of the p400-containing Tip60 complex [7, 9, 11].

Tip60 complexes have three enzymatic functions: (1) a histone H2A/H4 acetlytransferase activity, (2) an ATP-dependent H2AZ.H2B dimer exchange activity and (3) a helicase activity [7, 8]. Several studies have shown that Tip60/NuA4-type complexes are involved in diverse cellular processes including transcription, cell cycle control, apoptosis, cell proliferation and DNA repair [4]. Mammalian Tip60 has been described as a transcriptional co-activator complex that is supposed to mediate the action of large variety of transcription factors, including nuclear receptors, c-Myc, STAT3, NF-kappaB, E2F1, p53 and others [4]. Importantly, a mass-spectrometry based study demonstrated that the intact Tip60-p400 (NuA4) HAT complex interacts with Myc and suggested that histone 3 and 4 acetylation patterns may be generated in part by interactions of Myc with the Tip60-p400 complex through Tip60 in mouse embryonic stem cells (mESCs) [12].

Homozygous knockout of the *Tip60* gene in mouse results in pre-implantation lethality at embryonic day 3.5 [13]. Additionally, seven subunits of the Tip60 complex, including Tip60 and p400, have been further identified in an RNAi screen to be required for mESC maintenance [14]. Moreover, siRNA down-regulation of six other components of the Tip60-complex exhibited the same phenotypic defects in alkaline phosphatase activity, embryonic body formation and teratoma formation as Tip60. This indicates that the whole Tip60 complex is necessary for mESC maintenance and normal mESC identity [14]. Interestingly, siRNA-based depletion of Tip60 and p400 in mESCs resulted in an impaired expression of developmental regulators and expression of these affected genes significantly overlapped with that regulated by Nanog in mESCs [14]. Chromatin immunoprecipitation (ChIP) linked to hybridization to promoter tiling arrays indicated that p400 localization correlates with H3K4me3 at both active and silent genes in mESCs [14], though no anti-Tip60 ChIP or ChIP-seq was carried out in this study. Surprisingly, mRNA expression analyses identified that only about 800 genes were differentially regulated in both Tip60 and p400 knock-down mESCs [14, 15]. Moreover, a recent study demonstrated that Flag-Tip60-containing complexes bind to active and developmental genes in mESCs [14, 15].

Interestingly, an additional HAT, Mof (males absent on the first or KAT8) was shown to be required for early mouse development and mESC pluripotency [16, 17]. Recently, it has been shown that Mof-associated complexes have overlapping and distinct roles in mESCs [18, 19]. We hypothesise that there is a complex interplay between different transcriptional co-factors and that both Tip60- and Mof- containing complexes have distinct role in mESCs. To better characterize the genome-wide action of the Tip60 complex, we carried out an anti-Tip60 ChIP experiment coupled to high-throughput sequencing (ChIP-seq) in mESCs. Our data demonstrate that the Tip60 complex is present at all active promoters and a subset of well-defined mESC-specific enhancer sites, suggesting that mouse Tip60 complex plays a very broad role in regulating the gene expression programmes necessary for mESC maintenance.

Results

The Tip60 complex acts mainly in large molecular complexes and is enriched at active promoters in mESCs

In order to investigate whether Tip60 acts mainly in large molecular complexes in mESCs, nuclear extracts were prepared and subjected to gel filtration that allows separation of macromolecules of different sizes. The analysis of the gel filtration by western blot indicated that Tip60 is present mainly in fractions eluting around 2 MDa that may correspond to endogenous Tip60 complex (about 1.3 MDa) (Fig. 1a). Moreover, in these fractions Tip60 is present together with two other Tip60 complex subunits, Tip48 (or RuvBl2) and Baf53a [7]. Note that the three subunits are also present in smaller size fractions, but less abundantly (Fig. 1a). Importantly, Tip60 is only detectable at very low levels in fractions eluting around 60 kDa, suggesting that there is very little 'free' Tip60 in mESCs. These results indicate that Tip60 binding profiles will mainly represent the genome-wide binding of Tip60-containing complexes.

To gain more insights into the genome-wide function of Tip60 in mESCs, we have generated high quality ChIP-seq data using previously characterized purified polyclonal anti-Tip60 antibodies that specifically recognize endogenous Tip60 [20]. Using MACS14 algorithm [21], we determined high-confidence binding sites for Tip60. The validation of several randomly selected ChIP-seq

positive sites by ChIP-qPCR indicated specific Tip60 enrichments at these sites, when compared to control IgG ChIP signals and to background enrichment at an intergenic region negative for Tip60 binding (Fig. 1b).

Next, we verified our genome-wide ChIP-seq Tip60 binding data at known Tip60-regulated genes [11, 14]. We also compared Tip60 binding at these genes with available data for DNase I hypersensitive sites (DHSs), H3K4me3 and RNA polymerase II (Pol II) profiles that are markers of open chromatin and active transcription. Importantly, Tip60 is enriched at these previously described target gene promoters (Rps9, Nodal and Cdkn1 [11, 14]), together with DHSs, Pol II binding and histone H3K4me3 mark (Fig. 1c–e). These results, together with the ChIP-qPCR validation (Fig. 1b), indicate that the obtained anti-Tip60 ChIP-seq signal is specific.

To analyze Tip60 binding genome-wide around all mESC transcription start sites (TSSs), we compared the binding of Tip60 and Pol II, and the appearence of the H3K4me3 mark at all ENSEMBL TSSs [22] by k-means clustering. Interestingly, the resulting heatmap shows that Tip60 is enriched at virtually all Pol II and H3K4me3 positive promoters (11719) in mESCs (Fig. 2a).

To better characterize the genome-wide Tip60 binding around promoters, we selected all Pol II positive genes and analysed the global distribution of Tip60 around annotated TSSs (Fig. 2b). Surprisingly Tip60 peaks at about 200 bp downstream of the TSSs. Note that this Tip60 peak is slightly more downstream than the Pol II enrichment peak, which is known to be around 40–50 bps downstream of the TSSs (reviewed in [23]).

The distinct binding profiles of Tip60, MSL and NSL complexes at TSSs suggest specific roles for these HAT complexes in transcription regulation

Genome-wide binding studies in differentiated human cells show a global co-localization of HATs and acetylated histones at transcriptional active promoters [24, 25]. The mammalian HAT Mof is the catalytic subunit of the NSL (non-specific lethal) and MSL (male-specific lethal) complex [26, 27]. We have recently analysed the genome-wide binding of two complex specific subunits Nsl1 (NSL) and Msl1 (MSL) in mESCs [18]. Since both Mof and Tip60 deletions affect mESC pluripoteny [14, 17, 28] we were further interested in the genome-wide comparison of Tip60, NSL and MSL binding at promoters. Thus, we isolated 12304 Pol II positive ENSEMBL TSSs and conducted k-means clustering. The resulting heatmap in Fig. 2c indicates Tip60, Nsl1 and Msl1 enrichment in a subset of active promoters (upper cluster), while only Tip60 and Msl1 co-localize at the second subset of promoters (lower cluster). These results suggest that the function of Tip60, NSL and MSL complexes may overlap at certain, but not all promoters.

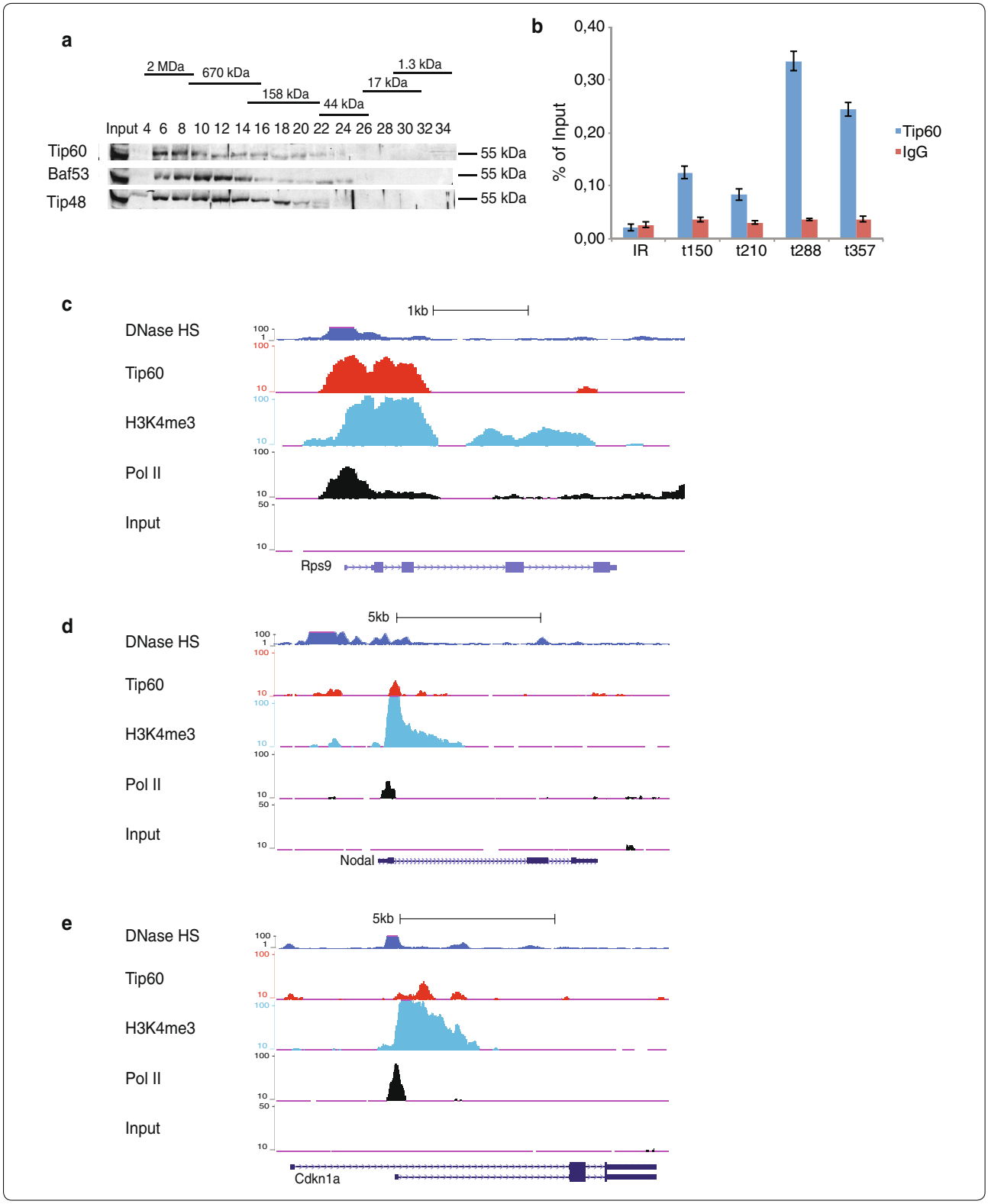
To better dissect the function of these complexes, we compared the binding distribution of Tip60, Nsl1 and Msl1 at all Pol II positive promoters (Fig. 2d). These analyses show that Nsl1 binds directly to the TSSs, Tip60 peaks about 200 bps downstream and Msl1 even more downstream of the TSSs (in the gene bodies [18]) showing that all three HAT complexes have distinct binding profiles at promoters. These data suggest that the Tip60- and Mof-containing (NSL and MSL) complexes may not have only redundant, but also specific roles in histone acetylation, histone variant exchange and/or transcriptional regulation. Additional genome-wide comparisons between Tip60 binding and available acetylated histone H3 and H4 profiles (H3K9ac, H3K27ac and H4K16ac) show that Tip60 overlaps with these marks, but peaks slightly upstream of the analysed acetylated H3 and H4 marks (Fig. 2e).

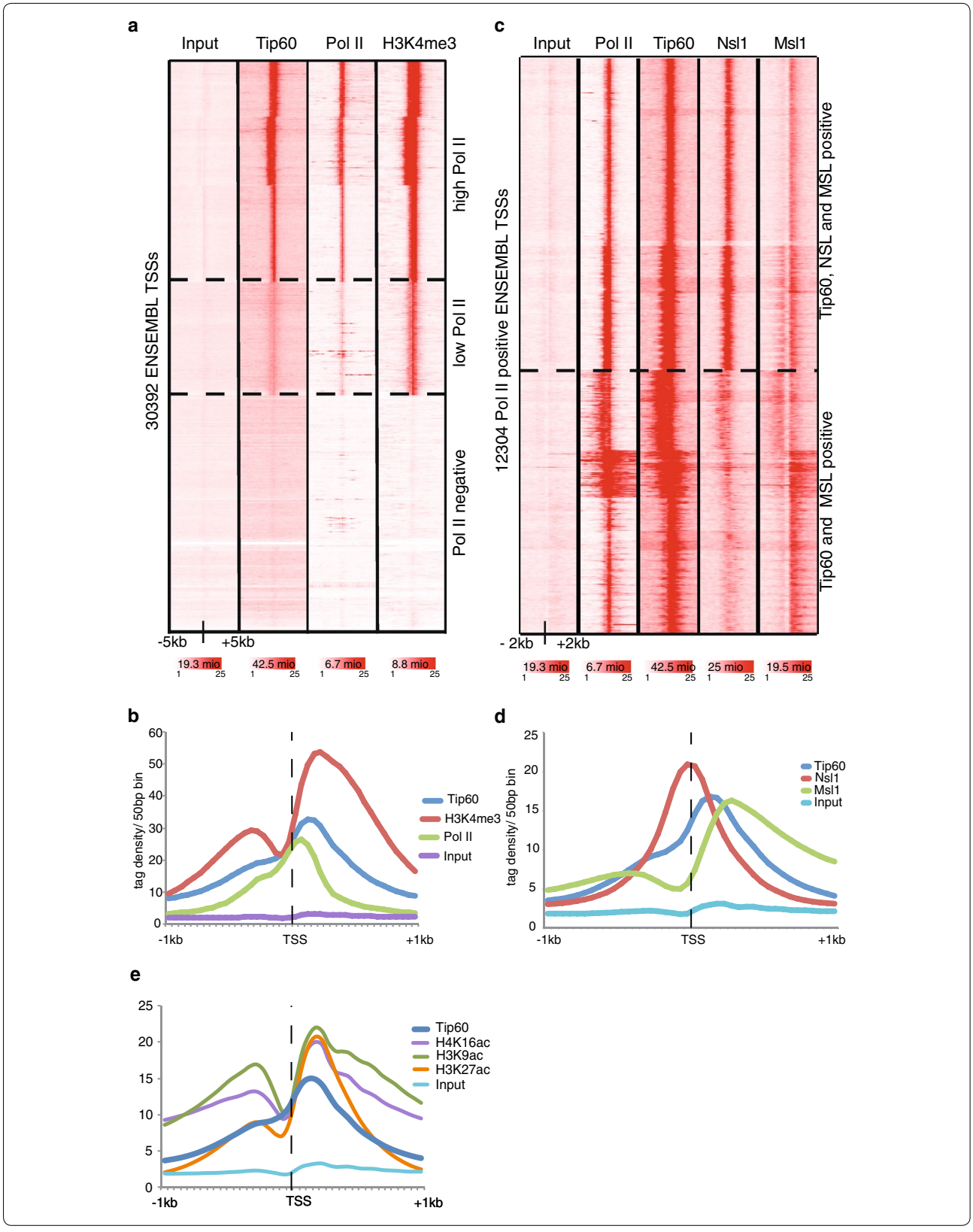
The large majority of Tip60 binding sites overlap with that of c-Myc

Tip60 is known to interact with and regulate various transcription factors as a transcriptional co-factor [4]. c-Myc is a transcription factor of the basic helix-loop-helix leucine zipper (bHLH-LZ) family, which dimerizes with another bHLH-LZ protein, Max [29]. Importantly, the oncoprotein c-Myc recruits Tip60 [20] and is regulated by the catalytic activity of Tip60 [30]. c-Myc is a somatic cell reprogramming factor (together with Oct4, Sox2, Klf4, Nanog and others) and a member of the so-called Myc regulatory module (together with n-Myc, Rex1, Zfx and E2f1) that is known to be involved in mESC self-renewal and cell metabolism [31–33]. Protein–protein interaction

(See figure on next page.)

Fig. 1 Tip60 binds to promoters as a complex in mESCs. **a** Gel filtration of mESC nuclear extracts. Every second fraction eluted from a Superose 6 column was analysed for the presence of Tip60, together with Tip49 and Baf53α by Western Blot. Native molecular weight markers eluting in the corresponding fractions are indicated on the *top of the panel*. **b** ChIP-qPCR validation of the ChIP-seq data in mESCs using purified anti-Tip60 antibodies [20] and a negative control IgG antibody. Primers were designed at randomly selected MACS14 peaks with different tag densities (*t*), as indicated. An intergenic region (IR) without Tip60 binding was selected as an additional negative control. The *graph* represents the results obtained in two biological replicates (with three technical qPCR replicates each). Standard deviations are indicated. **c**–**e** Tip60 binding profiles (GSE69671) together with DNAse I hypersensitive sites (DHS) (GSM1014154), H3K4me3 (GSM307618) and Pol II (GSM307623) binding are shown at three selected genes (*Rps9, Nodal and Cdkn1a*) as demonstrated by the UCSC genome browser. The input (GSM798320) serves as control


a
2 MDa
670 kDa
158 kDa
44 kDa
17 kDa
1.3 kDa
Input 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34
Tip60
Baf53
Tip48
55 kDa
55 kDa
55 kDa
b
% of Input
0,40
0,30
0,20
0,10
0,00
IR
t150
t210
t288
t357
Tip60
IgG
c
1kb
DNase HS
Tip60
H3K4me3
Pol II
Input
Rps9
d
5kb
DNase HS
Tip60
H3K4me3
Pol II
Input
Nodal
e
5kb
DNase HS
Tip60
H3K4me3
Pol II
Input
Cdkn1a


a
Input
Tip60
Pol II
H3K4me3
30392 ENSEMBL TSSs
high Pol II
low Pol II
Pol II negative
-5kb
+5kb
19.3 mio
42.5 mio
6.7 mio
8.8 mio
c
Input
Pol II
Tip60
Nsl1
Msl1
12304 Pol II positive ENSEMBL TSSs
Tip60, NSL and MSL positive
Tip60 and MSL positive
- 2kb
+2kb
19.3 mio
6.7 mio
42.5 mio
25 mio
19.5 mio
b
tag density/ 50bp bin
Tip60
H3K4me3
Pol II
Input
-1kb
TSS
+1kb
d
tag density/ 50bp bin
Tip60
Nsl1
Msl1
Input
-1kb
TSS
+1kb
e
Tip60
H4K16ac
H3K9ac
H3K27ac
Input
-1kb
TSS
+1kb

(See figure on previous page.)
Fig. 2 Tip60 locates to Pol II positive genes. **a** The heatmaps represents k-means clustering of Tip60, Pol II (GSM307623) and H3K4me3 (GSM307618) binding at all ENSEMBL TSSs. The Input (GSM798320) serves as control. Three main clusters are observed, as indicated. Tip60 is enriched at all Pol II and H3K4me3 positive promoters. **b** Average binding profiles of Tip60, H3K4me3 and Pol II in a region of ∓1 kb around active TSSs are depicted. Reads were normalized to Input. The input serves as control. **c** After extracting all Pol II positive ENSEMBL TSSs, *k*-means clustering was performed using Tip60 (GSE69671), Pol II (GSM307623), Nsl1 (GSM1300940) and Msl1 (GSM1300939) data sets. **d** Average binding profiles of Tip60, Nsl1 and Msl1 in a region of ∓1 kb around active TSSs are depicted. Reads were normalized to input. The input (GSM798320) serves as control. The *colour scale bars* under each data set in **a** and **c** reflect the read densities between 1 and 25 of the given dataset. The number of reads of each dataset is indicted in the *colour scale bars* in millions (mio) of reads. **e** Average binding profiles of Tip60, H4K16ac (GSM1056596), H3K9ac (GSM775313) and H3K27ac (GSM594578) in a region of ∓1 kb around active TSSs was calculated. Reads were normalized to input. To be able to better compare the datasets, the H3K27ac tag densities were divided by five. The Input (GSM798320) serves as control

networks revealed that Tip60 complex interacts with c-Myc in mESCs [12]. Moreover, c-Myc binds predominantly active genes in mESCs [34] and c-Myc is known to recruit Tip60 to target promoters [35]. To better understand the genome-wide interactions between Tip60 and c-Myc at the chromatin, we analysed the overlap between Tip60 and c-Myc high-confidence binding sites, which were identified by MACS14 peak calling algorithm [21]. When either all high-confidence (7693) Tip60 peaks are compared to published c-Myc sites (Fig. 3a), or when all 5318 c-Myc peaks are compared to Tip60 sites (Fig. 3b) using k-means clustering, about 50–65 % of all Tip60

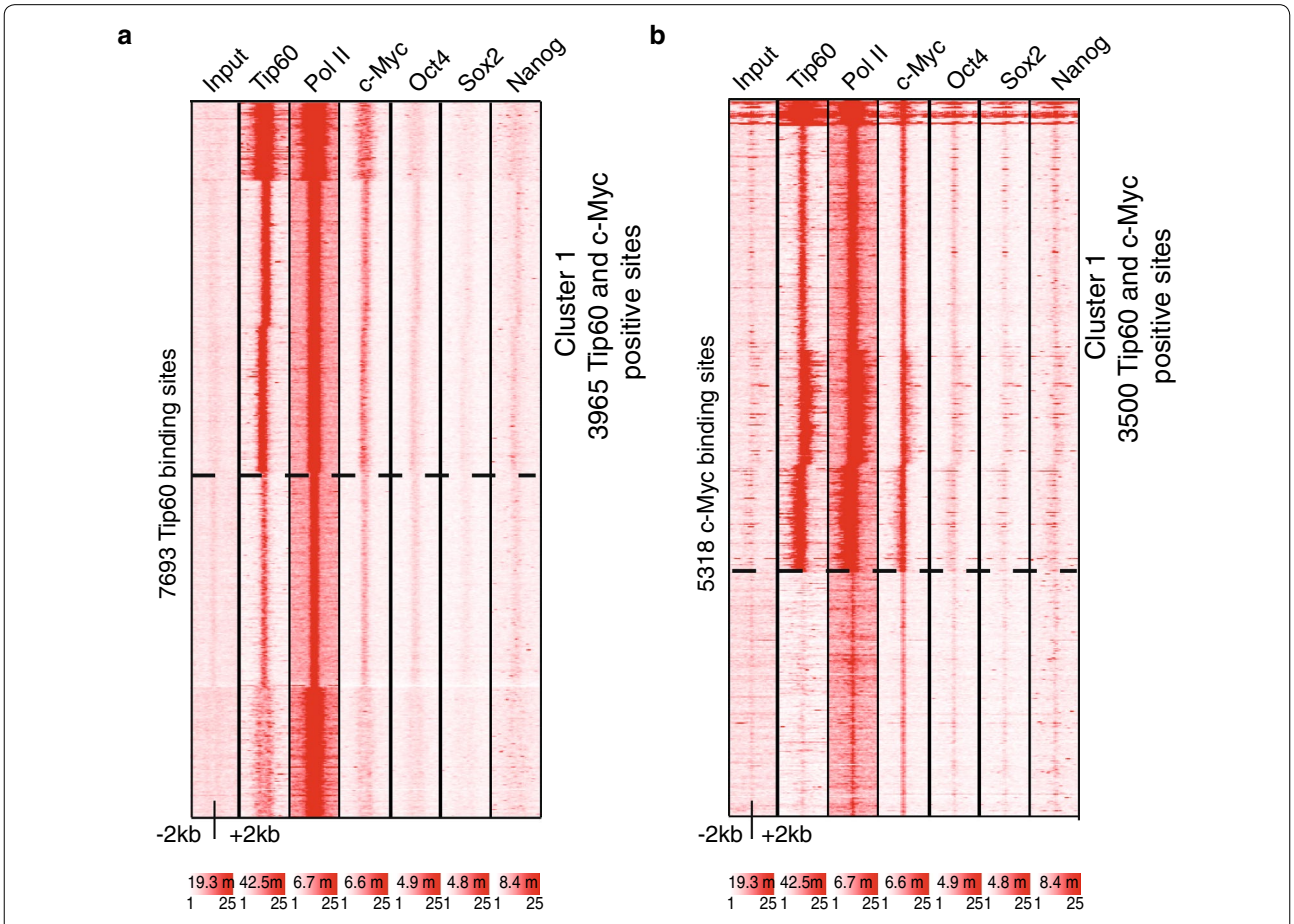


Fig. 3 Tip60 and c-Myc binding overlap. **a**–**b** Heatmap representing *k*-means clustering results of normalized Tip60 (GSE69671), Pol II (GSM307623), c-Myc (GSM288356), Oct4 (GSM288347), Sox2 (GSM288346) and Nanog (GSM288345) density profiles against all 7693 MACS14 Tip60 peaks (**a**), or against all 5318 MACS14 c-Myc peaks (**b**). Two main clusters are observed. Cluster 1 shows Tip60 and c-Myc overlap in *both panels*. The *colour scale bars* under each data set in **a** and **b** reflect the read densities between 1 and 25 of the given dataset. The *number of reads* of each data set is indicted in the *colour scale bars* in millions (m) of reads

binding sites are co-bound by c-Myc and vice versa. Importantly, other mESC pluripotency factors, such as Oct4, Sox2 and Nanog, are not enriched at the c-Myc and Tip60 co-bound sites (Fig. 3b), which is in agreement with the finding that the Myc-cluster appears to function independently from the core pluripotency network [32, 34]. These results suggest that the Tip60 complex is co-bound at about 50–65 % of Myc/Max sites in mESCs and that it is directly involved in regulating the c-Myc-dependent transcriptional network.

Tip60 locates to transcriptional active genes in mESCs

To further characterize Tip60 function, we categorized the binding of Tip60 to different genomic regions. When the MACS identified 7693 Tip60 binding peaks were annotated to promoter-TSS, 5′- or 3′-untranslated regions (UTRs), exons, introns and intergenic regions, about 42 % of all high-confidence Tip60 binding peaks were found at promoter-TSS regions (Fig. 4a). Interestingly, about 35 % of the binding sites were localized at either intronic (24 %) or intergenic regions (11 %),

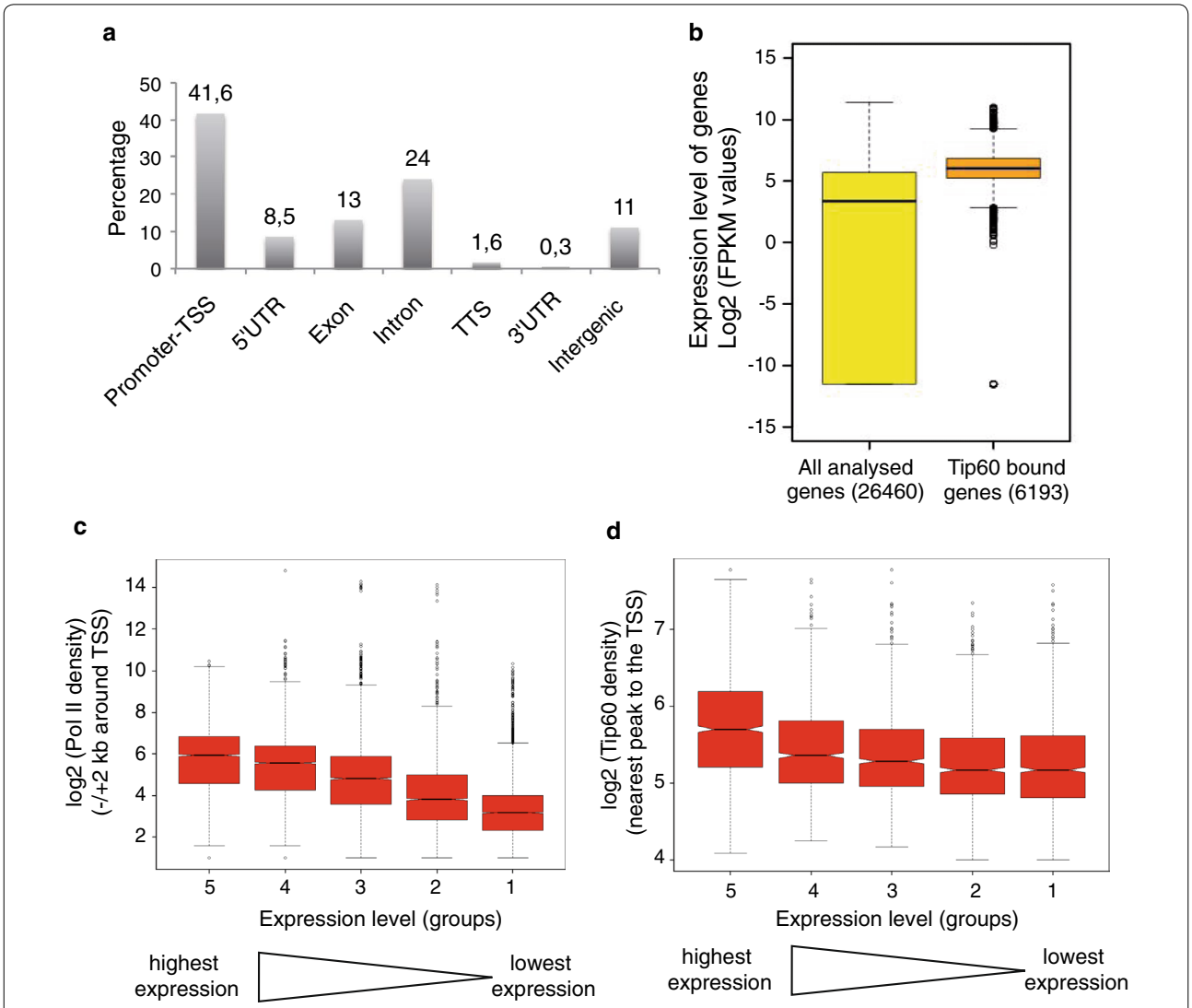


Fig. 4 Tip60 is a positive transcriptional co-activator. **a** All the identified Tip60 peaks (100 %) were annotated to different genomic elements: Promoter-TSS, UTRs (untranslated regions), exons, introns, TSS and intergenic regions and are represented in the % of their presence at these elements. **b** *Boxplots* showing the log2 of RNA fragments per kilobase per million mapped fragments (FPKM) expression values from ESCs (GSE34473) of all analysed genes and Tip60 bound genes. **c**–**d** RNA expression values are ranked into five groups with equal number of genes from highest to lowest RNA expression level as indicated. *Boxplots* represent the Pol II densities (**c**) or the Tip60 peak density (**d**) around the TSSs. Only density values higher than zero were taken. The median is significantly different between groups, if the notches do not overlap

suggesting that Tip60 may also play a role in regulating enhancer activity (see below).

Next, we were interested in expression levels of Tip60 enriched genes and how Tip60 binding correlates with gene expression. To isolate Tip60 bound ENSEMBL genes, each Tip60 peak was annotated to its closest gene in a region of 1 kb up- or downstream of TSSs. Thus, we defined 6193 genes bound by the Tip60 complex. Importantly, the average expression level of all Tip60 positive genes is significantly higher compared to that of all ENSEMBL genes (Fig. 4b), suggesting that Tip60 complexes bind predominantly to active genes. As it is known that the Pol II binding strength at promoters reflects the gene expression levels [36] (Fig. 4c), we wanted to analyse whether this would be the case for Tip60 complexes. Similarly to Pol II, the Tip60 peak tag density positively correlates with the gene expression level of bound genes (Fig. 4d). Altogether our data suggest that Tip60 locates to Pol II positive and transcriptional active genes.

Tip60 binds also to bivalent genes and active enhancer elements

To further address the global distribution and function of the Tip60 complex in mESCs, we compared high-confidence Tip60 binding sites with marks, which are either associated with active transcription at promoters (Pol II and H3K4me3) or enriched at enhancer sites (H3K4me1 and H3K27ac) [37, 38]. We also analysed H3K27me3, which establishes, together with H3K4me3, a bivalent chromatin state in mESCs at developmental genes [39]. After conducting k-means clustering using seqMINER against all identified Tip60 peaks [22], a large colocalization with Pol II, H3K27ac and H3K4me3 enriched sites was found (6136 peaks in Fig. 5a). Our GO term analyses of Tip60 enrichment at Pol II positive genes show that these Tip60 positive genes are involved not only in biological functions such as 'metabolic processes' and 'gene expression', but also in 'cell cycle and 'cellular response to stress' (Fig. 5b, upper panel). Importantly, some of these GO categories, i.e. metabolic processes', are identical than the ones defined for c-Myc [32], in agreement with our finding that at 50–65 % Tip60- and c-Myc-bound loci these factors cooperate in gene regulation.

The heatmap in Fig. 5a further indicates that Tip60 binds to about 400 sites, which are positive for H3K4me3 and H3K27me3 and are thus defined as bivalent sites (Fig. 5a). Tip60 binding at a representative bivalent gene is further illustrated in Fig. 5c. GO term analyses of all Tip60-bound bivalent genes resulted in a significant enrichment of GO terms with developmental functions (see Fig. 5b lower panel). In agreement with observations that bivalent genes are very weakly transcribed (reviewed in [40]), Tip60 tag density at bivalent sites (green line) was lower than Tip60 enrichment observed at 'Pol II positive peaks' (blue line) (Fig. 5d). Therefore, our analyses show that Tip60 complexes locate also to bivalent or developmental genes, as suggested previously [14, 15].

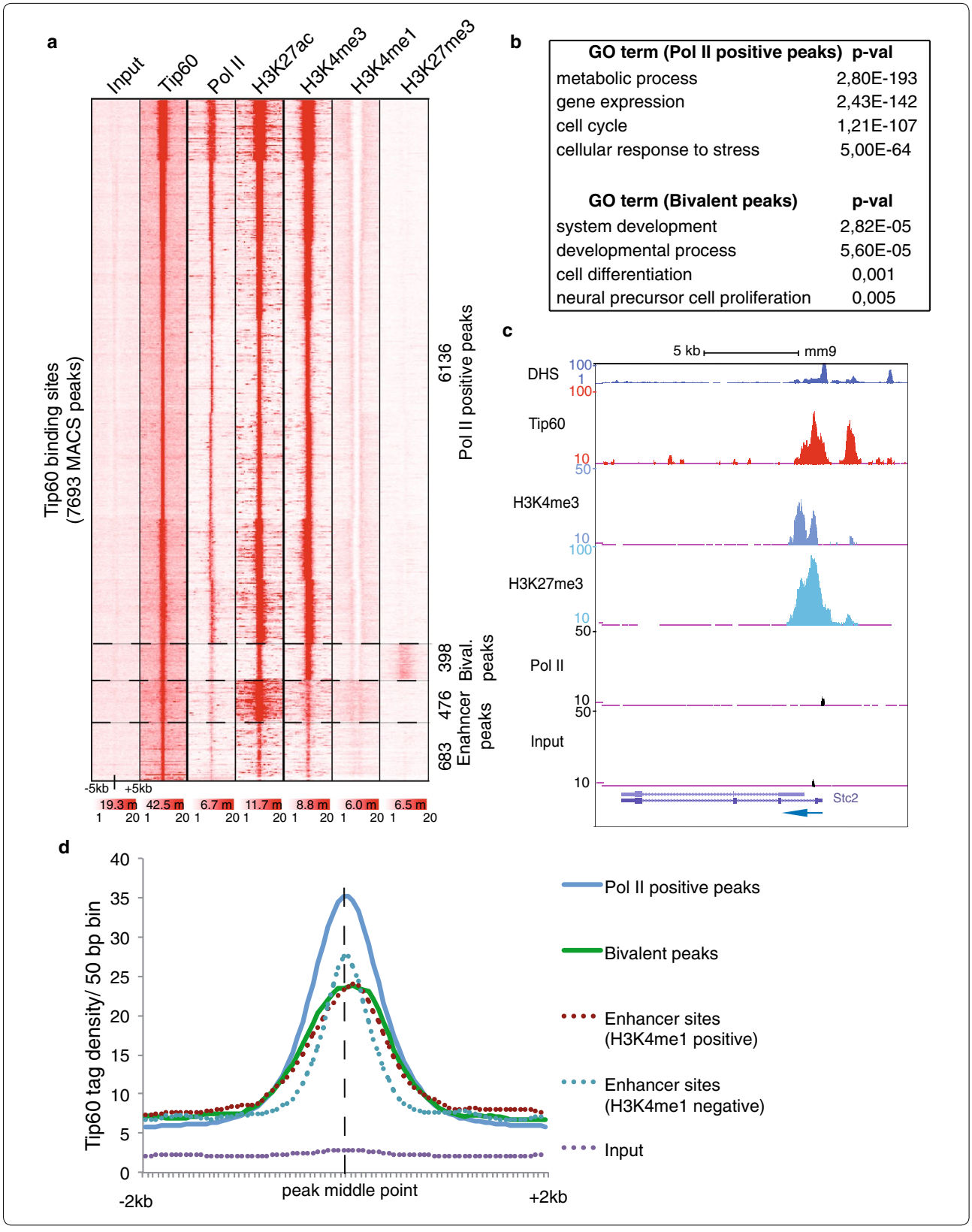
Interestingly, two of the last Tip60 positive clusters within the heatmap (Fig. 5a) show low Pol II enrichment levels. One of these Tip60 positive clusters, comprising 476 sites, contains high levels of H3K4me1 and H3K27ac marks, suggesting that these sites correspond to active enhancers [37, 38]. The comparison of H3K4me1/K3K27ac positive Tip60 peak distances to TSSs of annotated genes revealed that these peaks are located at distal regulatory regions, such as active enhancers (Additional file 1: Figure S1a, b). These findings are further illustrated at a known enhancer region, (see UCSC genome browser tracks at Additional file 1: Figure S1c). The second 'very low Pol II' cluster may represent less active enhancers and other not well-characterized genomic regions. Nevertheless, clustering of Tip60 with the above-described well-known chromatin marks and Pol II allowed us to suggest that Tip60-containing complexes act mainly at active Pol II promoters, at bivalent genes and at active enhancers.

Tip60 defines a subset of mESC-specific enhancer

As our above analyses suggested that Tip60 complexes could bind to enhancers; we wanted to examine the total enrichment of Tip60 at all known enhancer sites. To this end, we have taken annotated enhancer sites from mESCs [41]. Since active enhancers often have high H3K27ac levels [37], enhancers were sorted for H3K27ac signal intensity and analysed for enrichment of Tip60, p300, H3K4me1 and DHS. Interestingly, on these enhancer

(See figure on next page.)

Fig. 5 Tip60 binding defines active gene sets and enhancer regions. **a** Heatmap showing k-means clustering of Tip60 (GSE69671), Pol II (GSM307623), H3K27ac (GSM594578), H3K4me3 (GSM307618), H3K4me1 (GSM594577) and H3K27me3 (GSM307619) using 7693 high-confidence Tip60 binding sites as reference coordinates. Densities are represented in region of ∓5 kb around Tip60 binding sites. Four clusters are defined as indicted by the *dotted lines*. The two "enhancer" peaks are divided into H3K4me1 positive and H3K4me1 negative peaks. The *colour scale bars* under each data set in A reflect the read densities between 1 and 20 of the given dataset. The number of reads of each data set is indicted in the *colour scale bars* in millions (m) of reads. **b** GO term analyses of Pol II positive or bivalent peaks after gene annotation using Manteia [62]. **c** UCSC genome browser profiles of DHS, Tip60 binding, presence of H3K4me3 and H3K27me3, Pol II binding and input (negative control) at a randomly chosen bivalent gene, Stc2, are shown. An *arrow labels* the gene orientation. **d** Density profiles of Tip60 binding at the different clusters as defined in *panel* **a**


a
Input
Tip60
Pol II
H3K27ac
H3K4me3
H3K4me1
H3K27me3
Tip60 binding sites
(7693 MACS peaks)
6136
Pol II positive peaks
398
Bival. peaks
476
Enahncer peaks
683
-5kb
+5kb
19.3 m
42.5 m
6.7 m
11.7 m
8.8 m
6.0 m
6.5 m
b
GO term (Pol II positive peaks) p-val
metabolic process 2,80E-193
gene expression 2,43E-142
cell cycle 1,21E-107
cellular response to stress 5,00E-64
GO term (Bivalent peaks) p-val
system development 2,82E-05
developmental process 5,60E-05
cell differentiation 0,001
neural precursor cell proliferation 0,005
c
5 kb
mm9
DHS
Tip60
H3K4me3
H3K27me3
Pol II
Input
Stc2
d
Tip60 tag density/ 50 bp bin
peak middle point
-2kb
+2kb
Pol II positive peaks
Bivalent peaks
Enhancer sites
(H3K4me1 positive)
Enhancer sites
(H3K4me1 negative)
Input

sites Tip60 binding showed a co-occurence with p300, H3K4me1 and DHSs and a partial overlap with a subset of sites having high H3K27ac signals (Fig. 6a). Next, we divided these H3K27ac labelled enhancers into four equal clusters (from high to low H3K27ac signals) and average tag densities for Tip60, as well as for DHS profiles, around enhancer mid points of each cluster were calculated (Fig. 6b, c). Importantly, Tip60 and DHS enrichment have a positive correlation with each other and with that of H3K27ac. To define the number of enhancers that are positive for Tip60 and all the other enhancer defining marks (H3K4me1, H3K27ac and DHS), k-means clustering was performed to analyse the presence of Tip60, p300, H3K4me1, H3K27ac and DHS at these defined enhancer sites [41]. The heatmap analysis in Figure S2A clearly shows that 2305 enhancers defined in cluster 1 are positive for Tip60 and all the other enhancer chromatin marks (p300, H3K27ac and H3K4me1). In contrast, the 6489 enhancers in cluster 2 show almost no Tip60 enrichment (Additional file 2: Figure S2b, c). Moreover, GO terms of genes in the vicinity of the 2305 Tip60 bound enhancers are associated with developmental processes (Fig. 6d). Thus, it seems that Tip60 is recruited to about 26 % of all mESC-specific enhancers.

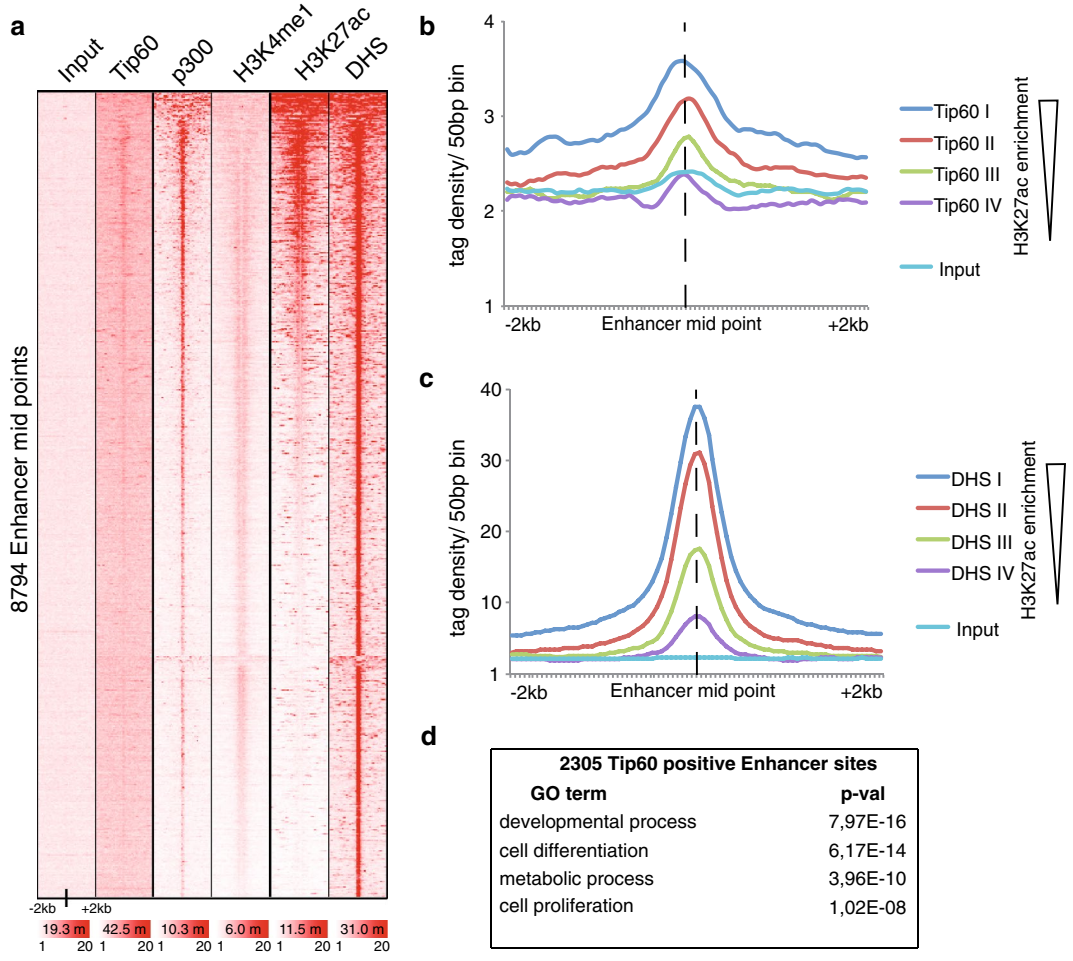


Fig. 6 Tip60 correlates with p300, H3K3me1 and DHS presence and with a subset of H3K27ac signals at ESC-specific enhancer sites. **a** Enhancers [41] were sorted for H3K27ac (GSM594578) signal intensities. Signals of Tip60 (GSE69671), p300 (GSM723018), H3K27ac (GSM594578), H3K4me1 (GSM594577), H3K27me3 (GSM307619) and DHS (GSM1014154) were calculated around enhancer mid points. The *colour scale bars* under each data set in **a** reflects the read densities between 1 and 20 of the given dataset. The number of reads of each data set is indicted in the *colour scale bars* in millions (m) of reads. **b**, **c** Enhancers were divided into four categories having equal number of genes based on their H3K27ac signal intensity from high to low (I–IV) as indicated on the *right of the panels*. *Graphs* show Tip60 enrichment (**b**) or DHS signals (**c**) around enhancer mid points of the different categories. (**d**) GO term category analysis by Manteia [62] of the closest genes to the Tip60 positive enhancers [41] identified by *k*-means clustering

As two of the so-called 'super enhancer' regions of the Nanog gene and the enhancer of the Klf4 gene are bound by Tip60, but not by Nsl1 and Msl1 (Fig. 7a, b), we further characterized Tip60 binding at enhancer regions in detail by analysing Tip60 presence at the previously defined 231 'super enhancer' regions [41]. We observed Tip60 enrichment (compared to the Input control) between the start and end positions of these 231 'super-enhancers' (Fig. 7c; Additional file 2: Figure S2d). However, global analyses of Mof-containing complexes at these 'super-enhancer' regions displayed Msl1, but no or very weak Nsl1 enrichment (Fig. 7d, e). Altogether our data show that Tip60 is recruited to a subset of active enhancers out of which certain have been defined as super enhancers in mESCs.

Discussion

In this study, we analysed the genome-wide binding of Tip60-containing complexes to understand their role in transcription regulation in mESCs. Our gel filtration analyses show that Tip60 incorporates into large molecular complexes in mESCs, as previously described in other systems, and that heterogeneous populations of Tip60-containing complexes exist, which might dynamically change their association with p400 and/or other subunits. However, our observations that there is very little free Tip60 within the nuclei of mESCs indicates that the ChIP binding profiles obtained with anti-Tip60 antibodies represent mostly the behaviour of the corresponding endogenous Tip60 complexes.

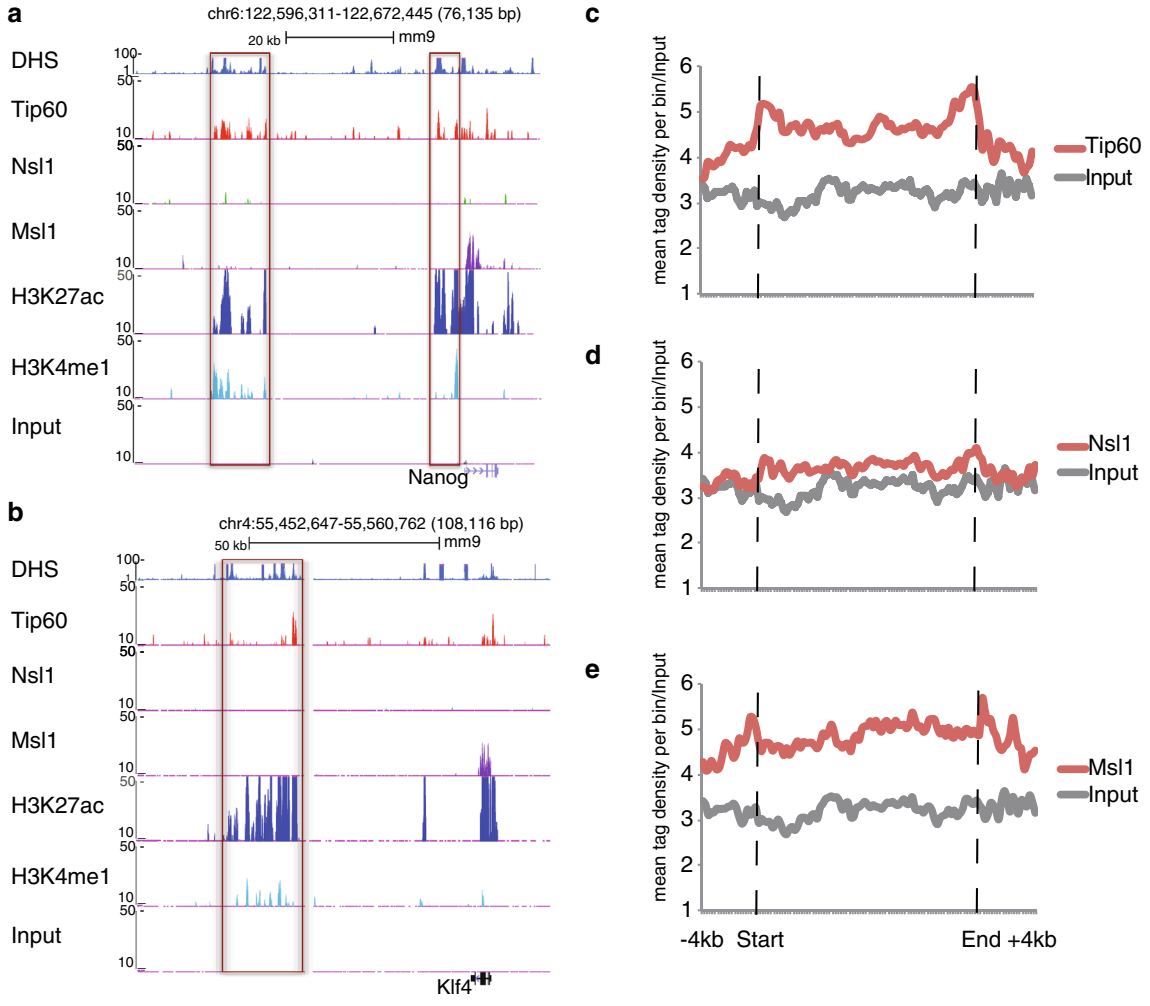


Fig. 7 Tip60 locates to super enhancer regions. **a** Tip60 (GSE69671) binding profiles together with DHS (GSM1014154), H3K4me1 (GSM594577), H3K27ac (GSM594578), Nsl1 (GSM1300940) and Msl1 (GSM1300939) are represented using the UCSC genome browser at the well-defined super enhancer regions (*boxed in red*) of the *Nanog* and *Klf4* locus. The Input (GSM798320) serves as negative control. Data were uploaded as wig files. **b**–**d** The 231 super enhancers [41] were divided into 80 bins and Tip60 (**b**), Nsl1 (**c**) and Msl1 (**d**) enrichment was calculated per bin between their start and end positions as well as up to 4 kb up- and downstream. Signals were normalized to the input (GSM798320)

Our study demonstrates that the majority of genome-wide Tip60 binding occurs at promoter regions, where it co-localizes with Pol II, H3K4me3 and DHS sites. Importantly, Tip60 bound genes are expressed and the Tip60 enrichment positively correlates with gene expression levels. Interestingly, using an unbiased clustering method of all mapped Tip60 reads against ENSEMBL TSS, and we observed that all transcriptionally active genes (Pol II positive) show Tip60 enrichment. Genome-wide binding and knock-down studies of Tip60-p400, NSL and MSL HAT complexes reported the binding of these complexes at Pol II positive genes and in mESCs [18, 19]. When directly comparing binding profiles of Tip60- and Mof-containing complexes around Pol II positive promoter regions, we have found a global overlap at Pol II positive mRNA coding genes, whereas each of the HAT complexes has a distinct binding profile around the TSS. Altogether, this suggests that (1) Tip60- and Mof-containing HAT complexes globally regulate gene expression though their presence at promoters (NSL and Tip60) and further downstream of promoters (i.e. MSL) and (2) that there might be a different function of each HAT complex in histone acetylation, histone exchange and transcriptional regulation.

The frequent overlap of Tip60 and c-Myc binding strongly indicates a role of Tip60 complexes within the Myc-centred regulatory pluripotency network in mESC [31, 33]. Moreover, Tip60 complexes bind and possibly regulate a subset of active enhancers and super enhancers, revealing an additional layer of regulation by which Tip60 complexes influence mESC maintenance.

Tip60 complexes preferentially acetylate histone H4 [8]. Acetylation of several histone H3 and H4 lysines was described to be up-regulated at the majority of Myc-target promoters [42]. Several studies in *Drosophila,* mESCs and human cells have revealed that c-Myc targets at least 10–15 % of all cellular promoters [35]. Nevertheless, it has been suggested that although Myc can bind to a large number of genes, it is critical for the regulation of a subset of those genes depending on (1) protein–protein interactions between Myc/Max dimers, such as chromatin-bound protein complexes [43] and (2) the cellular or physiological context [44]. Our finding that at least half of the c-Myc binding sites are co-bound by the Tip60 complex suggests that Myc/Max and the Tip60 complex may cooperate to stabilize each other's binding. Moreover, at these co-bound sites the Tip60 complex is thought to be involved in acetylation of promoter-associated nucleosomes that can participate in transcriptional activation [35].

Interestingly, the genome-wide Tip60 complex binding profile peaks at regions that are about 200 bps downstream of TSSs (Fig. 2b). It is, therefore, conceivable that this downstream binding may reflect an additional function of the Tip60 complex that may play a role in the early steps of transcription elongation processes, as suggested also in *Drosophila* [45]. Interestingly, c-Myc is believed to control the release of Pol II from promoter proximal transcriptional pause [44]. Thus, the Tip60 complexes bound 200 bps downstream of the TSS at many active genes may cooperate with c-Myc to co-regulate Pol II pause release and consequently, transcriptional elongation in mESCs.

Moreover, our results indicate that Tip60 complexes may be involved in gene expression regulation by occupying a set of previously identified ESC-specific enhancers [41]. Tip60 enrichment correlates with H3K27ac levels and chromatin accessibility (DHS sites). Genes associated with enhancers with increased Tip60 binding (determined by *k*-means clustering) are predicted to be involved in regulation of development, metabolism and proliferation of mESCs. Based on high Mediator occupancy, some of these enhancers can be clustered to 231 super-enhancer regions [41], which likely control important genes for stem cell maintenance (i.e. the Nanog or Klf4 gene). Importantly, these super-enhancers show Tip60 and Msl1 occupancy, but have only very little Nsl1 binding. The fact that Tip60 possibly regulates a subset of active enhancers and super enhancers reveals an additional layer of regulation by which Tip60 complexes influence mESC maintenance. Note that previously we have described enhancers, which were bound by the ATAC HAT complex but not by p300 in differentiated human cells [46]. Thus, it is possible that different enhancers are bound by a given combination of HAT complexes to regulate their activity.

Human Tip60 is known to play a wide role in transcriptional regulation [47]. The identified Tip60 bound genes belong to gene sets with housekeeping and developmental functions. Other studies further suggested that the Tip60 complex might have a repressive function at low expressed bivalent/developmental genes [14, 15]. Since the Tip60 complex has different catalytic activities by acetylating histones or exchanging H2A.Z, which further depend on each other, it will be challenging to understand the molecular mechanism by which the Tip60 complexes regulate transcription at these low-expressed genes in mESCs. Nevertheless, it is tempting to speculate that the Tip60 complex is recruited to these genes to poise these genes that will be important for their rapid activation during cellular differentiation.

Our genome-wide binding analyses suggest that the Tip60 complex has a very broad role in regulating transcription in mESCs, which is in agreement with the observation that Tip60 and several other Tip60-complex subunits are required for mESC maintenance [14]. In contrast, Tip60 knock-down studies identified less

than 900 differentially expressed genes when analysing steady-state mRNAs [14, 15]. The potential reason for this apparent contradiction may be that the previous Tip60 knock-down studies analysed steady-state mRNAs instead of newly synthetized transcripts. Recently, it has been demonstrated that the sole analysis of polyA + mRNAs may not give any information on the direct transcriptional output of a given transcription factor/co-activator complex, as cells can buffer global transcription changes by adjusting mRNA decay in parallel ([48–50] and refs therein). Thus, novel newly synthetized RNA detection methods are needed to address global effects of Tip60 on newly synthetized transcripts, such as in vivo RNA labelling with 4-thiouridine (4sU) [49] or analyses of profiles of transcriptional active Pol II (i.e. NET-seq; [51]) under ∓ Tip60 conditions. Furthermore, the understanding of how the different activities (acetylation, histone variant exchange and/or nucleosome remodelling activity) of the Tip60 complexes influence transcription at promoters and enhancers in mESCs will be an equally important task.

Conclusions

Overall, we establish that Tip60-complexes are present at promoter regions of active RNA polymerase II genes and that half of Tip60 binding sites overlap with binding of the somatic cell reprograming factor, c-Myc, known to regulate an ESC specific transcriptional module. Importantly, Tip60, NSL and MSL coactivator HAT complexes have a genome-wide overlap at many active genes, but their specific functions might be reflected in their distinct binding profiles around the TSSs. Moreover, Tip60 complexes define a subset of bivalent developmental genes and a subset of ESC-specific enhancers. Thus, our study suggests that the Tip60 complex is important for mESC pluripotency by acting as a global transcriptional co-factor at active genes and distal regulatory elements.

Methods

Cell culture

The ES cell line, E14 was derived from 129P2/OlaHsd strain blastocysts [52] and cells were cultivated on 0.1 % gelatine (Sigma, France) and CD1 feeder cells (37 °C, 5 % CO_2) in DMEM (4,5 g/L glucose/w-Glutamax) (Invitrogen, France), 15 % foetal calf serum (PAA), 5uM LIF, 50 mM ß-mercaptoethanol (Invitrogen, France), penicillin (10,000 U/ml) and streptomycin (10 mg/ml) (Invitrogen, France), 200 mM L-glutamine (Invitrogen, France) and 1× non-essential amino acids (GIBCO, France). To work under feeder-free conditions cells were treated with 1 mg/ml Collagenase and 2 mg/mL Dispase (GIBCO, France) and cultivated for one passage without feeder cells on 0.1 % gelatine plates. Experiments were conducted with E14.wt cells at passages 26–29.

Chromatin immunoprecipitation coupled to Solexa sequencing (ChIP-seq)

ChIP was carried out as described in [18]. For 500 µg of chromatin 6 µg of the well-characterized rabbit anti-Tip60 antibody mixture (RLPV, CLGT and CLHF purified polyclonal sera) was used [20]. 8 ng of precipitated DNA from Tip60 ChIP was used for Solexa sequencing. Rabbit anti-IgG antibody (ab37415) was purchased from Abcam. To create a genomic library the instructions of NEBNext protocol (E6240, Biolabs) was followed. The library was validated with the Agilent Bioanalyzer. Single reads run sequencing was done with Illumina HiSeq 2000. Image analysis and base calling were conducted with the Illumina pipeline (1.8.2). The July 2007 Mus musculus genome assembly (NCBI37/mm9) from NCBI was used for the sequence alignment by the software Bowtie (0.12.7) [53]. The analysis was conducted with unique reads. Read density (wig) files were created out of bed files by extending reads to 200 bp length and creating 25 bp bins. To detect ChIP-seq peaks the MACS14 peak-calling algorithm was applied using default parameters [21]. The Tip60 ChIP-seq data were deposited in the Gene Expression Omnibus (GEO) databank under the accession number GSE69671.

For further analyses the following ChIP-seq files were included from Gene Expression Omnibus(http://www.ncbi.nlm.nih.gov/geo/): Input (GSM798320) [54], RNA polymerase II (GSM307623), H3K4me3 (GSM307618), c-Myc (GSM288356), Oct4 (GSM288347), Sox2 (GSM288346) and Nanog (GSM288345) [31], H3K4me1 (GSM594577), H3K27ac (GSM594578) [37], H3K9ac (GSM775313) [54], H4K16ac (GSM1056596) [55], p300 (GSM723018) [56], H3K27me3 (GSM307619) [57], Nsl1 (GSM1300940) and Msl1 (GSM1300939) [18]. Tip60 data are deposited in GEO under the following accession number: GSE69671. Fastq files were generated from SRA lite format and aligned to the NCBI37/mm9 assembly using Bowtie (0.12.7) [53]. DNAse I hypersensitive sites were taken from Encode/UW (GSM1014154). Detailed information summarizing all the used ChIP-seq files is presented in Additional file 3: Table S1.

ChIP-qPCR

The Tip60 ChIP-seq was validated by ChIP coupled to quantitative PCR (ChIP-qPCR). Identified MACS14 Tip60 peaks were randomly taken based on different tag densitites (t). SYBR Green (Roche) was used according to the manufacturers protocol. Following primers were designed:

t150_fw	TGATCGGCGCAGAGACAAGA
t150_rv	ACAAAAGGCCCCTCCTTGCT
t210_fw	TCGCTTTGCAGCAGTGAGATG
t210_rv	TGGCCTCGGACCTTTCAATC
t288_fw	CGGCTTCGGGGTTTTCTTTT
t288_rv	TTATCCCATTCCGGGAGACG
t357_fw	ACCAGGTCCTCGGCGATAGTTT
t357_rv	CTTTCCTCGCGGATCGAAGA
intergenic_fw	TGATGCAACACATGGACATTTCTG
Intergenic_rv	TTCAGGGGTTGGGACAAAGTG

Gel filtration

The gel filtration experiment using a Superose 6 column was described in [18]. Input nuclear extract and every second fraction eluting from the column were tested by western blot assays using the Tip60 antibody mixture [20] at dilution 1:2000, the anti-Tip48 or anti-Baf53α antibodies [58] at dilution 1:500.

Bioinformatics analyses of Tip60 ChIP-seq in mESCs

Density profile calculation around TSSs as well as *K*-means linear clustering was conducted with seqMINER [22]. *K*-means clustering was performed with normalized read densities, while resulting heatmaps show total number of reads. Obtained MACS14 peaks were annotated using the software HOMER [59] based on the ENSEMBL 67 database (mm9).

To determine the Tip60 or Pol II enrichment at genes, the peak tag density of the nearest peak to the TSS (in a region of +2 kb) was taken and correlated with gene expression levels. We considered a total of 26,460 ENSEMBL TSSs based on the ENSEMBL 67 database (mm9). For this, raw RNA-seq data of mESCs from Gene Expression Omnibus (GSE34473) were processed using the software tools TopHat [60] and HTSeq with default parameters. FPKM (fragments per kilobase of exon per million fragments mapped) values were calculated with Cufflinks [61]. FPKM values were correlated with Tip60 and Pol II enrichment and taken to analyse average expression levels of Tip60 bound genes.

Gene ontology (GO) term analyses of Pol II positive or bivalent peaks as well as enhancer sites were conducted with Manteia [62]. Peaks were annotated to nearest promoters [59] prior GO term analysis.

ES-specific enhancers and super enhancer regions were taken from [41]. ES-specific enhancers were ranked according to H3K27ac signal intensities. Enhancer-midpoints were calculated for further analysis. The 231 total super enhancers were divided into 80 bins from start to end positions and the mean Tip60 and Input read densities were calculated for each bin. Moreover, read densities of regions 4 kb down-or upstream of super enhancer start or end positions were determined for each 50 bp bin. Total Number of reads is normalized to the Input.

Additional files

Additional file 1: Figure S1. Tip60 defines bivalent genes and enhancer sites. (**A** and **B**) Mapping of Tip60/Pol II positive (**A**) or Tip60/H3K27ac/H3K4me1 positive peaks to TSSs and more and more distal genomic regions. X-axis indicates distances of peaks to TSSs. (**C**) Tip60 enrichment at putative enhancer sites, which are H3K4me1 (GSM594577) and H3K27ac (GSM594578) positive at the UCSC genome browser. The Input (GSM798320) is the negative control.

Additional file 2: Figure S2. Tip60 is enriched at ESC-specific enhancer sites. (**A**) Heatmap showing k-means clustering of Tip60, p300 (GSM723018), H3K27ac (GSM594578) and H3K4me1 (GSM594577) using enhancer mid points [41] as reference coordinates. Two main clusters were defined: one with significant Tip60 binding (cluster 1), and a second where Tip60 cannot be detected (cluster 2). Densities are represented in region of ∓ 2 kb around enhancer mid points. The representation is as in Fig. 1a. (**B** and **C**) Tip60 enrichment around enhancer mid points of cluster 1 (**B**) and cluster 2 (**C**). Clusters were defined in panel A.

Additional file 3: Table S1. Summary of all ChIP-seq datasets used representing the number of unique reads and GEO accession numbers.

Abbreviations

Ac: acetyl; bHLH-LZ: basic helix-loop-helix leucine zipper; bp: base pairs; ChIP: chromatin immunoprecipitation; ChIP-seq: chromatin immunoprecipitation coupled to sequencing; DHS: DNase Hypersensitive Site; GEO: gene expression omnibus; H: histone; HAT: histone acetyltransferase; KAT: lysine acetyltransferase; K: lysine; kb: kilobase; kD: kilo Dalton; m: mouse; me: Methyl; mD: mega Dalton; mESCs: mouse embryonic stem cells; Mof: male absent on the first; MSL: male-specific lethal; mRNA: messenger RNA; NSL: non-specific lethal; Pol II: RNA polymerase II; qPCR: quantitative polymerase chain reaction; shRNA: small-hairin RNA; siRNA: small interfering RNA; Tip60: tat-interacting protein 60 kDa; TSS: transcription start site; UTR: untranslated-region; y: yeast; WB: western Blot.

Authors' contributions

SR, CY and MS performed experiments. SR and TY analysed data. SR and LT designed the study, analysed data and wrote the manuscript. All authors read and approved the final manuscript.

Author details

[1] Cellular Signalling and Nuclear Dynamics Programme, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U964, Université de Strasbourg (UdS), BP 10142, 1 Rue Laurent Fries, CU de Strasbourg, 67404 Illkirch Cedex, France. [2] Microarrays and Deep Sequencing Platform, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U964, UdS, BP 10142, CU de Strasbourg, 67404 Illkirch Cedex, France.

Acknowledgements

We are very grateful to B. Amati for providing the Tip60 antibodies. We would like to thank the IGBMC sequencing and bioinformatics platform for data generation and support, M. Hestin and G. Rossi from the IGBMC ES cell culture facility and F. Klein for the gelfiltration experiments. We thank J. Pellegrino, M. Gerard and D. Devys for critical reading the manuscript. SR was supported by a fellowship from Association pour la Recherche sur le Cancer (ARC). This work was supported by funds from CNRS, INSERM, Strasbourg University, the European Commission Marie Curie-ITN (NR-NET) and the Agence Nationale de Recherche (ANR-11-BSV5-010-02 Chromact; ANR-13-BSV6-0001-02 COREAC; ANR-13-BSV8-0021-03 DiscoverIID). LT is recipient of a European Research Council (ERC) Advanced Grant (ERC-2013-340551, Birtoaction).

Competing interests

The authors declare that they have no competing interests.

Received: 12 June 2015 Accepted: 29 October 2015
Published online: 06 November 2015

References

1. Squatrito M, Gorrini C, Amati B. Tip60 in DNA damage response and growth control: many tricks in one HAT. Trends Cell Biol. 2006;16(9):433–42. doi:10.1016/j.tcb.2006.07.007.
2. Voss AK, Thomas T. MYST family histone acetyltransferases take center stage in stem cells and development. BioEssays News Rev Mol Cell Develop Biol. 2009;31(10):1050–61. doi:10.1002/bies.200900051.
3. Doyon Y, Selleck W, Lane WS, Tan S, Cote J. Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Mol Cell Biol. 2004;24(5):1884–96.
4. Sapountzi V, Cote J. MYST-family histone acetyltransferases: beyond chromatin. Cell Mol Life Sci CMLS. 2011;68(7):1147–56. doi:10.1007/s00018-010-0599-9.
5. Boudreault AA, Cronier D, Selleck W, Lacoste N, Utley RT, Allard S, et al. Yeast enhancer of polycomb defines global Esa1-dependent acetylation of chromatin. Genes Dev. 2003;17(11):1415–28. doi:10.1101/gad.1056603.
6. Doyon Y, Cote J. The highly conserved and multifunctional NuA4 HAT complex. Curr Opin Genet Dev. 2004;14(2):147–54. doi:10.1016/j.gde.2004.02.009.
7. Ikura T, Ogryzko VV, Grigoriev M, Groisman R, Wang J, Horikoshi M, et al. Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell. 2000;102(4):463–73.
8. Auger A, Galarneau L, Altaf M, Nourani A, Doyon Y, Utley RT, et al. Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants. Mol Cell Biol. 2008;28(7):2257–70. doi:10.1128/MCB.01755-07.
9. Fuchs M, Gerber J, Drapkin R, Sif S, Ikura T, Ogryzko V, et al. The p400 complex is an essential E1A transformation target. Cell. 2001;106(3):297–307.
10. Samuelson AV, Narita M, Chan HM, Jin J, de Stanchina E, McCurrach ME, et al. p400 is required for E1A to promote apoptosis. J Biol Chem. 2005;280(23):21915–23. doi:10.1074/jbc.M414564200.
11. Park JH, Sun XJ, Roeder RG. The SANT domain of p400 ATPase represses acetyltransferase activity and coactivator function of TIP60 in basal p21 gene expression. Mol Cell Biol. 2010;30(11):2750–61. doi:10.1128/MCB.00804-09.
12. Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, et al. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell. 2010;143(2):313–24. doi:10.1016/j.cell.2010.09.010.
13. Hu Y, Fisher JB, Koprowski S, McAllister D, Kim MS, Lough J. Homozygous disruption of the Tip60 gene causes early embryonic lethality. Develop Dyn Off Publ Am Assoc Anat. 2009;238(11):2912–21. doi:10.1002/dvdy.22110.
14. Fazzio TG, Huff JT, Panning B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell. 2008;134(1):162–74. doi:10.1016/j.cell.2008.05.031.
15. Chen PB, Hung JH, Hickman TL, Coles AH, Carey JF, Weng Z, et al. Hdac6 regulates Tip60-p400 function in stem cells. eLife. 2013;2:e01557. doi:10.7554/eLife.01557.
16. Gupta A, Guerin-Peyrou TG, Sharma GG, Park C, Agarwal M, Ganju RK, et al. The mammalian ortholog of *Drosophila* MOF that acetylates histone H4 lysine 16 is essential for embryogenesis and oncogenesis. Mol Cell Biol. 2008;28(1):397–409. doi:10.1128/MCB.01045-07.
17. Thomas T, Dixon MP, Kueh AJ, Voss AK. Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture. Mol Cell Biol. 2008;28(16):5093–105. doi:10.1128/MCB.02202-07.
18. Ravens S, Fournier M, Ye T, Stierle M, Dembele D, Chavant V, et al. Mof-associated complexes have overlapping and unique roles in regulating pluripotency in embryonic stem cells and during differentiation. eLife. 2014;. doi:10.7554/eLife.02104.
19. Chelmicki T, Dundar F, Turley MJ, Khanam T, Aktas T, Ramirez F, et al. MOF-associated complexes ensure stem cell identity and Xist repression. eLife. 2014;3:e02024. doi:10.7554/eLife.02024.
20. Frank SR, Parisi T, Taubert S, Fernandez P, Fuchs M, Chan HM, et al. MYC recruits the TIP60 histone acetyltransferase complex to chromatin. EMBO Rep. 2003;4(6):575–80. doi:10.1038/sj.embor.embor861.
21. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. doi:10.1186/gb-2008-9-9-r137.
22. Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, et al. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. 2011;39(6):e35. doi:10.1093/nar/gkq1287.
23. Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet. 2012;13(10):720–31. doi:10.1038/nrg3293.
24. Wang ZB, Zang CZ, Cui KR, DE Schones, Barski A, Peng WQ, et al. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell. 2009;138(5):1019–31. doi:10.1016/J.Cell.2009.06.049.
25. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008;40(7):897–903. doi:10.1038/ng.154.
26. Cai Y, Jin J, Swanson SK, Cole MD, Choi SH, Florens L, et al. Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. J Biol Chem. 2010;285(7):4268–72. doi:10.1074/jbc.C109.087981.
27. Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, et al. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. Mol Cell. 2006;21(6):811–23. doi:10.1016/j.molcel.2006.02.007.
28. Li X, Li L, Pandey R, Byun JS, Gardner K, Qin Z, et al. The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network. Cell Stem Cell. 2012;11(2):163–78. doi:10.1016/j.stem.2012.04.023.
29. Amati B, Brooks MW, Levy N, Littlewood TD, Evan GI, Land H. Oncogenic activity of the c-Myc protein requires dimerization with Max. Cell. 1993;72(2):233–45.
30. Patel JH, Loboda AP, Showe MK, Showe LC, McMahon SB. Analysis of genomic targets reveals complex functions of MYC. Nat Rev Cancer. 2004;4(7):562–8. doi:10.1038/nrc1393.
31. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008;133(6):1106–17. doi:10.1016/j.cell.2008.04.043.
32. Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. Cell. 2008;132(6):1049–61. doi:10.1016/j.cell.2008.02.039.
33. Yeo JC, Ng HH. The transcriptional regulation of pluripotency. Cell Res. 2013;23(1):20–32. doi:10.1038/cr.2012.172.
34. Kidder BL, Yang J, Palmer S. Stat3 and c-Myc genome-wide promoter occupancy in embryonic stem cells. PLoS One. 2008;3(12):e3932. doi:10.1371/journal.pone.0003932.
35. Martinato F, Cesaroni M, Amati B, Guccione E. Analysis of Myc-induced histone modifications on target chromatin. PLoS One. 2008;3(11):e3650. doi:10.1371/journal.pone.0003650.
36. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823–37. doi:10.1016/j.cell.2007.05.009.
37. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA. 2010;107(50):21931–6. doi:10.1073/pnas.1016071107.
38. Heintzman ND, Stuart RK, Hon G, Fu YT, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39(3):311–8. doi:10.1038/Ng1966.
39. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125(2):315–26. doi:10.1016/j.cell.2006.02.041.
40. Voigt P, Tee WW, Reinberg D. A double take on bivalent promoters. Genes Dev. 2013;27(12):1318–38. doi:10.1101/gad.219626.113.
41. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307–19. doi:10.1016/j.cell.2013.03.035.

42. Eberhardy SR, D'Cunha CA, Farnham PJ. Direct examination of histone acetylation on Myc target genes using chromatin immunoprecipitation. J Biol Chem. 2000;275(43):33798–805. doi:10.1074/jbc.M005154200.
43. Sabo A, Amati B. Genome recognition by MYC. Cold Spring Harb Perspect Med. 2014;. doi:10.1101/cshperspect.a014191.
44. Perna D, Faga G, Verrecchia A, Gorski MM, Barozzi I, Narang V, et al. Genome-wide mapping of Myc binding and gene regulation in serum-stimulated fibroblasts. Oncogene. 2012;31(13):1695–709. doi:10.1038/onc.2011.359.
45. Kusch T, Mei A, Nguyen C. Histone H3 lysine 4 trimethylation regulates cotranscriptional H2A variant exchange by Tip60 complexes to maximize gene expression. Proc Natl Acad Sci USA. 2014;111(13):4850–5. doi:10.1073/pnas.1320337111.
46. Krebs AR, Karmodiya K, Lindahl-Allen M, Struhl K, Tora L. SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Mol Cell. 2011;44(3):410–23. doi:10.1016/j.molcel.2011.08.037.
47. Sapountzi V, Logan IR, Robson CN. Cellular functions of TIP60. Int J Biochem Cell Biol. 2006;38(9):1496–509. doi:10.1016/j.biocel.2006.03.003.
48. Bonnet J, Wang CY, Baptista T, Vincent SD, Hsiao WC, Stierle M, et al. The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes Dev. 2014;28(18):1999–2012. doi:10.1101/gad.250225.114.
49. Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Lariviere L, et al. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res. 2012;22(7):1350–9. doi:10.1101/gr.130161.111.
50. Helenius K, Yang Y, Tselykh TV, Pessa HK, Frilander MJ, Makela TP. Requirement of TFIIH kinase subunit Mat1 for RNA Pol II C-terminal domain Ser5 phosphorylation, transcription and mRNA turnover. Nucleic Acids Res. 2011;39(12):5025–35. doi:10.1093/nar/gkr107.
51. Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell. 2015;161(3):541–54. doi:10.1016/j.cell.2015.03.010.
52. Hooper M, Hardy K, Handyside A, Hunter S, Monk M. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature. 1987;326(6110):292–5. doi:10.1038/326292a0.
53. Langmead B. Aligning short sequencing reads with Bowtie. Current protocols in bioinformatics/editoral board, Andreas D Baxevanis [et al]. 2010; Chapter 11: Unit 11 7. doi:10.1002/0471250953.bi1107s32.
54. Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L. 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 Genom. 2012;13:424. doi:10.1186/1471-2164-13-424.
55. Taylor GC, Eskeland R, Hekimoglu-Balkan B, Pradeepa MM, Bickmore WA. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. Genome Res. 2013;23(12):2053–65. doi:10.1101/gr.155028.113.
56. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012;488(7409):116–20. doi:10.1038/nature11243.
57. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448(7153):553–60. doi:10.1038/nature06008.
58. Robert F, Hardy S, Nagy Z, Baldeyron C, Murr R, Dery U, et al. The transcriptional histone acetyltransferase cofactor TRRAP associates with the MRN repair complex and plays a role in DNA double-strand break repair. Mol Cell Biol. 2006;26(2):402–12. doi:10.1128/MCB.26.2.402-412.2006.
59. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576–89. doi:10.1016/j.molcel.2010.05.004.
60. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11. doi:10.1093/bioinformatics/btp120.
61. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics. 2011;27(17):2325–9. doi:10.1093/bioinformatics/btr355.
62. Tassy O, Pourquie O. Manteia, a predictive data mining system for vertebrate genes and its applications to human genetic diseases. Nucleic Acids Res. 2014;42(Database issue):D882–91. doi:10.1093/nar/gkt807.

Bibliography

Acker, J., de Graaff, M., Cheynel, I., Khazak, V., Kedinger, C., and Vigneron, M. (1997). Interactions between the human RNA polymerase II subunits. J Biol Chem *272*, 16815-16821.

Adam, R.C., Yang, H., Rockowitz, S., Larsen, S.B., Nikolova, M., Oristian, D.S., Polak, L., Kadaja, M., Asare, A., Zheng, D.*, et al.* (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. Nature *521*, 366-370.

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

Agalioti, T., Chen, G., and Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. Cell *111*, 381-392.

Aguilar-Gurrieri, C., Larabi, A., Vinayachandran, V., Patel, N.A., Yen, K., Reja, R., Ebong, I.O., Schoehn, G., Robinson, C.V., Pugh, B.F.*, et al.* (2016). Structural evidence for Nap1-dependent H2A-H2B deposition and nucleosome assembly. EMBO J *35*, 1465-1482.

Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol Cell *5*, 367-375.

Akhtar, W., and Veenstra, G.J.C. (2009). TBP2 is a substitute for TBP in Xenopus oocyte transcription. BMC biology *7*.

Akhtar, W., and Veenstra, G.J.C. (2011). TBP-related factors: a paradigm of diversity in transcription initiation. Cell and Bioscience *1*.

Albertini, D.F., and Gleicher, N. (2015). A detour in the quest for oogonial stem cells: methods matter. Nat Med *21*, 1126-1127.

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. EMBO J *18*, 5108-5119.

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

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat Rev Genet *17*, 487-500.

Alpern, D., Langer, D., Ballester, B., Le Gras, S., Romier, C., Mengus, G., and Davidson, I. (2014). TAF4, a subunit of transcription factor II D, directs promoter occupancy of nuclear receptor HNF4A during post-natal hepatocyte differentiation. Elife *3*, e03613.

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E.*, et al.* (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat Cell Biol *18*, 225-233.

Ancelin, K., Lange, U.C., Hajkova, P., Schneider, R., Bannister, A.J., Kouzarides, T., and Surani, M.A. (2006). Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. Nat Cell Biol *8*, 623-630.

Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54-59.

Andersson, R., Chen, Y., Core, L., Lis, J.T., Sandelin, A., and Jensen, T.H. (2015). Human Gene Promoters Are Intrinsically Bidirectional. Mol Cell *60*, 346-347.

Angelov, D., Bondarenko, V.A., Almagro, S., Menoni, H., Mongelard, F., Hans, F., Mietton, F., Studitsky, V.M., Hamiche, A., Dimitrov, S.*, et al.* (2006). Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. EMBO J *25*, 1669-1679.

Aramaki, S., Hayashi, K., Kurimoto, K., Ohta, H., Yabuta, Y., Iwanari, H., Mochizuki, Y., Hamakubo, T., Kato, Y., Shirahige, K.*, et al.* (2013). A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. Dev Cell *27*, 516-529.

Auger, A., Galarneau, L., Altaf, M., Nourani, A., Doyon, Y., Utley, R.T., Cronier, D., Allard, S., and Cote, J. (2008). Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants. Mol Cell Biol *28*, 2257-2270.

Baarends, W.M., Wassenaar, E., Hoogerbrugge, J.W., Schoenmakers, S., Sun, Z.W., and Grootegoed, J.A. (2007). Increased phosphorylation and dimethylation of XY body histones in the Hr6b-knockout mouse is associated with derepression of the X chromosome. J Cell Sci *120*, 1841-1851.

Babiarz, J.E., Halley, J.E., and Rine, J. (2006). Telomeric heterochromatin boundaries require NuA4-dependent acetylation of histone variant H2A.Z in Saccharomyces cerevisiae. Genes Dev *20*, 700-710.

Babu, A., and Verma, R.S. (1987). Chromosome structure: euchromatin and heterochromatin. Int Rev Cytol *108*, 1-60.

Banaszynski, L.A., Wen, D., Dewell, S., Whitcomb, S.J., Lin, M., Diaz, N., Elsasser, S.J., Chapgier, A., Goldberg, A.D., Canaani, E.*, et al.* (2013). Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell *155*, 107-120.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299-308.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell

Res *21*, 381-395.

Bao, J., and Bedford, M.T. (2016). Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. Reproduction *151*, R55-70.

Baptista, T., Grunberg, S., Minoungou, N., Koster, M.J.E., Timmers, H.T.M., Hahn, S., Devys, D., and Tora, L. (2017). SAGA Is a General Cofactor for RNA Polymerase II Transcription. Mol Cell *68*, 130-143 e135.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Herault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354*, 909-912.

Barber, C.M., Turner, F.B., Wang, Y., Hagstrom, K., Taverna, S.D., Mollah, S., Ueberheide, B., Meyer, B.J., Hunt, D.F., Cheung, P.*, et al.* (2004). The enhancement of histone H4 and H2A serine 1 phosphorylation during mitosis and S-phase is evolutionarily conserved. Chromosoma *112*, 360-371.

Bardot, P., Vincent, S.D., Fournier, M., Hubaud, A., Joint, M., Tora, L., and Pourquie, O. (2017). The TAF10-containing TFIID and SAGA transcriptional complexes are dispensable for early somitogenesis in the mouse embryo. Development *144*, 3808-3818.

Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., and Natoli, G. (2014). Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. Mol Cell *54*, 844-857.

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

Bartfai, R., Balduf, C., Hilton, T., Rathmann, Y., Hadzhiev, Y., Tora, L., Orban, L., and Muller, F. (2004). TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. Current Biology *14*, 593-598.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243-247.

Bauer, U.M., Daujat, S., Nielsen, S.J., Nightingale, K., and Kouzarides, T. (2002). Methylation at arginine 17 of histone H3 is linked to gene activation. EMBO Rep *3*, 39-44.

Bell, B., Scheer, E., and Tora, L. (2001). Identification of hTAF(II)80 delta links apoptotic signaling pathways to transcription factor TFIID function. Mol Cell *8*, 591-600.

Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N., Schwaiger, M., MacAlpine, D.M., Zilbermann, F., van Leeuwen, F., Bell, S.P., Imhof, A.*, et al.* (2007). Localized H3K36 methylation states define histone

H4K16 acetylation during transcriptional elongation in Drosophila. EMBO J *26*, 4974-4984.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090-1093.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. Nature *290*, 304-310.

Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes Dev *23*, 781-783.

Berk, A.J. (2000). TBP-like factors come into focus. Cell *103*, 5-8.

Berkes, C.A., Bergstrom, D.A., Penn, B.H., Seaver, K.J., Knoepfler, P.S., and Tapscott, S.J. (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Mol Cell *14*, 465-477.

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

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

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bettegowda, A., and Smith, G.W. (2007). Mechanisms of maternal mRNA regulation: implications for mammalian early embryonic development. Front Biosci *12*, 3713-3726.

Bhaumik, S.R., Raha, T., Aiello, D.P., and Green, M.R. (2004). In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. Genes Dev *18*, 333-343.

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. Cell *152*, 1270-1284.

Biddie, S.C., John, S., Sabo, P.J., Thurman, R.E., Johnson, T.A., Schiltz, R.L., Miranda, T.B., Sung, M.H., Trump, S., Lightman, S.L.*, et al.* (2011). Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. Mol Cell *43*, 145-155.

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.

Birck, C., Poch, O., Romier, C., Ruff, M., Mengus, G., Lavigne, A.C., Davidson, I., and Moras, D. (1998). Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical

evolutionary conserved motifs also found in the SPT3 family. Cell *94*, 239-249.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev *16*, 6-21.

Birling, M.C., Dierich, A., Jacquot, S., Herault, Y., and Pavlovic, G. (2012). Highly-efficient, fluorescent, locus directed cre and FlpO deleter mice on a pure C57BL/6N genetic background. Genesis *50*, 482-489.

Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. Science *281*, 60-63.

Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. (2004). Removal of promoter nucleosomes by disassembly rather than sliding in vivo. Mol Cell *14*, 667-673.

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

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. J Biol Chem *285*, 188-196.

Bogolyubov, D.S. (2018). Karyosphere (Karyosome): A Peculiar Structure of the Oocyte Nucleus. Int Rev Cell Mol Biol *337*, 1-48.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat Rev Genet *17*, 661-678.

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

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Boulard, M., Storck, S., Cong, R., Pinto, R., Delage, H., and Bouvet, P. (2010). Histone variant macroH2A1 deletion in mice causes female-specific steatosis. Epigenetics Chromatin *3*, 8.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536-2539.

Braunschweig, U., Hogan, G.J., Pagie, L., and van Steensel, B. (2009). Histone H1 binding is inhibited by histone variant H3.3. EMBO J *28*, 3635-3645.

Brind'Amour, J., Liu, S., Hudson, M., Chen, C., Karimi, M.M., and Lorincz, M.C. (2015). An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. Nat Commun *6*, 6033.

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.

Budry, L., Balsalobre, A., Gauthier, Y., Khetchoumian, K., L'Honore, A., Vallette, S., Brue, T., Figarella-Branger, D., Meij, B., and Drouin, J. (2012). The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. Genes Dev *26*, 2299-2310.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838-853.

Bulger, M., and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. Genes Dev *13*, 2465-2477.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol Cell *36*, 541-546.

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.

Buratowski, S., and Zhou, H. (1993). Functional domains of transcription factor TFIIB. Proc Natl Acad Sci U S A *90*, 5633-5637.

Burgess, R.J., and Zhang, Z. (2013). Histone chaperones in nucleosome assembly and human disease. Nat Struct Mol Biol *20*, 14-22.

Burke, K.A., Janke, A.M., Rhine, C.L., and Fawzi, N.L. (2015). Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. Mol Cell *60*, 231-241.

Burke, T.W., and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev *10*, 711-724.

Burke, T.W., and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev *11*, 3020-3031.

Burton, A., and Torres-Padilla, M.E. (2014). Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. Nat Rev Mol Cell Biol *15*, 723-734.

Burton, Z.F., Killeen, M., Sopta, M., Ortolan, L.G., and Greenblatt, J. (1988). RAP30/74: a general initiation factor that binds to RNA polymerase II. Mol Cell Biol *8*, 1602-1613.

Buschbeck, M., and Hake, S.B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. Nat Rev Mol Cell Biol *18*, 299-314.

Buschbeck, M., Uribesalgo, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., Morey, L., Guigo, R.,

Lopez-Schier, H., and Di Croce, L. (2009). The histone variant macroH2A is an epigenetic regulator of key developmental genes. Nat Struct Mol Biol *16*, 1074-1079.

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. Proc Natl Acad Sci U S A *100*, 6969-6973.

Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science *303*, 983-988.

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. Proc Natl Acad Sci U S A *108*, 15786-15791.

Cairns, B.R., Henry, N.L., and Kornberg, R.D. (1996). TFG/TAF30/ANC1, a component of the yeast SWI/SNF complex that is similar to the leukemogenic proteins ENL and AF-9. Mol Cell Biol *16*, 3308-3316.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

Cao, Z., Chen, C., He, B., Tan, K., and Lu, C. (2015). A microfluidic device for epigenomic profiling using 100 cells. Nat Methods.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C.*, et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Carone, B.R., Hung, J.H., Hainer, S.J., Chou, M.T., Carone, D.M., Weng, Z., Fazzio, T.G., and Rando, O.J. (2014). High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. Dev Cell *30*, 11-22.

Casadio, F., Lu, X., Pollock, S.B., LeRoy, G., Garcia, B.A., Muir, T.W., Roeder, R.G., and Allis, C.D. (2013). H3R42me2a is a histone modification with positive transcriptional effects. Proc Natl Acad Sci U S A *110*, 14894-14899.

Castrillon, D.H., Miao, L., Kollipara, R., Horner, J.W., and DePinho, R.A. (2003). Suppression of ovarian follicle activation in mice by the transcription factor Foxo3a. Science *301*, 215-218.

Catena, R., Argentini, M., Martianov, I., Parello, C., Brancorsini, S., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2005). Proteolytic cleavage of ALF into alpha- and beta-subunits that form homologous and heterologous complexes with somatic TFIIA and TRF2 in male germ cells. FEBS Lett *579*, 3401-3410.

Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.T., Sedelnikova, O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M.J.*, et al.* (2002). Genomic instability in mice lacking histone H2AX. Science *296*, 922-927.

Chalkley, G.E., and Verrijzer, C.P. (1999). DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. EMBO J *18*, 4835-4845.

Chang, C.C., Ma, Y., Jacobs, S., Tian, X.C., Yang, X., and Rasmussen, T.P. (2005). A maternal store of macroH2A is removed from pronuclei prior to onset of somatic macroH2A expression in preimplantation embryos. Dev Biol *278*, 367-380.

Changolkar, L.N., Costanzi, C., Leu, N.A., Chen, D., McLaughlin, K.J., and Pehrson, J.R. (2007). Developmental changes in histone macroH2A1-mediated gene regulation. Mol Cell Biol *27*, 2758-2764.

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 Res *32*, 35-44.

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

Chen, S., Rufiange, A., Huang, H., Rajashankar, K.R., Nourani, A., and Patel, D.J. (2015). Structure-function studies of histone H3/H4 tetramer maintenance during transcription by chaperone Spt2. Genes Dev *29*, 1326-1340.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003a). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol *23*, 5594-5605.

Chen, Y., Zheng, Y., Gao, Y., Lin, Z., Yang, S., Wang, T., Wang, Q., Xie, N., Hua, R., Liu, M.*, et al.* (2018b). Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. Cell Res *28*, 879-896.

Chen, Z., and Manley, J.L. (2000). Robust mRNA transcription in chicken DT40 cells depleted of TAF(II)31 suggests both functional degeneracy and evolutionary divergence. Mol Cell Biol *20*, 5064-5076.

Chen, Z., and Manley, J.L. (2003b). In vivo functional analysis of the histone 3-like TAF9 and a TAF9-related factor, TAF9L. J Biol Chem *278*, 35172-35183.

Cheng, Y., Buffone, M.G., Kouadio, M., Goodheart, M., Page, D.C., Gerton, G.L., Davidson, I., and Wang, P.J. (2007). Abnormal sperm in mice lacking the Taf7l gene. Mol Cell Biol *27*, 2582-2589.

Cheung, A.C., and Cramer, P. (2012). A movie of RNA polymerase II transcription. Cell *149*, 1431-1437.

Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev *11*, 3319-3326.

Choi, Y., Ballow, D.J., Xin, Y., and Rajkovic, A. (2008a). Lim homeobox gene, lhx8, is essential for mouse oocyte differentiation and survival. Biol Reprod *79*, 442-449.

Choi, Y., Yuan, D., and Rajkovic, A. (2008b). Germ cell-specific transcriptional regulator sohlh2 is essential for early mouse folliculogenesis and oocyte-specific gene expression. Biol Reprod *79*, 1176-1182.

Chong, J.A., Moran, M.M., Teichmann, M., Kaczmarek, J.S., Roeder, R., and Clapham, D.E. (2005). TATA-binding protein (TBP)-like factor (TLF) is a functional regulator of transcription: reciprocal regulation of the neurofibromatosis type 1 and c-fos genes by TLF/TRF2 and TBP. Mol Cell Biol *25*, 2632-2643.

Chotalia, M., Smallwood, S.A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., and Kelsey, G. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev *23*, 105-117.

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.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell *9*, 279-289.

Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., and Zaret, K.S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J *17*, 244-254.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273-304.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol *18*, 407-422.

Clift, D., and Schuh, M. (2013). Restarting life: fertilization and the transition from meiosis to mitosis. Nat Rev Mol Cell Biol *14*, 549-562.

Coin, F., Bergmann, E., Tremeau-Bravard, A., and Egly, J.M. (1999). Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. EMBO J *18*, 1357-1366.

Coleman, R.A., and Pugh, B.F. (1995). Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA. J Biol Chem *270*, 13850-13859.

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

Conaway, J.W., Reines, D., and Conaway, R.C. (1990). Transcription initiated by RNA polymerase II and purified transcription factors from liver. Cooperative action of transcription factors tau and epsilon in initial complex formation. J Biol Chem *265*, 7552-7558.

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. Proc Natl Acad Sci U S A *88*, 6205-6209.

Cook, A.J., Gurard-Levin, Z.A., Vassias, I., and Almouzni, G. (2011). A specific function for the histone chaperone NASP to fine-tune a reservoir of soluble H3-H4 in the histone supply chain. Mol Cell *44*, 918-927.

Cordeiro, M.H., Kim, S.Y., Ebbert, K., Duncan, F.E., Ramalho-Santos, J., and Woodruff, T.K. (2015). Geography of Follicle Formation in the Embryonic Mouse Ovary Impacts Activation Pattern during the First Wave of Folliculogenesis. Biol Reprod.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311-1320.

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

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

Crowley, T.E., Hoey, T., Liu, J.K., Jan, Y.N., Jan, L.Y., and Tjian, R. (1993). A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila. Nature *361*, 557-561.

Cuesta, I., Zaret, K.S., and Santisteban, P. (2007). The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure. Mol Cell Biol *27*, 7302-7314.

Culty, M. (2009). Gonocytes, the forgotten cells of the germ cell lineage. Birth Defects Res C Embryo Today *87*, 1-26.

Cvetesic, N., Leitch, H.G., Borkowska, M., Muller, F., Carninci, P., Hajkova, P., and Lenhard, B. (2018). SLIC-CAGE: high-resolution transcription start site mapping using nanogram-levels of total RNA. Genome Res.

D'Alessio, J.A., Ng, R., Willenbring, H., and Tjian, R. (2011). Core promoter recognition complex changes accompany liver development. Proc Natl Acad Sci U S A *108*, 3906-3911.

D'Alessio, J.A., Wright, K.J., and Tjian, R. (2009). Shifting players and paradigms in cell-specific transcription. Mol Cell *36*, 924-931.

Dahl, J.A., and Collas, P. (2008). A rapid micro chromatin immunoprecipitation assay (microChIP). Nat Protoc *3*, 1032-1045.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y.*, et al.* (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537*, 548-552.

Dang, X., Singh, A., Spetman, B.D., Nolan, K.D., Isaacs, J.S., Dennis, J.H., Dalton, S., Marshall, A.G., and Young, N.L. (2016). Label-Free Relative Quantitation of Isobaric and Isomeric Human Histone H2A and H2B Variants by Fourier Transform Ion Cyclotron Resonance Top-Down MS/MS. J Proteome Res *15*, 3196-3203.

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

Dantonel, J.C., Wurtz, J.M., Poch, O., Moras, D., and Tora, L. (1999). The TBP-like factor: an alternative transcription factor in metazoa? Trends Biochem Sci *24*, 335-339.

Daujat, S., Weiss, T., Mohn, F., Lange, U.C., Ziegler-Birling, C., Zeissler, U., Lappe, M., Schubeler, D., Torres-Padilla, M.E., and Schneider, R. (2009). H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. Nat Struct Mol Biol *16*, 777-781.

Davidson, I. (2003). The genetics of TBP and TBP-related factors. Trends Biochem Sci *28*, 391-398.

Dawson, M.A., Bannister, A.J., Gottgens, B., Foster, S.D., Bartke, T., Green, A.R., and Kouzarides, T. (2009). JAK2 phosphorylates histone H3Y41 and excludes HP1alpha from chromatin. Nature *461*, 819-822.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet *49*, 941-945.

De La Fuente, R. (2006). Chromatin modifications in the germinal vesicle (GV) of mammalian oocytes. Dev Biol *292*, 1-12.

De La Fuente, R., Viveiros, M.M., Burns, K.H., Adashi, E.Y., Matzuk, M.M., and Eppig, J.J. (2004). Major chromatin remodeling in the germinal vesicle (GV) of mammalian oocytes is dispensable for global transcriptional silencing but required for centromeric heterochromatin function. Dev Biol *275*, 447-458.

Deato, M.D., and Tjian, R. (2007). Switching of the core transcription machinery during myogenesis. Genes Dev *21*, 2137-2149.

Deato, M.D.E., Marr, M.T., Sottero, T., Inouye, C., Hu, P., and Tjian, R. (2008). MyoD Targets TAF3/TRF3 to Activate Myogenin Transcription. Molecular Cell *32*, 96-105.

Dekker, J., Belmont, A.S., Guttman, M., Leshyk, V.O., Lis, J.T., Lomvardas, S., Mirny, L.A., O'Shea, C.C., Park, P.J., Ren, B.*, et al.* (2017). The 4D nucleome project. Nature *549*, 219-226.

Deng, W., and Roberts, S.G. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev *19*, 2418-2423.

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

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

Dikstein, R., Zhou, S., and Tjian, R. (1996). Human TAFII 105 is a cell type-specific TFIID subunit related to hTAFII130. Cell *87*, 137-146.

Dogan, E.S., and Liu, C. (2018). Three-dimensional chromatin packing and positioning of plant genomes. Nature plants *4*, 521-529.

Donaghey, J., Thakurela, S., Charlton, J., Chen, J.S., Smith, Z.D., Gu, H., Pop, R., Clement, K., Stamenova, E.K., Karnik, R.*, et al.* (2018). Genetic determinants and epigenetic effects of pioneer-factor occupancy. Nat Genet *50*, 250-258.

Dorsett, D. (1999). Distant liaisons: long-range enhancer-promoter interactions in Drosophila. Curr Opin Genet Dev *9*, 505-514.

Drane, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev *24*, 1253-1265.

Dronamraju, R., Hepperla, A.J., Shibata, Y., Adams, A.T., Magnuson, T., Davis, I.J., and Strahl, B.D. (2018). Spt6 Association with RNA Polymerase II Directs mRNA Turnover During Transcription. Mol Cell *70*, 1054-1066 e1054.

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., He, J., Xiang, Y., Wang, Q., Li, Y.*, et al.* (2017). Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature *547*, 232-235.

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell *137*, 485-497.

Dutta, S., Burks, D.M., and Pepling, M.E. (2016). Arrest at the diplotene stage of meiotic prophase I is delayed by progesterone but is not required for primordial follicle formation in mice. Reprod Biol Endocrinol *14*, 82.

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

Ea, V., Baudement, M.O., Lesne, A., and Forne, T. (2015). Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. Genes *6*, 734-750.

Eckersley-Maslin, M.A., Alda-Catalinas, C., Blotenburg, M., Kreibich, E., Krueger, C., and Reik, W. (2018). Dppa2 and Dppa4 directly regulate the Dux driven zygotic transcriptional programme. bioRxiv.

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

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

El-Saafin, F., Curry, C., Ye, T., Garnier, J.M., Kolb-Cheynel, I., Stierle, M., Downer, N.L., Dixon, M.P., Negroni, L., Berger, I.*, et al.* (2018). Homozygous TAF8 mutation in a patient with intellectual disability results in undetectable TAF8 protein, but preserved RNA polymerase II transcription. Hum Mol Genet *27*, 2171-2186.

Elsasser, S.J., Noh, K.M., Diaz, N., Allis, C.D., and Banaszynski, L.A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature *522*, 240-244.

Erkek, S., Hisano, M., Liang, C.Y., Gill, M., Murr, R., Dieker, J., Schubeler, D., van der Vlag, J., Stadler, M.B., and Peters, A.H. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nat Struct Mol Biol *20*, 868-875.

Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. Curr Biol *11*, 1183-1187.

Fadloun, A., Kobi, D., Delacroix, L., Dembele, D., Michel, I., Lardenois, A., Tisserand, J., Losson, R., Mengus, G., and Davidson, I. (2008). Retinoic acid induces TGFbeta-dependent autocrine fibroblast growth. Oncogene *27*, 477-489.

Falender, A.E., Freiman, R.N., Geles, K.G., Lo, K.C., Hwang, K., Lamb, D.J., Morris, P.L., Tjian, R., and Richards, J.S. (2005a). Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID. Genes Dev *19*, 794-803.

Falender, A.E., Shimada, M., Lo, Y.K., and Richards, J.S. (2005b). TAF4b, a TBP associated factor, is required for oocyte development and function. Dev Biol *288*, 405-419.

Fei, J., Torigoe, S.E., Brown, C.R., Khuong, M.T., Kassavetis, G.A., Boeger, H., and Kadonaga, J.T. (2015). The prenucleosome, a stable conformational isomer of the nucleosome. Genes Dev *29*, 2563-2575.

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., and Hahn, S. (2012). Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. Mol Cell Biol *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. Proc Natl Acad Sci U S A *112*, 3961-3966.

FitzGerald, P.C., Sturgill, D., Shyakhtenko, A., Oliver, B., and Vinson, C. (2006). Comparative genomics of Drosophila and human core promoters. Genome Biol *7*, R53.

Flaus, A., Martin, D.M., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res *34*, 2887-2905.

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

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F., and Reinberg, D. (1991). The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc Natl Acad Sci U S A *88*, 9999-10003.

Flores, O., Lu, H., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. J Biol Chem *267*, 2786-2793.

Flores, O., Maldonado, E., Burton, Z., Greenblatt, J., and Reinberg, D. (1988). Factors involved in specific transcription by mammalian RNA polymerase II. RNA polymerase II-associating protein 30 is an essential component of transcription factor IIF. J Biol Chem *263*, 10812-10816.

Flores, O., Maldonado, E., and Reinberg, D. (1989). Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. J Biol Chem *264*, 8913-8921.

Flyamer, I.M., Gassler, J., Imakaev, M., Brandao, H.B., Ulianov, S.V., Abdennur, N., Razin, S.V., Mirny, L.A., and Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature *544*, 110-114.

Foltz, D.R., Jansen, L.E., Bailey, A.O., Yates, J.R., 3rd, Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-specific assembly of CENP-a nucleosomes is mediated by

HJURP. Cell *137*, 472-484.

Frank, S.R., Parisi, T., Taubert, S., Fernandez, P., Fuchs, M., Chan, H.M., Livingston, D.M., and Amati, B. (2003). MYC recruits the TIP60 histone acetyltransferase complex to chromatin. EMBO Rep *4*, 575-580.

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

Frontini, M., Soutoglou, E., Argentini, M., Bole-Feysot, C., Jost, B., Scheer, E., and Tora, L. (2005). TAF9b (formerly TAF9L) is a bona fide TAF that has unique and overlapping roles with TAF9. Mol Cell Biol *25*, 4638-4649.

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

Fuchs, M., Gerber, J., Drapkin, R., Sif, S., Ikura, T., Ogryzko, V., Lane, W.S., Nakatani, Y., and Livingston, D.M. (2001). The p400 complex is an essential E1A transformation target. Cell *106*, 297-307.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186-192.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science *361*, 1341-1345.

Fusauchi, Y., and Iwai, K. (1984). Tetrahymena histone H2A. Acetylation in the N-terminal sequence and phosphorylation in the C-terminal sequence. J Biochem *95*, 147-154.

Fussner, E., Strauss, M., Djuric, U., Li, R., Ahmed, K., Hart, M., Ellis, J., and Bazett-Jones, D.P. (2012). Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO Rep *13*, 992-996.

Gangloff, Y.G., Sanders, S.L., Romier, C., Kirschner, D., Weil, P.A., Tora, L., and Davidson, I. (2001). Histone folds mediate selective heterodimerization of yeast TAF(II)25 with TFIID components yTAF(II)47 and yTAF(II)65 and with SAGA component ySPT7. Mol Cell Biol *21*, 1841-1853.

Gaughan, L., Logan, I.R., Cook, S., Neal, D.E., and Robson, C.N. (2002). Tip60 and histone deacetylase 1 regulate androgen receptor activity through changes to the acetylation status of the receptor. J Biol Chem *277*, 25904-25913.

Gay, L., Karfilis, K.V., Miller, M.R., Doe, C.Q., and Stankunas, K. (2014). Applying thiouracil tagging to mouse transcriptome analysis. Nat Protoc *9*, 410-420.

Gay, L., Miller, M.R., Ventura, P.B., Devasthali, V., Vue, Z., Thompson, H.L., Temple, S., Zong, H., Cleary, M.D., Stankunas, K.*, et al.* (2013). Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. Genes Dev *27*, 98-115.

Gazdag, E. (2008). Functional studies of TBP2 (TATA-binding protein 2) : The vertebrate specific homolog of TBP. PhD thesis, University of Strasbourg, Strasbourg, FR.

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

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

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

Gegonne, A., Weissman, J.D., Lu, H., Zhou, M., Dasgupta, A., Ribble, R., Brady, J.N., and Singer, D.S. (2008). TFIID component TAF7 functionally interacts with both TFIIH and P-TEFb. Proc Natl Acad Sci U S A *105*, 5367-5372.

Gegonne, A., Weissman, J.D., and Singer, D.S. (2001). TAFII55 binding to TAFII250 inhibits its acetyltransferase activity. Proc Natl Acad Sci U S A *98*, 12432-12437.

Gegonne, A., Weissman, J.D., Zhou, M., Brady, J.N., and Singer, D.S. (2006). TAF7: a possible transcription initiation check-point regulator. Proc Natl Acad Sci U S A *103*, 602-607.

Geles, K.G., Freiman, R.N., Liu, W.L., Zheng, S., Voronina, E., and Tjian, R. (2006). Cell-type-selective induction of c-jun by TAF4b directs ovarian-specific transcription networks. Proc Natl Acad Sci U S A *103*, 2594-2599.

Gerard, M., Fischer, L., Moncollin, V., Chipoulet, J.M., Chambon, P., and Egly, J.M. (1991). Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. J Biol Chem *266*, 20940-20945.

Gevry, N., Chan, H.M., Laflamme, L., Livingston, D.M., and Gaudreau, L. (2007). p21 transcription is regulated by differential localization of histone H2A.Z. Genes Dev *21*, 1869-1881.

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

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Mol Cell *49*, 773-782.

Gill, M.E., Erkek, S., and Peters, A.H.F.M. (2012). Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming? Current Opinion in Cell Biology *24*, 387-396.

Ginsburg, M., Snow, M.H., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. Development *110*, 521-528.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X.*, et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678-691.

Goldberg, M.L. (1979). Sequence analysis of Drosophila histone genes. PhD thesis, Stanford University, Stanford, CA.

Golebiowski, F., and Kasprzak, K.S. (2005). Inhibition of core histones acetylation by carcinogenic nickel(II). Mol Cell Biochem *279*, 133-139.

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.

Goodrich, J.A., and Tjian, R. (1994). Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell *77*, 145-156.

Goodrich, J.A., and Tjian, R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet *11*, 549-558.

Goppelt, A., and Meisterernst, M. (1996a). Characterization of the basal inhibitor of class II transcription NC2 from Saccharomyces cerevisiae. Nucleic Acids Res *24*, 4450-4455.

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

Greber, B.J., Nguyen, T.H.D., Fang, J., Afonine, P.V., Adams, P.D., and Nogales, E. (2017). The cryo-electron microscopy structure of human transcription factor IIH. Nature *549*, 414-417.

Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet *13*, 343-357.

Griffith, G.J., Trask, M.C., Hiller, J., Walentuk, M., Pawlak, J.B., Tremblay, K.D., and Mager, J. (2011). Yin-yang1 is required in the mammalian oocyte for follicle expansion. Biol Reprod *84*, 654-663.

Griswold, M.D. (2016). Spermatogenesis: The Commitment to Meiosis. Physiological reviews *96*, 1-17.

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

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

Guermah, M., Ge, K., Chiang, C.M., and Roeder, R.G. (2003). The TBN protein, which is essential for early embryonic mouse development, is an inducible TAFII implicated in adipogenesis. Mol Cell *12*, 991-1001.

Guibert, S., Forne, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res *22*, 633-641.

Gumbs, O.H., Campbell, A.M., and Weil, P.A. (2003). High-affinity DNA binding by a Mot1p-TBP complex: implications for TAF-independent transcription. EMBO J *22*, 3131-3141.

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

Gurard-Levin, Z.A., Quivy, J.P., and Almouzni, G. (2014). Histone chaperones: assisting histone traffic and nucleosome dynamics. Annu Rev Biochem *83*, 487-517.

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

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

Hahn, S. (1998). Activation and the role of reinitiation in the control of transcription by RNA polymerase II. Cold Spring Harb Symp Quant Biol *63*, 181-188.

Haimovich, G., Choder, M., Singer, R.H., and 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.

Hainer, S.J., Boskovic, A., Rando, O.J., and Fazzio, T.G. (2018). Profiling of pluripotency factors in individual stem cells and early embryos. bioRxiv.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech Dev *117*, 15-23.

Hammond, C.M., Stromme, C.B., Huang, H., Patel, D.J., and Groth, A. (2017). Histone chaperone networks shaping chromatin function. Nat Rev Mol Cell Biol *18*, 141-158.

Hammoud, S.S., Low, D.H., Yi, C., Carrell, D.T., Guccione, E., and Cairns, B.R. (2014). Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. Cell Stem Cell *15*, 239-253.

Hanna, C.W., Demond, H., and Kelsey, G. (2018a). Epigenetic regulation in development: is the mouse a good model for the human? Hum Reprod Update *24*, 556-576.

Hanna, C.W., Taudt, A., Huang, J., Gahurova, L., Kranz, A., Andrews, S., Dean, W., Stewart, A.F., Colome-Tatche, M., and Kelsey, G. (2018b). MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. Nat Struct Mol Biol *25*, 73-82.

Hansen, S.K., Takada, S., Jacobson, R.H., Lis, J.T., and Tjian, R. (1997). Transcription properties of a cell type-specific TATA-binding protein, TRF. Cell *91*, 71-83.

Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res *21*, 396-420.

Hariharan, N., and Perry, R.P. (1990). Functional dissection of a mouse ribosomal protein promoter: significance of the polypyrimidine initiator and an element in the TATA-box region. Proc Natl Acad Sci U S A *87*, 1526-1530.

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

Hart, D.O., Raha, T., Lawson, N.D., and Green, M.R. (2007). Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3. Nature *450*, 1082-1085.

Hatzis, P., and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol Cell *10*, 1467-1477.

Hayashi, K., de Sousa Lopes, S.M., and Surani, M.A. (2007). Germ cell specification in mice. Science *316*, 394-396.

Heberlein, U., England, B., and Tjian, R. (1985). Characterization of Drosophila transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. Cell *41*, 965-977.

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. EMBO J *30*, 2843-2852.

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

Hendrickson, P.G., Dorais, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L.*, et al.* (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet *49*, 925-934.

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 Dev *8*, 2868-2878.

Henry, N.L., Sayre, M.H., and Kornberg, R.D. (1992). Purification and characterization of yeast RNA polymerase II general initiation factor g. J Biol Chem *267*, 23388-23392.

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem *279*, 48350-48359.

Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? Genes Dev *7*, 1291-1308.

Herrera, F.J., Yamaguchi, T., Roelink, H., and Tjian, R. (2014). Core promoter factor TAF9B regulates neuronal gene expression. Elife *3*, e02559.

Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nat Methods *14*, 1198-1204.

Hiller, M., Chen, X., Pringle, M.J., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B., Lin, T.Y., Marino, S., and Fuller, M.T. (2004). Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. Development *131*, 5297-5308.

Hiller, M.A., Lin, T.Y., Wood, C., and Fuller, M.T. (2001). Developmental regulation of transcription by a tissue-specific TAF homolog. Genes Dev *15*, 1021-1030.

Hiura, H., Obata, Y., Komiyama, J., Shirai, M., and Kono, T. (2006). Oocyte growth-dependent progression of maternal imprinting in mice. Genes Cells *11*, 353-361.

Hochheimer, A., Zhou, S., Zheng, S., Holmes, M.C., and Tjian, R. (2002). TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. Nature *420*, 439-445.

Hoiby, T., Mitsiou, D.J., Zhou, H., Erdjument-Bromage, H., Tempst, P., and Stunnenberg, H.G. (2004). Cleavage and proteasome-mediated degradation of the basal transcription factor TFIIA. EMBO J *23*, 3083-3091.

Hoiby, T., Zhou, H., Mitsiou, D.J., and Stunnenberg, H.G. (2007). A facelift for the general transcription factor TFIIA. Biochim Biophys Acta *1769*, 429-436.

Holmes, M.C., and Tjian, R. (2000). Promoter-selective properties of the TBP-related factor TRF1. Science *288*, 867-870.

Holstege, F.C., van der Vliet, P.C., and Timmers, H.T. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. EMBO J *15*, 1666-1677.

Hoopes, B.C., LeBlanc, J.F., and Hawley, D.K. (1992). Kinetic analysis of yeast TFIID-TATA box complex formation suggests a multi-step pathway. J Biol Chem *267*, 11539-11547.

Hosaka, T., Biggs, W.H., 3rd, Tieu, D., Boyer, A.D., Varki, N.M., Cavenee, W.K., and Arden, K.C. (2004). Disruption of forkhead transcription factor (FOXO) family members in mice reveals their functional diversification. Proc Natl Acad Sci U S A *101*, 2975-2980.

Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc Natl Acad Sci U S A *97*, 1148-1153.

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

Hsu, J.Y., Juven-Gershon, T., Marr, M.T., 2nd, Wright, K.J., Tjian, R., and Kadonaga, J.T. (2008). TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. Genes Dev *22*, 2353-2358.

Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C., and Zhao, K. (2013a). H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. Cell Stem Cell *12*, 180-192.

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., *et al.* (2013b). DNA methylation presents distinct binding sites for human transcription factors. Elife *2*, e00726.

Huang, H., Stromme, C.B., Saredi, G., Hodl, M., Strandsby, A., Gonzalez-Aguilera, C., Chen, S., Groth, A., and Patel, D.J. (2015). A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks. Nat Struct Mol Biol *22*, 618-626.

Huisinga, K.L., Brower-Toland, B., and Elgin, S.C. (2006). The contradictory definitions of heterochromatin: transcription and silencing. Chromosoma *115*, 110-122.

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.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Exp Mol Med *49*, e324.

Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell *102*, 463-473.

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.

Inostroza, J.A., Mermelstein, F.H., Ha, I., Lane, W.S., and Reinberg, D. (1992). Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription. Cell *70*, 477-489.

Inoue, A., Jiang, L., Lu, F., Suzuki, T., and Zhang, Y. (2017a). Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature *547*, 419-424.

Inoue, A., Jiang, L., Lu, F., and Zhang, Y. (2017b). Genomic imprinting of Xist by maternal H3K27me3. Genes Dev *31*, 1927-1932.

Ishii, H., Kadonaga, J.T., and Ren, B. (2015). MPE-seq, a new method for the genome-wide analysis of chromatin structure. Proc Natl Acad Sci U S A *112*, E3457-3465.

Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., and Tjian, R. (2007a). Transcription of histone gene cluster by differential core-promoter factors. Genes Dev *21*, 2936-2949.

Isogai, Y., Takada, S., Tjian, R., and Keles, S. (2007b). Novel TRF1/BRF target genes revealed by genome-wide analysis of Drosophila Pol III transcription. EMBO J *26*, 79-89.

Itoh, Y., Unzai, S., Sato, M., Nagadoi, A., Okuda, M., Nishimura, Y., and Akashi, S. (2005). Investigation of molecular size of transcription factor TFIIE in solution. Proteins *61*, 633-641.

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., and Zaret, K.S. (2016). The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol Cell *62*, 79-91.

Jack, A.P., Bussemer, S., Hahn, M., Punzeler, S., Snyder, M., Wells, M., Csankovszki, G., Solovei, I., Schotta, G., and Hake, S.B. (2013). H3K56me3 is a novel, conserved heterochromatic mark that largely but not completely overlaps with H3K9me3 in both regulation and localization. PLoS One *8*, e51765.

Jacobi, U.G., Akkers, R.C., Pierson, E.S., Weeks, D.L., Dagle, J.M., and Veenstra, G.J. (2007). TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation. EMBO J *26*, 3900-3909.

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.

Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., and Tora, L. (1994). Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. Cell *79*, 107-117.

Jacquet, K., Fradet-Turcotte, A., Avvakumov, N., Lambert, J.P., Roques, C., Pandita, R.K., Paquet, E., Herst, P., Gingras, A.C., Pandita, T.K.*, et al.* (2016). The TIP60 Complex Regulates Bivalent Chromatin Recognition by 53BP1 through Direct H4K20me Binding and H2AK15 Acetylation. Mol Cell *62*, 409-421.

Jagarlamudi, K., and Rajkovic, A. (2012). Oogenesis: transcriptional regulators and mouse models. Mol Cell Endocrinol *356*, 31-39.

Jallow, Z., Jacobi, U.G., Weeks, D.L., Dawid, I.B., and Veenstra, G.J. (2004). Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in Xenopus. Proc Natl Acad Sci U S A *101*, 13525-13530.

Jan, S.Z., Hamer, G., Repping, S., de Rooij, D.G., van Pelt, A.M., and Vormer, T.L. (2012). Molecular control of rodent spermatogenesis. Biochim Biophys Acta *1822*, 1838-1850.

Jang, C.W., Shibata, Y., Starmer, J., Yee, D., and Magnuson, T. (2015). Histone H3.3 maintains genome integrity during mammalian development. Genes Dev *29*, 1377-1392.

Jawhari, A., Uhring, M., De Carlo, S., Crucifix, C., Tocchini-Valentini, G., Moras, D., Schultz, P., and Poterszman, A. (2006). Structure and oligomeric state of human transcription factor TFIIE. EMBO Rep *7*, 500-505.

Jha, S., and Dutta, A. (2009). RVB1/RVB2: running rings around molecular biology. Mol Cell *34*, 521-533.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet *10*, 161-172.

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

Johnson, J., Canning, J., Kaneko, T., Pru, J.K., and Tilly, J.L. (2004). Germline stem cells and follicular renewal in the postnatal mammalian ovary. Nature *428*, 145-150.

Johnson, L.M., Du, J., Hale, C.J., Bischof, S., Feng, S., Chodavarapu, R.K., Zhong, X., Marson, G., Pellegrini, M., Segal, D.J.*, et al.* (2014). SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature *507*, 124-128.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet *13*, 484-492.

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

Joo, Y.J., Ficarro, S.B., Soares, L.M., Chun, Y., Marto, J.A., and Buratowski, S. (2017). Downstream promoter interactions of TFIID TAFs facilitate transcription reinitiation. Genes Dev *31*, 2162-2174.

Jullien, J., Miyamoto, K., Pasque, V., Allen, G.E., Bradshaw, C.R., Garrett, N.J., Halley-Stott, R.P., Kimura, H., Ohsumi, K., and Gurdon, J.B. (2014). Hierarchical molecular events driven by oocyte-specific factors lead to rapid and extensive reprogramming. Mol Cell *55*, 524-536.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol *339*, 225-229.

Kaltenbach, L., Horner, M.A., Rothman, J.H., and Mango, S.E. (2000). The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during C. elegans embryogenesis. Mol Cell *6*, 705-713.

Kamenova, I., Mukherjee, P., Conic, S., Mueller, F., El-Saafin, F., Bardot, P., Garnier, J.-M., Dembele, D., Capponi, S., Timmers, M.H.*, et al.* (2018). Co-translation drives the assembly of mammalian nuclear multisubunit complexes. bioRxiv.

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.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900-903.

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L.*, et al.* (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. Cell *170*, 367-381 e320.

Kebede, A.F., Schneider, R., and Daujat, S. (2015). Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. FEBS J *282*, 1658-1674.

Kedinger, C., Gniazdowski, M., Mandel, J.L., Jr., Gissinger, F., and Chambon, P. (1970). Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. Biochem Biophys Res Commun *38*, 165-171.

Kedmi, A., Zehavi, Y., Glick, Y., Orenstein, Y., Ideses, D., Wachtel, C., Doniger, T., Waldman Ben-Asher, H., Muster, N., Thompson, J.*, et al.* (2014). Drosophila TRF2 is a preferential core promoter regulator. Genes Dev *28*, 2163-2174.

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

Keogh, M.C., Mennella, T.A., Sawa, C., Berthelet, S., Krogan, N.J., Wolek, A., Podolny, V., Carpenter, L.R., Greenblatt, J.F., Baetz, K.*, et al.* (2006). The Saccharomyces cerevisiae histone H2A variant Htz1 is acetylated by NuA4. Genes Dev *20*, 660-665.

Khuong, M.T., Fei, J., Ishii, H., and Kadonaga, J.T. (2015). Prenucleosomes and Active Chromatin. Cold Spring Harb Symp Quant Biol *80*, 65-72.

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

Kim, J.M., Liu, H., Tazaki, M., Nagata, M., and Aoki, F. (2003). Changes in histone acetylation during mouse oocyte meiosis. J Cell Biol *162*, 37-46.

Kim, J.Y. (2012). Control of ovarian primordial follicle activation. Clin Exp Reprod Med *39*, 10-14.

Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993b). Crystal structure of a yeast TBP/TATA-box complex. Nature *365*, 512-520.

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.

Klejman, M.P., Zhao, X., van Schaik, F.M., Herr, W., and Timmers, H.T. (2005). Mutational analysis of BTAF1-TBP interaction: BTAF1 can rescue DNA-binding defective TBP mutants. Nucleic Acids Res *33*, 5426-5436.

Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T*., et al.* (2013). High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res *23*, 616-627.

Kokubo, T., Yamashita, S., Horikoshi, M., Roeder, R.G., and Nakatani, Y. (1994). Interaction between the N-terminal domain of the 230-kDa subunit and the TATA box-binding subunit of TFIID negatively regulates TATA-box binding. Proc Natl Acad Sci U S A *91*, 3520-3524.

Koleske, A.J., and Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. Nature *368*, 466-469.

Kolthur-Seetharam, U., Martianov, I., and Davidson, I. (2008). Specialization of the general transcriptional machinery in male germ cells. Cell Cycle *7*, 3493-3498.

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

Konev, A.Y., Tribus, M., Park, S.Y., Podhraski, V., Lim, C.Y., Emelyanov, A.V., Vershilova, E., Pirrotta, V., Kadonaga, J.T., Lusser, A*., et al.* (2007). CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo. Science *317*, 1087-1090.

Kong, S., Bohl, D., Li, C., and Tuan, D. (1997). Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. Mol Cell Biol *17*, 3955-3965.

Kopytova, D.V., Krasnov, A.N., Kopantceva, M.R., Nabirochkina, E.N., Nikolenko, J.V., Maksimenko, O., Kurshakova, M.M., Lebedeva, L.A., Yerokhin, M.M., Simonova, O.B*., et al.* (2006). Two isoforms of Drosophila TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes. Mol Cell Biol *26*, 7492-7505.

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

Koster, M.J., Snel, B., and Timmers, H.T. (2015). Genesis of Chromatin and Transcription Dynamics in the Origin of Species. Cell *161*, 724-736.

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.

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

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

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev *22*, 1617-1635.

Kuryan, B.G., Kim, J., Tran, N.N., Lombardo, S.R., Venkatesh, S., Workman, J.L., and Carey, M. (2012). Histone density is maintained during transcription mediated by the chromatin remodeler RSC and histone chaperone NAP1 in vitro. Proc Natl Acad Sci U S A *109*, 1931-1936.

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

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

Kwak, H., and Lis, J.T. (2013b). Control of transcriptional elongation. Annu Rev Genet *47*, 483-508.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049-1060.

Lacoste, N., Woolfe, A., Tachiwana, H., Garea, A.V., Barth, T., Cantaloube, S., Kurumizaka, H., Imhof, A., and Almouzni, G. (2014). Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. Mol Cell *53*, 631-644.

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 Dev *12*, 34-44.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat Rev Mol Cell Biol *18*, 548-562.

Langer, D., Martianov, I., Alpern, D., Rhinn, M., Keime, C., Dolle, P., Mengus, G., and Davidson, I. (2016). Essential role of the TFIID subunit TAF4 in murine embryogenesis and embryonic stem cell differentiation. Nat Commun *7*, 11063.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S.*, et al.* (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943-949.

Latrick, C.M., Marek, M., Ouararhni, K., Papin, C., Stoll, I., Ignatyeva, M., Obri, A., Ennifar, E., Dimitrov, S., Romier, C.*, et al.* (2016). Molecular basis and specificity of H2A.Z-H2B recognition and deposition by the histone chaperone YL1. Nat Struct Mol Biol *23*, 309-316.

Lau, P.N., and Cheung, P. (2011). Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. Proc Natl Acad Sci U S A *108*, 2801-2806.

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell *152*, 1021-1036.

Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M., Strahl, B.D., Patel, D.J., and Jacobsen, S.E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature *498*, 385-389.

Lawinger, P., Rastelli, L., Zhao, Z., and Majumder, S. (1999). Lack of enhancer function in mammals is unique to oocytes and fertilized eggs. J Biol Chem *274*, 8002-8011.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet *32*, 42-56.

Lawson, K.A., and Hage, W.J. (1994). Clonal analysis of the origin of primordial germ cells in the mouse. Ciba Found Symp *182*, 68-84; discussion 84-91.

Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Mol Cell Biol *25*, 9674-9686.

Lee, D.H., and Schleif, R.F. (1989). In vivo DNA loops in araCBAD: size limits and helical repeat. Proc Natl Acad Sci U S A *86*, 476-480.

Lee, H.J., Hore, T.A., and Reik, W. (2014). Reprogramming the methylome: erasing memory and creating diversity. Cell Stem Cell *14*, 710-719.

Lee, K., and Song, K. (2008). Basal c-Jun N-terminal kinases promote mitotic progression through histone H3 phosphorylation. Cell Cycle *7*, 216-221.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet *13*, 233-245.

Lesch, B.J., and Page, D.C. (2012). Genetics of germ cell development. Nat Rev Genet *13*, 781-794.

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

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13-25.

Lewandoski, M., Wassarman, K.M., and Martin, G.R. (1997). Zp3-cre, a transgenic mouse line for the activation or inactivation of loxP-flanked target genes specifically in the female germ line. Curr Biol *7*, 148-151.

Lewis, B.A., Kim, T.K., and Orkin, S.H. (2000). A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. Proc Natl Acad Sci U S A *97*, 7172-7177.

Li, E., and Zhang, Y. (2014). DNA methylation in mammals. Cold Spring Harb Perspect Biol *6*, a019133.

Li, J., and Gilmour, D.S. (2013a). Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. EMBO J *32*, 1829-1841.

Li, R., and Albertini, D.F. (2013b). The road to maturation: somatic cell interaction and self-organization of the mammalian oocyte. Nat Rev Mol Cell Biol *14*, 141-152.

Li, S., Lu, M.M., Zhou, D., Hammes, S.R., and Morrisey, E.E. (2007). GLP-1: a novel zinc finger protein required in somatic cells of the gonad for germ cell development. Dev Biol *301*, 106-116.

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

Liang, G., Lin, J.C., Wei, V., Yoo, C., Cheng, J.C., Nguyen, C.T., Weisenberger, D.J., Egger, G., Takai, D., Gonzales, F.A.*, et al.* (2004). Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. Proc Natl Acad Sci U S A *101*, 7357-7362.

Liang, X., Shan, S., Pan, L., Zhao, J., Ranjan, A., Wang, F., Zhang, Z., Huang, Y., Feng, H., Wei, D., *et al.* (2016). Structural basis of H2A.Z recognition by SRCAP chromatin-remodeling subunit YL1. Nat Struct Mol Biol *23*, 317-323.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new

core promoter element for transcription by RNA polymerase II. Genes Dev *18*, 1606-1617.

Lin, Y.C., Choi, W.S., and Gralla, J.D. (2005). TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. Nat Struct Mol Biol *12*, 603-607.

Lindell, T.J., Weinberg, F., Morris, P.W., Roeder, R.G., and Rutter, W.J. (1970). Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. Science *170*, 447-449.

Liu, W.L., Coleman, R.A., Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., and Tjian, R. (2009). Structures of three distinct activator-TFIID complexes. Genes Dev *23*, 1510-1521.

Liu, X., Bushnell, D.A., Wang, D., Calero, G., and Kornberg, R.D. (2010). Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. Science *327*, 206-209.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H.*, et al.* (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature *537*, 558-562.

Liu, Z., Scannell, D.R., Eisen, M.B., and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. Cell *146*, 720-731.

Liu, Z.W., Shao, C.R., Zhang, C.J., Zhou, J.X., Zhang, S.W., Li, L., Chen, S., Huang, H.W., Cai, T., and He, X.J. (2014). The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet *10*, e1003948.

Long, H.K., Blackledge, N.P., and Klose, R.J. (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans *41*, 727-740.

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.

Lovasco, L.A., Seymour, K.A., Zafra, K., O'Brien, C.W., Schorl, C., and Freiman, R.N. (2010). Accelerated ovarian aging in the absence of the transcription regulator TAF4B in mice. Biol Reprod *82*, 23-34.

Lucifero, D., Mann, M.R., Bartolomei, M.S., and Trasler, J.M. (2004). Gene-specific timing and epigenetic memory in oocyte imprinting. Hum Mol Genet *13*, 839-849.

Lue, N.F., and Kornberg, R.D. (1987). Accurate initiation at RNA polymerase II promoters in extracts from Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *84*, 8839-8843.

Luo, C., Hajkova, P., and Ecker, J.R. (2018). Dynamic DNA methylation: In the right place at the right time. Science *361*, 1336-1340.

Luo, G.Z., and He, C. (2017). DNA N(6)-methyladenine in metazoans: functional epigenetic mark or bystander? Nat Struct Mol Biol *24*, 503-506.

Lusser, A., Urwin, D.L., and Kadonaga, J.T. (2005). Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. Nat Struct Mol Biol *12*, 160-166.

Ma, P., Pan, H., Montgomery, R.L., Olson, E.N., and Schultz, R.M. (2012). Compensatory functions of histone deacetylase 1 (HDAC1) and HDAC2 regulate transcription and apoptosis during mouse oocyte development. Proc Natl Acad Sci U S A *109*, E481-489.

Ma, P., and Schultz, R.M. (2016). HDAC1 and HDAC2 in mouse oocytes and preimplantation embryos: Specificity versus compensation. Cell Death Differ *23*, 1119-1127.

Magnusdottir, E., Dietmann, S., Murakami, K., Gunesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Azim Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat Cell Biol *15*, 905-915.

Maile, T., Kwoczynski, S., Katzenberger, R.J., Wassarman, D.A., and Sauer, F. (2004). TAF1 activates transcription by phosphorylation of serine 33 in histone H2B. Science *304*, 1010-1014.

Malecova, B., Dall'Agnese, A., Madaro, L., Gatto, S., Coutinho Toto, P., Albini, S., Ryan, T., Tora, L., and Puri, P.L. (2016). TBP/TFIID-dependent activation of MyoD target genes in skeletal muscle cells. Elife *5*.

Mangia, F., and Epstein, C.J. (1975). Biochemical studies of growing mouse oocytes: preparation of oocytes and analysis of glucose-6-phosphate dehydrogenase and lactate dehydrogenase activities. Dev Biol *45*, 211-220.

Mao, Z., Pan, L., Wang, W., Sun, J., Shan, S., Dong, Q., Liang, X., Dai, L., Ding, X., Chen, S.*, et al.* (2014). Anp32e, a higher eukaryotic histone chaperone directs preferential recognition for H2A.Z. Cell Res *24*, 389-399.

Marmorstein, R., and Zhou, M.M. (2014). Writers and readers of histone acetylation: structure, mechanism, and inhibition. Cold Spring Harb Perspect Biol *6*, a018762.

Marshall, O.J., and Brand, A.H. (2015). damidseq_pipeline: an automated pipeline for processing DamID sequencing datasets. Bioinformatics *31*, 3371-3373.

Marshall, O.J., Southall, T.D., Cheetham, S.W., and Brand, A.H. (2016). Cell-type-specific profiling of protein-DNA interactions without cell isolation using targeted DamID with next-generation sequencing. Nat Protoc *11*, 1586-1598.

Martianov, I., Brancorsini, S., Gansmuller, A., Parvinen, M., Davidson, I., and Sassone-Corsi, P. (2002a). Distinct functions of TBP and TLF/TRF2 during spermatogenesis: requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. Development *129*, 945-955.

Martianov, I., Fimia, G.M., Dierich, A., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2001). Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. Mol Cell *7*, 509-515.

Martianov, I., Velt, A., Davidson, G., Choukrallah, M.A., and Davidson, I. (2016). TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. Sci Rep *6*, 32069.

Martianov, I., Viville, S., and Davidson, I. (2002b). RNA polymerase II transcription in murine cells lacking the TATA binding protein. Science *298*, 1036-1039.

Masternak, K., Peyraud, N., Krawczyk, M., Barras, E., and Reith, W. (2003). Chromatin remodeling and extragenic transcription at the MHC class II locus control region. Nat Immunol *4*, 132-137.

Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem *255*, 11992-11996.

Mattson, B.A., and Albertini, D.F. (1990). Oogenesis: chromatin and microtubule dynamics during meiotic prophase. Mol Reprod Dev *25*, 374-383.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C.*, et al.* (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

Maxon, M.E., Goodrich, J.A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev *8*, 515-524.

Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., and Drouin, J. (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nat Genet *50*, 259-269.

McLaren, A. (2003). Primordial germ cells in the mouse. Dev Biol *262*, 1-15.

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.

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

Mermelstein, F., Yeung, K., Cao, J., Inostroza, J.A., Erdjument-Bromage, H., Eagelson, K., Landsman, D., Levitt, P., Tempst, P., and Reinberg, D. (1996). Requirement of a corepressor for Dr1-mediated repression of transcription. Genes Dev *10*, 1033-1048.

Metzger, D., Scheer, E., Soldatov, A., and Tora, L. (1999). Mammalian TAF(II)30 is required for cell cycle progression and specific cellular differentiation programmes. EMBO J *18*, 4823-4834.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

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

Mitsiou, D.J., and Stunnenberg, H.G. (2000). TAC, a TBP-sans-TAFs complex containing the unprocessed TFIIAalphabeta precursor and the TFIIAgamma subunit. Mol Cell *6*, 527-537.

Mitsiou, D.J., and Stunnenberg, H.G. (2003). p300 is involved in formation of the TBP-TFIIA-containing basal transcription complex, TAC. EMBO J *22*, 4501-4511.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science *303*, 343-348.

Mizzen, C.A., Yang, X.J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T*., et al.* (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell *87*, 1261-1270.

Mohan, W.S., Jr., Scheer, E., Wendling, O., Metzger, D., and Tora, L. (2003). TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice. Mol Cell Biol *23*, 4307-4318.

Molyneaux, K.A., Stallock, J., Schaible, K., and Wylie, C. (2001). Time-lapse analysis of living mouse germ cell migration. Dev Biol *240*, 488-498.

Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A., and Kanduri, C. (2010). Characterization of the RNA content of chromatin. Genome Res *20*, 899-907.

Moore, P.A., Ozer, J., Salunek, M., Jan, G., Zerby, D., Campbell, S., and Lieberman, P.M. (1999). A human TATA binding protein-related protein with altered DNA binding specificity inhibits transcription from multiple promoters and activators. Mol Cell Biol *19*, 7610-7620.

Moreland, R.J., Tirode, F., Yan, Q., Conaway, J.W., Egly, J.M., and Conaway, R.C. (1999). A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. J Biol Chem *274*, 22127-22130.

Muller, F., Demeny, M.A., and Tora, L. (2007). New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. J Biol Chem *282*, 14685-14689.

Muller, F., Lakatos, L., Dantonel, J., Strahle, U., and Tora, L. (2001). TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. Curr Biol *11*, 282-287.

Muller, F., and Tora, L. (2004). The multicoloured world of promoter recognition complexes. EMBO J *23*, 2-8.

Muller, F., and Tora, L. (2009). TBP2 is a general transcription factor specialized for female germ cells. J Biol *8*, 97.

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

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

Munakata, T., Adachi, N., Yokoyama, N., Kuzuhara, T., and Horikoshi, M. (2000). A human homologue of yeast anti-silencing factor has histone chaperone activity. Genes Cells *5*, 221-233.

Murawska, M., and Brehm, A. (2011). CHD chromatin remodelers and the transcription cycle. Transcription *2*, 244-253.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat Genet *39*, 1507-1511.

Nair, D., Kim, Y., and Myers, L.C. (2005). Mediator and TFIIH govern carboxyl-terminal domain-dependent transcription in yeast extracts. J Biol Chem *280*, 33739-33748.

Nashun, B., Hill, P.W., Smallwood, S.A., Dharmalingam, G., Amouroux, R., Clark, S.J., Sharma, V., Ndjetehe, E., Pelczar, P., Festenstein, R.J.*, et al.* (2015). Continuous Histone Replacement by Hira Is Essential for Normal Transcriptional Regulation and De Novo DNA Methylation during Mouse Oogenesis. Mol Cell *60*, 611-625.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72-77.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.

Ng, J.H., Kumar, V., Muratani, M., Kraus, P., Yeo, J.C., Yaw, L.P., Xue, K., Lufkin, T., Prabhakar, S., and Ng, H.H. (2013). In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. Dev Cell *24*, 324-333.

Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G., and Burley, S.K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. Proc Natl Acad Sci U S A *93*, 4862-4867.

Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R.*, et al.* (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. Mol Cell *9*, 1201-1213.

Noonan, J.P., and McCallion, A.S. (2010). Genomics of long-range regulatory elements. Annual

review of genomics and human genetics *11*, 1-23.

North, J.A., Javaid, S., Ferdinand, M.B., Chatterjee, N., Picking, J.W., Shoffner, M., Nakkula, R.J., Bartholomew, B., Ottesen, J.J., Fishel, R.*, et al.* (2011). Phosphorylation of histone H3(T118) alters nucleosome dynamics and remodeling. Nucleic Acids Res *39*, 6465-6474.

Obata, Y., and Kono, T. (2002). Maternal primary imprinting is established at a specific time for each gene throughout oocyte growth. J Biol Chem *277*, 5285-5289.

Obri, A., Ouararhni, K., Papin, C., Diebold, M.L., Padmanabhan, K., Marek, M., Stoll, I., Roy, L., Reilly, P.T., Mak, T.W.*, et al.* (2014). ANP32E is a histone chaperone that removes H2A.Z from chromatin. Nature *505*, 648-653.

Ohbayashi, T., Kishimoto, T., Makino, Y., Shimada, M., Nakadai, T., Aoki, T., Kawata, T., Niwa, S., and Tamura, T. (1999a). Isolation of cDNA, chromosome mapping, and expression of the human TBP-like protein. Biochem Biophys Res Commun *255*, 137-142.

Ohbayashi, T., Makino, Y., and Tamura, T.A. (1999b). Identification of a mouse TBP-like protein (TLP) distantly related to the drosophila TBP-related factor. Nucleic Acids Res *27*, 750-755.

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. Cell *137*, 571-584.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A.*, et al.* (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207-213.

Ohkuma, Y., and Roeder, R.G. (1994). Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. Nature *368*, 160-163.

Ohkuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshi, M., and Roeder, R.G. (1991). Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE. Nature *354*, 398-401.

Ohler, U., Liao, G.C., Niemann, H., and Rubin, G.M. (2002). Computational analysis of core promoters in the Drosophila genome. Genome Biol *3*, RESEARCH0087.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet *19*, 219-220.

Oldfield, A.J., Yang, P., Conway, A.E., Cinghu, S., Freudenberg, J.M., Yellaboina, S., and Jothi, R. (2014). Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. Mol Cell *55*, 708-722.

Olins, A.L., Senior, M.B., and Olins, D.E. (1976). Ultrastructural features of chromatin nu bodies. J Cell Biol *68*, 787-793.

Olins, D.E., and Olins, A.L. (2003). Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol *4*, 809-814.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet *12*, 283-293.

Orisaka, M., Tajima, K., Tsang, B.K., and Kotsuji, F. (2009). Oocyte-granulosa-theca cell interactions during preantral follicular development. J Ovarian Res *2*, 9.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev *10*, 2657-2683.

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.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science *357*.

Oyama, T., Sasagawa, S., Takeda, S., Hess, R.A., Lieberman, P.M., Cheng, E.H., and Hsieh, J.J. (2013). Cleavage of TFIIA by Taspase1 activates TRF2-specified mammalian male germ cell programs. Dev Cell *27*, 188-200.

Ozer, J., Mitsouras, K., Zerby, D., Carey, M., and Lieberman, P.M. (1998). Transcription factor IIA derepresses TATA-binding protein (TBP)-associated factor inhibition of TBP-DNA binding. J Biol Chem *273*, 14293-14300.

Ozer, J., Moore, P.A., and Lieberman, P.M. (2000). A testis-specific transcription factor IIA (TFIIAtau) stimulates TATA-binding protein-DNA binding and transcription activation. J Biol Chem *275*, 122-128.

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. J Biol Chem *269*, 30101-30104.

Pangas, S.A., Choi, Y., Ballow, D.J., Zhao, Y., Westphal, H., Matzuk, M.M., and Rajkovic, A. (2006). Oogenesis requires germ cell-specific transcriptional regulators Sohlh1 and Lhx8. Proc Natl Acad Sci U S A *103*, 8090-8095.

Park, S.G., Hannenhalli, S., and Choi, S.S. (2014). Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. BMC Genomics *15*, 526.

Parker, C.S., and Topol, J. (1984). A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell *36*, 357-369.

Parry, T.J., Theisen, J.W., Hsu, J.Y., Wang, Y.L., Corcoran, D.L., Eustice, M., Ohler, U., and Kadonaga, J.T. (2010). The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. Genes Dev *24*, 2013-2018.

Patikoglou, G.A., Kim, J.L., Sun, L., Yang, S.H., Kodadek, T., and Burley, S.K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev *13*, 3217-3230.

Paule, M.R., and White, R.J. (2000). Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res *28*, 1283-1298.

Paulini, F., Silva, R.C., Rolo, J.L., and Lucci, C.M. (2014). Ultrastructural changes in oocytes during folliculogenesis in domestic mammals. J Ovarian Res *7*, 102.

Pchelintsev, N.A., McBryan, T., Rai, T.S., van Tuyn, J., Ray-Gallet, D., Almouzni, G., and Adams, P.D. (2013). Placing the HIRA histone chaperone complex in the chromatin landscape. Cell Rep *3*, 1012-1019.

Pedersen, T., and Peters, H. (1968). Proposal for a classification of oocytes and follicles in the mouse ovary. J Reprod Fertil *17*, 555-557.

Pehrson, J.R., Changolkar, L.N., Costanzi, C., and Leu, N.A. (2014). Mice without macroH2A histone variants. Mol Cell Biol *34*, 4523-4533.

Pepling, M.E. (2006). From primordial germ cell to primordial follicle: mammalian female germ cell development. Genesis *44*, 622-632.

Pepling, M.E., and Spradling, A.C. (2001). Mouse ovarian germ cell cysts undergo programmed breakdown to form primordial follicles. Dev Biol *234*, 339-351.

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.

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. Mol Cell Biol *21*, 7523-7534.

Perry, R.P. (2005). The architecture of mammalian ribosomal protein promoters. BMC Evol Biol *5*, 15.

Persengiev, S.P., Zhu, X., Dixit, B.L., Maston, G.A., Kittler, E.L., and Green, M.R. (2003). TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. Proc Natl Acad Sci U S A *100*, 14887-14891.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol Cell *23*, 297-305.

Peterson, M.G., Inostroza, J., Maxon, M.E., Flores, O., Admon, A., Reinberg, D., and Tjian, R. (1991). Structure and functional properties of human general transcription factor IIE. Nature *354*, 369-373.

Petryk, N., Dalby, M., Wenger, A., Stromme, C.B., Strandsby, A., Andersson, R., and Groth, A. (2018). MCM2 promotes symmetric inheritance of modified histones during DNA replication. Science *361*, 1389-1392.

Pfender, S., Kuznetsov, V., Pasternak, M., Tischer, T., Santhanam, B., and Schuh, M. (2015). Live imaging RNAi screen reveals genes essential for meiosis in mammalian oocytes. Nature *524*, 239-242.

Pikaard, C.S., Haag, J.R., Ream, T., and Wierzbicki, A.T. (2008). Roles of RNA polymerase IV in gene silencing. Trends Plant Sci *13*, 390-397.

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.

Pointud, J.C., Mengus, G., Brancorsini, S., Monaco, L., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2003). The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID subunit TAF7, is developmentally regulated during male germ-cell differentiation. J Cell Sci *116*, 1847-1858.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. Nat Rev Mol Cell Biol *16*, 245-257.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat Rev Mol Cell Biol *16*, 190-202.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.

Price, D.H., Sluder, A.E., and Greenleaf, A.L. (1987). Fractionation of transcription factors for RNA polymerase II from Drosophila Kc cell nuclear extracts. J Biol Chem *262*, 3244-3255.

Qiu, Y., and Gilmour, D.S. (2017). Identification of Regions in the Spt5 Subunit of DRB Sensitivity-inducing Factor (DSIF) That Are Involved in Promoter-proximal Pausing. J Biol Chem *292*, 5555-5570.

Rabenstein, M.D., Zhou, S., Lis, J.T., and Tjian, R. (1999). TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. Proc Natl Acad Sci U S A *96*, 4791-4796.

Racki, W.J., and Richter, J.D. (2006). CPEB controls oocyte growth and follicle development in the mouse. Development *133*, 4527-4537.

Rajkovic, A., Pangas, S.A., Ballow, D., Suzumori, N., and Matzuk, M.M. (2004). NOBOX deficiency disrupts early folliculogenesis and oocyte-specific gene expression. Science *305*, 1157-1159.

Ray-Gallet, D., Quivy, J.P., Scamps, C., Martini, E.M., Lipinski, M., and Almouzni, G. (2002). HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. Mol Cell *9*, 1091-1100.

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. J Biol Chem *262*, 3322-3330.

Rhee, H.S., Bataille, A.R., Zhang, L., and Pugh, B.F. (2014). Subnucleosomal structures and nucleosome asymmetry across a genome. Cell *159*, 1377-1388.

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

Richmond, T.J., and Davey, C.A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145-150.

Ricketts, M.D., Frederick, B., Hoff, H., Tang, Y., Schultz, D.C., Singh Rai, T., Grazia Vizioli, M., Adams, P.D., and Marmorstein, R. (2015). Ubinuclein-1 confers histone H3.3-specific-binding by the HIRA histone chaperone complex. Nat Commun *6*, 7711.

Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. Cell *155*, 39-55.

Robert, F., Douziech, M., Forget, D., Egly, J.M., Greenblatt, J., Burton, Z.F., and Coulombe, B. (1998). Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. Mol Cell *2*, 341-351.

Roberts, S.G., Choy, B., Walker, S.S., Lin, Y.S., and Green, M.R. (1995). A role for activator-mediated TFIIB recruitment in diverse aspects of transcriptional regulation. Curr Biol *5*, 508-516.

Robzyk, K., Recht, J., and Osley, M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science *287*, 501-504.

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. Mol Cell *63*, 433-444.

Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234-237.

Roeder, R.G., and Rutter, W.J. (1970). Specific nucleolar and nucleoplasmic RNA polymerases. Proc Natl Acad Sci U S A *65*, 675-682.

Roepcke, S., Zhi, D., Vingron, M., and Arndt, P.F. (2006). Identification of highly specific localized sequence motifs in human ribosomal protein gene promoters. Gene *365*, 48-56.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem *273*, 5858-5868.

Rossetto, D., Truman, A.W., Kron, S.J., and Cote, J. (2010). Epigenetic modifications in double-strand break DNA damage signaling and repair. Clin Cancer Res *16*, 4543-4552.

Roy, A.L., and Singer, D.S. (2015). Core promoters in transcription: old problem, new insights. Trends Biochem Sci *40*, 165-171.

Ruppert, S., and Tjian, R. (1995). Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. Genes Dev *9*, 2747-2755.

Sachs, M., Onodera, C., Blaschke, K., Ebata, K.T., Song, J.S., and Ramalho-Santos, M. (2013). Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. Cell Rep *3*, 1777-1784.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 129-143.

Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature *493*, 437-440.

Saitou, M., Barton, S.C., and Surani, M.A. (2002). A molecular programme for the specification of germ cell fate in mice. Nature *418*, 293-300.

Saitou, M., Kagiwada, S., and Kurimoto, K. (2012a). Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Development *139*, 15-31.

Saitou, M., and Miyauchi, H. (2016). Gametogenesis from Pluripotent Stem Cells. Cell Stem Cell *18*, 721-735.

Saitou, M., Payer, B., Lange, U.C., Erhardt, S., Barton, S.C., and Surani, M.A. (2003). Specification of germ cell fate in mice. Philos Trans R Soc Lond B Biol Sci *358*, 1363-1370.

Saitou, M., and Yamaji, M. (2012b). Primordial germ cells in mice. Cold Spring Harb Perspect Biol *4*.

Sammons, M.A., Zhu, J., Drake, A.M., and Berger, S.L. (2015). TP53 engagement with the genome occurs in distinct local chromatin environments via pioneer factor activity. Genome Res *25*, 179-188.

Sanchez, F., and Smitz, J. (2012). Molecular control of oogenesis. Biochim Biophys Acta *1822*, 1896-1912.

Sandaltzopoulos, R., and Becker, P.B. (1998). Heat shock factor increases the reinitiation rate from potentiated chromatin templates. Mol Cell Biol *18*, 361-367.

Sanders, S.L., Garbett, K.A., and Weil, P.A. (2002). Molecular characterization of Saccharomyces cerevisiae TFIID. Mol Cell Biol *22*, 6000-6013.

Sansoni, V., Casas-Delucchi, C.S., Rajan, M., Schmidt, A., Bonisch, C., Thomae, A.W., Staege, M.S., Hake, S.B., Cardoso, M.C., and Imhof, A. (2014). The histone variant H2A.Bbd is enriched at sites of DNA synthesis. Nucleic Acids Res *42*, 6405-6420.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Santos, F., Peat, J., Burgess, H., Rada, C., Reik, W., and Dean, W. (2013). Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. Epigenetics Chromatin *6*, 39.

Sapountzi, V., and Cote, J. (2011). MYST-family histone acetyltransferases: beyond chromatin. Cell Mol Life Sci *68*, 1147-1156.

Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet *9*, 129-140.

Sato, M., Kimura, T., Kurokawa, K., Fujita, Y., Abe, K., Masuhara, M., Yasunaga, T., Ryo, A., Yamamoto, M., and Nakano, T. (2002). Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. Mech Dev *113*, 91-94.

Sawadogo, M., and Roeder, R.G. (1985). Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. Proc Natl Acad Sci U S A *82*, 4394-4398.

Sayre, M.H., Tschochner, H., and Kornberg, R.D. (1992). Reconstitution of transcription with five purified initiation factors and RNA polymerase II from Saccharomyces cerevisiae. J Biol Chem *267*, 23376-23382.

Schagdarsurengin, U., and Steger, K. (2016). Epigenetics in male reproduction: effect of paternal diet on sperm quality and offspring health. Nature reviews Urology *13*, 584-595.

Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., and Cramer, P. (2017). Structures of transcription pre-initiation complex with TFIIH and Mediator. Nature *551*, 204-209.

Schluesche, P., Stelzer, G., Piaia, E., Lamb, D.C., and Meisterernst, M. (2007). NC2 mobilizes TBP on core promoter TATA boxes. Nat Struct Mol Biol *14*, 1196-1201.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods.

Schmidt, D., Ovitt, C.E., Anlag, K., Fehsenfeld, S., Gredsted, L., Treier, A.C., and Treier, M. (2004). The murine winged-helix transcription factor Foxl2 is required for granulosa cell differentiation and ovary maintenance. Development *131*, 933-942.

Schmitt, A.D., Hu, M., and Ren, B. (2016). Genome-wide mapping and analysis of chromosome architecture. Nat Rev Mol Cell Biol *17*, 743-755.

Schneider, M., Hellerschmied, D., Schubert, T., Amlacher, S., Vinayachandran, V., Reja, R., Pugh, B.F., Clausen, T., and Kohler, A. (2015). The Nuclear Pore-Associated TREX-2 Complex Employs Mediator to Regulate Gene Expression. Cell *162*, 1016-1028.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev *18*, 1251-1262.

Schubeler, D. (2015). Function and information content of DNA methylation. Nature *517*, 321-326.

Schuller, R., Forne, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol Cell *61*, 305-314.

Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H.*, et al.* (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature *523*, 212-216.

Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Mol Cell *58*, 1101-1112.

Seifart, K.H., and Sekeris, C.E. (1969). Alpha-amanitin, a specific inhibitor of transcription by mammalian RNA-polymerase. Z Naturforsch B *24*, 1538-1544.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell *48*, 849-862.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev Biol *278*, 440-458.

Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development *134*, 2627-2638.

Serandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R.*, et al.* (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome Res *21*, 555-565.

Serizawa, H., Makela, T.P., Conaway, J.W., Conaway, R.C., Weinberg, R.A., and Young, R.A. (1995). Association of Cdk-activating kinase subunits with transcription factor TFIIH. Nature *374*, 280-282.

Seto, E., and Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb Perspect Biol *6*, a018713.

Sexton, T., and Cavalli, G. (2015). The role of chromosome domains in shaping the functional genome. Cell *160*, 1049-1059.

Shalem, O., Groisman, B., Choder, M., Dahan, O., and Pilpel, Y. (2011). Transcriptome kinetics is governed by a genome-wide coupling of mRNA production and degradation: a role for RNA Pol II. PLoS Genet *7*, e1002273.

Shao, H., Revach, M., Moshonov, S., Tzuman, Y., Gazit, K., Albeck, S., Unger, T., and Dikstein, R. (2005). Core promoter binding by histone-like TAF complexes. Mol Cell Biol *25*, 206-219.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Shinagawa, T., Takagi, T., Tsukamoto, D., Tomaru, C., Huynh, L.M., Sivaraman, P., Kumarevel, T., Inoue, K., Nakato, R., Katou, Y.*, et al.* (2014). Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells. Cell Stem Cell *14*, 217-227.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-286.

Shoemaker, J., Saraiva, M., and O'Garra, A. (2006). GATA-3 directly remodels the IL-10 locus independently of IL-4 in CD4+ T cells. J Immunol *176*, 3470-3479.

Siklenka, K., Erkek, S., Godmann, M., Lambrot, R., McGraw, S., Lafleur, C., Cohen, T., Xia, J., Suderman, M., Hallett, M.*, et al.* (2015). Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. Science *350*, aab2006.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu Rev Biochem

72, 449-479.

Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet *43*, 811-814.

Smith, E.Y., Futtner, C.R., Chamberlain, S.J., Johnstone, K.A., and Resnick, J.L. (2011). Transcription is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman syndrome locus. PLoS Genet *7*, e1002422.

Smith, S., and Stillman, B. (1989). Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. Cell *58*, 15-25.

Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science *331*, 1036-1040.

Soni, S., Pchelintsev, N., Adams, P.D., and Bieker, J.J. (2014). Transcription factor EKLF (KLF1) recruitment of the histone chaperone HIRA is essential for beta-globin gene expression. Proc Natl Acad Sci U S A *111*, 13337-13342.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994-1004.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555-568.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat Rev Mol Cell Biol *19*, 262-274.

Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C., and Werner, M. (2011). Direct interaction of RNA polymerase II and mediator required for transcription in vivo. Science *331*, 1451-1454.

Soyal, S.M., Amleh, A., and Dean, J. (2000). FIGalpha, a germ cell-specific transcription factor required for ovarian follicle formation. Development *127*, 4645-4654.

Spedale, G., Timmers, H.T., and Pijnappel, W.W. (2012). ATAC-king the complexity of SAGA during evolution. Genes Dev *26*, 527-541.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet *13*, 613-626.

Spradling, A., Fuller, M.T., Braun, R.E., and Yoshida, S. (2011). Germline stem cells. Cold Spring Harb Perspect Biol *3*, a002642.

Stewart, K.R., Veselovska, L., and Kelsey, G. (2016). Establishment and functions of DNA

methylation in the germline. Epigenomics *8*, 1399-1413.

Strahl, B.D., Ohba, R., Cook, R.G., and Allis, C.D. (1999). Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. Proc Natl Acad Sci U S A *96*, 14967-14972.

Strauss, T.J., Castrillon, D.H., and Hammes, S.R. (2011). GATA-like protein-1 (GLP-1) is required for normal germ cell development during embryonic oogenesis. Reproduction *141*, 173-181.

Suka, N., Suka, Y., Carmen, A.A., Wu, J., and Grunstein, M. (2001). Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. Mol Cell *8*, 473-479.

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.

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 Res *22*, 1350-1359.

Sykes, S.M., Mellert, H.S., Holbert, M.A., Li, K., Marmorstein, R., Lane, W.S., and McMahon, S.B. (2006). Acetylation of the p53 DNA-binding domain regulates apoptosis induction. Mol Cell *24*, 841-851.

Szerlong, H.J., and Hansen, J.C. (2011). Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. Biochem Cell Biol *89*, 24-34.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Takada, S., Lis, J.T., Zhou, S., and Tjian, R. (2000). A TRF1:BRF complex directs Drosophila RNA polymerase III transcription. Cell *101*, 459-469.

Takeda, S., Chen, D.Y., Westergard, T.D., Fisher, J.K., Rubens, J.A., Sasagawa, S., Kan, J.T., Korsmeyer, S.J., Cheng, E.H., and Hsieh, J.J. (2006). Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. Genes Dev *20*, 2397-2409.

Talbert, P.B., and Henikoff, S. (2017). Histone variants on the move: substrates for chromatin dynamics. Nat Rev Mol Cell Biol *18*, 115-126.

Tam, P.P., and Snow, M.H. (1981). Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos. J Embryol Exp Morphol *64*, 133-147.

Tamaru, H. (2010). Confining euchromatin/heterochromatin territory: jumonji crosses the line.

Genes Dev *24*, 1465-1478.

Tan, S., Aso, T., Conaway, R.C., and Conaway, J.W. (1994). Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. J Biol Chem *269*, 25684-25691.

Tan, S.K., Lin, Z.H., Chang, C.W., Varang, V., Chng, K.R., Pan, Y.F., Yong, E.L., Sung, W.K., and Cheung, E. (2011). AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. EMBO J *30*, 2569-2581.

Tang, M.C., Jacobs, S.A., Mattiske, D.M., Soh, Y.M., Graham, A.N., Tran, A., Lim, S.L., Hudson, D.F., Kalitsis, P., O'Bryan, M.K*., et al.* (2015a). Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice. PLoS Genet *11*, e1004964.

Tang, M.C.W., Jacobs, S.A., Mattiske, D.M., Soh, Y.M., Graham, A.N., An, T., Lim, S.L., Hudson, D.F., Kalitsis, P., O'Bryan, M.K*., et al.* (2015b). Contribution of the Two Genes Encoding Histone Variant H3.3 to Viability and Fertility in Mice. PLoS genetics *11*.

Tang, W.W., Kobayashi, T., Irie, N., Dietmann, S., and Surani, M.A. (2016). Specification and epigenetic programming of the human germ line. Nat Rev Genet *17*, 585-600.

Tang, Y., Luo, J., Zhang, W., and Gu, W. (2006). Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. Mol Cell *24*, 827-839.

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.

Teichmann, M., Wang, Z., Martinez, E., Tjernberg, A., Zhang, D., Vollmer, F., Chait, B.T., and Roeder, R.G. (1999). Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. Proc Natl Acad Sci U S A *96*, 13720-13725.

Telfer, E.E., Gosden, R.G., Byskov, A.G., Spears, N., Albertini, D., Andersen, C.Y., Anderson, R., Braw-Tal, R., Clarke, H., Gougeon, A*., et al.* (2005). On regenerating the ovary and generating controversy. Cell *122*, 821-822.

Tessarz, P., Santos-Rosa, H., Robson, S.C., Sylvestersen, K.B., Nelson, C.J., Nielsen, M.L., and Kouzarides, T. (2014). Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. Nature *505*, 564-568.

Theisen, J.W., 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.

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

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D*., et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082-1086.

Timmers, H.T., Meyers, R.E., and Sharp, P.A. (1992). Composition of transcription factor B-TFIID. Proc Natl Acad Sci U S A *89*, 8140-8144.

Timmers, H.T.M., and Tora, L. (2018). Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation. Mol Cell *72*, 10-17.

Tirode, F., Busso, D., Coin, F., and Egly, J.M. (1999). Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. Mol Cell *3*, 87-95.

Tolstorukov, M.Y., Goldman, J.A., Gilbert, C., Ogryzko, V., Kingston, R.E., and Park, P.J. (2012). Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. Mol Cell *47*, 596-607.

Tonna, S., El-Osta, A., Cooper, M.E., and Tikellis, C. (2010). Metabolic memory and diabetic nephropathy: potential role for epigenetic mechanisms. Nature reviews Nephrology *6*, 332-341.

Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. Genes Dev *16*, 673-675.

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

Torres-Padilla, M.E., and Tora, L. (2007). TBP homologues in embryo transcription: who does what? Embo Reports *8*, 1016-1018.

Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Buhler, M*., et al.* (2013). Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. Cell *152*, 859-872.

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, W.W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.Y., Shi, X., Schwarzer, D., Plunkett, W*., et al.* (2010). TRIM24 links a non-canonical histone signature to breast cancer. Nature *468*, 927-932.

Tsunaka, Y., Fujiwara, Y., Oyama, T., Hirose, S., and Morikawa, K. (2016). Integrated molecular mechanism directing nucleosome reorganization by human FACT. Genes Dev *30*, 673-686.

Tyler, J.K., Adams, C.R., Chen, S.R., Kobayashi, R., Kamakaka, R.T., and Kadonaga, J.T. (1999). The RCAF complex mediates chromatin assembly during DNA replication and repair. Nature *402*,

555-560.

Uda, M., Ottolenghi, C., Crisponi, L., Garcia, J.E., Deiana, M., Kimber, W., Forabosco, A., Cao, A., Schlessinger, D., and Pilia, G. (2004). Foxl2 disruption causes mouse ovarian failure by pervasive blockage of follicle development. Hum Mol Genet *13*, 1171-1181.

Upadhyaya, A.B., Lee, S.H., and DeJong, J. (1999). Identification of a general transcription factor TFIIAalpha/beta homolog selectively expressed in testis. J Biol Chem *274*, 18040-18048.

Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. Mol Cell Biol *26*, 9185-9195.

van Ingen, H., van Schaik, F.M., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A.C., Kruijzer, J.A., Liskamp, R.M., Timmers, H.T., and Boelens, R. (2008). Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure *16*, 1245-1256.

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.

Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell *45*, 439-446.

Vardabasso, C., Gaspar-Maia, A., Hasson, D., Punzeler, S., Valle-Garcia, D., Straub, T., Keilhauer, E.C., Strub, T., Dong, J., Panda, T*., et al.* (2015). Histone Variant H2A.Z.2 Mediates Proliferation and Drug Sensitivity of Malignant Melanoma. Mol Cell *59*, 75-88.

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

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol *16*, 178-189.

Verma, N., Hung, K.H., Kang, J.J., Barakat, N.H., and Stumph, W.E. (2013). Differential utilization of TATA box-binding protein (TBP) and TBP-related factor 1 (TRF1) at different classes of RNA polymerase III promoters. J Biol Chem *288*, 27564-27570.

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.

Vernimmen, D., and Bickmore, W.A. (2015). The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. Trends Genet *31*, 696-708.

Veselovska, L., Smallwood, S.A., Saadeh, H., Stewart, K.R., Krueger, F., Maupetit-Mehouas, S., Arnaud, P., Tomizawa, S., Andrews, S., and Kelsey, G. (2015). Deep sequencing and de novo

assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome Biol *16*, 209.

Vincent, S.D., Dunn, N.R., Sciammas, R., Shapiro-Shalef, M., Davis, M.M., Calame, K., Bikoff, E.K., and Robertson, E.J. (2005). The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. Development *132*, 1315-1325.

Vo Ngoc, L., Cassidy, C.J., Huang, C.Y., Duttke, S.H., and Kadonaga, J.T. (2017a). The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. Genes Dev *31*, 6-11.

Vo Ngoc, L., Wang, Y.L., Kassavetis, G.A., and Kadonaga, J.T. (2017b). The punctilious RNA polymerase II core promoter. Genes Dev *31*, 1289-1301.

Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nat Protoc *2*, 1467-1478.

Volk, A., and Crispino, J.D. (2015). The role of the chromatin assembly complex (CAF-1) and its p60 subunit (CHAF1b) in homeostasis and disease. Biochim Biophys Acta *1849*, 979-986.

Voronina, E., Lovasco, L.A., Gyuris, A., Baumgartner, R.A., Parlow, A.F., and Freiman, R.N. (2007). Ovarian granulosa cell survival and proliferation requires the gonad-selective TFIID subunit TAF4b. Dev Biol *303*, 715-726.

Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018a). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature *560*, 607-612.

Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018b). Structure of paused transcription complex Pol II-DSIF-NELF. Nature *560*, 601-606.

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

Walker, A.K., Rothman, J.H., Shi, Y., and Blackwell, T.K. (2001). Distinct requirements for C.elegans TAF(II)s in early embryonic transcription. EMBO J *20*, 5269-5279.

Wang, C., Liu, X., Gao, Y., Yang, L., Li, C., Liu, W., Chen, C., Kou, X., Zhao, Y., Chen, J.*, et al.* (2018). Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat Cell Biol *20*, 620-631.

Wang, G., Balamotis, M.A., Stevens, J.L., Yamaguchi, Y., Handa, H., and Berk, A.J. (2005a). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. Mol Cell *17*, 683-694.

Wang, H., Huang, Z.Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B.D., Briggs, S.D., Allis, C.D., Wong, J., Tempst, P.*, et al.* (2001a). Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. Science *293*, 853-857.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.

Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001b). An abundance of X-linked genes expressed in spermatogonia. Nat Genet *27*, 422-426.

Wang, P.J., and Page, D.C. (2002). Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis. Hum Mol Genet *11*, 2341-2346.

Wang, Q., Carroll, J.S., and Brown, M. (2005b). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. Mol Cell *19*, 631-642.

Wang, Y.L., Duttke, S.H., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J., and Kadonaga, J.T. (2014). TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes Dev *28*, 1550-1555.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897-903.

Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., and Hahn, S. (2017). Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. Mol Cell *68*, 118-129 e115.

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

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Mol Cell *53*, 819-830.

Weber, S., Eckert, D., Nettersheim, D., Gillis, A.J., Schafer, S., Kuckenberg, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L.H.*, et al.* (2010). Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. Biol Reprod *82*, 214-223.

Wegel, E., and Shaw, P. (2005). Gene activation and deactivation related changes in the three-dimensional structure of chromatin. Chromosoma *114*, 331-337.

Weil, P.A., Luse, D.S., Segall, J., and Roeder, R.G. (1979). Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. Cell *18*, 469-484.

Weinmann, R., and Roeder, R.G. (1974). Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes. Proc Natl Acad Sci U S A *71*, 1790-1794.

Werten, S., Mitschler, A., Romier, C., Gangloff, Y.G., Thuault, S., Davidson, I., and Moras, D. (2002). Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). J Biol Chem *277*, 45502-45509.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522-525.

Western, P.S., Miles, D.C., van den Bergen, J.A., Burton, M., and Sinclair, A.H. (2008). Dynamic regulation of mitotic arrest in fetal male germ cells. Stem Cells *26*, 339-347.

Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. Nat Genet *49*, 935-940.

White, Y.A., Woods, D.C., Takai, Y., Ishihara, O., Seki, H., and Tilly, J.L. (2012). Oocyte formation by mitotically active germ cells purified from ovaries of reproductive-age women. Nat Med *18*, 413-421.

Woodcock, C.L., and Ghosh, R.P. (2010). Chromatin higher-order structure and dynamics. Cold Spring Harb Perspect Biol *2*, a000596.

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 U S A *103*, 12347-12352.

Wu, D., Sunkel, B., Chen, Z., Liu, X., Ye, Z., Li, Q., Grenade, C., Ke, J., Zhang, C., Chen, H.*, et al.* (2014). Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer. Nucleic Acids Res *42*, 3607-3622.

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W.*, et al.* (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. Nature *534*, 652-657.

Wu, J., Parkhurst, K.M., Powell, R.M., Brenowitz, M., and Parkhurst, L.J. (2001). DNA bends in TATA-binding protein-TATA complexes in solution are DNA sequence-dependent. J Biol Chem *276*, 14614-14622.

Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? Cell *155*, 990-996.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet *18*, 517-534.

Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., and Wu, C.

(2001). Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. Mol Cell *8*, 531-543.

Xiao, L., Kim, M., and DeJong, J. (2006). Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice. Gene Expr Patterns *6*, 409-419.

Xie, X., Kokubo, T., Cohen, S.L., Mirza, U.A., Hoffmann, A., Chait, B.T., Roeder, R.G., Nakatani, Y., and Burley, S.K. (1996). Structural similarity between TAFs and the heterotetrameric core of the histone octamer. Nature *380*, 316-322.

Xu, J., Watts, J.A., Pope, S.D., Gadue, P., Kamps, M., Plath, K., Zaret, K.S., and Smale, S.T. (2009). Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. Genes Dev *23*, 2824-2838.

Xu, Q., and Xie, W. (2018). Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. Trends Cell Biol *28*, 237-253.

Yadav, T., Quivy, J.P., and Almouzni, G. (2018). Chromatin plasticity: A versatile landscape that underlies cell fate and identity. Science *361*, 1332-1336.

Yamaguchi, Y., Shibata, H., and Handa, H. (2013). Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. Biochim Biophys Acta *1829*, 98-104.

Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet *40*, 1016-1022.

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. J Biol Chem *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.

Yang, J., Zhang, X., Feng, J., Leng, H., Li, S., Xiao, J., Liu, S., Xu, Z., Xu, J., Li, D*., et al.* (2016). The Histone Chaperone FACT Contributes to DNA Replication-Coupled Nucleosome Assembly. Cell Rep *14*, 1128-1141.

Yang, Y., Cao, J., Huang, L., Fang, H.Y., and Sheng, H.Z. (2006). Regulated expression of TATA-binding protein-related factor 3 (TRF3) during early embryogenesis. Cell Res *16*, 610-621.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F*., et al.* (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*.

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 Dev *7*, 2235-2245.

Yoshida, S. (2010). Stem cells in mammalian spermatogenesis. Dev Growth Differ *52*, 311-317.

Young, R.A. (1991). RNA polymerase II. Annu Rev Biochem *60*, 689-715.

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.

Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. Nature *408*, 225-229.

Zawel, L., Kumar, K.P., and Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. Genes Dev *9*, 1479-1490.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet *39*, 1512-1516.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.

Zenk, F., Loeser, E., Schiavo, R., Kilpert, F., Bogdanovic, O., and Iovino, N. (2017). Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. Science *357*, 212-216.

Zentner, G.E., and Henikoff, S. (2013). Mot1 redistributes TBP from TATA-containing to TATA-less promoters. Mol Cell Biol *33*, 4996-5004.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q.*, et al.* (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537*, 553-557.

Zhang, C., and Burton, Z.F. (2004). Transcription factors IIF and IIS and nucleoside triphosphate substrates as dynamic probes of the human RNA polymerase II mechanism. J Mol Biol *342*, 1085-1099.

Zhang, D., Penttila, T.L., Morris, P.L., and Roeder, R.G. (2001a). Cell- and stage-specific high-level expression of TBP-related factor 2 (TRF2) during mouse spermatogenesis. Mech Dev *106*, 203-205.

Zhang, D., Penttila, T.L., Morris, P.L., Teichmann, M., and Roeder, R.G. (2001b). Spermiogenesis deficiency in mice lacking the Trf2 gene. Science *292*, 1153-1155.

Zhang, H., Ma, Z.Y., Zeng, L., Tanaka, K., Zhang, C.J., Ma, J., Bai, G., Wang, P., Zhang, S.W., Liu,

Z.W.*, et al.* (2013a). DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci U S A *110*, 8290-8295.

Zhang, H., Wang, Z., and Zhang, Z. (2013b). PP1alpha, PP1beta and Wip-1 regulate H4S47 phosphorylation and deposition of histone H3 variant H3.3. Nucleic Acids Res *41*, 8085-8093.

Zhang, Z., and Gilmour, D.S. (2006). Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Mol Cell *21*, 65-74.

Zhao, X., and Herr, W. (2002). A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition. Cell *108*, 615-627.

Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X.*, et al.* (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. Mol Cell *63*, 1066-1079.

Zhou, H., Grubisic, I., Zheng, K., He, Y., Wang, P.J., Kaplan, T., and Tjian, R. (2013a). Taf7l cooperates with Trf2 to regulate spermiogenesis. Proc Natl Acad Sci U S A *110*, 16886-16891.

Zhou, H., Kaplan, T., Li, Y., Grubisic, I., Zhang, Z., Wang, P.J., Eisen, M.B., and Tjian, R. (2013b). Dual functions of TAF7L in adipocyte differentiation. Elife *2*, e00170.

Zhou, H., Spicuglia, S., Hsieh, J.J., Mitsiou, D.J., Hoiby, T., Veenstra, G.J., Korsmeyer, S.J., and Stunnenberg, H.G. (2006). Uncleaved TFIIA is a substrate for taspase 1 and active in transcription. Mol Cell Biol *26*, 2728-2735.

Zhou, H., Wan, B., Grubisic, I., Kaplan, T., and Tjian, R. (2014). TAF7L modulates brown adipose tissue formation. Elife *3*.

Zhou, Q., Li, T., and Price, D.H. (2012). RNA polymerase II elongation control. Annu Rev Biochem *81*, 119-143.

Zhu, H., Wang, G., and Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet *17*, 551-565.

Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E.*, et al.* (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477-481.

Zlatanova, J., Seebart, C., and Tomschik, M. (2007). Nap1: taking a closer look at a juggler protein of extraordinary skills. FASEB J *21*, 1294-1310.

Zuccotti, M., Piccinelli, A., Giorgi Rossi, P., Garagna, S., and Redi, C.A. (1995). Chromatin organization during mouse oocyte growth. Mol Reprod Dev *41*, 479-485.

Zylber, E.A., and Penman, S. (1971). Products of RNA polymerases in HeLa cell nuclei. Proc Natl Acad Sci U S A *68*, 2861-2865.

Changwei YU

Analysis of the composition and the function of oocyte-specific TBP2-containing transcription machinery during mouse oogenesis

Résumé

La synthèse d'ARN au cours de la différenciation des ovocytes est essentielle à la fécondation et à l'initiation du développement précoce. La nature de la machinerie basale de transcription pendant la croissance ovocytaire n'est pas connue mais la protéine TBP est remplacée par une protéine semblable spécifique des vertébrés, TBP2. Pour comprendre le rôle de TBP2 dans l'initiation de la transcription, nous avons effectué un RNA-seq à partir d'ovocytes contrôles et *Tbp2-/-* et montré que l'expression des gènes les plus transcrits ainsi celle des éléments rétroviraux endogènes de type MaLR est diminuée. Par immunoprécipitation couplée à la spectrométrie de masse à partir d'ovaires, nous avons montré que TBP2 ne forme pas un complexe TFIID, mais est associé à TFIIA dans les ovocytes. Globalement nos données montrent qu'une machinerie d'initiation de la transcription spécifique différente du complexe canonique TFIID contrôle la transcription dans les ovocytes de souris.

Mots clés : TFIID, TBP2, RNA polymerase II, TFIIA, MaLR, oocytes

Résumé en anglais

Mammalian oocytes go through consecutive differentiation process, during which the synthesis and accumulation of RNAs are essential for oocyte growth, maturation, fertilization and early embryogenesis. Little is known about the nature and function of the oocyte Pol II transcription machinery. During oocyte growth TBP is replaced by a vertebrate specific paralog, TBP2, and *Tbp2-/-* females are sterile. To understand whether and how TBP2 is controlling transcription initiation during oogenesis, we carried out RNA-seq analyses from wild-type and *Tbp2-/-* oocytes from primary and secondary follicles. These analyses show a main decrease in the expression of the most abundant genes as well as specific down-regulation of the expression of the MaLR-type endogenous retroviral elements. To identify the nature of the complex associated with TBP2 in the oocytes, we carried out immunoprecipitation followed by mass spectrometry. We demonstrate that, in the oocytes, TBP2 associates with TFIIA, but does not assemble into a TFIID-type complex. Altogether, our data show that a specific TBP2-TFIIA-containing transcription machinery, different from canonical TFIID, drives transcription in mouse oocytes.

Keywords : TFIID, TBP2, RNA polymerase II, TFIIA, MaLR, oocytes