UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ DE STRASBOURG
IGBMC - CNRS UMR 7104 - Inserm U 1258

THÈSE présentée par :

Fang WANG

soutenue le : **05 Décembre 2020**

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

Discipline/ Spécialité : Développement et physiologie

The role of SAGA deubiquitinase module in transcriptional regulation

Le rôle du module de deubiquitination SAGA dans la régulation de la transcription

THÈSE dirigée par :

Dr. Laszlo TORA	DR, IGBMC, université de Strasbourg

RAPPORTEURS :

Dr. Matthieu GERARD	DR, Institut de Biologie Intégrative de la Cellule (I2BC), Paris
Dr. Tilman BORGGREFE	Professor, Institute of Biochemistry, University of Giessen

AUTRES MEMBERS DU JURY

Dr. Izabela SUMARA	DR, IGBMC, université de Strasbourg
Dr. Christel BROU	DR, Institut Pasteur, Paris

Acknowledgements

First of all I would like to thank Laszlo for being my supervisor and for giving me the opportunity to work in his Lab. Thank you for your patience, guidance, encouragement and full support over the last three years. With all your guidance, I learn how to think in a scientific way and am really grateful to have you as my supervisor. Besides, I would like to thank both Stéphane and Didier for the scientific discussions and supportive comments for my project.

I would like to thank the other members of the jury, Matthieu Gérard, Tilman Borggrefe, Christel Brou and Izabela Sumara for accepting to be members for my thesis committee. Thank you for your time and effort to read and evaluate my thesis.

A special thanks goes to both the past and current numbers of the Tora Lab. I appreciate all your helpful discussions and suggestions during the past years. Among all, I would like to thank Kenny and Vincent who started around the same time as me. Thanks for the companionship and the French translation work offered for me during the past years. I want to thank Farrah, Pooja, Sascha, Changwei, Gizem, Veronique, Emmanuel and Attila for all your help about my project and for all the crazy moments we have spent together inside and outside of the Lab. I would like to thank Matthieu and Eli for all the general help in the Lab.

I would like to thank everyone who has important technical and scientific support on my work. I want to thank all the IGBMC facilities, especially Tao of the Genomeast platform for the kind help and discussions, Luc of the mass spectrometry facility for the help with my MS experiment, Betty of the cell culture facility for preparing medium, Claudine of flow cytometry facility for the help with FACS, and the animal facility for taking care of my mice.

I would also thank all my friends inside and outside of IGBMC for the help, guidance and joyful moment we shared together. I would to thank my friend Shicheng for helping me to improve my written English. Most of all, I would thanks to my family for their encouragement and constant support to me.

Table of contents

List of abbreviations

3C	Chromosome conformation capture
7SK-P-TEFb	7SK small nuclear ribonucleoprotein (snRNP)-associated P-TEFb
ADs	Activation domains
AID	The auxin-inducible degron
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related
ATXN7	Ataxin 7
ATXN7L3	Ataxin 7 like 3
AVE	Anterior visceral endoderm
BAF	Brahma-associated factor
BD	Bromodomain
BMP	β/bone morphogenetic protein
BPTF	Bromodomain PHD-finger transcription factor
BRD	Bromodomain-containing protein
BRD4-P-TEFb	Protein 4 (BRD4)-associated P-TEFb
BREs	TFIIB recognition elements
BRG1	Brahma-related gene 1
β-Tubulin III	Neuron-specific class III β-tubulin
CAGE	Cap analysis gene expression
CDK9	Cyclin-dependent kinase 9
CENP-A	Histone H3 like centromeric protein A
CERF	CECR2-containing remodelling factor
CHD1	Chromatin-helicase DNA-binding protein 1
COMPASS	Complex proteins-associated with SET1
cTnT	Cardiac muscle troponin T
CTCF	CCCTC-binding factor
CTD	C-terminal domain
CTF	CCAAT box-binding transcription factor
CTs	Chromosome territories

DBDs	DNA-binding domains
DCEs	Downstream core elements
DDR	DNA damage response
DNA/PKcs	DNA/dependent protein kinase catalytic subunit
DPE	Downstream promoter element
DRB	5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole
DSB	Double strand break
DSIF	DRB-sensitivity-inducing factor
DUBm	Histone deubiquitinase module
DVE	Distal visceral endoderm
E-cadherin	Epithelial cadherin
EMT	Epithelial-mesenchymal transition
ENY2	Enhancer of yellow 2
Epi	Epiblast
EpiLCs	Epiblastlike cells
EpiSCs	Epiblast stem cells
eRNA	Enhancer-originating RNAs
Exe	Extraembryonic ectoderm
E-ribbon	Zinc ribbon domain
ExE	Extraembryonic ectoderm
FACT	Facilitates chromatin transcription
FBP1	Far upstream element (FUSE)-binding protein 1
Fgf	Fibroblast growth factor
FgfR1	Fgf receptor 1
GCN5	General control of amino acid synthesis protein 5
GSK3	Glycogen synthase kinase-3
GTFs	General transcription factors
HATm	Histone acetyltransferase module
HCP	High-CpG-density promoters
HDAC	Histone deacetylase

HFD	Histone-fold domain
HP1	Heterochromatin protein 1
Hsp genes	Heat shock genes
HSS	Carboxy-terminal HAND-SANT-SLIDE
IF	Immunofluorescence
ICM	Inner cell mass
ICs	Interchromatin compartments
Id	Inhibitor of differentiation
IDRs	Disordered regions
INO80	SWI2/SNF2 related (SWR)
Inr	Initiator
ISWI	Imitation switch
JAK	Janus kinase
KO	Know-out
LADs	Lamin-associated chromatin domains
LIF	Leukaemia inhibitory factor
MAPK	Mitogen-activated protein kinase
MBD	Methyl-CpG-binding domain
MED12L	Mediator subunit 12 like protein
MED13L	Mediator subunit 13 like protein
Mediator	Mediator of RNA polymerase II transcription
MEF	Mouse Embryonic Fibroblasts
Mesp1	Basic-helix-loop-helix transcription factors MESoderm Posterior 1
MLL1	Mixed lineage leukaemia protein 1
MTA	Metastasis-associated
MTE	The motif ten element
mTOR	Mammalian target of rapamycin
MYOD1	Myoblast determination protein 1
NegC	Negative regulator of coupling
NELF	Negative elongation factor

NL	Nuclear lamina
NoRC	Nucleolar remodelling complex
NPC	Nuclear pore complex
NURD	Nucleosome-remodelling and histone deacetylase
NURF	Nucleosome-remodelling factor
PAF1	Polymer aseassociated factor 1
PCAF	p300/CBP-associated factor
PIC	Preinitiation complex
PI	Propidium iodide
PIKK	Phosphoinositide 3 kinase related kinase
Pol II	RNA polymerases II
PRC1	Polycomb repressive complex 1
PrE	Primitive endoderm
PS	Primitive streak
P-TEFb	Positive transcription elongation factor b
RbBP	Retinoblastoma-associated-binding protein
SCA7	Spinocerebellar ataxia type 7
SAGA	Spt-Ada-Gcn5 acetyltransferase
SEC	Super elongation complex
SEP	Shp1-eyc-p47 domain
SF3B3	Splicing Factor 3b Subunit 3
SF3B5	Splicing Factor 3b Subunit 5
snRNP	U2 small nuclear ribonucleoprotein particle
SSRP1	Structure specific recognition protein 1
SWI/SNF	Switch/sucrose non-fermentable
TADs	Topologically associating domains
TAF	TATA-binding protein (TBP)-associated factors
TAF1	TBP-associated factors
TBP	TATA-box-binding protein
TC-NER	Transcription-coupled nucleotide excision repair

TE	Trophectoderm
TFIIA	Transcription factor II A
TFIIB	Transcription factor II B
TFIIC	Transcription factor II C
TFIID	Transcription factor II D
TFIIE	Transcription factor II E
TFIIF	Transcription factor II F
TFIIH	Transcription factor II H
TGF	Transforming growth factor
TREX-2	Nuclear pore-associated transcription export complex 2
TRF1	Telomeric repeat binding factor 1
TRRAP	Transformation/transcription-domain-associated protein
TSS	Transcription start site
uaRNA	Upstream antisense RNA
VE	Visceral endoderm
WD40	WD40 β-propeller domain
WICH	WSTF ISWI chromatin remodelling
Wnt	Wingless type
WT	Wild type
ZF	Zinc finger domain

Abstract

In the nucleus, DNA is wrapped around histone proteins to form nucleosomes. The histone globular domains and the tails which extend from the nucleosome, are the substrates for a vast variety of enzymes carrying out diverse post-translational modifications (PTMs). Coactivator complexes regulate chromatin accessibility by dynamically depositing or removing PTMs on histones. SAGA (Spt-Ada-Gcn5 acetyltransferase) is an evolutionary conserved multi-subunit co-activator complex with a modular organization. The deubiquitylation module (DUBm) of SAGA is composed of the ubiquitin-specific protease 22 (USP22) and three adaptor proteins, ATXN7, ATXN7L3 and ENY2, which are all required for the removal of mono-ubiquitin (ub1) from histone H2B. To better understand the role of the DUBm in a physiological context during development, we generated null *Atxn7l3*$^{-/-}$ mouse embryos. We found that *Atxn7l3*$^{-/-}$ embryos were developmentally delayed as early as E8.5 and died around E12.5. For further analyses, we derived mESC from *Atxn7l3*$^{-/-}$ blastocysts. Our results showed that ATXN7L3 facilitated mESC self-renewal but had no obvious effect on the expression of pluripotency genes. To better characterize the function of ATXN7L3, we carried out *in vitro* mESC differentiation assays. Surprisingly, we found that ATXN7L3 was required for cardiomyocyte differentiation, but not for ectoderm neural precursor development. This observation suggests that ATXN7L3 might function in a tissue-specific manner. To understand the molecular mechanisms underlying these phenotypes, we performed transcriptomic and anti-histone H2Bub1 ChIP-seq analyses from *Atxn7l3*$^{-/-}$ mESC and wild type ESCs. Unexpectedly, although H2Bub1 levels significantly increased, the genome-wide occupancy of Pol II was only modestly changed in *Atxn7l3*$^{-/-}$ ESCs. Therefore, H2Bub1 deubiquitination did not directly regulate global Pol II transcription and the embryonic phenotypes of the *Atxn7l3*$^{-/-}$ embryo could be a consequence of the activity of the DUBm on other proteins that remains to be identified.

Thesis summary in French

Dans un contexte nucléaire, l'unité de base de la chromatine est le nucléosome qui contient ~147 paires de bases d'ADN entourant un octamère d'histone en forme de tonneau. Cet octamère contient deux copies de chacun de ces histones : H2A, H2B, H3 et H4. Le domaine globulaire central et la partie N-terminal de chaque histone peuvent servir de plateforme pour une variété de modification post-traductionnelle (PTMs), tel que l'acétylation, la phosphorylation, la méthylation ou encore l'ubiquitination. Ces modifications ont pour rôle de moduler l'accessibilité et la compaction de la chromatine via deux grands mécanismes. Premièrement, ils influencent le recrutement de certaines protéines effectrices telles que des protéines modificatrices de la chromatine ou des facteurs de transcription. Deuxièmement, ils perturbent le contact des nucléosomes ou des interactions histone-ADN. Par conséquence, les PTMs d'histone régulent des processus essentiels tels que la transcription, la réparation des dommages à l'ADN ou encore la compaction et la ségrégation des chromosomes.

L'histone H2B peut être modifiée par l'ajout ou l'élimination dynamique d'une seule molécule d'ubiquitine (ub1) sur la lysine 123 chez le modèle levure ou sur la lysine 120 chez les modèles mammifères (H2Bub1). La mise en place de cette mono-ubiquitine sur les histones H2B est catalysée by la ligase Bre1 chez la levure et par le complexe RNF20/RNF40 chez les mammifères. L'ubiquitination de H2B n'entraine pas sa dégradation, jouant néanmoins un rôle dans plusieurs processus moléculaires. Il a été reporté que H2Bub1 peut favoriser l'accessibilité de l'ADN en favorisant sa décompaction. De plus, des expériences d'immunoprécipitations de chromatine couplées à du séquençage haut-débit (ChIP-seq) ont révélé que H2Bub1 est trouvé au niveau des corps des gènes exprimés et absent au niveau des régions non transcrites, suggérant que H2Bub1 pourrait être important pour l'élongation de la transcription. Cependant, perturber la mise en place de H2Bub1 par le *knock-down* RNF20 ou *knock-out* de RNF40 n'affecte l'expression que de quelques gènes. Le rôle de H2Bub1 n'est pas donc pas encore clairement défini. H2Bub1 est aussi impliqué dans les intermodulations de PTMs d'histones. Il est supposé requis pour la tri méthylation de H3K4 et de H3K79 à la fois chez la levure et chez les mammifères. Cependant, durant la différentiation en cardiomyocyte, la tri méthylation d'un ensemble de gènes a lieu bien

que H2Bub1 ne soit pas détectable. Les cellules musculaires pourraient par conséquence constituer un nouveau modèle pour étudier la tri méthylation de H3K4.

H2B peut être déubiquitiné par le module de déubiquitination (DUB) du complexe SAGA *(Spt-Ada-Gcn5 acetyltransferase)*. Chez les mammifères, le module DUB de SAGA est composé de l'enzyme déubiquitinante USP22 et des protéines adaptatrices ATXN7, ATXN7L3 et ENY3. Dans le modèle cellulaire humain, la déplétion soit d'ENY2 ou d'ATXN7L3 empêche le fonctionnement d'USP22 et de ce fait empêche la déubiquitination de l'H2Bub1. Il a aussi été décrit que deux autres protéines voisines de USP22, appelées USP27X et USP51, peuvent interagir avec ATXN7L3 et ENY2 pour déubiquitiner H2Bub1 indépendamment du complexe SAGA. En résumé, la mise en place de H2Bub1 sur le génome dépend de sa mise en place par le complexe RNF20/RNF40 et sa suppression par trois différents modules DUB, chacun contenant une enzyme déubiquitinase différente : USP22, USP27X ou USP51. Ces trois modules semblent ne pas être complétement redondants. En effet, la mutation induisant la perte de fonction d'USP22 est létale chez la souris, les embryons ne pouvant se développer au-delà de E14.5 (14.5 jours embryonnaires). Les trois modules, ou du moins celui contenant USP22, pourraient avoir des fonctions particulières.

Plusieurs cancers humains présentent une dérégulation de la quantité à la fois de H2Bub1 ainsi que des facteurs impliqués dans sa mise en place et sa suppression. Ceci suggère que H2Bub1 jouerait un rôle important dans le maintien de l'homéostasie cellulaire. De plus, il a été rapporté qu'un changement dynamique et précis dans le temps des marques épigénétiques H2Bub doit avoir lieu pour une différentiation optimale des cellules souches embryonnaires murines (mESC).

Pour mieux comprendre le rôle du module DUB de SAGA dans un contexte physiologique et durant le développement embryonnaire, nous avons premièrement généré des lignées de souris dans lesquels les gènes *Usp22* ou *Atxn7l3* ont été éteints, respectivement appelés mutants *Usp22*$^{-/-}$ et *Atxn7l3*$^{-/-}$. Nous avons découvert que les mutants *Atxn7l3*$^{-/-}$ montrent un retard de développement dès E8.5 alors que les mutant *Usp22*$^{-/-}$ sont normaux à ce stade mais meurent à E14.5. Ces résultats indiquent qu'USP22 et ATXN7L3 sont tous les deux essentiels pour un développement embryonnaire normal mais qu'ils n'ont en revanche pas le même niveau

d'importance. Des analyses plus poussées ont montré que la quantité d'H2Bub1 n'est que faiblement modifiée dans les mutants *Usp22*$^{-/-}$ alors que le mutant *Atxn7l3*$^{-/-}$ présente une forte augmentation de la quantité H2Bub1 dans les cellules souches pluripotentes murines (mESCs) et dans les cellules fibroblastiques murines (MEFs) issues de la dérivation des embryons *Atxn7l3*$^{-/-}$ cultivé *in vitro.* Nos analyses du transcriptomique suggèrent que l'activité déubiquitination liée à ATXN7L3 régule seulement un ensemble de gènes, ces gènes n'étant pas les même dans les cellules mutantes mESCs et MEFs. De plus, la faible modification de la répartition des ARN polymèrases II (pol II) sur le gènome ne se corrèle pas aux régions présentant une forte augmentation de H2Bub1 dans les mESCs et les MEFs *Atxn7l3*$^{-/-}$. Par conséquence, la déubiquitination de H2Bub1 n'est pas impliquée dans la régulation de la transcription.

La deuxième partie de mon travail est de tester le rôle du module DUB de SAGA dans l'auto-renouvèlement et le maintien de la capacité de différenciation des mESC. Premièrement nous avons découvert que la perte d'ATXN7L3 impactait la croissance de la population cellulaire. Deuxièmement, l'analyse du cycle cellulaire des cellules mutantes indiquent un fort retard dans la transition de la phase G1 à la phase S, la transition de la phase S/G2 à G1 n'étant pas/peu affectée. Ces résultats indiquent le rôle particulier de d'ATXN7L3 dans la prolifération cellulaire. Cependant, mes résultats suggèrent que la perte d'ATXN7L3 n'affecte pas l'état de pluripotence des mESCs.

La létalité précoce des embryons *Atxn7l3*$^{-/-}$ pose problème pour détailler les processus moléculaires. Nous avons donc utilisé le modèle mESCs précédant pour réaliser des expériences de différenciation et étudier les voies de signalisation dans lesquelles la protéine ATXN7L3 pourrait être impliquée. Avec ces expériences il a été montré que ATXN7L3 est important pour l'acquisition du destin cellulaire, acquis par les cellules subissant la gastrulation. La gastrulation est une étape du développement embryonnaire durant lequel les trois feuillets embryonnaires sont acquis. En particulier ATXN7L3 est requis pour la différentiation en cardiomyocyte mais ne semble pas essentiel pour la spécification en progéniteur neuraux, suggérant la fonction spécifique de ATXN7L3.

La troisième partie de mon travail fut d'analyser le rôle du module DUB de SAGA dans la réparation de l'ADN. Les dommages non réparés sont des barrières pour l'élongation de la transcription. La Pol II bloquée par le dommage bloque ainsi les Pol II qui suivent mais aussi empêchent le recrutement des facteurs impliqués dans la réparation de l'ADN. La réparation de l'ADN couplée à l'excision de nucléotide (TC-NER) permet la réparation de l'ADN dans ces conditions particulières. En effet, la pol II bloquée peut être ubiquitinée et dégradée si le TC-NER échoue à réparer le dommage. L'ubiquitination de pol II est rapportée être un processus à plusieurs étapes qui débute avec une mono ou pluri ubiquitination au niveau de la lysine 63. Par la suite, les pol II poly-ubiquitinées sont déubiquitinées pour une forme mono-ubiquitinée et sont alors ciblées pour être dégradées. Pour identifier si ATXN7L3 est impliqué dans le mécanisme précédemment décrit, nous avons réalisé une expérience de GST-DSK2 pull-down qui permet d'isoler les protéines ubiquitinées, le domaine UBA de DSK2 se liant à l'ubiquitine (Tufegdzic Vidakovic et al., 2019). Cette expérience montre, qu'après irradiation aux UVs, ATXN7L3 facilite l'ubiquitination des pol II en élongation. Après traitement des cellules avec du DRB, un inhibiteur bloquant l'élongation de la transcription, nous avons découvert que ATXN7L3 était aussi requis pour maintenir l'ubiquitination des Pol II en élongation.

Pour résumer, mon travail de thèse est focalisé sur l'analyse de la fonction du module DUBs de SAGA durant le développement embryonnaire et la différenciation cellulaire à la fois dans la régulation de la transcription et la réparation de l'ADN.

Publication:

Farrah El-Saafin*, Fang Wang*, Tao Ye, Matthieu Stierle, Matej Durik, Veronique Fischer, Didier Devys, Stéphane D. Vincent, and László Tora. Histone H2Bub1 deubiquitylation is essential for mouse development, but does not regulate global RNA polymerase II transcription. (Submitted, Cell Death & Differentiation, 2020 July 13); * equal first authors

Workshops and Conference:

06/2018 Participation in EMBO Workshop: Chromatin dynamics and nuclear organization in genome maintenance, Illkirch, France.

08/2018 Participation in Conference: Stem Cell Dynamics Throughout Life: From Development to the Adult, Basel, Switzerland.

10/2019 Participation Conference: 6th Tri-Regional stem cell & developmental biology meeting, Illkirch, France.

09/2020 Participation in CSH meeting: Epigenetics & Chromatin (Virtual). September 2020, New York, United States. Submit abstract “The role of SAGA deubiquitinase module in transcriptional regulation”

1. Introduction

1.1 The basal transcriptional machinery

Temporospatial gene expression is a highly complex process that contributes to the identification of cell- and tissue-specific transcription in multicellular organisms. The mechanisms that regulate gene expression comprise orchestrated cooperation of diverse dedicated components. Among them, the basal transcriptional machinery plays a vital role in regulating gene expression.

The discovery of eukaryotic RNA polymerases (Pol I, Pol II, and Pol III) in sea urchin embryo nuclei (Roeder and Rutter, 1969) has triggered a huge amount of innovative inquiries. The type of RNA polymerases can be determined based on their differential sensitivities to the mushroom toxin called α-amanitin. This toxin selectively inhibits the activity of Pol II and Pol III but has no effect on Pol I (Kedinger et al., 1970; Lindell et al., 1970). Extensive research has shown that these polymerases transcribe their specific RNA types. Pol I primarily synthesizes rRNAs (28S rRNA, 18S rRNA, and 5.8S rRNA), and Pol II transcribes all the protein-coding genes and considerable noncoding RNAs. Pol III is involved in transcribing 5S rRNA, tRNAs, and adenovirus VA RNAs (Roeder and Rutter, 1970; Weinmann et al., 1974; Weinmann and Roeder, 1974).

Besides RNA polymerases, several general transcription factors (GTFs) are needed to induce site-specific transcription. *In vitro* transcriptional reaction assay has shown that the purified Pol II could accurately transcribe the DNA template only if supplemented with the crude HeLa cell extracts (Weil and Blatti, 1976). This result suggests the existence of crucial factors that facilitate the transcription process. Further studies unveiled that six Pol II-associated factors, including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, are essential for efficient transcription initiation (Matsui et al., 1980; Orphanides et al., 1996; Roeder, 1996). Detailed information on Pol II and the GTFs are discussed as following.

1.1.1 RNA polymerase II (Pol II)

Pol II is the central component of the basal transcriptional machinery. This evolutionally conserved 12-subunit complex contains a 10-subunit catalytic core as well as two subunits (RPB4 and RPB7) that form the polymerase stalk (Table 1).The first X-ray crystal structure of yeast Pol II was reported by Roger Kornberg's group (Cramer et al., 2000). They proposed the backbone model of

a 10-subunit "core" holoenzyme. Subsequently, the 12-subunit "complete" Pol II structure was described by Cramer's group and Kornberg's group (Armache et al., 2003; Bushnell and Kornberg, 2003). These studies provide new insights into the structure of Pol II. They showed that Pol II contained a 10-polypeptide catalytic core and a two-subunit Rpb4/7 complex that was critical for transcription initiation. Pol II is also highly conserved across species. Yeast and human Pol II exhibited 53% overall sequence identity (Cramer et al., 2001). However, due to the inability to obtain large quantities of the purified complex, the first 3D structure of human Pol II complex was not characterized until 2006 by Nogales' group (Kostek et al., 2006). Specifically, the sequences of Pol II catalytic core are conserved between yeast and human, which may reflect the similar mechanism of Pol II function at the DNA template (Cramer et al., 2001). Whereas the sequences at the exterior/surface residues are more divergent suggesting that they have distinct interfaces with other factors (Hahn, 2004; Kostek et al., 2006; Schier and Taatjes, 2020). Moreover, the three eukaryotic nuclear RNA polymerases also share several common subunits. It has been reported that RPB5, RPB6, RPB8, RPB10, and RPB12 were present in both Pol I and Pol III; RPB1, RPB2, RPB3 and RPB11 had homologous proteins in Pol I and Pol III. Only RPB4, RPB7, RPB9, and the C-terminal domain of RPB1 were unique to RNA polymerase II (Thomas and Chiang, 2006; Woychik et al., 1993). These results indicate that the three polymerases have significant structural and functional relationships.

Table 1 | Subunits of RNA polymerase II. (Sainsbury et al., 2015)

‡: Subunits shared among RNA polymerase I, RNA polymerase II, and RNA polymerase III.

Based on the X-ray crystallography structure (Cramer et al., 2001), Pol II contains four mobile modules: A core module, a clamp module, a shelf module and a jaw-lob module. The core module represents half of the Pol II mass. It contains part regions of RPB1 and RPB2 that form the active center and the subunits of RPB3, RPB10, RPB11 and RPB12. The clamp module is comprised of RPB1, RPB2, and RPB6. The shelf module contains RPB5, RPB6 and a part of RPB1. The jaw-lob module includes RPB1, RPB9 and the "lobe" region of RPB2. Additionally, several flexible domains facilitate Pol II functions, including the cleft region, the wall region, the stalk region and the trigger loop (Gout et al., 2017; Kaplan et al., 2008). RPB1 is the largest and catalytic subunit of Pol II (Figure 1-1) (Cramer et al., 2001; Meinhart et al., 2005). It can make up a variety of Pol II functional domains. Notably, the disordered RPB1 C-terminal domain (CTD) has been intensively investigated due to its tendency for phosphorylation.

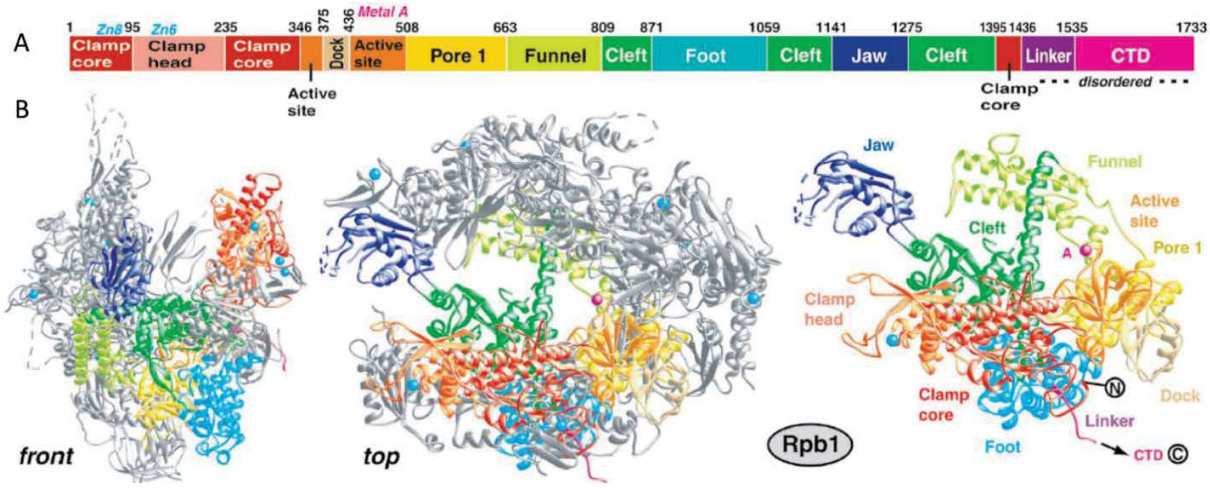


Figure 1-1: The structure of Rpb1 A. Domains and domain-like regions of Rpb1. B. Ribbon diagrams shows the location of Rpb1 within Pol II. (Cramer et al., 2001)

The tail-like RPB1 CTD contains a consensus hepta-amino acid repeat (Tyr1–Ser2–Pro3–Thr4–Ser5–Pro6–Ser7). The number of this repeat varies from 26 to 52 according to the complexity of organism. For example, budding yeast had a 26 hepta-amino acid repeats CTD whereas vertebrate species typically had 52 hepta-amino acid repeats (Chapman et al., 2008; Corden et al., 1985; Liu et al., 2010). The CTD was phosphorylated by transcription-associated kinases including CDK7 (TFIIH kinase) and CDK9 (P-TEFb kinase). TFIIH phosphorylated Ser-5 was mainly located at the promoter proximal region of active genes, and it gradually declined at the gene body due to the action of the phosphatases Ssu72 and Rtr1 (Mosley et al., 2009; Rosado-Lugo and Hampsey, 2014). Whereas CDK9 phosphorylated Ser-2 was primarily along the gene body region and was associated with transcript elongation and termination (Wilson et al., 2013). In addition, these phosphorylation markers promoted the binding of RNA processing factors (like capping enzymes, splicing, and termination factors) when Pol II left the promoter-proximal region and transcribed through gene bodies (Hsin and Manley, 2012).

Moreover, recent study revealed that both human and yeast CTDs undergo liquid phase separation. Indeed, the highly disordered CTD sequence promotes the formation of molecular condensates at active genes. These condensates are dissolved by CDK7 mediated phosphorylation of CTD. These observations suggest that the liquid phase separation property of CTD might be a key aspect of transcription regulation (Boehning et al., 2018; Lu et al., 2018).

RPB1 also undergoes ubiquitination specifically in response to DNA damage. The dynamics of Pol II pool is important for transcription regulation during DNA damage process (Tufegdzic Vidakovic et al., 2020). In eukaryotic cells, the genomic DNA is continuously damaged, and unrepaired DNA lesions interfere with the transcription. Upon DNA damage, elongating RNA polymerases are stalled at DNA lesions. Transcription-impeding DNA lesions are rapidly removed by transcription-coupled nucleotide excision repair (TC-NER). If TC-NER fails to repair the lesion, the stalled Pol II can be ubiquitylated and is subsequently degraded as a 'last resort' solution. Interestingly, Pol II ubiquitination has been reported to be a multi-step process in yeast. Firstly, RPB1 is monoubiquitylated or polyubiquitylated with Lys63-linked ubiquitin chains by the E3 ubiquitin protein ligase NEDD4. Then these polyubiquitin chains are shortened to monoubiquitylated forms by deubiquitylating enzymes (DUBs), which generate a substrate for the E3 ubiquitin ligase complex to add Lys48-linked polyubiquitin chains (Wilson et al., 2013). At last, the proteasome targets this K48-linked polyubiquitined RPB1 for degradation. However, the mechanism of RPB1 degradation is still unclear in mammalian cells. Extensive research has shown that RPB1 could be ubiquitinated at multiple sites, such as K1268, K163, K177, K758, K853, and at K1350 in HeLa cells (Nakazawa et al., 2020; Tufegdzic Vidakovic et al., 2020) (Figure 1-2).

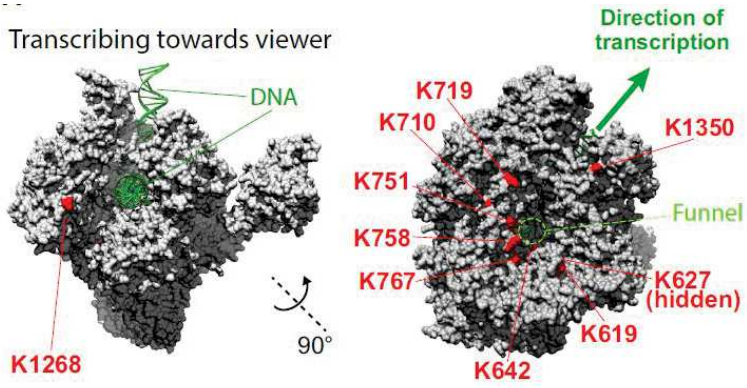


Figure 1-2: UV-induced RPB1 ubiquitylation sites on the mammalian RNAPII structure (Bernecky et al., 2016; Tufegdzic Vidakovic et al., 2020). At the catalytic subunit of Pol II, ubiquitination (K1268) site is very close to the DNA entry path.

Interestingly, only robust RPB1-K1268ub was detected upon UV treatment, suggesting that RPB1-K1268ub enables the stalled Pol II to undergo polyubiquitylation and following degradation after DNA damage (Tufegdzic Vidakovic et al., 2020). Another study found that RPB1-K1268ub also facilitated DNA repair by recruiting TFIIH to the DNA damage site in human cells (Nakazawa et

al., 2020). However, lacking RPB1 polyubiquitylation has divergent consequences between yeast and human. For example, the site analogous to human RPB1 K1268 (i.e., yeast Rpb1 K1246) locates near the entrance of the Pol II active site and affects mRNA splicing in yeast (Milligan et al., 2017). Conversely, human K1268R mutation has little or no effect on splicing (Tufegdzic Vidakovic et al., 2020). Besides, human K1268R cells are UV-sensitive whereas yeast Rpb1 K1246R cells are not. Causal factors leading to the differences remain unknown.

1.1.2 The preinitiation complex (PIC)

Although Pol II is a highly regulated complex, it cannot initiate transcription without the assistance of other transcriptional factors. To initiate transcription, general transcription factors include TFIID, TFIIA, TFIIB, TFIIF, TFIIE and TFIIH, as well as Mediator complex interacting with Pol II to form the preinitiation complex (PIC). This complex is required for targeting and melting the promoter DNA, loading Pol II onto the DNA, and for phosphorylating the CTD of Pol II.

1.1.2.1 TFIID and TFIIA recognize the promoter

TFIID is composed of the TATA-box binding protein (TBP) and 13 evolutionarily conserved TBP-associated factors (TAF1 to TAF13). Six TAFs (TAF4, TAF5, TAF6, TAF9, TAF10, and TAF12) of TFIID are present in two copies (Kolesnikova et al., 2018; Patel et al., 2018). Structural analysis has shown that TFIID was organized into a horseshoe-shaped architecture with three flexibility lobes: A, B, and C (Figure 1-3) (Brand et al., 1999; Louder et al., 2016). This structural organization enables TFIID to recognize the core promoter and nucleate the assembly of the rest of the PIC components (Buratowski et al., 1989).

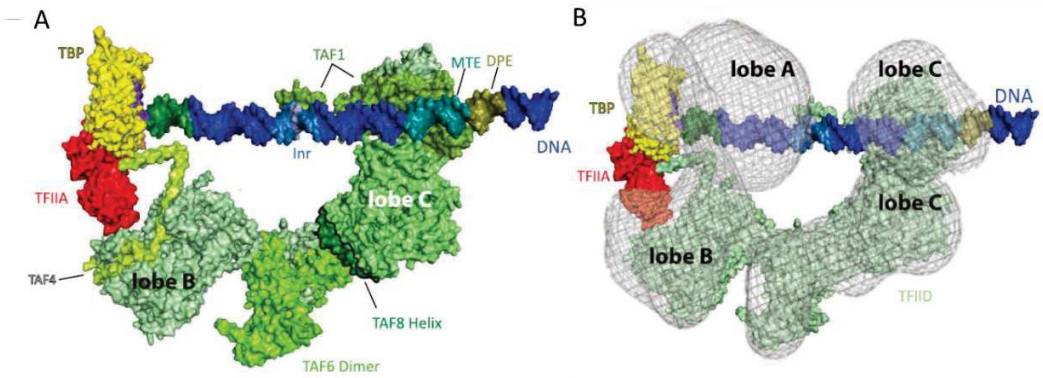


Figure 1-3: Structure of TFIID bound to promoter DNA A. TFIID structural model in the canonical state. B. The structure in A includes only structured domains (Schier and Taatjes, 2020).

The central DNA binding subunit of TFIID is a relatively small protein TBP. It is saddle-shaped and contains two highly conserved lobes (N-terminal and C-terminal lobes) that bind to the TATA-box DNA sequence in the gene promoter (Chasman et al., 1993; Louder et al., 2016). Many factors interact with TBP (Figure 1-4). For instance, the C-terminal lobe of TBP-specifically interact with TFIIB, or with multiple TFIIB paralogs, including Rrn7p, TFB, and Brf1/2 (Colbert and Hahn, 1992; Engel et al., 2017; Kosa et al., 1997). Furthermore, the N-terminal lobe of TBP can interact with TFIIA, TFIID (TAF1 subunit) and the TBP evicting factor BTAF1/Mot1p and NC2 (Anandapadamanaban et al., 2013; Bleichenbacher et al., 2003; Butryn et al., 2015; Kamada et al., 2001; Wollmann et al., 2011). In vitro studies suggested that TBP would initially bind the unbent TATA element, then it would interact with the minor groove and bend DNA about 90° (Kim et al., 1993a; Kim et al., 1993b). Besides, the TBP-dependent bending of DNA could be essential for the subsequent recruitment of TFIIB, as TFIIB binds both TBP and bent DNA on either side of TATA box (Kosa et al., 1997; Nikolov et al., 1995).

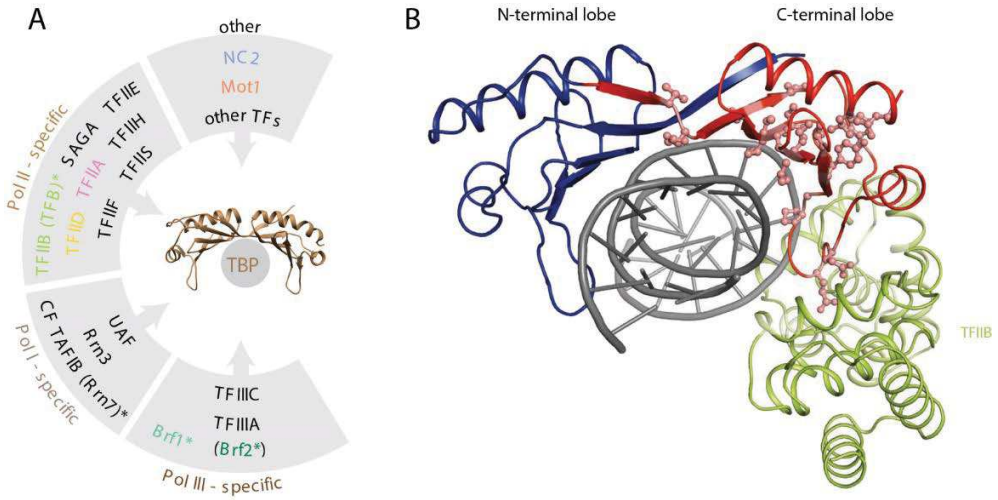


Figure 1-4: The interaction overview of TBP. A. The structure of TBP and its interactions with various components of the transcription initiation complexes. B. The C-terminal lobe of TBP interaction with TFIIB homologs. (Ravarani et al., 2020)

To recognize the promoter, the concave surface of TBP and TAFs binds the minor groove of the site of conserved TATA box (Geiger et al., 1996; Kim et al., 1993b; Tan et al., 1996). However, the TATA-box motif is only found in a small fraction (approximately 10–20%) of all Pol

II promoters (Yang et al., 2007). Therefore, in the TATA-less promoters, the TAF mediated promoter recognitions may also be essential for transcriptional initiation. Moreover, analysis of the promoter sequence leads to the identification of several additional core promoter elements that are recognized by TAFs. For example, TAF1 and TAF2 can specifically bind to the initiator element (Inr) (Chalkley and Verrijzer, 1999). TAF6–TAF9 recognizes both the motif ten element (MTE) and the downstream promoter element (DPE) (Burke and Kadonaga, 1997; Theisen et al., 2010). Moreover, the TAFs can also direct the recruitment of TFIID complex through interacting with histone post-translational modifications. Particularly, the tandem bromodomains within TAF1 selectively bind to multiple H4 acetylated peptides (H4K5ac/K12ac) (Jacobson et al., 2000), and the PHD finger domain in TAF3 specifically interacts with H3K4me3 (Lauberth et al., 2013). As H4 acetylation and H3K4me3 are associated with active transcription, their ability to bind TFIID may contribute to this function.

Although the affinity is approximately 1000 times lower compared with the TATA motif, TBP can still bind to nonspecific DNAs and form non-productive PIC (Coleman and Pugh, 1995). It has been reported that MOT1 and NC2 factors could interact with the concave surface of TBP to block TFIIA or TFIIB from binding to TBP, thereby inhibiting the formation of non-productive PICs (Gilfillan et al., 2005; Kamada et al., 2001).

TFIID also undergoes structural rearrangement after binding to the promoter, which is a critical determinant of PIC assembly. Based on the position of lobe A, TFIID has two significant conformations, including the canonical TFIID and rearranged TFIID. In the canonical state, lobe A of the free-TFIID interacts with lobe C and TBP, and the TAND motif of TAF1 blocks TFIID binding to DNA. Notably, the presence of promoter DNA and TFIIA stimulates TFIID rearrangement. In the rearranged TFIID state, TFIIA represses the inhibitory effect of TAF1 and drives the lobe A to bind the lobe B (Nogales et al., 2017) (Figure 1-5). The rearrangement of TFIID and its interaction with TFIIA are likely to be coupled events that induce the shifting of TBP to the upstream promoter DNA. Together, these structural rearrangement enhances the process of PIC assembly (Orphanides et al., 1996).

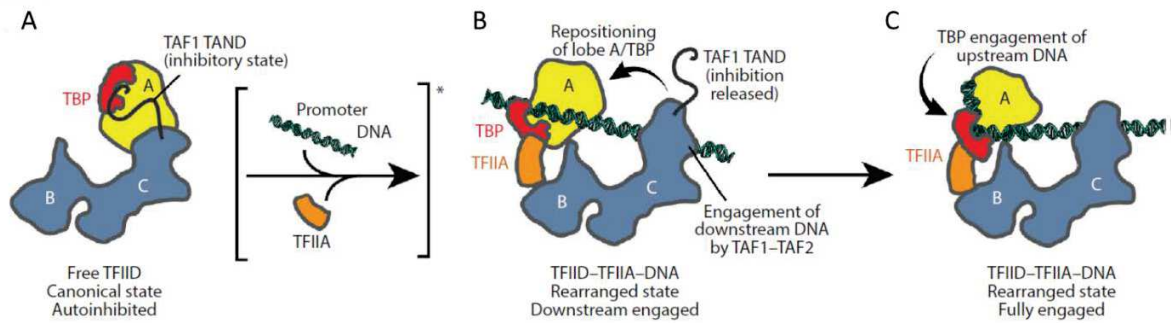


Figure 1-5: A model of promoter binding by TFIID A. TFIID is in the autoinhibitory canonical state. B. TFIID interacts with promoter DNA and TFIIA. C. Interactions between TFIID and the promoter are probably initiated by TAF1-TAF2 in the downstream promoter region, placing the upstream promoter DNA in position to be engaged by TBP (Nogales et al., 2017).

1.1.2.2 TFIIB interacts with RNA polymerase II

TFIIB is the only GTF composed of a single polypeptide, which can be divided into several functional domains, including a B-ribbon, a B-reader, a B-linker, and two B-cores. Structural studies have revealed how TFIIB specifically selected the TSS sites on the promoter regions. Firstly, the B-core domain binds the wall at the end of the Pol II cleft, which positions the promoter DNA at the Pol II active center cleft (Bushnell et al., 2004). Subsequently, the promoter DNA is opened with the assistance of the B-linker domain. Finally, the DNA template strand escapes into the cleft, where the B-reader domain reads the DNA sequence and facilitates the TSS selection (Kostrewa et al., 2009) (Figure 1-6).

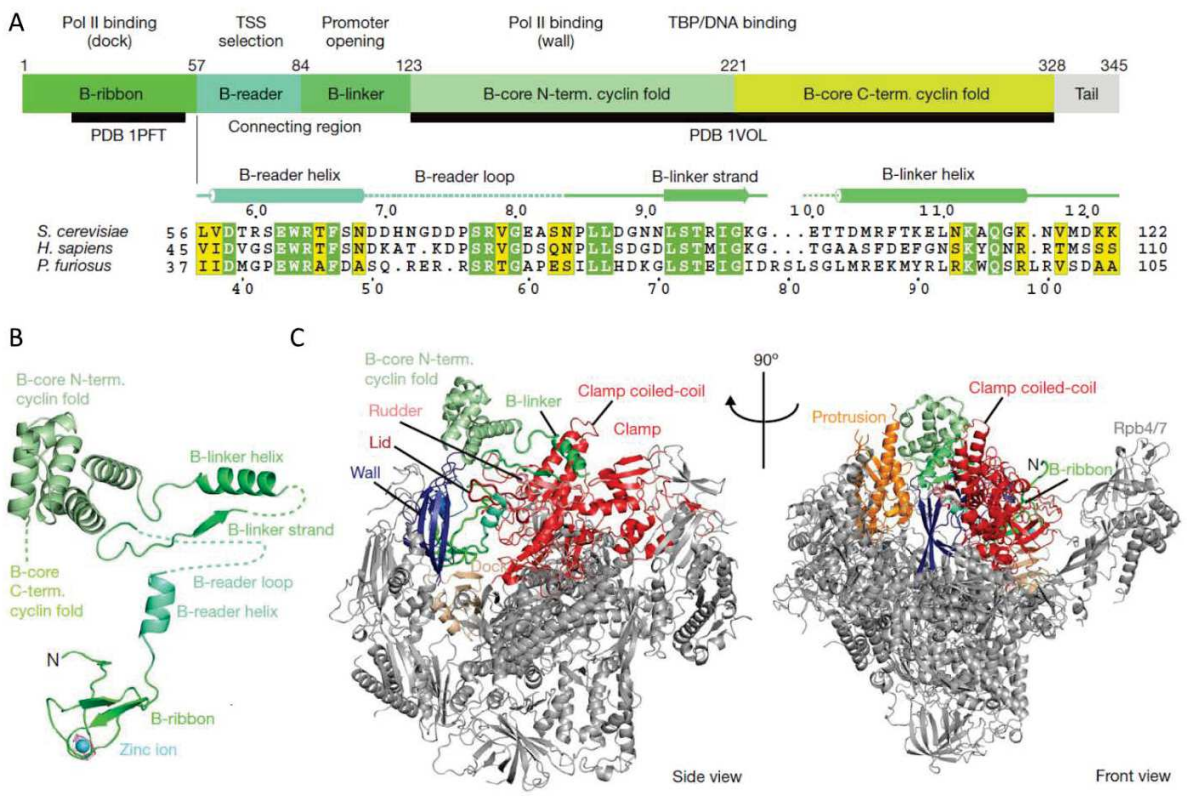


Figure 1-6: Structure of Pol II–TFIIB complex. A. The domains of TFIIB among different species. B. Ribbon module of TFIIB-Pol II. C. Overview of the Pol II–B structure. (Kostrewa et al., 2009)

TFIIB in turn also stabilizes the TFIID complex that has been recruited at the promoter region. As introduced above, once TBP is bound to the promoter, TFIIB seems to be the next GTF to enter the PIC assembling pathway and will interact with the promoter-bound TBP. This process facilitates the formation of a more stable complex composing of TBP-TFIIB-DNA or TBP-TFIIA-TFIIB-DNA (Lagrange et al., 1996). On the other hand, TFIIB acts as a bridge to connect Pol II with promoter DNA. As the B-ribbon domain of TFIIB associates with Pol II meanwhile the B-core domain binds to the promoter DNA (Nikolov et al., 1995). Additionally, TFIIB is also involved in the transcriptional initiation-to-elongation transition. It has been reported that the B-linker of TFIIB promoted DNA opening and maintained the transcription bubble whereas the synthesis of the RNA chain and rewinding of upstream DNA released the B-linker (Kostrewa et al., 2009; Sainsbury et al., 2015; Sainsbury et al., 2013). This observation suggests that the release of B-linker might trigger the formation of elongation complex.

1.1.2.3 TFIIF and its role in transcriptional initiation

The heterodimer TFIIF contains two subunits, TFIIFα and TFIIFβ, that correspond to Tfg1 and Tfg2 in yeast, respectively (Flores et al., 1988). The structure of TFIIF shows that the N-terminal regions of TFIIFα and TFIIFβ form a dimerization module. The C-terminal winged-helix domain connects to the dimerization module by a charged region in TFIIFα and a linker region in TFIIFβ (Chen et al., 2010; Eichner et al., 2010; Gaiser et al., 2000).

Structural studies have elucidated that TFIIF facilitates transcriptional initiation from various aspects. Firstly, TFIIF prevents Pol II from nonspecifically interacting with DNA by interacting with Pol II at the RPB2 lobe and protrusion domains (Conaway et al., 1991). Secondly, TFIIF facilitates the association of Pol II with promoter DNA. Upon connecting to a promoter-bound Pol II–TBP–TFIIB–TFIIA complex, TFIIF induces structural changes to the PIC. This process enables the Pol II subunits (RPB2 and RPB5) to bind to the promoter DNA that had positioned at the upstream and downstream of the TSS (He et al., 2013; He et al., 2016). Thirdly, TFIIF also trappes the double-stranded DNA above the Pol II cleft domain, which sets a stage for promoter melting and transcription initiation (Plaschka et al., 2016; Schilbach et al., 2017).

1.1.2.4 TFIIE and TFIIH facilitate promoter DNA opening

TFIIE and TFIIH are required for promoter DNA opening to form a transcriptionally competent PIC. TFIIE is a heterodimer complex comprising of two subunits TFIIEα and TFIIEβ (Ohkuma et al., 1990; Peterson et al., 1991). TFIIEα contains an N-terminal WH domain and a central zinc ribbon domain (E-ribbon) (Figure 1-7A, B). The N-terminal of TFIIEα is essential for its connection with TFIIEβ. TFIIEβ is comprised of two WH domains and a basic C-terminal region (E-tether) (Okamoto et al., 1998; Plaschka et al., 2016; Sainsbury et al., 2015). Site-specific cleavage analysis showed that the three WH domains of TFIIE were close within the PIC. The TFIIEα WH domain anchors the complex to the Pol II clamp while the other two WH domains in TFIIEβ surround the promoter DNA (Grunberg et al., 2012). In addition, TFIIE also interacts with the TFIIFβ WH domain. Therefore, four WH domains, one from TFIIF and three from TFIIE, span over the Pol II cleft that contains the loaded DNA (He et al., 2013) (Figure 1-7C). Consequently,

TFIIF plays a critical role in positing the promoter DNA over the Pol II cleft where the double-strand DNA will be melting by TFIIH.

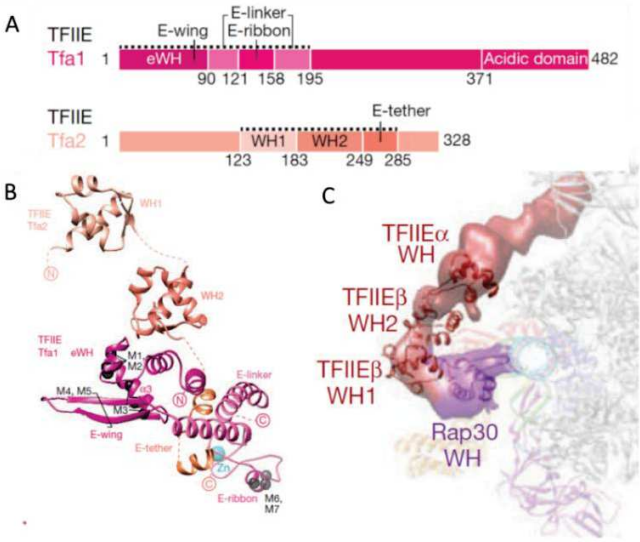


Figure 1-7: TFIIE architecture and interactions (A, B) Domain organization of yeast TFIIE. (C) A chain of four WH domains formed by the C-terminus of RAP30 and subunits of TFIIE (He et al., 2013).

TFIIH is reported to regulate transcription process by triggering promoter DNA opening (Holstege et al., 1996) and Pol II escaping (Goodrich and Tjian, 1994; Moreland et al., 1999). TFIIH is a 10-subunit complex containing a six-subunit core module, a dissociable three-subunit kinase module and a XPD subunit. Among them, the core module includes XPB, p62, p52, p34, p8 and p44. CDK7–cyclin H–MAT1 complex constitutes the kinase module. Moreover, the XPD subunit connects the core module and the kinase module by interacting with p44 and MAT1 (Gibbons et al., 2012; Murakami et al., 2012) (Figure 1-8).

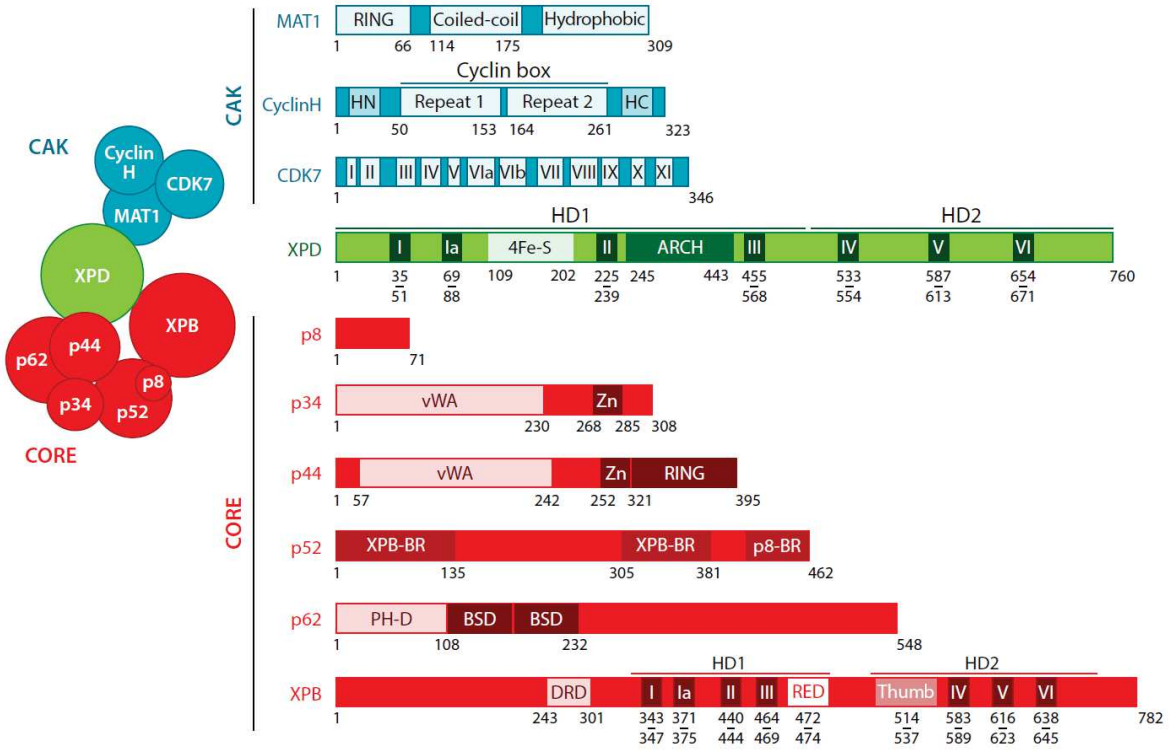


Figure 1-8: Composition of human transcription factor IIH (TFIIH). The CORE (red) of TFIIH contains six subunits, including XPB, p62, p52, p44, p34, and p8/TTDA; CAK (blue) is composed of CDK7, CyclinH, and MAT1. (Compe and Egly, 2016)

XPB was initially characterized as a helicase but could also function as a 5'-3' DNA translocase. It was proposed that the ATPase activity of XPB but not the helicase role of XPB initiated a conformational change in the PIC, which is required for promoter opening during transcriptional process (Holstege et al., 1996; Lin et al., 2005). However, another study suggested that Ssl2 (XPB in human) promoted DNA opening by functioning as a dsDNA translocase (Fishburn et al., 2015). Based on these results, it can be speculated that the ATPase Ssl2/XPB tracks along the template strand DNA in the 3'-5' direction and places downstream DNA into the active center cleft of Pol II, whereas the upstream DNA remains fixed to promot DNA opening (Sainsbury et al., 2015).

Besides the function on transcriptional regulation, the ATPase activity of XPB is also involved in the DNA damage response. The enzymatic subunits, XPD and XPB, both possess two RecA-like helicase domains, HD1 and HD2. They are supposed to function together during the DNA damage repair process. It has been investigated that the ATP hydrolysis of XPB induced a large XPB

conformational change, which promoted XPB to separate the two DNA strands around the lesion, creating an enviroment that favored the DNA binding of XPD (Fan et al., 2006). Then, the translocase function of XPD induced DNA opening at the DNA damage sites, which facilitated DNA damage repair (Coin et al., 2007; Moreland et al., 1999). During this process, p52 and p8 subunits stimulated the ATPase activity of XPB (Coin et al., 2007; Coin et al., 2006), while p44 subunit regulated the helicase activity of XPD (Dubaele et al., 2003).

The CDK kinase of TFIIH also plays essential roles in transcriptional regulation. CDK7 is initially identified as a catalytic subunit of CAK (Roy et al., 1994). As part of the CAK subcomplex, MAT1 stabilizes the association of CDK7 and Cyclin H (Adamczewski et al., 1996; Devault et al., 1995; Rossignol et al., 1997). CDK7 specifically phosphorylates the Pol II C-terminal at serine 5 and 7 residues (Feaver et al., 1994; Glover-Cutter et al., 2009; Lu et al., 1992), which facilitates the release of Pol II from the PIC. Consequently, without TFIIH, Pol II tend to abortive transcription and stalled at the promoter-proximal region (Thomas and Chiang, 2006).

Two different PIC assembling models have been described: One is the sequential assemble model, in which GTFs join the PIC in a stepwise manner, except for TFIIA, which can enter the PIC at any step after TFIID binding. In the stepwise assembly model, TFIID first binds the TATA element of the promoter. TFIIB is next to bind. TFIIF then facilitates Pol II recruitment at the promoter. The preinitiation complex is completed by the binding of TFIIE and TFIIH (Orphanides et al., 1996). The second model, the RNA Pol II holoenzyme pathway model, is based on the observation that Pol II could be purified as a preassembled holoenzyme containing also several GTFs, Mediator, and chromatin remodeler proteins. In this model, the authors suggested that TFIID would bind first to the core promoter and would promote the recruitment of the pre-assembled holoenzyme (Thomas and Chiang, 2006). Both models are supported by *in vitro* studies. However, there is no conclusive evidence of which one is used *in vivo*.

1.1.3 Promoter-proximal pausing of RNA polymerase II

Following the assembly of a pre-initiation complex at the gene promoter, Pol II is typically stalled within the promoter-proximal region (Core et al., 2008; Muse et al., 2007) (Figure 1-9). The phenomenon of promoter-proximal Pol II pausing was first described at the *Drosophila* heat shock

genes (*Hsp* genes), where Pol II accumulates just downstream of the promoter regions and is associated with 20–60 nucleotides nascent RNA (Gilmour and Lis, 1986; Rasmussen and Lis, 1993, 1995; Rougvie and Lis, 1988). Further studies revealed that the promoter-proximal pausing was a widespread phenomenon, as the majority of active genes in metazoan showed Pol II peaking near promoters and underwent a rate-limiting step from the transcriptional initiation to productive elongation (Guenther et al., 2007; Kim et al., 2005; Levine, 2011; Muse et al., 2007). These paused polymerases either terminate or are released into productive elongation (Brannan and Bentley, 2012; Brannan et al., 2012; Wagschal et al., 2012). Moreover, the paused Pol II can also transcribe upstream antisense RNA (uaRNA), enhancer-originating RNAs (eRNA), and long noncoding RNAs (Bunch et al., 2016; Core et al., 2014; Tome et al., 2018), which indicates its function in regulating noncoding RNA species.

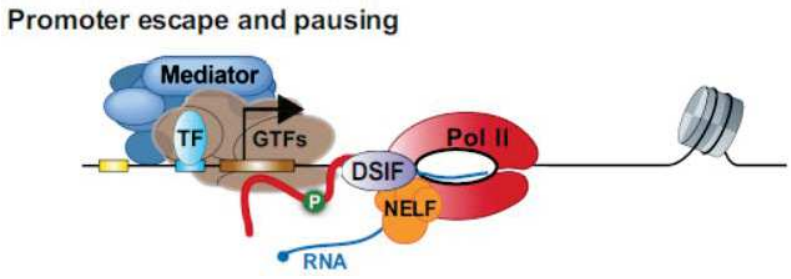


Figure 1-9: Escape and pausing of RNA polymerase II. Recruitment of Pol II by general transcription factors (GTFs) results in the formation of a pre-initiation complex (Lepoivre et al.). After rapid Pol II initiation and entry into the pause site, NELF and DSIF facilitate Pol II pause (Core and Adelman, 2019).

Pol II promoter-proximal pausing depends on the binding of two factors: DRB-sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) (Adelman and Lis, 2012; Kwak and Lis, 2013). DSIF is comprised of Spt5 and Spt4. Structural studies showed that is Spt5 docked on Pol II near the RNA exit channel (Sainsbury et al., 2015; Vos et al., 2018a), whereas Spt5 stablizes the pasuing of Pol II and facilitates the capping of the nascent RNA (Pei and Shuman, 2002; Yamaguchi et al., 1999b). Another factor NELF is composed of the four subunits NELF-A, -B, -C and -E (Narita et al., 2003; Yamaguchi et al., 1999b). NELF is suggested to recognize the Pol II–Spt5 interface. Thereby, NELF can restrain Pol II mobility and prevent the binding of the anti-pausing transcription elongation factor IIS (TFIIS) that is required for pause release (Vos et al., 2018b).

In addition to NELF and DSIF which stabilize promoter-paused Pol II on most genes, TFs are involved in enhancing Pol II pausing in a gene- and sequence-specific manner. For example, in mammals, SP1, myoblast determination protein 1 (MYOD1) and CCAAT box-binding transcription factor (CTF) are considered to be DNA sequence-specific TFs that recruite Pol II to the promoter region without stimulating the release of paused Pol II, thereby increasing the levels of paused Pol II (Blau et al., 1996; Krumm et al., 1995). In some cases, nucleosome composition can also influence pausing. It has been suggested that the histone variant H2A.Z negatively correlates with the establishment of pausing (Hu et al., 2013; Weber et al., 2014). Interestingly, the paused Pol II is not likely to be fixed in one position but tends to undergo persistent rounds of transcription, pausing, termination and/or backtracking (Krebs et al., 2017; Nechaev et al., 2010; Weber et al., 2014). In turn, the backtracking Pol II may contribute to the longevity of the paused state (Sheridan et al., 2019). TFIIS can cleave the RNA attched with the backtracking, or pausing Pol II, to realign the RNA with the Pol II active site, which enables Pol II to be released into productive elongation upon inducing the kinase activity of positive transcription elongation factor b (P-TEFb) (Cheung and Cramer, 2011; Izban and Luse, 1992; Kettenberger et al., 2003). Above all, Pol II promoter-proximal pausing is dynamicly regulated by various factors.

Different hypotheses for the function of Pol II pausing have been proposed, including establishing permissive chromatin, severing as pausing framework for rapid and synchronous gene activation in response to developmental or enviromental cues, integrating multiple regulatory signals, acting as a checkpoint for coupling elongation and RNA processing (Adelman and Lis, 2012; Levine, 2011).

1.1.4 Release of paused Pol II

Release of paused Pol II into productive RNA synthesis is triggered by the activity of positive transcription P-TEFb (Figure 1-10). P-TEFb is comprised of cyclin T1 and cyclin-dependent kinase 9 (CDK9) (Peterlin and Price, 2006; Zhou et al., 2012). It is recruited to promoters through direct or indirect interacting with specific TFs, Mediator and cofactors (Peterlin and Price, 2006; Takahashi et al., 2011). The kinase activity of P-TEFb can phosphorylate the CTD of Pol II at Ser2, as well as DSIF and NELF (Kwak and Lis, 2013), leading to the dissociation of NELF and

the conversion of DSIF into a positive transcription elongation factor (Cheng and Price, 2007; Guo et al., 2000; Jonkers and Lis, 2015; Yamada et al., 2006). Besides, P-TEFb also directly regulates the initial recruitment of PAF1 complex (PAF1C) that is a critical regulator of paused Pol II release to genes (Yu et al., 2015). Together, P-TEFb enables Pol II reactivation and resumption of elongation.

Notably, P-TEFb is part of several larger complexes, such as the super elongation complex (SEC) (Luo et al., 2012), bromodomain containing protein 4 (BRD4)-associated P-TEFb (BRD4-P-TEFb) (Yang et al., 2005) and 7SK small nuclear ribonucleoprotein (snRNP)-associated P-TEFb (7SK-P-TEFb) (Yang et al., 2001) (Table 2). Thereby the activity of P-TEFb is highly regulated by the subunits of these complexes. For example, 7SK small nuclear RNA binds to and inhibits the activity of P-TEFb, whereas the bromodomain protein Brd4 positively regulates P-TEFb and stimulates Pol II-dependent transcription (Jang et al., 2005; Nguyen et al., 2001; Yang et al., 2001; Zhou et al., 2012). Consequently, the level of Pol II pausing depends on the balance between pausing factors (such as NELF, DSIF and nucleosome) and activating factors (that either recruit P-TEFb to paused Pol II, or regulate the activity of P-TEFb).

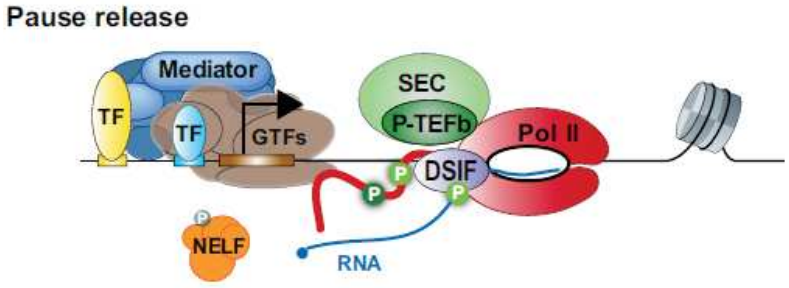


Figure 1-10: Pause release of RNA polymerase II. SEC complex contains most of the active P-TEFb which promotes rapid release of the paused Pol II into productive elongation by phosphorylating the CTD of Pol II at Ser2, as well as DSIF and NELF (Core and Adelman, 2019).

Table 2 | Pausing-related factors (Chen et al., 2018)

Pausing- related factors	Subunits	Occupancy	Function in pausing
NELF	• NELF-A • NELF-B • NELF-C or NELF- D • NELF-E	Promoter	Stabilizes paused Pol II by preventing premature promoter-proximal termination
DSIF	• SPT4 • SPT5	• Promoter • Gene body	Promotes the recruitment of NELF and capping factors
PAF1C	• PAF1 • CTR9 • LEO1 • Parafibromin • WDR61 • RTF1	• Enhancer • Promoter • Gene body	Modulates enhancer activity and maintains paused Pol II by hindering its release into productive elongation
Gdown1[a]	–	Promoter	Blocks TFIIF recruitment and prevents early termination of promoter-proximal Pol II
PARP1	–	• Enhancer • Promoter	ADP-ribosylates NELF and inhibits its function in pausing
P-TEFb	• CDK9 • CCNT1 or CCNT2	• Enhancer • Promoter • Gene body	Phosphorylates the Pol II CTD, NELF and the SPT5 CTR to promote release from pausing
SEC	• AFF1 or AFF4 • ELL2 • AF9 or ENL • EAF1 or EAF2 • P-TEFb	• Enhancer • Promoter • Gene body	Most active P-TEFb-containing complex; promotes rapid release of paused Pol II into productive elongation
BRD4–P-TEFb	• BRD4 • P-TEFb	• Enhancer • Promoter • Gene body	Stimulates P-TEFb activity and promotes pause release
7SK–P-TEFb	• 7SK snRNP • MEPCE • LARP7 • HEXIM1 or HEXIM 2 • P-TEFb	Promoter	Sequesters P-TEFb and prevents pause release

1.1.5 Transcription elongation

After releasing from the promoter-proximal region, Pol II enters productive elongation (Figure 1-11). The transcription rate is variable and can be different as much as threefold in different genes (Danko et al., 2013; Jonkers et al., 2014; Saponaro et al., 2014; Veloso et al., 2014). Moreover, in mammalian cells, productive elongation is not very efficient within the first kilobase, and is increased from approximately 0.5 kb per min within the first few kilobases, to 2–5 kb per minute after approximately 15 kb (Jonkers et al., 2014; Jonkers and Lis, 2015). Moreover, mRNA

cleavage, the presence of exons and polyadenylation sites can slowdown Pol II trascriptional rate (Jonkers and Lis, 2015).

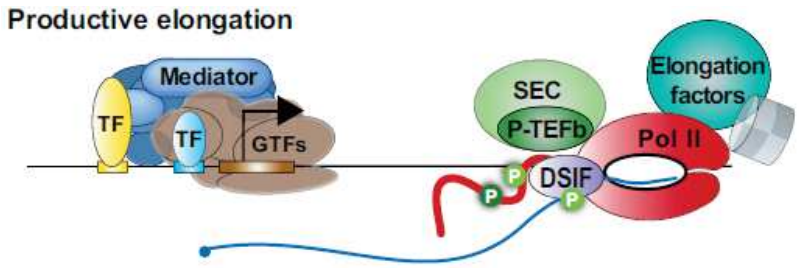


Figure 1-11: Productive elongation of RNA polymerase II. Elongation factors, such as Spt6, FACT, elongin, TFIIS and polymerase associated factor 1 (PAF1), facilitate productive elongation (Core and Adelman, 2019).

One of the hallmarks for transcription elongation is the nucleosome dynamics that occurs during the passage of Pol II. For example, nucleosomes are evicted in front of transcribing Pol II and rapidly reassembled behind the elongating Pol II (Bernstein et al., 2004; Dion et al., 2007). Histone chaperone FACT binds and displaces the H2A/H2B dimer from the core nucleosomes, which enhance nucleosome breathing to facilitate the passage of Pol II (Belotserkovskaya et al., 2003; Hondele et al., 2013; Kemble et al., 2015). Factors implicated in Pol II pausing also facilitate transcription elongation. For instance, after releasing Pol II into productive elongation, PAF1C travels with the elongating Pol II and acts as platforms by recruiting a variety of factors to promote elongation (Ng et al., 2003; Pavri et al., 2006; Simic et al., 2003; Wood et al., 2003b). As another example, pausing factor DSIF can change into an elongation factor upon phosphorylation by P-TEFb. This phosphorylated DSIF promotes productive elongation by interacting with elongating Pol II and reinforcing the closed conformation of the Pol II clamp for the passage of the template DNA through Pol II (Doamekpor et al., 2014; Grohmann et al., 2011; Klein et al., 2011). Interestingly, recent study reveals that the CTD-S2P can also be incorporated into phase-separated condensate formed by a disordered region in P-TEFb at gene body regions (Lu et al., 2018). This condensate in turn facilitates elongation and cotranscriptional RNA processing (Cramer, 2019; Lu et al., 2018).

Besides, many histone modifications deposited at the gene body are supposed to associate with transcription elongation. For example, histone H2Bub1, H3K36me3 and H3K79me2 are supposed

to regulate Pol II elongation by serving as platforms for the binding of histone chaperones or chromatin remodellers that regulate nucleosome disassembly and reassembly in the wake of the elongating Pol II (Venkatesh and Workman, 2015; Zentner and Henikoff, 2013). Histone chaperones, including FACT, SPT6 and ASF1, as well as chromatin-remodelling complexes like chromatin-helicase DNA-binding protein 1 (CHD1) are also involved in productive transcriptional elongation (Venkatesh and Workman, 2015). Therefore, transcription elongtion is a highly regulated process which is inviolved in a variety of factors.

1.2 *Cis*-acting DNA elements

Transcriptional initiation occurs following the recruitment of PIC at the core promoter (Thomas and Chiang, 2006). *Cis*-acting DNA elements that can be separated into promoter and distal regulatory elements, act as platforms for the assembly of PIC. The distal regulatory elements contain locus control region, silencer, enhancer, and insulator DNA elements. The core promoter and proximal promoter elements comprise the promoter which typically spans less than 1 kilo base (bp) pairs. However, the distance between promoter and distal regulatory elements can be up to 1 million base (Mb) pairs (Figure 2-1) (Maston et al., 2006).

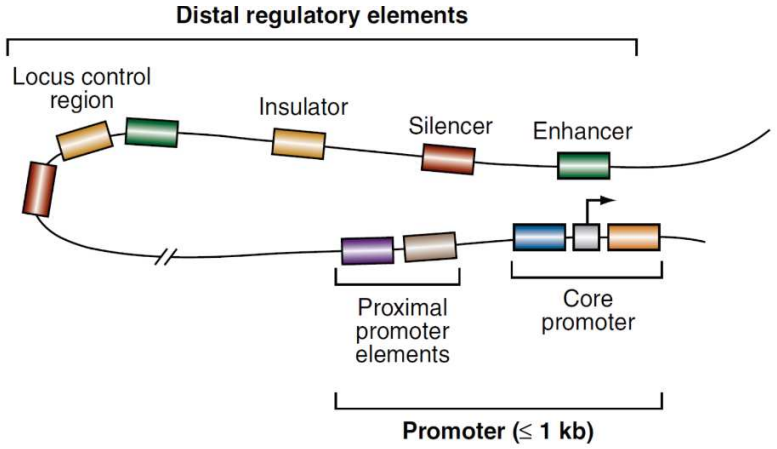


Figure 2-1: Schematic of a typical gene regulatory region. The promoter is composed of a core promoter and proximal promoter elements. Distal regulatory elements include enhancers, silencers, insulators, and locus control regions, which is located up to 1 Mb pairs from the promoter. (Maston et al., 2006)

1.2.1 Core promoter

The core promoter is the minimum docking site that is required to assemble the transcriptional initiation complex. It encompasses the transcription start site (TSS) and the 40-50 bp of upstream and downstream DNA that extent from TSS (Butler and Kadonaga, 2002; Kadonaga, 2012). Some of the identified core promoter motifs are shown in table 3. However, there are no universal core promoter motifs and some core promoters even lack any of these identified motifs during transcription.

Table 3 | Core promoter motifs. (Haberle and Stark, 2018)

Core-promoter motif	Sequence logo	Consensus sequence[a]	Position relative to TSS	Bound by	Fly	Human
TATA-box		TATAWAWR	–31 to –24	TBP	+	+
Inr (fly)		TCAGTY	–5 to –2	TAF1 and TAF2	+	–
Inr (human)		YR	–1 to +1	NA	–	+
		BBCABW	–3 to +3			
DPE		RGWCGTG	+28 to +34	TAF6 and TAF9 and 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	+	+
BREu		SSRCGCC	–38 to –32	TFIIB	+	+
BREd		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	+	?

The well-known TATA-box core promoter motif is located at about 30 bp upstream of TSS and is recognized by the TBP protein (Patikoglou et al., 1999). In the past, the TATA-box was assumed to be a universal core promoter element. Nowadays, however, growing evidence suggest that this is overestimated, as considerable percentage of promoters do not contain a conserved TATA-box (TATAWAWR) motif (Yang et al., 2007). For example, the core promoters that contain a TATA-box only constitute about 17% of the total promoters in *S. cerevisiae*, 14% of the total promoters in *D. melanogaster*, 10% of the total promoters in zebrafish, 9% of the total promoters in *C.elegans*, and 3% of the total promoters in human and mice (Yella and Bansal, 2017).

The initiator (Inr) is another widely used core promoter motif. It is located directly overlaps with the TSS (Smale and Baltimore, 1989). This conserved Inr motif serves as a binding site for TAF1 and TAF2 that are the subunits of TFIID (Chalkley and Verrijzer, 1999; Louder et al., 2016). In the TATA-less promoters, the Inr motif is often accompanied with the downstream promoter element (DPE)(Burke and Kadonaga, 1996). Moreover, the spacing between the Inr and the DPE motif is reported to facilitate the deposition of TFIID at the DPE motif (Burke and Kadonaga, 1997; Louder et al., 2016). Interestingly, as the DPE and the TATA-box motif are rarely co-occurring in flies, they are suggested to be associated with functionally distinct groups of genes (FitzGerald et al., 2006; Kutach and Kadonaga, 2000). In addition to these three most abundant core promoter motifs, other defined motifs, including the motif ten element (MTE) (Lim et al., 2004), TFIIB recognition elements (BREs) (Deng and Roberts, 2005) and downstream core elements (DCEs) (Lewis et al., 2000), are bound by specific GTFs *in vitro* (Lee et al., 2005b). Therefore, they are suggested to mediate PIC recruitment and assembly. Overall, the discovery of core-promoter motifs significantly contributes to the complexity of transcription regulation in eukaryotic genes.

The development of the high-throughput-sequencing technologies, such as cap analysis of gene expression (CAGE), have allowed comprehensive promoter analysis (Shiraki et al., 2003). Based on the properties of transcription initiation pattern, DNA sequence composition and histone modifications, the core promoters are supposed to be separated into three main types (Figure 2-2) (Haberle and Stark, 2018; Lenhard et al., 2012). Type I core promoter is associated with active transcription in terminally differentiated cells. These core promoters tend to have a sharp initiation

pattern with a TATA-box and Inr motif near the TSS (Roider et al., 2009). H3K4me3 and H3K27ac histone modifications are also deposited at the type I promoters (Rach et al., 2011). However, the type I promoter lacks CpG islands. Besides, the nearby nucleosome is imprecisely positioned at this type of promoters (Rach et al., 2011). In contrast to type I core promoter, the type II core promoter is mainly found in the broadly expressed housekeeping genes. This core promoter is associated with dispersed transcription initiation and has a well-defined nucleosome depleted region flanked by precisely positioned nucleosome (Rach et al., 2011). Furthermore, the type II promoter also overlaps with individual CGIs in mammals (Carninci et al., 2006). Different from the type II promoter, the type III core promoter is associated with key developmental transcription factor genes, which contain a dispersed TSS pattern and a precisely poised nucleosome. In embryonic stem cells, type III core promoter is distinctly marked with both the active chromatin modification H3K4me3 and the repressive chromatin modification H3K27me3. These bivalent histone modifications guarantee the quick activation of silent genes in specific cell lineages during differentiation process. This type III promoter is associated with multiple CGIs that extend to the gene bodies. However, the mechanisms by which CGIs confer core promoter function are still unclear (Akalin et al., 2009).

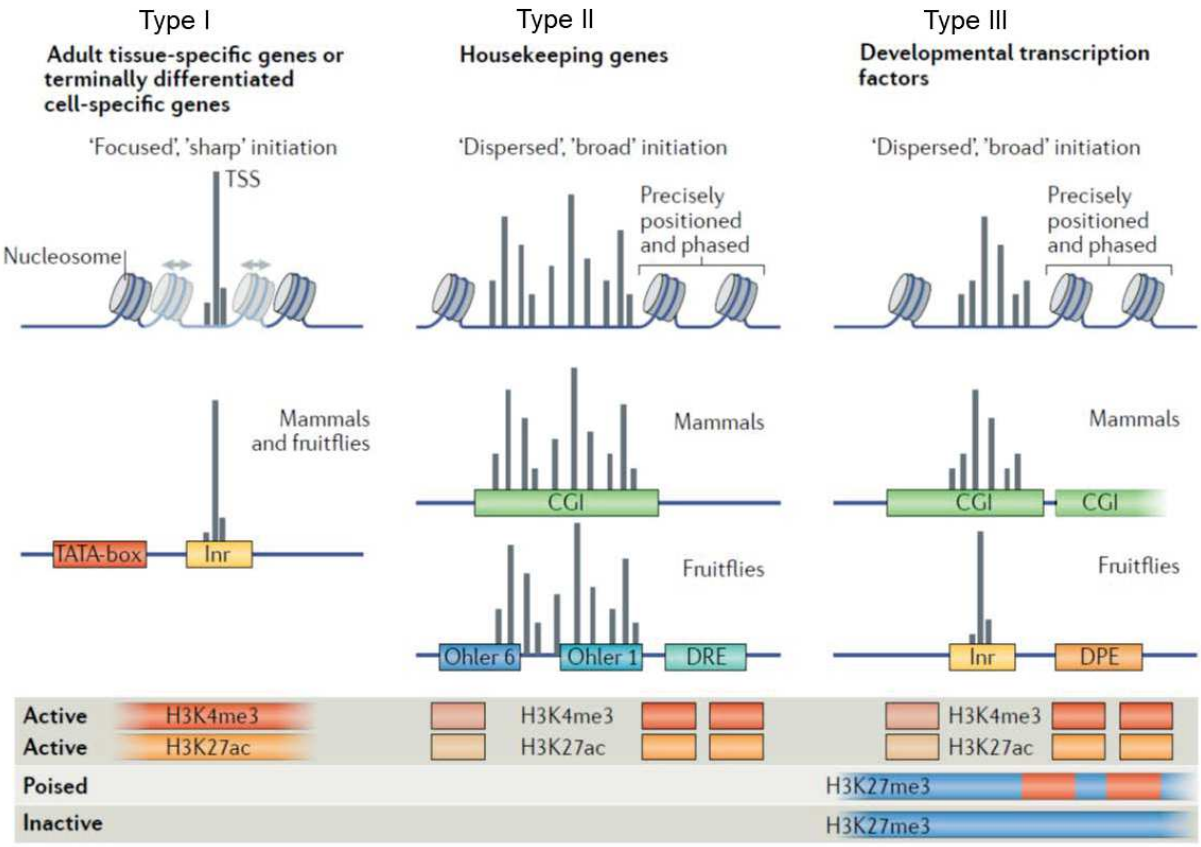


Figure 2-2: Transcription initiation patterns and core promoter types. This model showed three types of core promoters. Adapted from (Haberle and Stark, 2018).

To sum up, the core promoter acts as the minimum docking site for PIC assembly at the TSS. However, the core promoter bound PIC just stimulates a low transcriptional level. To achieve highly active and precisely regulated gene transcription, other *cis*-DNA elements, including proximal promoter, enhancer, silence, and insular, are required. These elements can act as a platform for the binding of DNA-associated transcription factors that further regulate the transcriptional process.

1.2.2 Proximal promoter

Proximal promoter is a transcription-activating sequence located up to 250 bp upstream of the TSS. It contains several binding sites for sequence-specific transcription factors (Haberle and Stark, 2018). Interestingly, the upstream antisense RNA (uaRNA) is generated at the proximal

promoter regions and is associated with genes related to transcriptional regulation during developmental process (Lepoivre et al., 2013; Sigova et al., 2013). Moreover, the CpG islands located at the proximal promoter region is another important factor for transcriptional regulation (Stefansson et al., 2017).

1.2.3 Enhancer

In contrast to proximal promoter, enhancer is a distal DNA element that regulates transcription in a distance- and orientation-independent manner (Banerji et al., 1981; Levine, 2010). Enhancers are reported to be the most dynamically utilized part of the genome (Consortium, 2012). One prominent feature of enhancer is that it contains a cluster of TF binding sites for regulating cell type-specific or condition-specific gene expression (Spitz and Furlong, 2012). Enhancers also have certain chromatin features, including H3.3/H2A.Z-containing nucleosomes, histone H3K4me1 and H3K27ac modifications.

1.2.3.1 Enhancer-associated chromatin

H3.3/H2A.Z-containing nucleosomes play a critical role in maintaining the accessible chromatin structure in enhancer regions. It has been reported that the occupancy of TFs at enhancers is accompanied with regions of nucleosomal depletion (Gross and Garrard, 1988). These regions are associated with nucleosomes containing histone variants H3.3 and H2A.Z that are hyperdynamic and are easily to be displaced from DNA than the canonical nucleosomes (Goldberg et al., 2010; Henikoff et al., 2009; Jin and Felsenfeld, 2007). In contrast to TFs binding regions, the nucleosomes directly flanking TF binding regions are less mobile and decorated with specific histone modifications, including, but not limited to, H3K4me1 and H3K27ac (Creyghton et al., 2010; Heintzman et al., 2007; Zentner et al., 2011).

H3K4me1 is the first histone modification found to be globally linked to enhancers through genomic studies (Heintzman et al., 2007). However, that presence of H3K4me1 is not unique to enhancers, as it is also detected at parts of actively transcribed genes and noncoding sequences. In addition, the presence of H3K4me1 often precedes nucleosomal depletion and H3K27 acetylation, which suggests that this modification exists before enhancer activation and might promote

enhancer activation by maintaining nucleosomal mobility or binding of pioneer TFs (Creyghton et al., 2010; Heintzman et al., 2007; Rada-Iglesias et al., 2011; Zentner et al., 2011). Moreover, unlike H3K27ac, H3K4me1 is not tightly linked to enhancer activity. As H3K4me1 appears to persist binding at enhancers after loss of enhancer activation potential (Bogdanovic et al., 2012; Bonn et al., 2012).

Based on these chromatin features, enhancer has three distinct states, including "primed", "activated" and "poised" (Figure 2-3). Before activation, enhancer exists in a primed state that is characterized by the presence of histone variants H3.3 and H2A.Z, H3K4me1, pioneer TFs, and DNA 5mC hypomethylation and hydroxylation (Calo and Wysocka, 2013). Upon activation, H3K4me1 and H3K27ac mark the chromatin landscape of active enhancers that are bound by GTFs and Pol II. This process leads to the production of enhancer-originating RNAs termed eRNAs (Natoli and Andrau, 2012). Similar to the proximal promoter produced uaRNA, the function of eRNAs await further investigation (Natoli and Andrau, 2012). In addition, these poised enhancers tend to locate near key early developmental genes and share most of the properties of active enhancers, such as nucleosomal depletion and H3K4me1. Conversely, they are marked with H3K27me3 and are bound by the Polycomb complex PRC2, but lack of H3K27ac (Rada-Iglesias et al., 2011; Zentner et al., 2011). Even though poised enhancers are unable to drive gene expression in pluripotent cells (Rada-Iglesias et al., 2011), they are already looped to their target promoters in human ESCs (Calo and Wysocka, 2013; Sanyal et al., 2012).

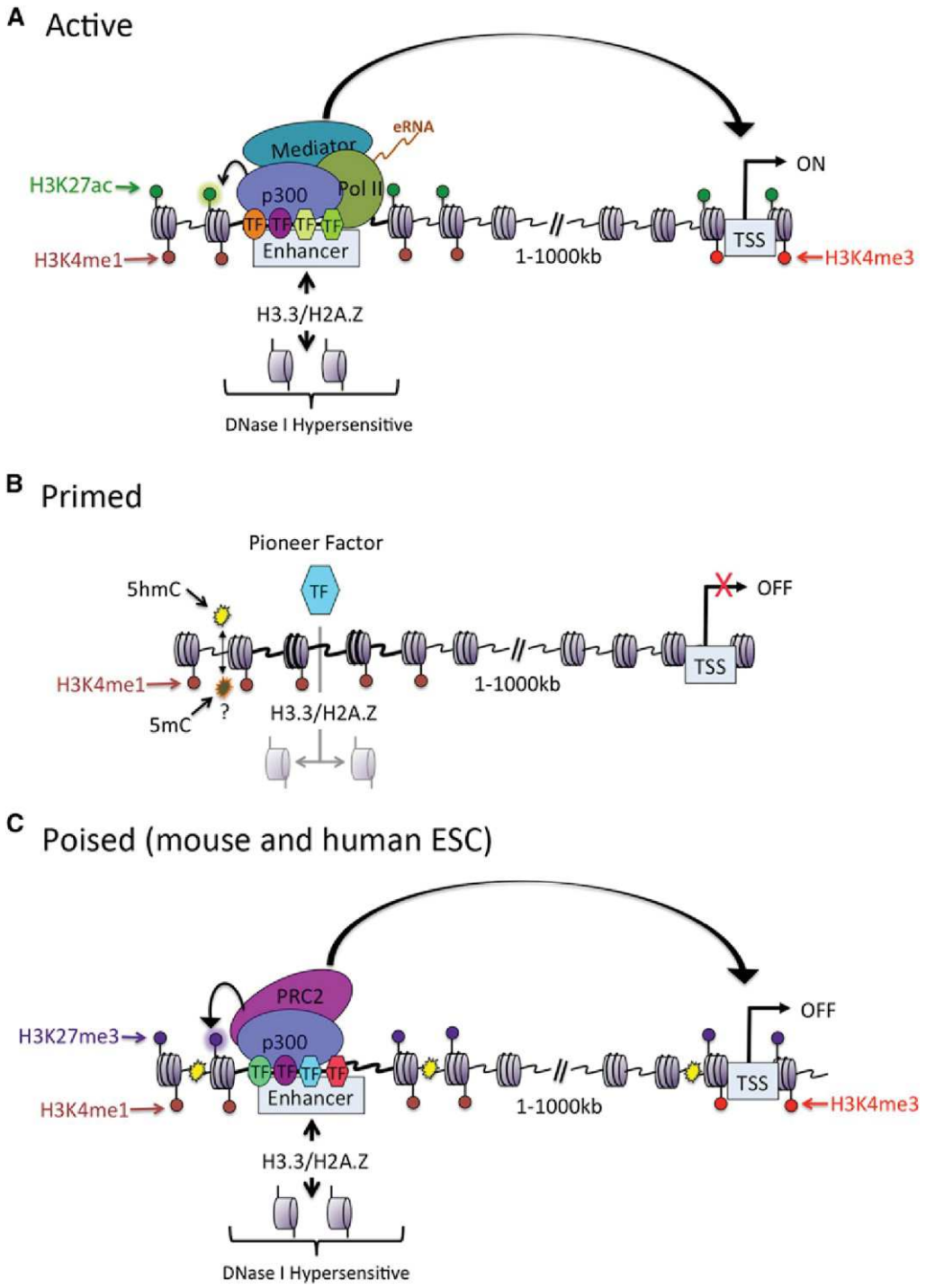


Figure 2-3: Epigenetic features of active, primed, and poised enhancers (Calo and Wysocka, 2013). (A) Schematic representation of the major chromatin features found at active enhancers. (B) Before activation, enhancers can exist in a primed state. (C) Schematic representation of the chromatin landscape surrounding poised enhancers found in human and mouse ESCs.

1.2.3.2 Enhancer and promoter communication

The capacity of enhancer to regulate transcription is independent of distance and orientation. However, how remote enhancers express regulatory information to their target promoters? Various models for enhancer-promoter communication have been proposed, including tracking, looping, linking, and tracking-looping models (Figure 2-4).

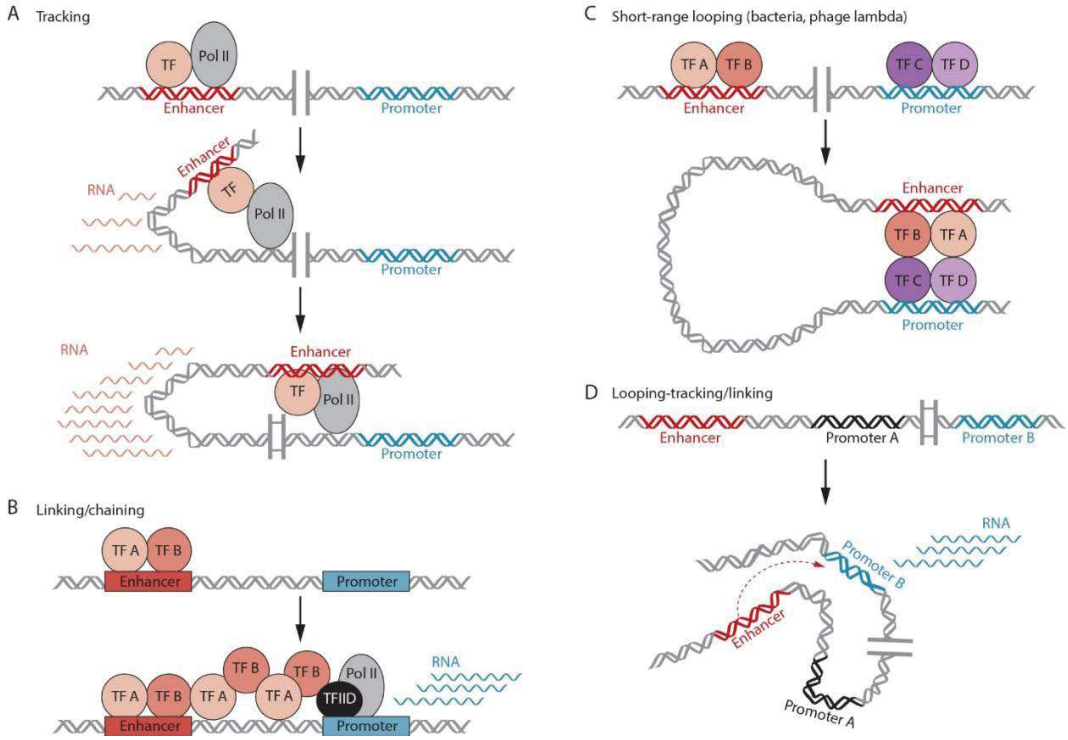


Figure 2-4: Models of enhancer-promoter communication. (A) Pol II binds to an enhancer and tracks along chromatin to associate with the enhancer. (B) TFs bind to the enhancer region and elongate to the promoter. (C) Looping (in bacteria, lambda) requires protein-protein interactions between factors on the same face of the helix. (D) Long-range loops can bring enhancers close to a promoter, but not in direct proximity. Tracking or linking could bridge the distance (Furlong and Levine, 2018).

The tracking model suggests that the upstream regulatory elements bound by Pol II (or another factor) can move along the DNA, ultimately pulling the enhancer to contact with a proximal promoter (Hatzis and Talianidis, 2002; Kong et al., 1997). While the looping model proposes that the protein factors deposited at promoter and enhancer could physically interact with each other,

resulting in extrusion of the intervening DNA. In this case, the intervening DNA is passive during the formation of loops (Furlong and Levine, 2018; Vernimmen and Bickmore, 2015). Besides, the linking model proposes that protein-protein oligomers can bridge the distal enhancer and the target promoter, when the distance between enhancers to their target promoter is at short range (Bulger and Groudine, 1999; Vernimmen and Bickmore, 2015). The looping-tracking/linking model seems to combine the above discussed three models. According to this model, firstly the long-range loop brings enhancers close to a promoter, but not in direct proximity. Subsequently, the remaining distance between the enhancer and promoters is bridged in a tracking or linking manner (Furlong and Levine, 2018). Among these proposed models, the looping model has gained extensive support with the emergence of technologies such as chromosome conformation capture (3C) and its derivatives 4C and 5C (de Wit and de Laat, 2012; Gibcus and Dekker, 2013). Moreover, emerging evidence suggests that phase-separated condensates also play essential role in promoter-enhancer interaction. In this model, the disordered transcription activation region of TFs can recruit Pol II (Chong et al., 2018) and Mediator (Boija et al., 2018), thereby driving the formation of a dynamic "promoter condensates". In turn, this condensate was suggested to support Pol II phosphorylation, PIC assembly and RNA synthesis (Cramer, 2019). Following promoter-proximal pausing, CDK7-mediated CTD phosphorylation counteracts the establishment of "promoter condensates" (Boehning et al., 2018). Whereas, this phosphorylated CTD can be incorporated into the "gene-body condensate" formed by a disordered region in P-TEFb (Lu et al., 2018). When the Pol II reaches the end of the gene, dephosphorylation of Pol II can liberate it from the gene-body condensate (Parua et al., 2018; Proudfoot, 2016).

1.2.4 Silencers

Similar to enhancer, silencer is a *cis*-regulatory DNA element which function in a position- and orientation-independent manner (Ogbourne and Antalis, 1998). It acts as a platform for repressive transcription factors to inactivate gene expression (Gilbert and Muller-Hill, 1966; Ptashne, 1967; Zinn et al., 1983). However, silencer lacks the unique chromatin signature to aid their genome-wide identification.

The mating-type loci study in yeast first identified a distal silencer element named HMRE that could repress non-mating-type gene expression (Brand et al., 1985). Later, a silencer located at

intronic region was reported to control CD4 gene expression during lineage specificity both in human and mouse cells (Donda et al., 1996; Sawada et al., 1994). Moreover, several studies have identified a variety of mammalian silencers in the genomic sequences (Baniahmad et al., 1987; Bergeron et al., 2015). However, the characteristic of silencers is still understudied, possibly due to the poor understanding of those elements with non-promoter locations.

1.2.5 Insulators

Insulator is long-range *cis*-regulatory element that contains a clustered binding sites for sequence specific DNA-binding proteins. This feature of insulator enables it to prevent interactions between adjacent chromatin domains (Yang and Corces, 2011). Thus, insulator can block the inappropriate enhancer-promoter interaction or protect chromatin from the spreading of repressive histone modifications (Dhillon et al., 2009; Gaszner and Felsenfeld, 2006; Huang et al., 2007).

The ability of the Insulator to regulate gene expression depends on recruiting relevant *trans*-acting proteins. The transcriptional repressor CCCTC-binding factor (CTCF) is the main insulator *trans*-acting protein described in vertebrates. This protein contains a highly conserved DNA-binding domain and usually colocalized with cohesin at the intergenic region (Cuddapah et al., 2009), which creates boundaries between topologically associating domains in chromosomes (Ong and Corces, 2014).

1.3 Dynamic regulation of transcriptional states

Cell fate decision is regulated by the complex and precise gene expression, which is central to the developmental process of multicellular organisms (Davidson, 2010). In response to environmental or cellular signals, DNA-binding transcription factors (TFs) interact with enhancers to control the promoter activity in a cell-type-specific manner. The interaction between TFs and the chromatin landscape that they encounter is the central mechanism of transcriptional regulation (Spitz and Furlong, 2012; Voss and Hager, 2014).

Compacted chromatin is supposed to restrict TFs to gain access to DNA-binding sites (Johnson and Dent, 2013). Numerous factors regulate chromatin dynamics, including histone variants, histone chaperones, chromatin remodelers and chromatin epigenetic modifications. To overcome the structural barriers that are intrinsic to nucleosome arrays, TFs must induce the reorganization of local nucleosome structures by cooperating with these chromatin related components in the spatially organized genome (Figure 3-1). In brief, control of transcription programs is mediated by three major mechanisms. The first one is gene regulation by higher-order chromatin organization (Dekker, 2008; Fraser and Bickmore, 2007) (discussed in section 1.3.1). The second regulatory mechanism involves the "histone code" that modulates the cell fate decision (Mohn and Schubeler, 2009) (discussed in section 1.3.2 and 1.3.3). The third major mechanism is based on TFs that occupancy at specific sequence motifs to regulate particular sets of genes (Welstead et al., 2008) (discussed in section 1.3.4).

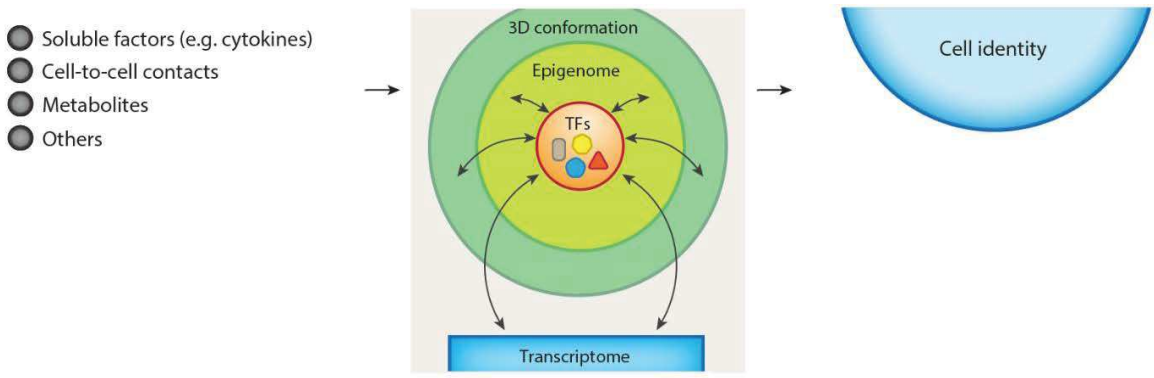


Figure 3-1: Signal transduction modulates the activity of TFs in a cell-specific manner. External cues activate TFs to interact with chromatin landscape within the nucleus. Through the recruitment of

epigenetic modifiers and the transcriptional machinery, TFs regulates the cell's gene expression program transcriptome. Ultimately, the interplay between these nuclear components, orchestrated by transcription factors, results in the adoption of a specific cellular identity.(Stadhouders et al., 2019)

1.3.1 The hierarchically organized chromatin

In eukaryotes, genomes are more than linear sequences. Actually, the DNA is hierarchically packaged insides the nucleus. The well-organized chromatin structure includes chromosome territories (CTs), interchromatin compartments (ICs) and topologically associating domains (TADs), that are essential for transcriptional regulation and genome maintenance (Erkek et al., 2013; Zheng and Xie, 2019).

1.3.1.1 The spatial genome structure

The territorial organization of interphase chromosome is a basic feature of the three-dimensional nuclear architecture (Cremer and Cremer, 2001). The discovery of the phenomenon that chromosomes decondensate on the exit of mitosis and subsequently form confined nuclear territories at interphase, have triggered a huge amount of innovative scientific inquiry (Cremer and Cremer, 2001; Cremer et al., 1993). Chromosomes occupy distinct territories in the cell nucleus. These separated territories organize chromosomes into two interchromosomal contact hubs: gene-dense segments of active (euchromatic) chromatin and Po II-depleted inactive (heterochromatic) chromatin (Quinodoz et al., 2018). The active chromatin tends to associate with Pol II clustering and locate around the nuclear speckle, whereas the inactive chromatin which usually contains centromeric chromatin and the genes coding for ribosomal RNA, resides near the nucleolus (Quinodoz et al., 2018). Moreover, upon inducing transcription, the chromatin often loops out of its chromosome territory and intermingles with the neighboring chromosome territory, resulting in potentially functional interchromosomal interactions (Branco and Pombo, 2006; Chambeyron and Bickmore, 2004; Volpi et al., 2000). Together, these observations revealed a functional association between the 3D genome architecture and gene expression.

With the help of Hi-C technology, further analysis point out two major levels of topological organization in the genome (Cavalli and Misteli, 2013; Denker and de Laat, 2016). At the megabase scale, the first level segregates the genome into two subnuclear compartments: the A

compartment that corresponds to active chromatin, and the B compartment that represents inactive chromatin. These two compartments are characterized according to the spatial segregation of open and closed chromatin (Lieberman-Aiden et al., 2009). Similar to the territories discussed above, these compartments are also specifically associated with various nuclear structures. For example, the compartment A tends to occupy at the nuclear interior region and accompanies with active histone modifications, while compartment B is preferentially associated with either the nuclear lamina (van Steensel and Belmont, 2017) or the nucleolus (Bickmore and van Steensel, 2013) (Figure 3-2).

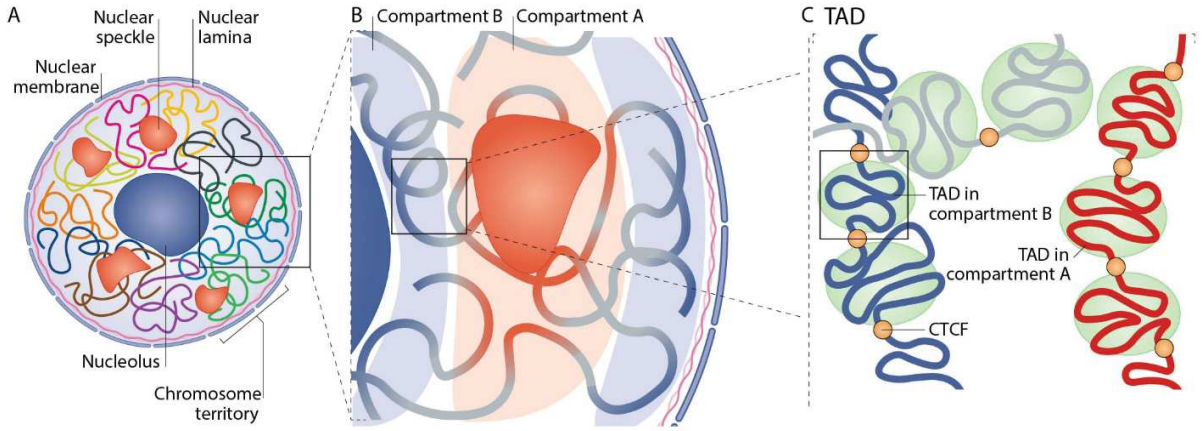


Figure 3-2: The hierarchical organization of the 3D genome. A. Individual chromosomes (indicates by different colors) occupy separate territories in the interphase nucleus. B. At a smaller scale, transcriptionally active regions prefer to interact with other active regions to form compartment A. Inactive regions tend to interact with other inactive regions to form compartment B. C. Locally genomic domains that forms the TADs. Adapted from (Zheng and Xie, 2019).

Secondly, the sub mega base level of genomic structure consists of TADs and chromatin loops (Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014; Sexton et al., 2012). TADs play a role in restricting the nuclear search space, as many promoters and enhancers predominantly communicate within individual TADs (Dowen et al., 2014). Besides, the boundary of TADs are typically enriched with CTCF and cohesin complex in mammalian cells. This finding gives rise to the loop extrusion model. In this model, loop-extruding factors (likely cohesins) engage the chromatin to initiate extrusion of a chromatin loop, until they are stalled at the extrusion boundaries due to interactions with boundary proteins (like CTCF) (Alipour and Marko, 2012; Fudenberg et

al., 2016; Sanborn et al., 2015). Notably, several studies have shown that TADs can compete with compartments. For example, the deletion of chromatin-associated cohesin not only decreases TADs formation but also increases compartmentalization (Haarhuis et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017). These observations indicate that cohesin alters the compartmentalization by regulating the TAD states. It is worth to note that the removal of CTCF or cohesin from chromatin only affects the expression of ~1000 genes. This unexpected observation suggests TAD boundaries either fine-tune the cellular transcriptome or play a role in regulating only a subset of genes (Nora et al., 2017; Schwarzer et al., 2017).

1.3.1.2 The links between genome conformation and cell fate decision

The TADs are found to correspond to lamin-associated chromatin domains (LADs) in nuclei (Guelen et al., 2008). Nuclear lamina (NL) modulates the position of chromosomes by interacting with DNA and many different proteins, such as heterochromatin protein 1 (HP1) and histones (Prokocimer et al., 2009; Zuleger et al., 2013). Moreover, nuclear architecture regulates gene-expression programs during cell-fate specification. For example, the genome NL interactions can regulate cardiac stem cell lineage restriction (Poleshko et al., 2017). Moreover, the genome regions between compartments A and B also switch with each other during cellular differentiation or reprogramming (Bonev et al., 2017; Dixon et al., 2015; Stadhouders et al., 2018). Together, these studies suggest that the 3D genome structure provides a distinct layer of gene regulation during cell fate decision.

1.3.1.3 A phase separation model for transcriptional control

Phase separation is implicated in proteins that contain intrinsically disordered regions (IDRs). These IDRs are classified by their low complex amino acid profile, such as acidic, proline, serine/threonine, or glutamine rich (van der Lee et al., 2014). They generally not amenable to crystallography due to lacking bulky hydrophobic amino acids (Uversky, 2002; Uversky et al., 2000). However, the IDR can self-organize into liquid-like droplets that act as a membrane-less organelle (Hnisz et al., 2017; Hyman et al., 2014; Shin et al., 2019).

The process of phase separation provides a plausible mechanism for intra- and interchromosomal compartmentalization. IDRs from various nuclear proteins, including RNAPII, Mediator, HP1, polycomb, cyclin T1, bromodomain-containing protein 4 (BRD4) and various TFs, can phase separate into liquid condensates (Boehning et al., 2018; Boija et al., 2018; Cho et al., 2018; Chong et al., 2018; Larson et al., 2017; Lu et al., 2018; Schoenfelder et al., 2010). In addition, many of these proteins also possess targeting ‘‘reader’’ motifs, such as the bromodomain of BRD4 can target the phase separation-prone protein and drive it to histones exhibiting acetylated lysine resides (Dey et al., 2003).

Moreover, phase separation plays essential roles in transcriptional regulation during differentiation process. In this process, many TFs that contain disordered protein regions at the activation domains, can form condensates with the transcriptional co-activator Mediator or Pol II (Chong et al., 2018). For example, either OCT4 or GCN4 can form phase-separated droplet with Mediator, which regulates the expression of genes in a IDRs dependent manner. Moreover, the size and the number of condensates are decreased upon mESC differentiating into epiblast like cells (EpiLCs) (Boija et al., 2018). These results suggest that the phase separation directed condensates might be cell type specific.

1.3.2 Nucleosome structure and variability

In the nucleus, nucleosome consists of approximately 146-base pair genomic DNA wrapped around the lateral surface of an octamer comprising histone proteins H2A, H2B, H3 and H4 (Kornberg, 1974). Histone H1 that binds to the outside of the octamer was suggested to stabilize the higher-order chromatin structures (Luger et al., 1997; Szerlong and Hansen, 2011). In addition, each histone protein had a histone fold domain that allows for heterodimerization (H2A with H2B and H3 with H4) (Luger et al., 1997) (Figure 3-3).

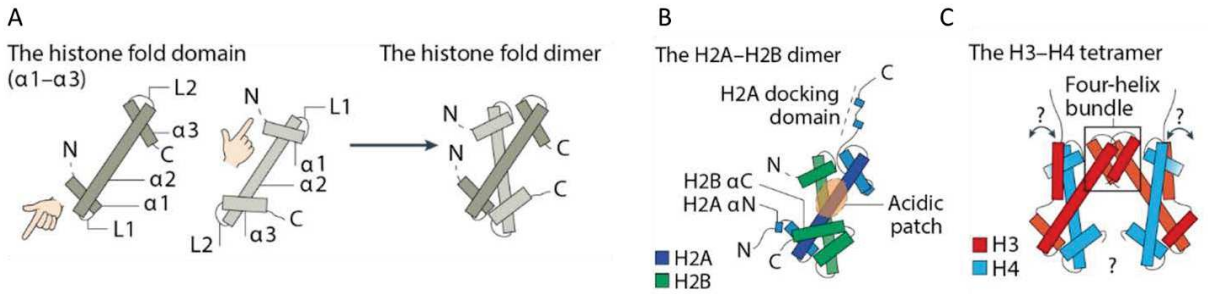


Figure 3-3: Nucleosome assembly intermediates. (A) The structure of histone fold domains consisting of three helices (α1–α3) linked together by short loops (L1 and L2). (B) The structure of H2A-H2B dimer. H2A has a carboxy-terminal extension and a short amino-terminal helix (H2A αN); H2B contains a C-terminal α-helix (H2B αC). (C) The structure of H3-H4 tetramer containing a Four-helix bundle domain. (Hammond et al., 2017)

The nucleosome is very dynamic and undergoes assembly and disassembly cycles during transcription process. It has been reported that the occupancy of H3–H4 tetramer-initiated nucleosome assemble process. Subsequently, two H2A–H2B dimers wrapped the remaining DNA and each H2A–H2B dimer associated with the H3–H4 tetramer via the four-helix bundle (Arents et al., 1991; Smith and Stillman, 1991). Conversely, the disassembly of nucleosomes was thought to occur through a reversal of these processes (Mazurkiewicz et al., 2006). The first steps in nucleosome disassembly was opening up the interface between the H2A−H2B dimers and the $(H3-H4)_2$ tetramer, which is followed by removing either one or both of the H2A−H2B dimers (Gansen et al., 2007; Li et al., 2005). Finally, the $(H3-H4)_2$ tetramer complex can be further dissociated from the DNA (Bohm et al., 2011; Tagami et al., 2004).

The arrangement of the core histones within the histone octamer produces a highly contoured and negatively charged binding interface on the nucleosome surface. This nucleosome surface possesses a cluster of eight acidic residues (E56, E61, E64, D90, E91, E92 of H2A and E102, E110 of H2B) that forms a negatively charged 'acidic patch' domain (Kalashnikova et al., 2013). X-ray crystallography studies have found that this acidic patch domain was bound to the basic patch of the H4 N-terminal tail of a neighboring nucleosome (Kalashnikova et al., 2013). This interaction between the acidic patch and H4 tail might promotes the higher order chromatin folding (Kalashnikova et al., 2013). Unexpectedly, the nucleosome acidic patch is necessary for maximum activity of CHD and SWI/SNF family remodellers (Dann et al., 2017). Moreover, modifications that close to the acidic patch domain, such as H2BK120ub, H2BK108ac, H2BK120ac,

H2BS112GlcNAc, H2BK116ac and histone variant H2A.Z, can regulate remodeling activity *in vitro* (Dann et al., 2017). All of these results suggest that the acidic patch domain act as a tunable interaction hotspot for ATP dependent chromatin remodellers and related chromatin effectors.

Aside from the 'canonical' histones discussed above, evolution drove the emergence of histone variants. Eight variants of H2A (H2A.X, H2A.Z.1, H2A.Z.2.1, H2A.Z.2.2, H2A.B, macroH2A1.1, macroH2A1.2 and macroH2A2) and six variants of H3 (H3.3, CENP-A, H3.1T, H3.5, H3.X and H3.Y) have been identified in human cells. Moreover, two testis-specific variants of histone H2B (H2BFWT and TSH2B) are also identified (Buschbeck and Hake, 2017) (Figure 3-4).

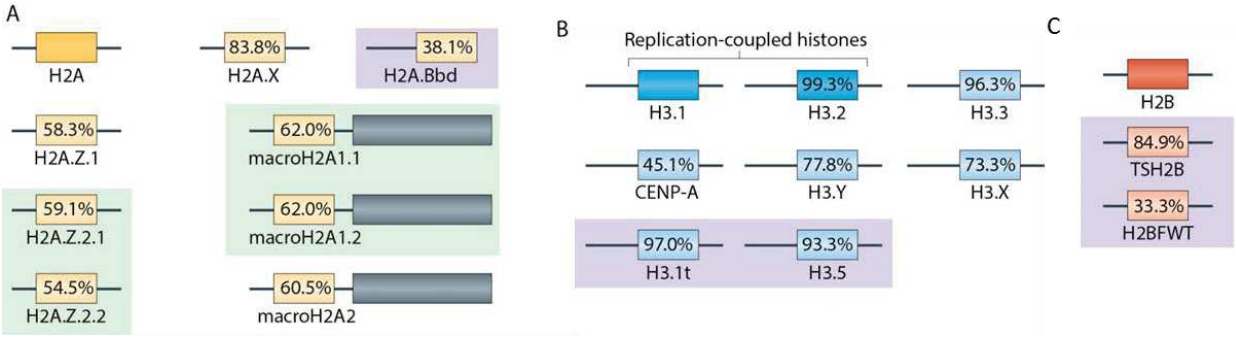


Figure 3-4: A depiction of human histone variants. (A) Variants of histone H2A (yellow) with variants shown in pale yellow. (B) Variants of histone H3 (blue) with variants shown in pale blue. (C) Variants of histone H2B (orange) with variants shown in pale orange. Percentages indicate total amino acid sequence conservation (% sequence identity) of the variants relative to their canonical histone counterparts (for H3, two replication-coupled isoforms are present (H3.1 and H3.2). Adapted from (Buschbeck and Hake, 2017).

There are some differences between canonical histones and histone variants. For example, genes encoding canonical histones usually lack introns. Moreover, their mRNAs are not polyadenylated but instead have a unique 3' stem-loop structure that regulates mRNA stability and translation (Dominski and Marzluff, 1999; Pandey et al., 1990). In contrast to canonical histones, histone variant coding genes contain introns and are polyadenylated (Marzluff et al., 2002). This specific transcription pattern of canonical histones might enable their quickly protein synthesis. Consequently, most canonical histones can assemble into nucleosomes behind the replication fork to timely package newly synthesized DNA. In contrast to canonical histones, histone variants are incorporated into nucleosomes throughout the cell cycle (Ahmad and Henikoff, 2002).

Replacement of canonical histones with histone variants adds a distinct way of modulating chromatin function. It has been reported that the structural difference introduced by a histone variant can affect the accessibility of chromatin. For example, histone variants H2A.Z and H3.3 are mainly linked with an open chromatin conformation, whereas macroH2A tends to associate with a repressive chromatin state (Biterge and Schneider, 2014; Chakravarthy et al., 2005; Thakar et al., 2009). These distinct chromatin properties in turn affect many cellular processes, such as DNA replication, repair and transcription.

Crystal structure study found that H2A.Z-containing nucleosomes displayed an extended acidic patch on their surface than the core histone H2A, which caused slightly destabilizing the interaction between the H2A.Z–H2B dimer and with the H3–H4 tetramer (Suto et al., 2000; Thakar et al., 2009). In line with this, H2A.Z is enriched at the nucleosome-depleted region (NDR) of active transcriptional start sites (TSS) (Nekrasov et al., 2012). Therefore, H2A.Z is suggested to be necessary for the binding of the transcriptional machinery by facilitating to establish NDR. On the other hand, the capacity of H2A.Z to regulate chromatin dynamic is also dependent on H2A.Z posttranscriptional modifications. For example, five lysine residues (K4, K7, K11, K13 and K15) of H2A.Z are potential acetylation sites (Bonenfant et al., 2006; Ku et al., 2012). Generally, the acetylation of H2A.Z destabilizes the nucleosome and in turn more competent to recruit the transcriptional machinery (Bruce et al., 2005; Ishibashi et al., 2009). In contrast to these acetylated residues, three lysine residues (K120, K121, or K125) of H2A.Z can be ubiquitinated (Ku et al., 2012; Sarcinella et al., 2007). However, H2A.Zub is mainly localized at heterochromatin regions and associated with repressed transcription (Sarcinella et al., 2007). Therefore, H2A.Z posttranscriptional modification can greatly change the effect of H2A.Z on chromosome.

Another histone H2A variant, H2A.X, plays an important role in DNA double-strand break repair. Upon damage, DNA-dependent protein kinases (ATR/ATM) phosphorylates H2A.X at serine 139, forming a γH2A.X foci (Rogakou et al., 1998). These foci facilitate the recruitment of the damage repair proteins (NBS1, 53BP1, MDC1 and BRCA1), as well as chromatin remodelers (INO80 and SWR1) (Celeste et al., 2002; van Attikum et al., 2007; Wu et al., 2011). In addition, H2A.X ubiquitination might crosstalk with H2A.X phosphorylation. For example, during DNA damage,

PRC1 complex can ubiquitinate H2A.X at K119 (H2A.XK119ub). Then, the H2A.XK119ub recruits the ATM kinase towards the damaged site, which allows rapid γH2AX formation (Bergink et al., 2006; Wu et al., 2011).

Genomic studies have revealed that H3.3 is enriched at enhancers and active gene bodies, as well as repeat regions such as telomeres and regions adjacent to centromeres (Goldberg et al., 2010; Shi et al., 2017; Wong et al., 2010). Moreover, H3.3 is colocalized with H2A.Z at the nucleosome-depleted regions and marks the promoter of active gene (Jin et al., 2009). Recent studies have showed that H3.3 contains a specific serine residue (Ser31) that is not present at H3.1 and H3.2. Phosphorylated H3.3S31 (H3.3S31P) is initiated identified as a mitosis-specific modification which is present only in late prometaphase and metaphase (Hake et al., 2005; Wong et al., 2009). Later study shows that H3.3S31P promotes new enhancers formation during differentiation by stimulating p300 histone acetyltransferase activity (Martire et al., 2019). Moreover, in mouse macrophages, H3.3S31P enables rapid stimulation-induced transcription through recruiting the active transcription related histone methyltransferase SETD2 and ejecting the elongation corepressor ZMYND11 (Armache et al., 2020). Therefore, these observations indicate that H3.3S31P plays an essential role in regulating gene expression and cell fate decision.

In conclusion, nucleosome structures are dynamic during transcription (Erdel et al., 2011). Partial histone disassembly or integration with histone variants participate in nucleosome reorganization process (Cairns, 2007; Clapier and Cairns, 2009; Glatt et al., 2011). Moreover, histone chaperones and chromatin remodelers also modulate histone exchange.

1.3.2.1 Histone chaperones

Histone chaperones are broadly defined as histone-interacting proteins that are involved in histone storage, transport, nucleosome assembly and disassembly. Histones do not have the intrinsic ability to form nucleosomes; rather they tend to randomly associate with DNA and form aggregates (D'Arcy et al., 2013). To avoid spurious interaction with DNA, the free histone oligomer is stabilized via binding to histone chaperones (Hondele et al., 2013; Luk et al., 2007). Thereby, histone chaperones play essential roles in nucleosomes reconstruction associated events, such as DNA replication, repair and transcription processes (Adam et al., 2015; Alabert and Groth, 2012;

Venkatesh and Workman, 2015). Table 4 showed several specific chaperones and their roles in histone exchange. In addition to regulate chromatin exchange, histone chaperones also promote PTMs at the globular domain of histones where are normally inaccessible for enzymes. This chaperone-aided PTMs can either activate or repress transcription. For example, Rtt109 and Asf1-dependent H3K56 acetylation enhances transcription (Williams et al., 2008), whereas Spt6-assisted H3K36me3 tends to restrict transcription initiation (Carrozza et al., 2005; Kaplan et al., 2003).

Table 4 | Histone chaperones involved in transcription-associated exchange: their targets, modulators and functions. Adapted from (Venkatesh and Workman, 2015)

Histone chaperone*	Histone or variant bound	Factors modulating exchange	Function
HIRA (humans)	H3.3	H4S47 phosphorylation by PAK2	Promotes increased association of H3.3 with HIRA, disfavouring its interaction with CAF1, thereby facilitating H3.3 exchange
NAP1 (humans and yeast)	H2A–H2B or H2A.Z–H2B	RSC remodelling complex	Provides H2A.Z to SWR and stabilizes the hexameric nucleosome during elongation by accepting the H2A–H2B dimer
Chz1 (yeast)	H2A.Z–H2B	Unknown	Provides H2A.Z to SWR
FACT (humans and yeast)	H2A–H2B	H2BK120 ubiquitylation by Rad6–Bre1 and H3K36 methylation by Set2	Facilitates transcription elongation by removing the ubiquitylated H2A–H2B dimers and resetting of chromatin by replacing histones after passage of Pol II. Methylation of H3.3K36 limits continuous H3 exchange
Spt6 (yeast)	H3–H4	H3K36 methylation by Set2	Involved in resetting of chromatin after the passage of Pol II, thereby preventing the initiation of aberrant transcription; methylation of H3K36 limits continuous H3 exchange
Asf1 (yeast)	H3–H4	H3K56 acetylation by Rtt109 and Swi/Snf remodelling complex, and H3K36 methylation by Set2	Facilitates the removal of histones from the promoter and gene bodies to aid transcription; methylation of H3K36 limits continuous H3 exchange
Rtt106 (yeast)	H3–H4	H3K56 acetylation by Rtt109	Rtt106 facilitates resetting of chromatin and prevents the initiation of aberrant transcription
ANP32E (humans)	H2A.Z–H2B	Unknown	Removes H2A.Z from the nucleosome

1.3.3 Histone post-translational modifications (PTMs)

PTMs are another important factor to regulate the architecture of chromatin. Histones have two structurally and functionally distinct domains: the globular domain that forms the nucleosomal core, and the unstructured N-terminal tail domain. Both of them can sever as a platform for various PTMs, such as acetylation, phosphorylation, methylation and ubiquitylation (Figure 3-5 and Table 5). Except for methylation, histone modifications can alter the net charge of nucleosomes, which affects the chromosomal accessibility. In line with this, it has been reported that acetylated histones

are easier to displace from chromatins (Chandy et al., 2006; Reinke and Horz, 2003; Shogren-Knaak et al., 2006).

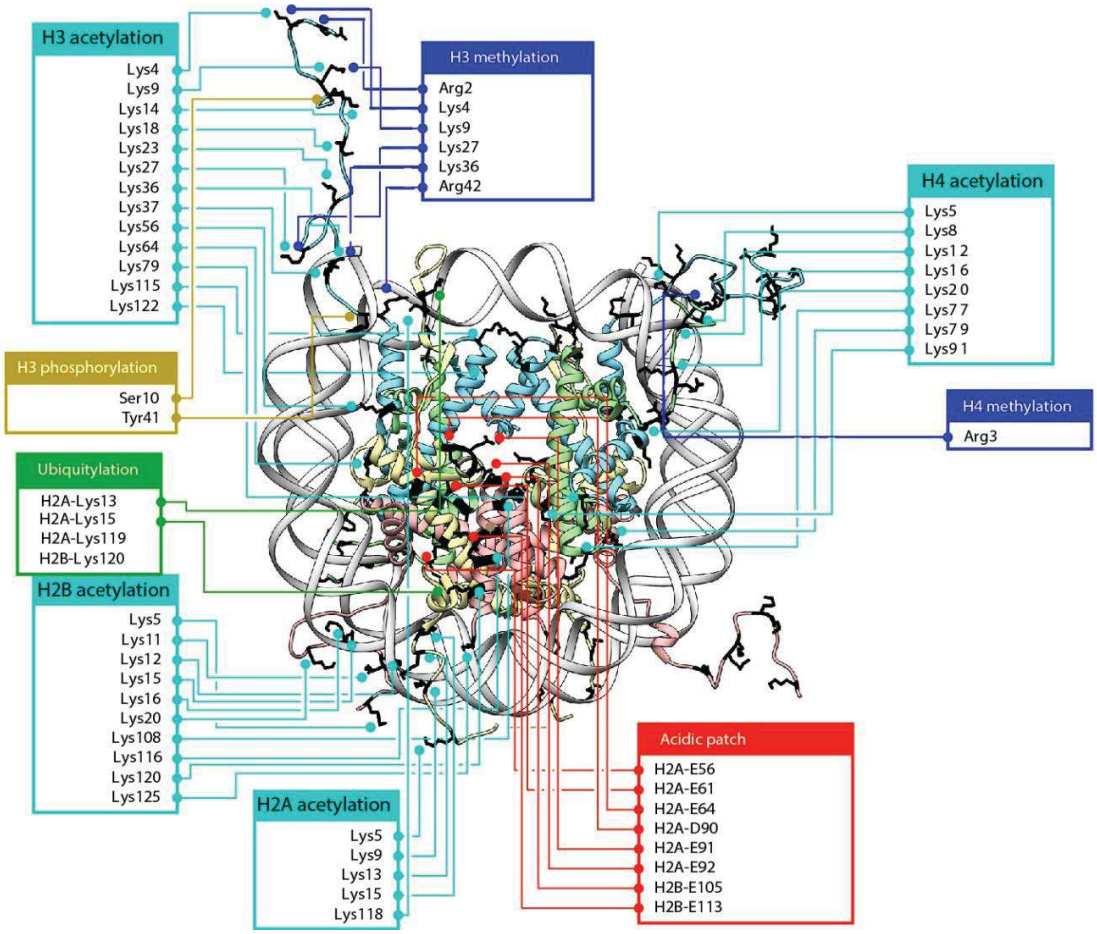


Figure 3-5: Major sites of histone for PTMs and the acidic patch related residues. Diagram showed modifications on histone H2A (light yellow), H2B (light red), H3 (light blue) and H4 (light green). Adapted from (Dann et al., 2017)

Interestingly, PTMs provide the binding sites for many specific protein motifs. For example, bromodomains interacts with acetylated histones. The plant homeodomain (PHD) domain, Tudor domain, and chromo domain selectively bind methylated histones (Smith and Shilatifard, 2010; Yun et al., 2011). Therefore, histone modifications can recruit effector proteins, such as chromatin modifying enzymes, chromatin remoderllers and TFs, which have corresponded binding domains (Clements et al., 2003; Lee et al., 2007; Vettese-Dadey et al., 1996).

Table 5 | Histone modifications involved in transcription regulation. (Li et al., 2007)

Modifications	Position		Enzymes				Recognition Module(s) [a]	Functions in Transcription
			S. cerevisiae	S. pombe	Drosophila	Mammals		
Methylation	H3	K4	Set1	Set1	Trx, Ash1	MLL, ALL-1, Set9/7, ALR-1/2, ALR, Set1	PHD, Chromo, WD-40	Activation
		K9	n/a	Clr4	Su(var)3-9, Ash1	Suv39h, G9a, Eu-HMTase I, ESET, SETBD1	Chromo (HP1)	Repression, activation
		K27				E(Z)	Ezh2, G9a	Repression
		K36	Set2			HYPB, Smyd2, NSD1	Chromo(Eaf3), JMJD	Recruiting the Rpd3S to repress internal initiation
		K79	Dot1			Dot1L	Tudor	Activation
	H4	K20		Set9	PR-Set7, Ash1	PR-Set7, SET8	Tudor	Silencing
Arg Methylation	H3	R2				CARM1		Activation
		R17				CARM1		Activation
		R26				CARM1		Activation
	H4	R3				PRMT1	(p300)	Activation
Phosphorylation	H3	S10	Snf1				(Gcn5)	Activation
Ubiquitination	H2B	K120/123	Rad6, Bre1	Rad6		UbcH6, RNF20/40	(COMPASS)	Activation
	H2A	K119				hPRC1L		Repression
Acetylation	H3	K56					(Swi/Snf)	Activation
	H4	K16	Sas2, NuA4		dMOF	hMOF	Bromodomain	Activation
	Htz1	K14	NuA4, SAGA					Activation

[a] The proteins that are indicated within the parentheses are shown to recognize the corresponding modifications but specific domains have yet to be determined.

PTMs have been closely linked with transcriptional process. Depending on their effect on transcription, PTMs are classified as activating or repressing marks (Smolle and Workman, 2013). Actively transcribed genes are usually associated with high levels of histone H3/H4 acetylation as well as H3K4me3, H3K36me3, H3K79me3 and H2BK120ub1, whereas H3K9me3, H3K27me3 and H2AK119ub1 modifications are often deposited on inactive genes or regions (Zhang et al., 2015). The landscape of histone marks are established through a dynamic interplay between histone readers, writers, and erasers. Below, we discuss some of the better-described histone modifications.

1.3.3.1 Histone H3K4 methylation

H3K4me3 is detected at the promoter and TSS regions of active genes (Piunti and Shilatifard, 2016). In mamlian cells, H3K4me3 is deposited by COMPASS-like complexes that contain six related homologs of the yeast SET1 (SETD1A, SETD1B, MLL1, MLL2, MLL3, and MLL4) (Sims et al., 2003). ESCs that lose the core components of COMPASS-like complexes, results in a range of phenotypes, including reduced self-renewal and impaired differentiation (Ang et al., 2011; Jiang et al., 2011). Mice embryos deletion of the core component of COMPASS-like complexes *Ash2l* can survive pre-implantation stage but die after gastrulation (Bertero et al., 2015; Stoller et al., 2010). Overall these observations suggest that histone methyltransferase complex are important for ESC differntiation and embryonic development.

However, the roles of H3K4me3 in gene expression are still not clear. Biochemically, H3K4me3 facilitates assembly of the transcriptional machinery and mediates more efficient induction of gene expression in response to environmental cues. For example, SAGA complex binds to H3K4me3 via a double Tudor-domain in the C terminus of Sgf29 (Vermeulen et al., 2010). The TFIID directly interacts with the H3K4me3 mark via the PHD finger of TAF3 (Vermeulen et al., 2007). However, in zebrafish, the deposition of H3K4me3 does not predict the level of gene expression. Conversely, it might even mark a subset of inactive genes during the maternal-to-zygotic transition (Vastenhouw et al., 2010). Moreover, in mESCs, the chromatin of bivalent genes possess both activating (H3K4me3) and repressive (H3K27me3) markers, which poise the silent development genes for rapid activation upon differentiation (Bannister and Kouzarides, 2005). Together, these results suggest that the deposition of H3K4me3 might only fine-turning the transcriptional activity although there is a general correlation between H3K4me3 and gene expression (Howe et al., 2017).

H3K4me1 and H3K4me2 markers are mainly detected at intergenic sites and function as general enhancer marks (Heintzman et al., 2009; Lee et al., 2013; Li et al., 2016b). MLL4 is a major enhancer H3K4me1/2 methyltransferase with functional redundancy with MLL3 (Piunti and Shilatifard, 2016). *Mll4* deletion does not affect the self-renewal of mouse ESCs but strongly suppresses their potential for differentiation (Wang et al., 2016). This suggests that H3K4me1 is not required for maintaining cellular identity under steady-state conditions when enhancers have been established, but that H3K4me1 becomes important for establishing *de novo* lineage-specific

enhancers when cells are triggered for differentiation. Furthermore, the important cellular role of H3K4 methylation might primarily relay on its signaling functions. For example, MLL4 is required for binding of H3K27 acetyltransferase p300 at enhancers, which plays an important role in enhancer activation (Wang et al., 2016).

1.3.3.2 Histone H3K27 acetylation

Histone H3K27 acetylation is predominantly located at promoters and/or enhancers, which assist to distinguish active enhancers from inactive/poised enhancer elements (Creyghton et al., 2010; Rada-Iglesias et al., 2011). The presence of H3K27ac distinguishes active enhancer states from those poised for activation enhancers and primed enhancers. As a consequence, enhancer bound H3K27ac shows a high degree of cell-type specificity (Bonn et al., 2012; Creyghton et al., 2010; Heintzman et al., 2007; Rada-Iglesias et al., 2011; Zentner et al., 2011). In addition, CBP/p300 specifically acetylates H3K27. However, in human ESCs, poised developmental enhancers are bound by p300 but lack H3K27ac (Rada-Iglesias et al., 2011). These results suggest the existence of counteracting mechanisms to prevent the preloaded p300 function, such as turnover by deacetylases, direct inhibition of p300 enzymatic activity or mutually exclusive relationship with H3K27me3 (Jin and Felsenfeld, 2007; Pasini et al., 2010; Tie et al., 2009).

1.3.3.3 Histone H2Bub1

In yeast, histone H2BK123ub1 is deposited by the E3 ubiquitin ligase Bre1 (an orthologue of RNF20/RNF40 proteins in human cells), together with the E2 ubiquitin-conjugating enzyme Rad6 and the E1 ubiquitin-activating enzyme Uba1 (Hwang et al., 2003; Kim et al., 2009). Generally, the transcription elongation factor PAF complex promotes Rad6 to deposit H2Bub1 at actively transcribed regions (Wood et al., 2003a; Xiao et al., 2005). Genome wide approaches revealed a nonrandom distribution of H2Bub1 within active gene bodies, as H2Bub1 is significantly reduced following the first internal exon (Huff et al., 2010). Biochemical study reveals that the deposition of H2Bub1 is highly sensitive to H2A.Z and H2A modifications. This crosstalk might contribute to the spatial organization of H2Bub1 on gene bodies (Wojcik et al., 2018). Besides, the deubiquitinase module of SAGA can efficiently remove H2Bub1 (Bonnet et al., 2014). More

recently, it was reported that the histone H4 basic patch affects global H2Bub1 levels by regulating the SAGA deubiquitinase activity in yeast (Meriesh et al., 2020).

H2Bub1 modulated specific groups of genes rather than the whole genome, as the depletion of the H2B ubiquitin ligases RNF20 or RNF40 altered the expression of only a subset of genes (Shema et al., 2008; Xie et al., 2017). Moreover, H2Bub1 has been associated with the regulation of inducible genes, such as HOX genes that involved in cell differentiation (Zhu et al., 2005) and relatively long genes induced by retinoic acid (Fuchs et al., 2012). Consequently, H2Bub1 was suggested to play an important role in ESC differentiation process (Fuchs et al., 2012; Karpiuk et al., 2012). These observations raised the possibility that H2Bub1 may primarily regulate inducible genes, while having no obvious effect on constitutive transcription.

Genome wide approaches have revealed that H2Bub1 regulated transcriptional elongation process. For example, H2Bub1 was supposed to be coupled with the elongation rate of RNA polymerase II (Fuchs et al., 2014; Minsky et al., 2008). This is in line with the preferential deposition of H2Bub1 at the intron 1 region of gene body (Jung et al., 2012), which is also the region where Pol II elongation is still slow and possibly requires the presence of elongation factors and histone marks that could increase elongation efficiency (Danko et al., 2013; Jonkers et al., 2014; Saponaro et al., 2014; Veloso et al., 2014). Besides, H2Bub1 was supposed to facilitate nucleosome reassembly in the wake of elongating Pol II via regulating the localization of Spt16, a subunit of the histone chaperone FACT (Fleming et al., 2008). In addition to influence nucleosome dynamics, H2Bub1 also facilitates di- and tri-methylation of H3K4 and H3K79 through the recruitment of relevant enzymes, Set1 and Dot1 (Lee et al., 2007). Each of these histone modifications has been widely linked to actively transcribed genes by direct recruitment of various chromatin-modifying factors (Ruthenburg et al., 2007). Therefore, H2Bub1 seem to promote efficient transcription elongation by recruiting transcriptional elongation factors and by a crosstalk with other histone modifications.

H2Bub1 was also suggested to regulate promoter and enhancer activities. Even though H2Bub1 within highly active gene bodies promotes transcription elongation, H2Bub1 inhibits the occupancy of Pol II at normally quiescent promoters by assisting nucleosome reassembly in yeast (Batta et al., 2011). In agreement with the repressive role of H2Bub1, a series of biochemical

analyses showed that nucleosome stability is enhanced when H2Bub1 levels increase (Chandrasekharan et al., 2009). This feature of H2Bub1 was also suggested to affect enhancer activity. For example, one study suggested that H2Bub1 inhibits the activity of inducible enhancer by impairing the chromatin access to INO80 which is a chromatin remodeller protein promoting histone H2A.Z eviction (Segala et al., 2016). Together, these results, contrary to the above studies, may suggest that H2Bub1 have a repressive function at the promoter and enhancer regions.

Above all, H2Bub1 regulates transcription at both enhancer, promoter and gene body. However, compared with other histone modifications, such as H3K36me3 and H3K79me2, H2Bub1 is highly dynamic during transcription process (Fuchs et al., 2014). It has been reported that H2Bub1 was erased by the DUBm of SAGA within 10 mins (Bonnet et al., 2014). However, as described above the function of this dynamic feature of H2Bub1 is still unclear. Future studies addressing how H2B deubuiquitination influences transcription will be important for understanding the role of H2B dynamic.

1.3.4 Transcription factors (TFs)

TFs occupied at specific DNA sequence motifs to regulate particular sets of genes, which is a major mechanism for cell fate decision (Welstead et al., 2008). TFs typically recognize 6-12 bp degenerated DNA sequence at promoter-proximal and/or enhancer regions (Koster et al., 2015). It contains DNA-binding domain and activation domain. TFs are grouped into classes based on their DNA-binding domains that can attach to a specific sequence of DNA (Mitchell and Tjian, 1989; Ptashne and Gann, 1997). Besides, considerable activation domains of TFs usually have an intrinsically disordered regions (IDRs). Recent studies found that this region enables transcription factor to form phase-separated condensates with Mediator complex at super enhancer regions (Boija et al., 2018). For example, the OCT4 transcription factor can form phase-separated droplets with Mediator *in vitro* and activate genes *in vivo,* which are dependent on the same amino acid residues (Boija et al., 2018). These results suggest that the IDR-mediated phase separation with activator domains is a mechanism by which TFs activates gene transcription.

Multiple factors were suggested to facilitate TFs to overcome the nucleosomal barriers (Bossard and Zaret, 2000; Cirillo et al., 1998; Laganiere et al., 2005; Xu et al., 2009). Among them the so-

called ‘‘pioneer’’ factors can directly associate with nucleosomal DNA to enable the occupancy of other TFs. Moreover, several TFs that have been reported to have pioneer activity (Vernimmen and Bickmore, 2015) (Table 6). For example, the pioneer factor PU.1 is shown to promote H3K4me1 recruitment at enhancers in macrophage and B-cell differentiation process (Ghisletti et al., 2010; Heinz et al., 2010). Another typical pioneer factors are Foxa1 and Foxa2 (forkhead box proteins A1 and A2) that have the capacity to access their binding sites in nucleosomal DNA by opening compacted chromatin structures of the target enhancers during liver specification process (Lee et al., 2005a). Therefore, pioneer TFs regulate cell type-specific transcriptional programs (Bossard and Zaret, 2000; Cirillo et al., 1998; Laganiere et al., 2005; Xu et al., 2009).

Moreover, the capacity of TFs to regulate gene expression is also dependent on cooperating with coactivators (Weake and Workman, 2010). It has been reported that the coactivator can facilitate TFs function through acting as histone modifiers, ATP-dependent chromatin remodelers (Calo and Wysocka, 2013). Together, the establishment and maintenance of cell-type-specific gene-expression programs result from the interaction between transcription factors and the chromatin landscape that they encounter.

Table 6 | Pioneer TFs and their DNA binding domains (Vernimmen and Bickmore, 2015)

Pioneer TF	DNA Binding Domain
AP-1	Basic leucine zipper
AP-2γ (TFAP2C)	Basic helix–span–helix
FOXA1 (HNF-3α)	Forkhead
FOXA2 (HNF-3β)	Forkhead
FOXE1	Forkhead
FOXD3	Forkhead
GATA2	2X GATA-type zinc fingers
GATA3	2X GATA-type zinc fingers
GATA4	2X GATA-type zinc fingers
KLF4	3X C2H2-type zinc fingers
NF-Y (CBF)	NF-YA/HAP2
OCT4	POU-specific + POU-Homeodomain
OTX2	Homeodomain
PAX7	Paired + Homeodomain
PBX1	Homeodomain
PU.1	Ets
SOX2	Hmg box
SOX9	Hmg box
TP53	p53
P63	p53
RFX	Rfx-type winged helix

1.4 Coactivator complexes

Binding of TFs at enhancers is not enough to stimulate transcription. Following the deposition of TFs, coactivators are recruited to the regulator elements. The recruitment of coactivators can further regulate chromatin accessibility via enhancing the interaction with the core transcription machinery or modulating histone epigenetic modifications. In the following part, chromatin remodellers, Mediator and SAGA coactivator complexes are particularly introduced.

1.4.1 ATP-dependent chromatin remodellers

ATP-dependent chromatin remodellers regulate chromatin dynamics by driving histone sliding and ejection with their DNA translocase (Becker and Workman, 2013). Based on the similarity sequence between their ATPase domains, remodellers can be divided into four subfamilies: imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), switch/sucrose non-fermentable (SWI/SNF) and INO80 (Bartholomew, 2014; Clapier and Cairns, 2009; Narlikar et al., 2013) (Figure 4-1). All the four subfamilies contain an ATPase–translocase domain (Tr) with two RecA-like lobes. However, they also contain specific domains. For example, ISWI subfamily remodellers harbor a carboxy-terminal HAND–SANT–SLIDE (HSS) domain as well as a negative regulator of coupling (NegC) domain; CHD proteins uniquely contain a tandem N-terminal chromodomains; SWI/SNF proteins are defined by the presence of an N-terminal helicase-SANT domain and a C-terminal bromodomain; INO80 subfamily contains a large insertion between the RecA-like lobes (Clapier et al., 2017).

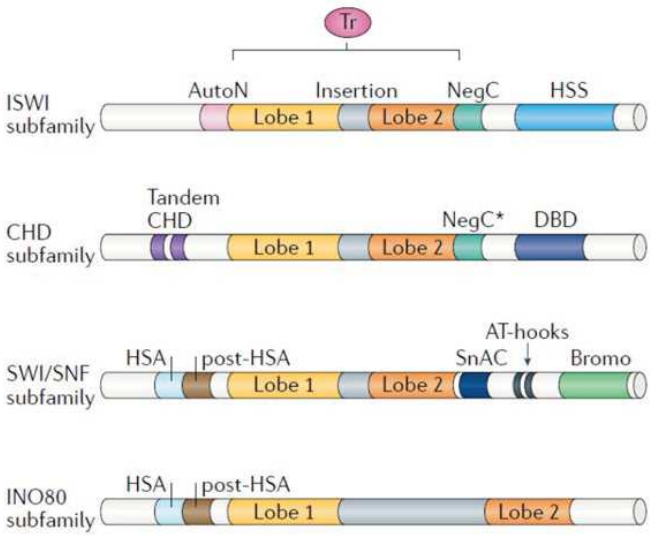


Figure 4-1: Domain organization of chromatin remodelers. The ATPase–translocase domain (Tr) of all the remodellers play an important role in DNA translocation. The Tr domain is comprised of two RecA-like lobes (lobe 1 and lobe 2). (Clapier et al., 2017)

The chromatin remodellers are specialized to conduct mainly three functions: chromatin assembly, chromatin access, and nucleosome editing. For example, ISWI and CHD subfamily remodelers tend to involve the nucleosomes assembly (Fei et al., 2015; Ito et al., 1997; Torigoe et al., 2011). The SWI/SNF remodellers have the ability to regulate chromatin accessibility by ejecting histone octamers or dimers (Boeger et al., 2004). In addition, INO80 subfamily has been reported to change nucleosome composition through exchanging canonical and variant histones (Clapier et al., 2017). Notably, many tissue-specific chromatin remodelers have been identified, indicating that they might be involved in tissue-specific gene expression. Moreover, many chromatin remodellers play important roles in regulating embryonic development process (Table 7).

Table 7 | Developmental roles of chromatin remodellers. (Ho and Crabtree, 2010)

SWI/SNF complex subunit (synonym)	Gene-family members	Developmental phenotype in mammals
BRM or BRG1	NA	Brg1knockout is peri-implantation lethal in mice . BRG1 is required for zygotic genome activation and for differentiation into neurons , lymphocytes , adipose tissue and heart tissue . BRG1 is essential in T-cell development, in which it suppressesCd4 expression and activates Cd8expression . BRG1 is also essential during embryonic erythropoiesis for activation of expression of theβ-globin gene .
		Brm-knockout mice are normal, with greater body mass .
BAF250-family member (ARID1)	BAF250A, BAF250B and BAF250C	Baf250a-knockout mice die at E6.5. Baf250a-knockout mouse ESCs have reduced self-renewal capacity and defective mesodermal differentiation .
		Baf250b-knockout mouse ESCs have a propensity for spontaneous differentiation in culture .
BAF155 and/or BAF170	NA	Baf155knockout is peri-implantation lethal in mice. Heterozygotes (Baf155$^{+/-}$) have exencephaly owing to failure of neural tube closure .
BAF47 (INI1, SNF5)	NA	Baf47 knockout is peri-implantation lethal in mice. Heterozygotes (*Baf47*$^{+/-}$) develop sarcomas of the neural and soft tissues .
BAF60-family member	BAF60A, BAF60B and BAF60C	BAF60C is expressed in the mouse heart and somites and is required for normal heart morphogenesis and establishment of left–right asymmetry , and ectopic expression of BAF60C outside developing mouse heart regions is sufficient to specify development into cardiomyocytes .
Actin	NA	The contribution of actin has been difficult to analyse because of its essential roles as a component of the cytoskeleton.
BAF53-family member	BAF53A and BAF53B	BAF53A is required for neuronal stem-cell proliferation in mice .
		BAF53B is neuron specific and is required for activity-dependent dendritic outgrowth in mice .
BAF57	NA	A dominant-negative mutant of BAF57 prevents T-cell development in mice .
BAF200 (ARID2)	NA	Not reported
Polybromo (BAF180)	NA	Polybromo is required for cardiac chamber maturation and coronary development in mice.
BAF45-family member	BAF45A, BAF45B, BAF45C and BAF45D	BAF45A is necessary and sufficient for neuronal progenitor proliferation in mice .
		BAF45C is required for heart and muscle development in zebrafish .
BRD7 or BRD9	NA	BRD7 is essential for mouse ESC proliferation .
CHD family		
CHD1	SSRP1	Chd1knockdown in mouse ESCs renders them defective in multilineage differentiation, and they undergo global heterochromatinization of euchromatin .
CHD2	Unknown	CHD2-null mouse embryos have retarded growth and die before birth .
CHD3 or CHD4	NURD complex: HDAC1 or HDAC2; MTA1, MTA2 or MTA3; RbBP4 and/or RbBP7; MBD2 or MBD3; P66	CHD4 is required for the development of T cells in the mouse thymus and for the self-renewal of haematopoietic stem cells and differentiation along the myeloid lineage in the bone marrow.
		MBD3-null mouse embryos die mid-gestation, owing to a failure of the inner cell mass to develop into a late epiblast and to the misregulation of several genes during the transition from pre-implantation to post-implantation .
CHD5	Unknown	CHD5 is a tumour-suppressor protein associated with human malignancies such as neuroblastomas
CHD7	Unknown	CHD7 is mutated in CHARGE syndrome in humans . CHD7-null mice show perinatal lethality and widespread tissue defects. CHD7 is required for the proliferation and differentiation of olfactory stem cells .
CHD9	Unknown	CHD9 might be required for differentiation of osteogenic cells .
ISWI family		
SNF2H	NoRC complex: TIP5	SNF2H-null mouse embryos implant but die between E5.5 and E7.5, owing to the failure of both the inner cell mass and the trophoblast to survive and grow.
	WICH complex: WSTF	The NoRC complex regulates cell growth by regulating the transcription of ribosomal DNA.
		WSTF resides in the haploinsufficient region of human chromosome 7, which is responsible for Williams–Beuren syndrome. WSTF-null mice have cardiovascular defects similar to those of patients with Williams–Beuren syndrome .
SNF2L	NURF complex: BPTF, and RbBP4 or RbBP7	SNF2L knockdown in human cells leads to reduced expression of engrailed genes SNF2L expression in a neuroblastoma cell line potentiates neurite outgrowth.
	CERF complex: CECR2	BPTF-null mouse embryos die between E7.5 and E8.5, owing to defects in gastrulation, the absence of an anteroposterior axis and primitive streak, and lack of differentiation of mesoderm and definitive endoderm. BPTF-null ESCs are viable but defective in mesodermal and endodermal differentiation.
		CECR2-null mouse embryos develop exencephaly and defects in neurulation.
INO80 family		
p400	TIP60–p400 complex: TIP60 and TRRAP (and others as listed in ref. 85)	Depletion of TIP60, p400 or TRRAP from mouse ESCs by using RNAi results in altered (differentiated) ESC morphology.
		TRRAP-null mouse embryos die at peri-implantation, owing to the failure of the blastocyst to proliferate

1.4.2 Mediator complex

1.4.2.1 Mediator compositions

Mediator is an evolutionarily conserved complex that contains 25 subunits in budding yeast and up to 30 subunits in humans (Figure 4-2). Mediator can be divided into four modules, including the head module, middle module, tail module, and CDK8 kinase module (Verger et al., 2019). The head module together with the middle module form the active core that is essential for transcription regulation, whereas the tail module and CDH8 kinase module serve regulator function (Cevher et al., 2014; Plaschka et al., 2015; Soutourina, 2018). Due to conformational heterogeneity, the structure of the Tail module is still unresolved (Harper and Taatjes, 2018). Structural study revealed that Med14 subunit acts as a scaffold protein to unit all three main Mediator modules in budding yeast (Robinson et al., 2015). Notably, the CDK8 kinase module is transiently associated with the Mediator complex (Kornberg, 2005) and the dissociation of this module is required for Mediator to join the PIC (Tsai et al., 2014).

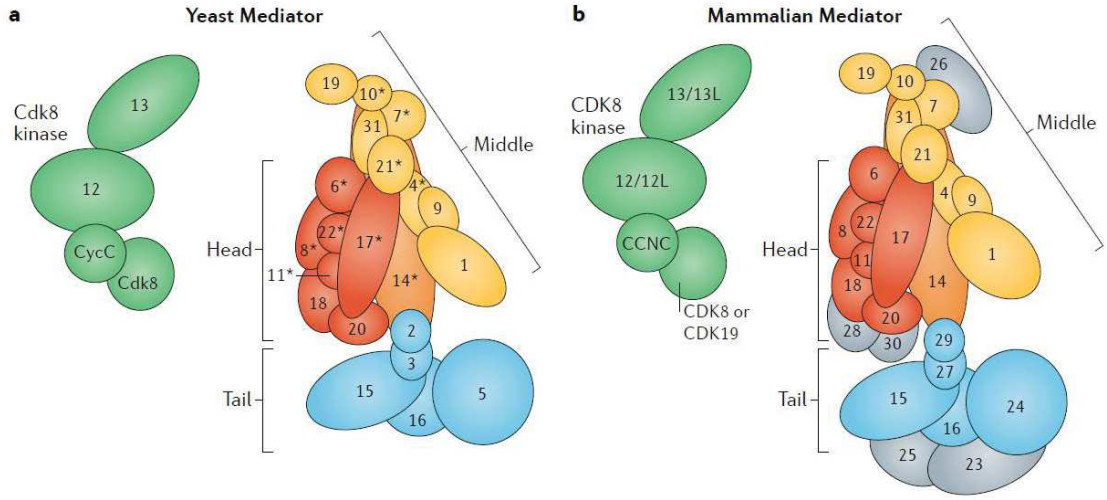


Figure 4-2: Subunit composition of the Mediator complex. Schematics representing the modular organization of the budding yeast Mediator complex (part a) and the mammalian Mediator complex (part b). Mediator comprises four distinct modules: a head module (in red), middle module (in yellow) and tail module (in blue) and the CDK8 kinase module (in green). In metazoan Mediator, MED24, MED27 and MED29 are orthologous to Med5, Med3 and Med2 in yeast, respectively. CDK8, MED12 and MED13 (components of the CDK8 kinase module) also have paralogues (CDK19, Mediator subunit 12-like protein (MED12L) and Mediator subunit 13-like protein (MED13L), respectively) in vertebrates (Soutourina, 2018).

1.4.2.2 Mediator functions

The main function of Mediator is to transduce signals from the transcription activators bound at enhancer regions to the transcriptional machinery located at the promoter regions (Soutourina et al., 2011; Thompson and Young, 1995). Many transcription activators directly interact with the Mediator (Poss et al., 2013). These Mediator–transcription factors interaction frequently involves Mediator tail module (Allen and Taatjes, 2015; Malik and Roeder, 2010; Poss et al., 2013). Nevertheless, it remains to be determined how Mediator is able to control hundreds or even thousands of different transcription factors. Recently, it suggested that considerable activation domains of transcription factors usually contain the intrinsically disordered regions (IDRs), which enables transcription factor to form phase-separated condensates with Mediator complex (Boija et al., 2018).

The recruited Mediator can further modulate PIC formation by interacting with various PIC components, including TFIIB, TFIID, TFIIH and Pol II (Cai et al., 2010; Esnault et al., 2008; Eychenne et al., 2016; Soutourina et al., 2011). The transient interaction between Mediator and GTFs in turn stimulates the dissociation of Cdk8 kinase module from Mediator (Soutourina, 2018). Besides, the binding of Pol II to Mediator induces Med14 conformational change, which further promotes Mediator–Pol II complex formation by altering the orientation between the Mediator head and middle modules (Tsai et al., 2017). After PIC formation, the Mediator stimulates the enzymatic activity of CDK7, a subunit of TFIIH, which phosphorylates Pol II CTD and subsequently induces Pol II release from promoters (Boeing et al., 2010; Kim et al., 1994; Nair et al., 2005).

1.4.3 SAGA complex

SAGA (Spt-Ada-Gcn5 acetyltransferase) is an evolutionary conserved multi-subunit co-activator complex with a modular organization. The SAGA complex contains 18 to 20 subunits in yeast (Grant et al., 1997). Based on structural and functional characteristics, SAGA can be separated into four distinct modules, including core structural module, Tar1 transcription factor binding module, histone acetyltransferase module (HATm), and histone deubiquitinase module (DUBm) (Table 8).

Table 8 | SAGA is a conserved transcriptional co-activator complex organized into well-defined structural and functional modules. (Helmlinger and Tora, 2017)

Functional tools	Orthologous SAGA complexes				Chromatin 'reader' domains	Structural domains
	Saccharomyces cerevisiae	Schizosaccharomyces pombe	Drosophila melanogaster	Homosapiens		
HAT module	Gcn5	Gcn5	KAT2 (GCN5)	KAT2A/KAT2B (GCN5/PCAF)	Bromo	
	Ada2	Ada2	Ada2b	TADA2b	SANT, SWIRM	
	Ngg1 (Ada3)	Ngg1 (Ada3)	Ada3	TADA3		
	Sgf29	Sgf29	Sgf29	SGF29	Tudor(X2)	
DUB module	Ubp8	Ubp8	dNonstop	USP22 (UBP22)		
	Sgf11	Sgf11	dSgf11	ATXN7L3	Nucleosome binding	
	Sgf73	Sgf73	dATXN7	ATXN7/ATXN7L1/L2		
	Sus1	Sus1	dE(y)2	ENY2		
Core structural module	Taf5	Taf5	WDA/TAF5L	TAF5L		WD40
	Taf6	Taf6	SAF6/TAF6L	TAF6L		HFD
	Taf9	Taf9	TAF9	TAF9/TAF9b		HFD
	Taf10	Taf10	TAF10b	TAF10		HFD
	Taf12	Taf12	TAF12	TAF12		HFD
	Spt7	Spt7	Spt7	SUPT7L (STAF65G)	Bromo	HFD
	Hfi1 (Ada1)	Hfi1 (Ada1)	Ada1	TADA1		HFD
	Spt20	Spt20	Spt20	SUPT20H		
TBP binding	Spt3	Spt3	Spt3	SUPT3H		HFD (×2)
	Spt8	Spt8	–	–		WD40
TF-binding module	Tra1	Tra1	Nipped-A(Tra1)	TRRAP		
Splicing module	–		SF3B3	SF3B3		
	–		SF3B5	SF3B5		

1.4.3.1 The core structural module of SAGA

The core structural module of SAGA servers as a scaffold to assemble other modules. It consists of subunits Taf5, Sgf73 and Spt20, and a histone octamer-like fold (Wang et al., 2020a). The

octamer-like fold contains three pairs of subunits (Taf6-Taf9, Taf10-Spt7, and Taf12-Ada1), of which each contributes one histone fold, and one Spt3 subunit contributing another two histone folds (Figure 4-3). In contrast to a canonical twofold symmetry histone octamer, the SAGA octamer-like fold is fully asymmetric (Papai et al., 2020; Wang et al., 2020a). This deformed octamer is suggested to establish a peripheral site for the binding of TBP (Papai et al., 2020). Moreover, two subunits, Taf12 and Spt20, form a flexible connection between the core module and the Tra1 TF-binding module, while Sgf73 bridges the core module to the DUB module (Kohler et al., 2010; Samara et al., 2010).

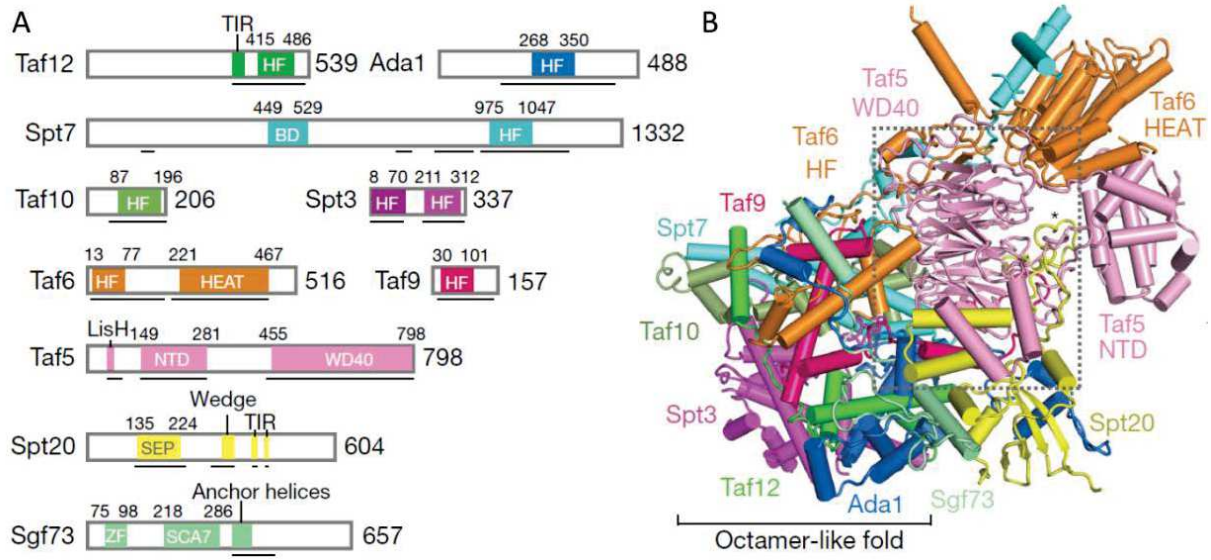


Figure 4-3: Structure of the SAGA core-module. A. Subunits architecture. Residues at domain boundaries are indicated. BD, bromodomain; HF, histone fold; HEAT, HEAT repeat domain; NTD, N-terminal domain; WD40, WD40 β-propeller domain; SEP, shp1–eyc–p47 domain; ZF, zinc finger domain; SCA7, SCA7 domain. B. Ribbon model shows the arrangement and interactions between these subunits. (Wang et al., 2020a)

Similar to the function of TFIID, SAGA can also deliver TBP to gene promoters (Bhaumik and Green, 2002; Larschan and Winston, 2001) and regulates global RNA polymerase II transcription in yeast (Baptista et al., 2017; Warfield et al., 2017). Besides, nine TAFs of TFIID and seven subunits of SAGA contain the histone-fold domain (HFD) that can mediate the interaction between different subunits (Gangloff et al., 2001; Trowitzsch et al., 2015). Moreover, SAGA and TFIID share a subset of TAFs that are crucial functional elements for both complexes. For example, in most species, TAF9, TAF10, and TAF12 are shared between SAGA and TFIID (Helmlinger and Tora, 2017). However, in metazoans, SAGA and TFIID have their specific proteins that are unique

to each complex, such as TAF5 and TAF6 are specific to TFIID, whereas TAF5L and TAF6L are unique to SAGA. These observations suggest that the function of SAGA might be various in different species even though they have some similar structures.

As mentioned above, both SAGA and TFIID contain an octamer-like fold domain. In TFIID complex, the Lobe A domain harbors an octamer-like fold that is similar to the fold found in SAGA. However, they are different in the composition of subunits. Unlike TFIID, SAGA does not have Taf3 and Taf4 subunits but contains Spt7 and Ada1 subunits instead. Moreover, the two histone-fold domains of Spt3 in SAGA are exchanged to the histone fold pair (Taf11–Taf13) in TFIID. Despite these changes, these two octamer-like folds can separately recruit TBP at the same relative position (Wang et al., 2020a). It worth to note that SAGA can recruit TBP to promoter region via Spt3 and Spt8 subunits in *S.cerevisiae* (Hahn and Young, 2011; Han et al., 2014). However, in mammalian cells, the function of SUPT3H (Spt3 in yeast) remains to be determined, and the orthologous of Spt8 does not exist. Therefore, it remains unknown whether a TBP-binding activity exists in mammalian SAGA.

1.4.3.2 The splicing module

The splicing module of SAGA contains two subunits: SF3B3 (Splicing Factor 3b Subunit 3) and SF3B5 (Splicing Factor 3b Subunit 5). SF3B3/SF3B5 also form part of the U2 small nuclear ribonucleoprotein particle (snRNP) that plays a well-defined role in splicing (Fabrizio et al., 2009; Golas et al., 2003). However, the splicing module of SAGA has a splicing-independent function in *drosophila* (Stegeman et al., 2016). Given SF3B3 shows 50.7% sequence similarity and has a similar predicted structure to the UV-damaged DNA-binding protein (DDB1, p127) that function in DNA repair pathway, several studies suggest that SF3B3 may also play a role in DNA damage recognition (Brand et al., 2001; Martinez et al., 2001). Further studies will be needed to define the role of these spliceosomal factors in SAGA.

1.4.3.3 Tra1/TRRAP transcription factor binding module

In mammalians, TRRAP (homolog of Tar1 in *S. cerevisiae*), the largest component of SAGA (420 kDa), is an evolutional conserved multidomain protein. TRRAP belongs to the phosphoinositide

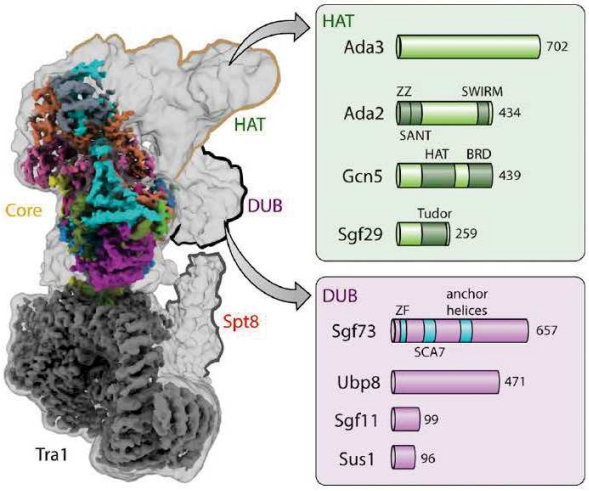
3-kinase-related kinases (PIKK) family that contains six Ser/Thr-protein kinases, including ataxia-telangiectasia mutated (ATM), ataxia- and Rad3-related (ATR), DNA-dependent protein kinase catalytic subunit (DNA-PKcs), mammalian target of rapamycin (mTOR), suppressor of morphogenesis in genitalia (SMG-1) and transformation/transcription domain-associated protein (TRRAP) (Lempiainen and Halazonetis, 2009). In contrast to the five other members, TRRAP is a pseudokinase due to the fact that its kinase domain lacks the catalytic residues required for kinase activity. However, TRRAP contains a HEAT □-helical motifs and two tetratricopeptide repeat (TPR) motifs which are all critical for protein-protein interactions (Perry et al., 2004; Sikorski et al., 1990). Consequently, Tra1 plays a crucial role in recruiting SAGA to gene-specific promoters by interacting with TFs.

As a TF coactivator, TRRAP is originally identified as an interacting partner of c-Myc (McMahon et al., 1998). Besides, several other transcription factors, including Gal4 and Gcn4 in yeast and E2F1 in human, also interact with TRRAP (Brown et al., 2001; Herbig et al., 2010; Murr et al., 2007; Reeves and Hahn, 2005). TRRAP is also a component of the NuA4/Tip60 complex that is another conserved transcriptional co-activator with histone acetyltransferase activity. In contrast to the SAGA HAT module that preferentially acetylates histone H3, the NuA4/TIP60 complex tends to acetylate H4, H2A and H2A.Z (Allard et al., 1999; Babiarz et al., 2006; Keogh et al., 2006).

1.4.3.4 The histone acetyltransferase module (HATm) of SAGA

The HAT module of SAGA complex contains the acetyltransferase enzyme Gcn5 together with Ada2, Ada3, and Sgf29 subunits (Figure 4-4). Ada2 connects the HAT module to the rest of the SAGA (Balasubramanian et al., 2002; Lee et al., 2011). Gcn5 is first identified as a transcription-related HAT in the *ciliate Tetrahymena thermophile* (Brownell et al., 1996). Histone acetylation promotes transcription by assisting to form an open chromatin structure, therefore increases the accessibility of DNA for transcription factors (Choi and Howe, 2009; Nagy and Tora, 2007; Suganuma and Workman, 2011). The BRD domain of Gcn5 binds to acetylated lysine residues and acetylates histone H3 preferentially on residues K9 and K14 (Bonnet et al., 2014; Hassan et al., 2002). Additionally, Spt7, which is the subunit of the central SAGA module, also contains a BRD domain that can interact with histone H3K9ac *in vitro* (Hassan et al., 2007). These findings

suggest that SAGA can read the products of the HATm to further stabilize itself on the nucleosome. This feedforward loop between Spt7 and Gcn5 perhaps maintain robust HAT activity on that nucleosome or neighboring nucleosomes (Strahl and Briggs, 2020). Moreover, Gcn5 also targets other nonhistone protein. For example, GCN5 is reported to act as an inhibitor of autophagy and lysosome biogenesis by targeting TFEB in mammalian and drosophila cells (Wang et al., 2020b).



Modules and subunits	Chromatin reader and writer domains	Histone recognition function
Ada3	None	?
Ada2	ZZ, SANT	Histone binding?
Gcn5	HAT, BRD	Histone acetylation, H3K14ac binding
Sgf29	Tudor	H3K4me2/3 binding
Sgf73	ZF	H2A/H2B acid patch binding
Ubp8	H2B deubiquitination	H2Bub1 binding/removal
Sgf11	None	None
Sus1	?	Phospho-CTD interaction

Figure 4-4: HAT and DUB subunits within SAGA. Left, Cryo-EM structure of the SAGA complex with HAT and DUB modules. Arrows extending from the SAGA structure show the individual subunits of the HAT and DUB modules with their functional domains. Right, Table highlighting the functional domains found in each DUB or HAT subunit with the documented histone or RNA polymerase II interactions. (Strahl and Briggs, 2020)

1.4.3.5 The histone deubiquitinase module (DUBm) of SAGA

The DUBm is composed of the ubiquitin-specific protease 22 (USP22) and adaptor proteins, including ATXN7, ATXN7L3 and ENY2. In human cells, depletion of either ENY2 or ATXN7L3 resulted in a non-functional USP22 enzyme, and consequently H2Bub1 was not removed from the genome (Atanassov et al., 2016). The DUBm is also associated with a wide array of paralogues and variants. For example, two novel DUBm variants are found in human cells, namely USP27X and USP51, which are associated with ATXN7L3 and ENY2, but not with ATXN7 (Atanassov et al., 2016). This work indicates that these DUBms might have redundant function in H2Bub1 deubiquitination and they might compete for the limited ATXN7L3 and ENY adaptor proteins.

Interestingly, the adaptor protein ENY2 is shared between SAGA and nuclear pore-associated transcription export complex 2 (TREX-2) (Gonzalez-Aguilera et al., 2008; Rodriguez-Navarro et al., 2004). TREX-2 was initially characterized in yeast. It can interact with the inner face of the nuclear pore complex (NPC) via the basket nucleoporin Nup1. Moreover, in yeast, deletion of any TREX-2 subunits results in mRNA export defects (Fischer et al., 2004; Fischer et al., 2002; Wilmes et al., 2008). Therefore, TREX-2 is suggested to play an essential role in mRNA export. Besides, Sus1 (homolog of human ENY2) physically bridges these two complexes in yeast indicating that SAGA-dependent transcription might be coupled with the TREX-2 mediated mRNA export process (Kohler et al., 2008; Rodriguez-Navarro et al., 2004). However, in human cells, ENY2 separately interacts with either SAGA, or TREX-2 complex, and no other TREX-2 subunit is part of SAGA and *vice versa* (Umlauf et al., 2013). Thereby, it still unclear whether the SAGA complex is involved in TREX-2-mediated mRNA export process in mammalian cells.

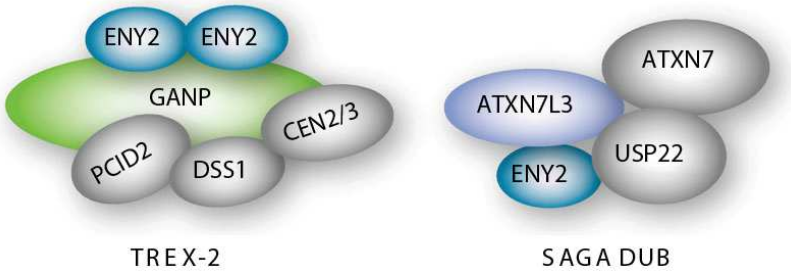


Figure 4-5: TREX-2 and SAGA DUB complexes. ENY2 is shared between TREX-2 and SAGA. Two protein molecules of ENY2 binds to the large GANP subunit of TREX-2 complex and ENY2 is also a part of the deubiquitination module of the SAGA complex (Kamenova et al., 2019).

ENY2 also interacts with ATXN7L3B which is a paralog of ATXN7L3 in humans. The N-terminal region between ATXN7L3B and ATXN7L3 is very similar and both contain a "Sus1-binding" motif that interacts with ENY2 (Figure 4-6A) (Ellisdon et al., 2010; Li et al., 2016a). Despite their sequence similarity, ATXN7L3B predominantly localizes to the cytoplasm, whereas ATXNL3 is mostly located in the nuclear (Li et al., 2016a). This observation suggests that the ATXN7L3B-ENY2 interaction might regulate the SAGA DUB activity by sequestering ENY2 in the cytoplasm and limiting the ENY2-ATXN7L3 interaction in the nucleus (Li et al., 2016a).

Although the ATXN7 does not contain Sus1-binding domain, it contains a ZnF domain and a SAC domain instead (Figure 4-6). The ZnF-Sgf73 domain in ATXN7 associates the DUBm with SAGA

and also integrates the three components within the DUBm. Meanwhile, the ZnF-Sgf11 domain in ATXN7L3 further stimulates USP22 activity (Ellisdon et al., 2010; Kohler et al., 2008; Lang et al., 2011; Samara et al., 2010). The SCA7 domain in ATXN7 is also found in ATXN7L3, but not in ATXN7L3 yeast orthologue Sgf11 (Zhao et al., 2008). Besides, the ATXN7-SCA7 can bind to nucleosomes (Bonnet et al., 2010).

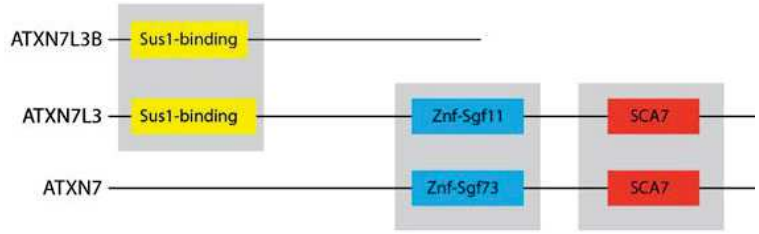


Figure 4-6: Comparison of the protein structures of ATXN7L3, ATXN7L3B and ATXN7.

The N-termininal region between ATXN7L3B and ATXN7L3 shares 74% identity, however, ATXN7L3B lacks the Znf-Sgf11 domains and SCA7 domain that are present in ATXN7L3. Adapted from (Li et al., 2016a)

The crystal structure of the DUBm showed that it contained two distinct functional lobes, including assembly lobe and catalytic lobe in yeast. The N-terminal region of Sgf73 (human ATXN7 homolog) connects these two lobes. In the assembly lobe, Sgf11 (human ATXN7L3 homolog) N-terminal helix is clamped onto the Ubp8 (human USP22 homolog) ZnF-UBP motif via the assistance of Sus1 (human ENY2 homolog). Meanwhile, in the catalytic lobe, the C-terminal Znf region of Sgf11 links Ubp8 catalytic domain (Figure 4-7A). Moreover, the well-positioned ZnF domains in Sgf11 and Sgf73 was reported to be required for inducing the enzymatic activate of Ubp8 (Kohler et al., 2010). Notably, the Sgf11 arginine cluster on Sgf11 zinc finger domain can dock on the conserved H2A/H2B heterodimer acidic patch. Besides, the Ubp8 catalytic domain mediates additional interactions with the C-terminal helix of H2B, as well as with the conjugated ubiquitin (Morgan et al., 2016). Therefore, these observations suggest that the capacity of the DUBm to bind the nucleosome is partly dependent on Sgf11 and Usp8.

More recently, another crystal structure study reveals that the nucleosome binding of the SAGA complex can displace the HATm and DUB modules from the core module in yeast (Figure 4-7B). In this case, these two catalytic modules can move around or downstream of the TSS, whereas the core module and Spt8 subunit recruit TBP at the promoter (Wang et al., 2020a). This finding suggests that the flexibility between SAGA modules is important for it to fulfill multiple functions

at different regions. In line with this observation, there is a hypothesis that the DUBm can function independent of SAGA complex. Since deletion of S*pt20*, which destroys the core module of SAGA, accumulates less H2Bub1 than the condition lacking *Ubp8*. Together, this finding indicates that the residual DUBm activity still exists in SAGA-deficient cells (Henry et al., 2003).

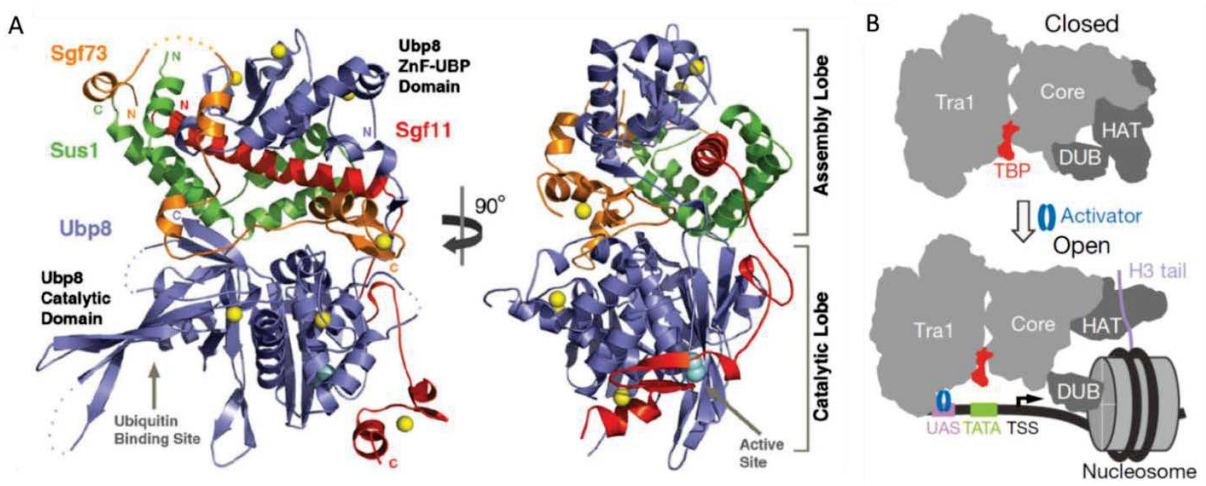


Figure 4-7: Overall view of the DUB module in the yeast. A. Two orthogonal views of the DUB module complex (Kohler et al., 2010). B. Nucleosome binding displaces the HAT and DUB modules from the SAGA core module (Wang et al., 2020a).

Roles of H2B deubiquitination in transcription

Histone H2Bub1 peaks slightly downstream of the transcription start site and slowly tail off across the gene body (Bonnet et al., 2014; Minsky et al., 2008). The DUBm of SAGA can quickly remove H2Bub1 (Bonnet et al., 2014). It has been reported that the dynamic of H2Bub1 is important for transcriptional regulation. Actually, in *S. cerevisiae*, H2Bub1 also acts as a barrier for depositing Ctk1 at the coding region of active genes (Cho et al., 2001). Ubp8 can timely deubiquitinate H2Bub1, which triggers Ctk1 recruitment and in turn facilitates productive elongation by phosphorylating the Pol II serine 2 (Wyce et al., 2007). Therefore, the efficient H2Bub1 deubiquitination promotes transcription elongation by recruiting Ctk1 in yeast. However, analyzing the separately roles of the DUBm at promoter and gene body is still a challenging task for future analysis. In addition to regulate transcriptional elongation, H2Bub1 represses some inducible enhancers. For example, upon activating some enhancers, H2Bub1 is deubiquitinated at these enhancers and then H2A.Z is evicted by INO80, which allows additional transcriptional

activators to gain access to the DNA (Segala et al., 2016). This finding suggests that the DUBm may act as an activator at some enhancers. In line with this hypothesis, studies in *Drosophila* have shown that many SAGA-bound genes require the SAGA ubiquitin protease activity for full expression. These genes tend to be expressed at higher levels in muscle than other tissues (Weake et al., 2011).

Roles of H2Bub1 deubiquitination in nucleosome remodelling

H2Bub1 plays an essential role in nucleosome dynamics via regulating the localization of Spt16, a subunit of the histone chaperone FACT (Fleming et al., 2008). Moreover, the FACT complex in turn promotes H2Bub1 deubiquitination by cooperating with Ubp10, but not Ubp8 in yeast (Nune et al., 2019). However, it is still unknown whether SAGA and FACT can act on a nucleosome simultaneously or sequentially, as structural studies still could not find a common docking site for both FACT and DUB module on the H2A/H2B histone octamer (Hondele et al., 2013; Kemble et al., 2015; Marciano and Huang, 2016).

Roles of H2Bub1 deubiquitination in histone crosstalk

H2Bub1 has been implicated in histone crosstalk in both yeast and mammalian cells (Dover et al., 2002; Ng et al., 2002). Notably, H2Bub1 is generally regarded to be a prerequisite for methylation of H3K4me3 and H3K79. Structural studies revealed that H2Bub1 was a conformational plastic epitope that can be recognized in structurally distinct ways. For example, in yeast, the H3K4 methyltransferase Set1 recognized H2Bub1 on one face of the nucleosome and the methylated H3 on the opposing face (Worden et al., 2020) (Figure 4-8A). MLL1 was another H3K4 histone methyltransferase containing RbBP5, WDR5, and ASH2. H2Bub1 was reported to orient the association between MLL1 and the nucleosome by directly binding to the RBBP5 subunit (Worden et al., 2019) (Figure 4-8B). Whereas, the H3K9 histone methyltransferase Dot1L engaged the nucleosome acidic patch and occupies a conformation poised for methylation. In this conformation, H2Bub1 and Dot1L interact directly through the complementary hydrophobic surfaces (Anderson et al., 2019) (Figure 4-8C). However, whether H2Bub1 deubiquitylation has a function in histone crosstalk is still unclear.

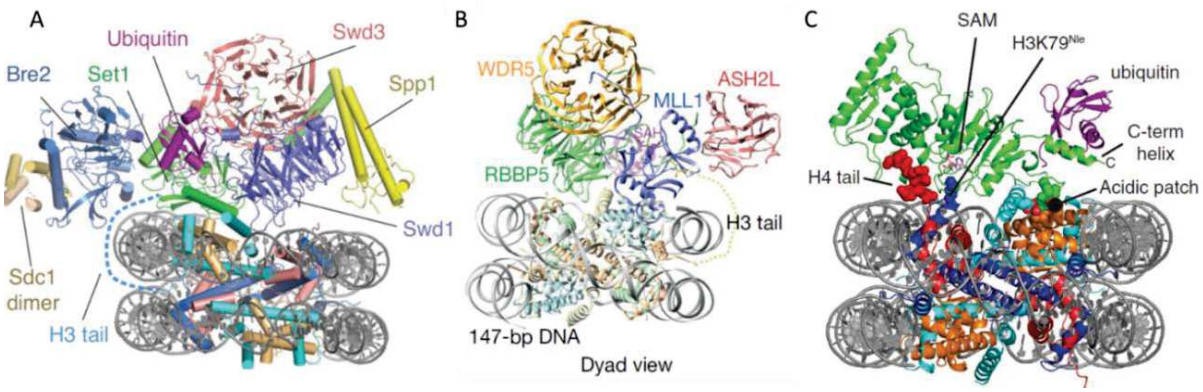


Figure 4-8: Overview of the structure of the H2Bub1 complex. A. The model of the COMPASS-nucleosome complex (Worden et al., 2020). B. Schematic of the domain organizations of the human MLL1 catalytic module (Xue et al., 2019) . C. Structures of Dot1L Bound to H2B-Ubiquitin Nucleosome (Worden et al., 2019)

Notably, the Set1, which is a trimethylase of H3K4, can efficiently recruited to a large subset of genes in myotubes even in the absence of detectable H2Bub1. This unexpected finding suggests that muscle cells may represent a novel set of histone crosstalk that the deposition of H3K4 methylation is independent of H2Bub1 (Vethantham et al., 2012).

The DUBm deubiquitinates other proteins

In addition to H2Bub1, a multitude of substrates have been identified as targets of the ubiquitin protease activity of the SAGA complex. For example, the DUBm can deubiquitinate monoubiquitinated histone H2A (H2Aub1) in human cells (Lang et al., 2011; Zhao et al., 2008). H2Aub1 is deposited by the Polycomb repressive complex 1 (PRC1) that is linked with gene silencing and X chromosome inactivation (Fang et al., 2004; Wang et al., 2004; Zhang et al., 2004). Moreover, H2Aub1 also inhibits transcription elongation (Eskeland et al., 2010; Zhou et al., 2008). These observations raise the hypothesis that the DUBm-mediated H2A deubiquitination might enhance gene expression by counteracting the repressive effects of Polycomb-mediated gene silence. In addition to transcriptional regulation, the DUBm is also associated with other processes, such as telomere maintenance and cell-cycle regulation. For example, the mammalian telomeric repeat binding factor 1 (TRF1), CCNB1, CCND1 and the far upstream element (FUSE)-binding protein 1 (FBP1) are all the targets of the DUBm (Atanassov and Dent, 2011; Atanassov et al., 2009; Gennaro et al., 2018; Lin et al., 2015). The DUBm is also associated with genomic integrity maintenance. For example, USP22 has been associated with the DNA damage response (DDR)

through regulation of class switch recombination and double strand break (DSB) repair in B cells (Li et al., 2018; Ramachandran et al., 2016). Moreover, USP22 is involved in viral infection-triggered signaling through deubiquitinating and stabilizing KPNA2, which facilitated virus-triggered nuclear translocation and subsequent expression of downstream genes (Cai et al., 2020).

1.4.2.6 The recruitment of SAGA on chromatin

SAGA can be deposited at gene promoters to stimulate transcription (Lang et al., 2011; Nagy et al., 2009; Sellam et al., 2009). How SAGA is specifically recruited to its target genes? As discussed before, Tra1 is suggested to play a crucial role in recruiting SAGA to gene-specific promoters. However, the interaction between Tra1 and transcription activators is not sufficient to recruit SAGA to all its target genes. Several studies suggest that SAGA can directly interact with the transcription machinery and chromatin PTM marks, which promote SAGA recruitment. For instance, two SAGA core subunits, Spt3 and Spt8, could interact with TBP; the bromodomains of Spt7 and Gcn5 can interact with acetylated nucleosomes. Both of them are important for SAGA recruitment at promoters (Hassan et al., 2002). Besides, the Zn-binding fold within SCA7 domain of ATXN7 can bind to the H2A–H2B dimers (Bonnet et al., 2010) and the double Tudor domain of Sgf29 interacts with H3K4me2/3 (Bian et al., 2011). Above all, SAGA recruitment or retention at promoters is regulated through multiple interactions.

1.5 Embryonic development

Embryogenesis is the development process from fertilized egg to entire embryo. At the early phase of embryonic development, the mouse embryo generates multiple cell lineages and body axes. Gastrulation plays essential roles in transforming a single layer of epithelial cells into the three germ layers, including ectoderm, mesoderm and endoderm, which contribute to all of the organs. Embryonic development process and its related pathways will be discussed as follows.

1.5.1 Cell fate decisions in the early mouse embryos

Mouse embryonic development is involved in several lineage specification events. Before implantation, the fertilized egg turns into eight-cell stage after three times of cell division. Then it undergoes compaction to increase cell-cell contacts and apical-basal polarity. Following this, the

morula cells undergo further cleavages to reach the blastocyst stage. By E3.5, the first cell fate decision is the choice of inner cell mass versus trophectoderm (TE) fate. *Cdx2* is a functional marker for TE, whereas *Pou5f1* and *Nanog* are markers for the ICM. Moreover, *Cdx2* can repress *Pou5f1* and *Nanog* in the TE, which is essential for segregation of the ICM and TE lineages (Strumpf et al., 2005). By E4.5, the second fate decision is the differentiation of ICM into primitive endoderm (PrE) versus epiblast. The PrE forms one layer of cells on the surface of the ICM and is positive for *Gata6* and *Gata4*, whereas the epiblast is located inside the ICM and is marked with *Nanog* and *Pou5f1* (Chazaud et al., 2006; Takaoka and Hamada, 2012). Shortly after implantation, the anterior visceral endoderm (AVE) is formed from the distal visceral endoderm (DVE). Nodal can antagonist signals secreted from the AVE, including Lefty1 and Cer1, influence the nearby epiblast and specify it to the anterior identity. However, the epiblast located far from the AVE, escapes from the AVE-derived signals and forms the primitive streak on the opposite side of the embryo (Beddington and Robertson, 1998; Mesnard et al., 2006; Takaoka and Hamada, 2012). Thus, the anterior-posterior (AP) polarity is established by E6.5 (Figure 5-1).

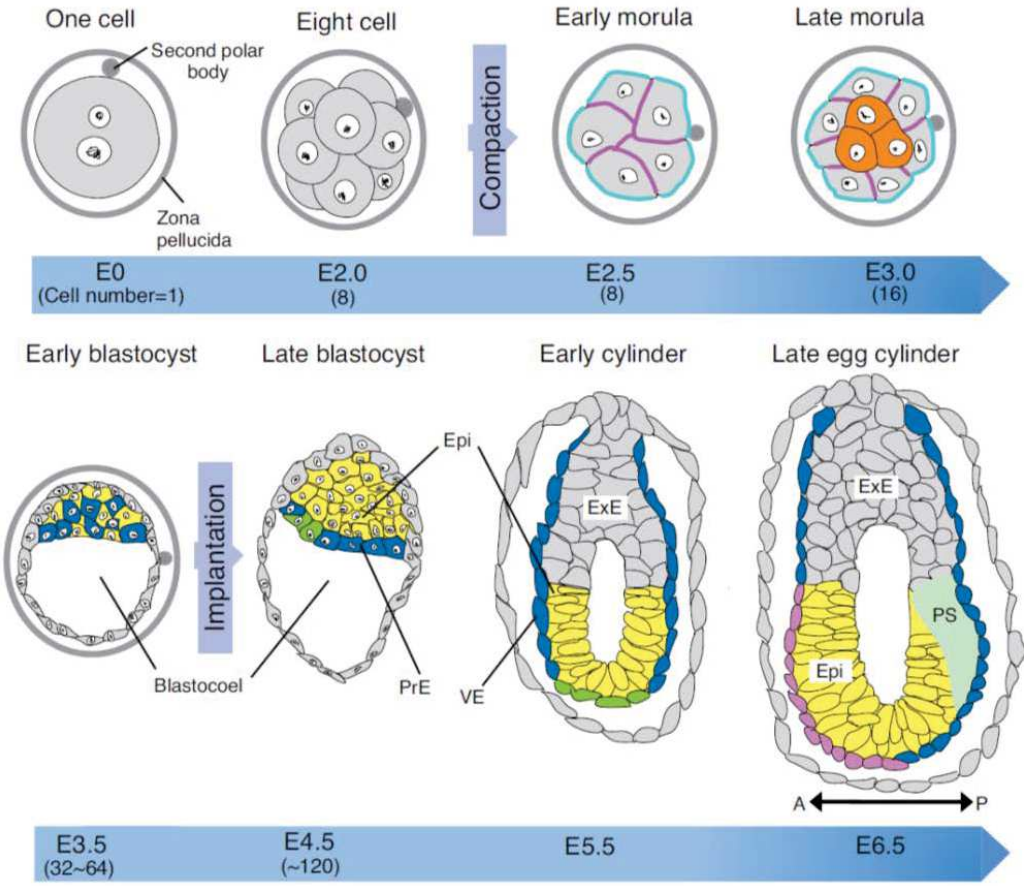


Figure 5-2: Mouse embryo development from fertilization to gastrulation. AVE, anterior visceral endoderm; DVE, distal visceral endoderm; Epi, epiblast; Exe, extraembryonic ectoderm; PrE primitive ectoderm; PS, primitive streak; VE, visceral endoderm. (Takaoka and Hamada, 2012)

1.5.2 Mouse primitive streak

The primitive streak is initially induced at the proximal posterior pole of epiblast. It contains the epiblast layer and the cells that are making transition into the mesenchymal layer. Gastrulation occurs at the primitive streak. During this process, the gradient of signals at the streak temporally regulates the cells differentiation potential. Therefore, cells adopt various fates depending on which position they occupy and when they leave the primitive streak.

1.5.2.1 Factors regulate the primitive streak

Canonical Wnt and Nodal signaling are essential to induce the primitive streak (PS) (Funa et al., 2015). By E6.25, *Wnt3* and *Nodal* regulates the initiation of primitive streak formation on the posterior side of the embryo (Figure 5-3A). Nodal is derived from its secreted precursor proNodal. The subtilisin-like proprotein convertases (SPC), Furin and PACE4, stimulate this Nodal maturation process (Beck et al., 2002). In return, the Nodal precursor maintains the expression of Furin and PACE4 in extraembryonic ectoderm (ExE) by binding and activating activin receptors (Ben-Haim et al., 2006). Besides, proNodal also mediates the expression of BMP4 to induce *Wnt3*, which amplifies *Nodal* expression and stimulates mesoderm (Ben-Haim et al., 2006). Meanwhile, *Wnt3* positively regulates the levels of proNodal to initiates the feedback loop of *Nodal* signal (Ben-Haim et al., 2006). *Nodal* also promotes its own expression (Norris et al., 2002). However, *Nodal* induced *Lefty2* negatively regulates the activity of *Nodal* (Chen and Shen, 2004). Thus the initiation loop established by the *Nodal* signaling is reduced. All of these signals corporately induce cells to move through the streak and generate the extraembryonic mesoderm (Ramkumar and Anderson, 2011).

Wnt and fibroblast growth factor (Fgf) pathways establish the positive feedback loops to maintain the streak (Ramkumar and Anderson, 2011) (Figure 5-3B). As the streak progresses, *Wnt3A* replaces *Wnt3* and induces the expression of the T box family genes, including *Brachyury* and *Tbx6* at the streak (Arnold and Robertson, 2009; Yamaguchi et al., 1999a). *Fgf4* and *Fgf8* trigger

the activity of Fgf receptor 1 (*FgfR1*) to maintain the expression level of *Brachyury* and *Tbx6*. Moreover, *FgfR1* promotes mesoderm cell fate by controlling SNAIL and E-cadherin expression (Ciruna and Rossant, 2001). In the tail bud and presomitic mesoderm, Wnt signals also regulate transcription of the Notch ligand *Dll1* to control somite formation and patterning (Hofmann et al., 2004). On the other hand, *Tbx6* negatively regulates the expression of *Sox2* by inactivating enhancer N1 to inhibit the neural fate, which is important for the specification of paraxial mesoderm from the axial stem cells (Takemoto et al., 2011).

Epithelial-mesenchymal transition (EMT) occurs in the streak to from mesoderm and specification of definitive endoderm. The EMT process involves the loosening of epithelial adherens junctions, disassociation with the basement membrane and rearrangement of the cytoskeletal architecture (Yang and Weinberg, 2008). Before EMT, the epithelial cells are connected to the basement membrane and display apical–basal polarity, which is organized by polarity complexes that depend on the cell junction architecture (Huang et al., 2012; Yang et al., 2020). As EMT progresses, the expression of junction proteins is transcriptionally repressed, which in turn promotes the loss of epithelial junctions (De Craene and Berx, 2013). In the meantime, two transcriptional repressor, *Snail1* and *Snail2*, inhibit the expression of E-Cadherin through binding to E-box DNA sequences with their carboxy-terminal zinc-finger domains (Cano et al., 2000; Peinado et al., 2007). Besides, the basic-helix-loop-helix transcription factors MESoderm Posterior 1 and 2 (*Mesp1* and *Mesp2*) enhance the expression of Snai1 (Lindsley et al., 2008), which participate this feedback loop (Figure 5-3C). Following the disassembly of epithelial cell–cell contacts, the epithelial actin architecture remodeling, cells become mobile and gain invasive capacities (Lamouille et al., 2014). Upon cells ingression through the streak, actin expression is changed from the apical side to the entire cell periphery, which enables cell elongation and motility to migrate away from the streak (Thiery and Sleeman, 2006; Yilmaz and Christofori, 2009, 2010). At later stages, the primitive streak is replaced by the tail bud at mid-somite stages (E9.25–E9.5, P22 somites)(Beddington, 1983), and EMT continues late into elongation of the anterior-posterior axis between E12.5 and E13.5(Cunningham et al., 2011).

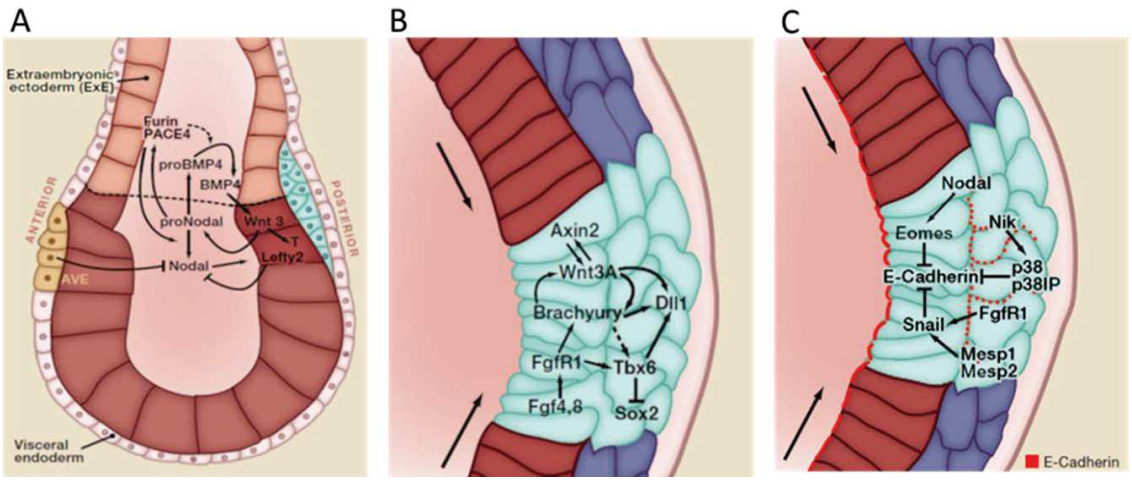


Figure 5-3: Mouse Primitive Streak. Signals and transcription factors, which are required for the establishment (A), maintenance of the primitive streak (B) and down-regulation of the E-Cadherin. (Ramkumar and Anderson, 2011)

Finally, EMT results in the generation of a variety of tissues, such as mesoderm, neural crest cells, heart valves (Murry and Keller, 2008). Epiblast cells that do not migrate through the streak, will give rise to the neurectoderm and eventually the central nervous system as well as the ectoderm (Lawson, 1999). Moreover, loss of either the BMPR1A receptor (Di-Gregorio et al., 2007) or Nodal (Camus et al., 2006) results in precocious neuronal differentiation and premature loss of pluripotency within the epiblast. All of these suggests that anterior neurectoderm represents the default state of epiblast differentiation (Camus et al., 2006).

1.5.2.2 Heart development

The heart is the first organ to function during vertebrate embryogenesis. Heart formation via several well-established transitions. In mammals, a first wave of heart progenitors migration through the node/organizer and primitive streak (Garcia-Martinez and Schoenwolf, 1993), and take a lateral migratory path towards the cranial and cranio-lateral parts of the embryo to form the cardiac crescent (Redkar et al., 2001)(Figure 5-4). This event requires Fgf8, as well as the basic helix–loop–helix (bHLH) transcription factors *Mesp1* and *Mesp2* (Kitajima et al., 2000; Sun et al., 1999). Subsequently, the linear heart tube is formed that is a transient structure composed of an inner endothelial tube shrouded by a myocardial layer. Meanwhile, the elongating heart begins to adopt a rightward spiral form, in a process called cardiac looping. During looping, the future

ventricles become distinct and balloon outwards, and the atrial region and systemic venous tributaries are forced dorsally and cranially (Dehaan, 1963; Harvey, 2002) (Figure 5-4). Moreover, correct differentiation of embryonic endoderm is crucially required for this hear tube formation, and several mutations affecting endoderm in zebrafish and mouse embryos partially disrupt the process, leading to various degrees of severity of cardia bifida (Narita et al., 1997; Stainier, 2001).

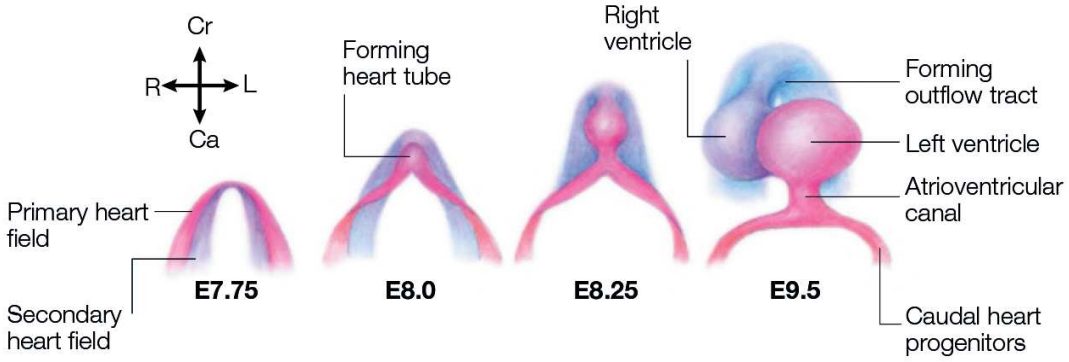


Figure 5-4: Primary and secondary heart fields. Drawings showed the relative position of secondary heart field cells (blue) relative to the primary heart field during cardiac crescent through the looping stages of heart development in the mouse. The compass indicates the body axes. Ca, caudal; Cr, cranial; L, left; R, right. (Harvey, 2002)

Two sources of multipotent cardiovascular progenitors (MCPs) contribute to the formation of the heart in mammals (Buckingham et al., 2005). The primary heart field MCPs give rise to the left ventricle and cells of both atria, whereas the second heart field MCPs give rise to the right ventricle, atrial cells, and cells of the vascular outflow tract (Srivastava, 2006). Thus, different cell lineages constitute the mature heart, including cardiomyocytes, pacemaker cells, vascular cells, and smooth muscle cells (Martin-Puig et al., 2008). Lineage-tracing studies indicate that *Mesp1* marks the earliest cardiovascular progenitors of both heart fields (Saga et al., 1996; Saga et al., 1999). Moreover, both heart fields are marked by the expression of *Flk-1* and *Nkx2-5*, whereas *Isl1* mainly express in the secondary heart field (Ema et al., 2006; Moretti et al., 2006; Wu et al., 2006). In line with the discovery in mice embryos, the Flk-1^{+} cardiovascular progenitors were able to generate cardiac, endothelial, and vascular smooth muscle cells *in vitro* (Kattman et al., 2006).

Pluripotent ESCs can also be induced to undergo stepwise differentiation to differentiate into cardiomyocytes (Murry and Keller, 2008). Early differentiating ESCs can differentiated into a

transition PS-like (primitive streak-like) stage that will adopt either a mesoderm or an endoderm fate depending on the concentration of Wnt and Activin (Kattman et al., 2007). Mimicking gene expression during mouse gastrulation, *Mesp1* is expressed soon after the onset of *Brachyury* expression during mESC differentiation (Asahina et al., 2009; Liu et al., 2007; Ueno et al., 2007). Transcription factors that related to the core gene regulatory network of cardiovascular differentiation, such as *Nkx2-5*, *Gata4*, *Hand2*, and *Mef2c,* are expressed after *Mesp1* (Bondue et al., 2008; Lindsley et al., 2008; Liu et al., 2007; Ueno et al., 2007). Subsequently, the cardiac structural genes, such as *Myh6*, *Myl1*, *Myl2*, *Myl7*, and *Tnnt2* are also expressed (Lindsley et al., 2008). Meanwhile, *Mesp1* represses the expression of several genes that regulate the early steps of PS formation and early endoderm cell fate specification (Bondue and Blanpain, 2010; Bondue et al., 2008) (Figure 5-5). Wnt/β-catenin signaling was reported to promote the differentiation of mouse ESCs into mesoderm (Gadue et al., 2006; Lindsley et al., 2006; Ueno et al., 2007). However, this signaling inhibits cardiac differentiation after mesoderm is induced (Naito et al., 2006; Ueno et al., 2007). Therefore, Wnt signaling has a biphasic role in cardiac differentiation in mouse ESCs.

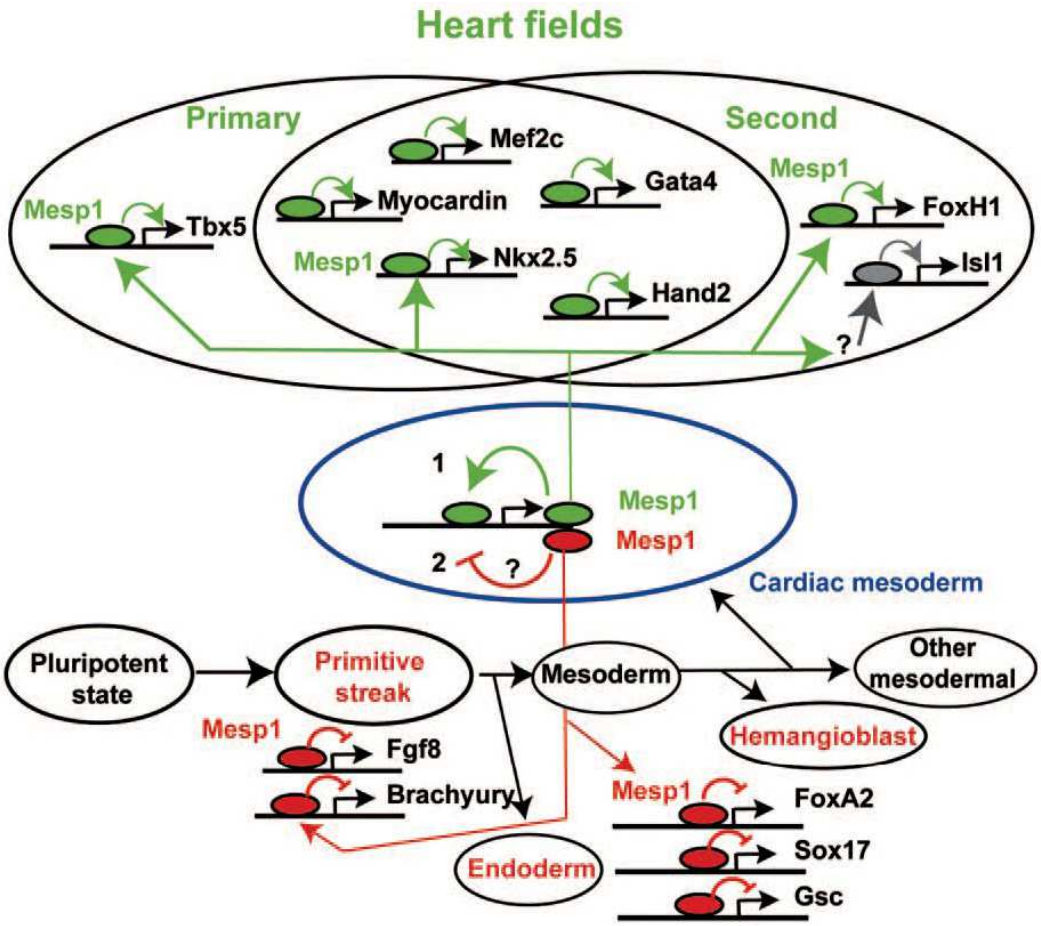


Figure 5-5: Model of Mesp1 regulates the cardiovascular transcriptional network. *Mesp1* promotes the expression of many key transcription factors promoting cardiac cell fate, on the other hand, *Mesp1* also represses several key genes that promote mesendoderm cell fate. (Bondue and Blanpain, 2010)

1.5.3 The pluripotency of mice embryonic stem cells (mESCs)

The molecular analysis of embryonic signalling is often limited by the small size and heterogeneity of embryonic tissues. The generation of ESCs and the development of ESC differentiation technique that mimic embryonic cell differentiation have made a progress to solve this problem (Martin, 1981; Murry and Keller, 2008).

1.5.3.1 Native and primed pluripotency states

Pluripotency cells have the ability to develop into the three primary germ cell layers of the early embryo and possibly primordial germ cells (PGCs), but not extra-embryonic tissues (Hanna et al.,

2010).Mouse ESCs derived from pre-implantation blastocyst represent "naive" pluripotency (Nagy et al., 1993). In contrast, epiblast-derived stem cells (EpiSCs) or epiblast-derived stem cell-like cells (EpiSCLCs) represent "primed" state pluripotency, which resemble the early and late stage post-implantation epiblast cells (Brons et al., 2007; Tesar et al., 2007) (Figure 5-6).

ESCs cultured in a serum-free 2i (GSK3 and MEK inhibitors) medium with LIF (2i ESCs) exhibit greater level pluripotent gene expression than ESCs cultured in serum with LIF (serum ESCs)(Sim et al., 2017). Serum ESCs are heterogeneous and have different transcriptional and epigenetic compared with pre-implantation embryo derived cells (Habibi et al., 2013; Marks et al., 2012). By contrast, 2i ESCs are more resemble the pre-implantation epiblast derived cells (Habibi et al., 2013; Marks et al., 2012; Ying et al., 2008). Nonetheless, both serum and 2i-cultured ESCs contribute to chimaera formation when injected into a blastocyst or when used in tetraploid complementation assays, therefore, they represent two types of "naive" pluripotency (Atlasi and Stunnenberg, 2017). Whereas the "primed" state pluripotency EpiSCs and EpiSCLCs do not contribute to chimaera formation (Weinberger et al., 2016). Similar to ESCs, EpiSCs express core pluripotency factors, including *POU5F1*, *Sox2*, and *Nanog*. However, they also express several differentiated transcripts which indicate the primed state. Notably, the two phases of pluripotent cells are reversely changeable. ESCs can be differentiated into EpiSCs by exposing to activin and Fgf factors in culture. Meanwhile, EpiSCs can be reprogrammed to naive pluripotency by transfection of a single factor, *Klf4* (Guo et al., 2009).

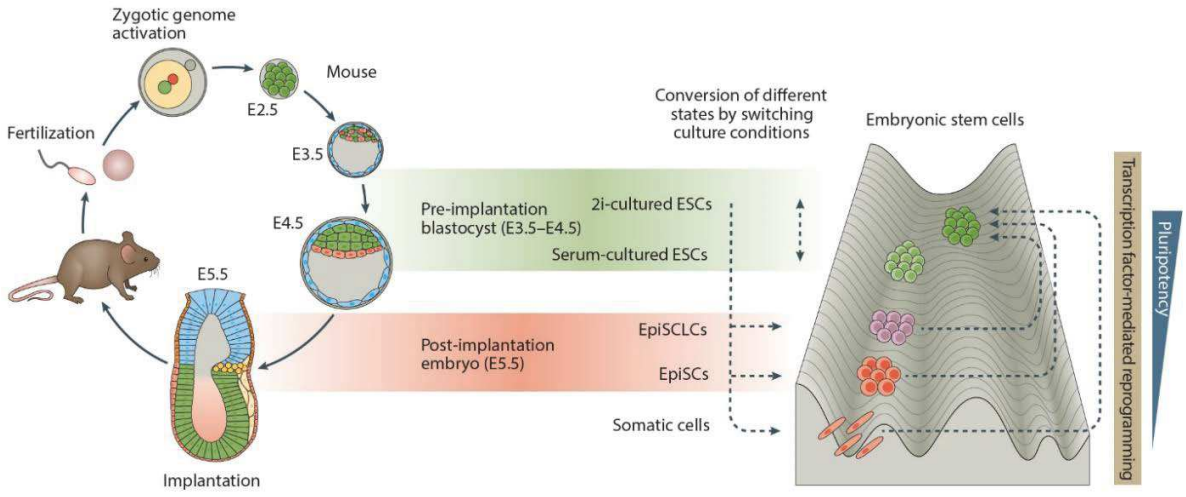


Figure 5-6: Two Phases of Pluripotency. Ground state naive pluripotency is established in the epiblast of pre-implantation blastocyst. Shortly after implantation, the epiblast transforms into a cup-shaped epithelium and becomes primed for lineage specification. Primed-state pluripotency has also been captured *in vitro* as cultured EpiSCs or EpiSCLCs resemble the early and late stage post-implantation epiblast cells, respectively. (Atlasi and Stunnenberg, 2017)

1.5.3.2 Molecular pathways involved in the maintenance of pluripotency

The core transcription factors, such as *Pou5f1*, *Sox2* and *Nanog*, are essential for ESCs to maintain a stable pluripotent state and inhibit differentiation (Avilion et al., 2003; Boyer et al., 2005; Kagey et al., 2010; Lee et al., 2006; Masui et al., 2007). Besides, another transcription factor c-Myc also plays important roles in maintaining ESC self-renewal by binding to E-box elements at core promoter sites (Cartwright et al., 2005) or recruiting transcription elongation factor p-TEFb (Rahl et al., 2010). These findings suggest that the core transcription factors choose the genes that will be actively transcribed, while c-Myc mainly regulates the full transcriptional efficiency of these selected genes. Furthermore, *Tcf3*, *Smad1*, *Stat3*, *Esrrb*, *Sall4*, *Tbx3*, *Zfx, Ronin, Klf2, Klf4, Klf5,* and *Prdm14* transcription factors are also involved in the control of ESC state (Young, 2011). In conclusion, a network of various factors regulates the pluripotency of mESCs.

Several signaling pathways, such as the leukaemia inhibitory factor (LIF), Wingless type (Wnt), and transforming growth factor (TGF)-β/bone morphogenetic protein (BMP) signaling pathways, are found to modulate mESC stem-cell pluripotency (Brandenberger et al., 2004; Sato et al., 2004; Sato et al., 2003; Williams et al., 1988; Ying et al., 2003) (Figure 5-7).

As described before, conventional mouse ESCs are cultured in serum medium supplemented with LIF. LIF signaling is not essential for pluripotency *in vivo* (Stewart et al., 1992), but supports the self-renewal of mouse ESCs (Darnell, 1996; Niwa et al., 1998). In the presence of LIF, STAT3 binds to phosphor-tyrosine residues on activated LIFR–gp130 heterodimer receptors and undergoes phosphorylation and dimerization itself (Darnell, 1996). Then phosphorylated STAT3 dimers translocate to the nucleus and function as transcription factors. Moreover, ESCs cultured with LIF can induce the phosphorylation of extracellular signal-regulated protein kinases (Burdon

et al., 1999), and increase mitogen-activated protein kinase (MAPK) activity (Boeuf et al., 1997) through activation of tyrosine phosphatase tyrosine phosphatase-2 (SHP2) protein (Auernhammer et al., 2000).

BMP4 is an essential anti-neurogenesis factor in the embryo, since ESCs differentiate into neurons in the absence of BMP4 (Ying et al., 2003). Interestingly, BMP4 displays distinct functions according to the statement of LIF. For example, in the presence of LIF, BMP4 promotes LIF cascade. Then SMAD4 further activates members of *inhibitor of differentiation (Id)* gene to enhance the self-renewal and pluripotency of ESCs (Ying et al., 2003). By contrast, in the absence of LIF, BMP4 counteracts the LIF cascade via interacting with different SMAD transcription factors that have an inhibitory effect on the *Id* gene. Above all, the balance between LIF and BMP4 is jointly responsible for maintaining the undifferentiated state of mouse ESCs (Boiani and Scholer, 2005).

Moreover, WNT are secreted glycosylated proteins that have widespread roles in tissue differentiation (Cadigan and Nusse, 1997). The canonical WNT pathway is activated upon binding of the WNT protein to the Frizzled receptor at the cell membrane. Activated WNT pathway leads to inhibition of glycogen-synthase kinase-3 (GSK3), subsequent nuclear accumulate β-catenin and express targeted pluripotent transcription factors. Similarly, small molecule inhibitor (CHIR99021) inhibits GSK3 has been essential in the maintenance of embryonic stem cells (ESCs)(Doble and Woodgett, 2003). To block ESC commitment, another small-molecule inhibitor (PD0325901) was used to inhibit the mitogen-activated protein kinase (MAPK)/extracellular signal-related kinase (ERK1/2) (Ying et al., 2008).

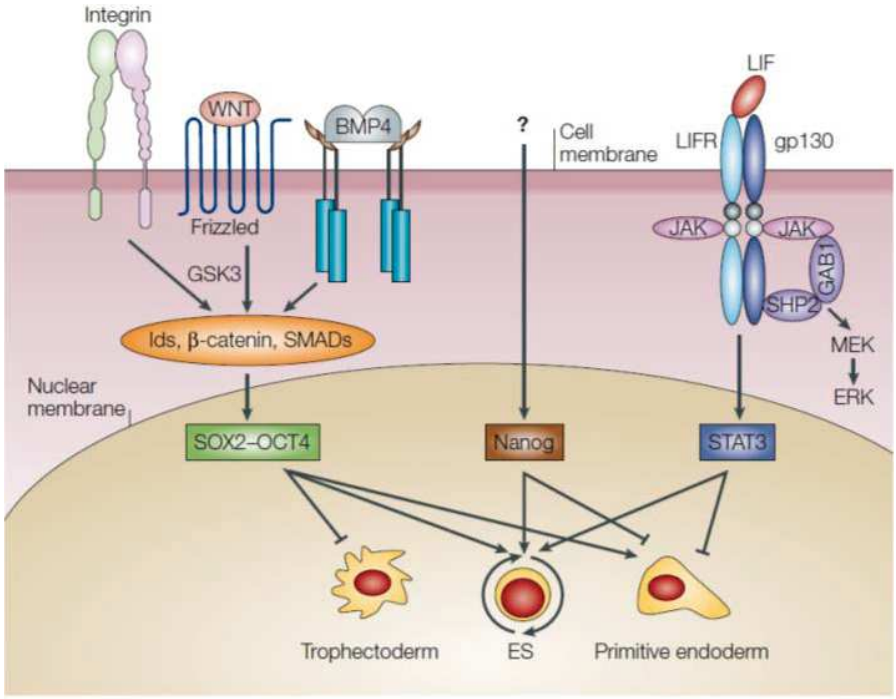


Figure 5-7: Combinatorial signalling pathways involved in maintaining mouse ESC pluripotency. Cell-surface receptors initiate signals to the nucleus and affect key pluripotency transcription factors such as OCT4 and Nanog, and self-renewal transcription factors such as STAT3 (Boiani and Scholer, 2005).

1.5.4 Epigenetic modification regulates development

Epigenetic modification of chromatin provides the necessary plasticity for cells to respond to environmental and developmental cues. In this section, we discuss the dynamics epigenetic in the early embryo and the *in vitro*-cultured ESCs.

Multiple waves of epigenetic resetting take place during early embryo development. The first wave occurs after fertilization when the epigenomes of differentiated gametes (sperm and oocyte) undergo reprogramming (Lee et al., 2014). The second wave occurs during blastocyst formation. At this time in female cells, both X chromosomes are reactivated with erased DNA methylation and increased chromatin accessibility (Tang et al., 2016; Weinberger et al., 2016). The third wave takes place after implantation, in which the chromatin was deposited of repressive epigenetic markers and showed less chromatin accessibility. Moreover, one X chromosome is randomly inactivated in female cells (Brons et al., 2007; Weinberger et al., 2016). Similar to preimplantation

epiblast, female mESCs share the epigenetic features of two active X chromosomes in female cells (Heard, 2004). Whereas EpiSCs, which are the counterpart of primed epiblast, epigenetically silenced one copy of the X chromosome in female cells (Guo et al., 2009).

The chromatin of pluripotent stem cells has a unique epigenetic plasticity that enables cells to undergo a wide range of lineage specifications. For example, pluripotent cells have open chromatin configuration associated with reduced DNA methylation and reduced H3K27me3 levels, which become progressively restricted during development (Buecker et al., 2014; Gafni et al., 2013). ESCs also contain many bivalent genes. The chromatin of bivalency genes accumulates both H3K4me3 and H3K27me3 histone modifications. As ESCs differentiate, bivalent loci lose one of the two histone marks. However, bivalency is not an essential feature for ESC pluripotency and self-renewal. For example, mESCs cultured in serum-free 2i (GSK3 and MEK inhibitors) medium mostly lose promoter bivalency, whereas they maintain normal self-renewal and lineage differentiation potential (Galonska et al., 2015; Smith and Meissner, 2013).

In addition to the bivalency histone modification, the HAT and DUB modules of SAGA are also play essential roles for embryonic development. Individual loss of the SAGA subunits *Gcn5, Ada2b, Ada3* and *Sgf11* results in developmental defects and larval lethality in *drosophila* (Pankotai et al., 2005; Qi et al., 2004; Weake et al., 2009). Similarly, *Gcn5* deletion in mice leads to defect in mesoderm development and embryonic lethality (Xu et al., 2000). Moreover, *Nonstop* that is the homology of mammalian *Usp22*, controls the development of neuronal connectivity visual system by regulating H2B deubiquitination (Weake et al., 2008). Whereas *Usp22* deletion results in mice embryonic death around E14 due to defect in vasculature formation (Koutelou et al., 2019). In addition, histone H2Bub1 that is the target of the DUBm of SAGA is required for the optimal ESC differentiation (Fuchs et al., 2012). GCN5 regulates proper expression of FGF signaling pathway-related genes during early embryoid bodies differentiation (Wang et al., 2018). Interestingly, SAGA seems to regulate genes in a tissue specific manner. For example, *Ada2b* interacted with more transcription factors in muscle than neurons in *Drosophila* embryos. Consistently, *Ada2b* occupied more genes specifically in muscle than in neurons (Weake et al., 2011). Moreover, the DUBm of SAGA is more important for the expression of muscle-specific

developmental genes relative to the whole embryo in *Drosophila* (Weake et al., 2011). However, the mechanism for this is still unclear.

2. The aim of the project

Monoubiquitylation of histone H2B (H2Bub1) plays a non-degradative role in multiple molecular and biological processes, including transcription activation, elongation, mRNA splicing, mRNA export, as well as DNA damage repair (Fuchs and Oren, 2014). Compared with other histone modifications, such as H3K36me3 and H3K79me2, H2Bub1 is highly dynamic during transcription process (Fuchs et al., 2014). Notably, the DUBm of SAGA can erase H2Bub1 within 10 mins in the wake of elongating Pol II (Bonnet et al., 2014). SAGA is an evolutionary conserved multi-subunit co-activator complex with a modular organization. The DUBm of SAGA is composed of USP22 and three adaptor proteins, ATXN7, ATXN7L3 and ENY2, which are all needed for the full enzymatic activity of USP22 to remove monoubiquitin (ub1) from histone H2B. However, the function of H2Bub1 deubiquitination is still not fully clarified.

To better understand the role of histone H2Bub1 deubiquitination by the SAGA DUBm and the two other related DUBm-s, during my PhD work I focused on analyzing the function of these DUBm-s in three aspects:

(1) Uncover the role of the SAGA DUBm and the two other related DUBm-s in the processes of mouse embryonic development and mESC differentiation;

(2) Test the link between H2Bub1 deubiquitination and transcriptional regulation genome-wide;

(3) Find the novel protein targets of the ATXN7L3-related DUBm-s and start to analyze their roles.

3. Results

3.1 Submitted publication

Histone H2Bub1 deubiquitylation is essential for mouse development, but does not regulate global RNA polymerase II transcription

Farrah El-Saafin[1,2,3,4,*,∞] , Fang Wang[1,2,3,4,*], Tao Ye[1,2,3,4,5], Matthieu Stierle[1,2,3,4], Matej Durik[1,2,3,4], Veronique Fischer[1,2,3,4], Didier Devys[1,2,3,4], Stéphane D. Vincent[1,2,3,4,], and László Tora[1,2,3,4,#]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France;

[2]Centre National de la Recherche Scientifique (CNRS), UMR7104, 67404 Illkirch, France;

[3]Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, 67404 Illkirch, France;

[4]Université de Strasbourg, 67404 Illkirch, France;

[5]Plateforme GenomEast, infrastructure France Génomique; 67404 Illkirch, France.

*These authors contributed equally to this work

[∞]Present address: Olivia Newton-John Cancer Research Institute, Melbourne, Victoria, Australia

[#]Corresponding author: László Tora; Development and stem cells Department
Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), UMR 7104 CNRS, INSERM U1258, Université de Strasbourg (Unistra), 1, rue Laurent Fries, 67404 ILLKIRCH Cedex, FRANCE; Tel +33 3 88 65 34 44, Fax: +33 3 88 65 32 01; e-mail: laszlo@igbmc.fr

Conflict of interest: The authors declare no conflict of interest.

Running Title: H2Bub DUBs do not regulate transcription directly

Key words: histone, ubiquitin, deubiquitylase, SAGA (Spt-Ada-Gcn5 acetyltransferase) complex, RNA polymerase II, transcription, ubiquitin-specific protease 22 (USP22), knock-out, mouse embryo, development, mESC, MEF, RNA-seq, ChIP-seq.

Funding: This study was supported by grants from European Research Council (ERC) (ERC-2013-Advanced grant 340551, Birtoaction), Agence Nationale de la Recherche (ANR) PICen-19-CE11-0003-02 and EpiCAST-19-CE12-0029-01 grants, NIH 1R01GM131626-01 grant (to LT) and ANR-18-CE12-0026 grant (to DD), fellowships by the IdEx-University of Strasbourg international PhD program and by the 'Fondation pour la Recherche Médicale' (FRM) association (FDT201904008368) (to VF) and an ANR-10-LABX-0030-INRT grant, a French State fund managed by the ANR under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02.

Abstract

Co-activator complexes dynamically deposit post-translational modifications (PTMs) on histones, or remove them, to regulate chromatin accessibility and/or to create/erase docking surfaces for proteins that recognize histone PTMs. SAGA (Spt-Ada-Gcn5 Acetyltransferase) is an evolutionary conserved multisubunit co-activator complex with modular organization. The deubiquitylation module (DUB) of mammalian SAGA complex is composed of the ubiquitin-specific protease 22 (USP22) and three adaptor proteins, ATXN7, ATXN7L3 and ENY2, which are all needed for the full activity of the USP22 enzyme to remove monoubiquitin (ub1) from histone H2B. Two additional USP22-related ubiquitin hydrolases (called USP27X or USP51) have been described to form alternative DUBs with ATXN7L3 and ENY2, which can also deubiquitylate H2Bub1. Here we report that USP22 and ATXN7L3 are essential for normal embryonic development of mice, however their requirements are not identical during this process, as *Atxn7l3*$^{-/-}$ mutant embryos show developmental delay already at embryonic day (E) 8.5, while *Usp22*$^{-/-}$ mutant embryos are normal at this stage, but die at E14.5. Global histone H2Bub1 levels were only slightly affected in *Usp22* null embryos, in contrast H2Bub1 levels were strongly increased in *Atxn7l3* null embryos and derived cell lines. Our transcriptomic analyses carried out from wild type and *Atxn7l3* null mutant mouse embryonic stem cells (mESCs), or primary embryonic fibroblasts (MEFs) suggest that the ATXN7L3-related DUB activity regulates only a subset of genes in both cell types. However, the gene sets and the extent of their deregulation were different in mESCs and MEFs. Interestingly, the strong increase of H2Bub1 levels observed in the *Atxn7l3*$^{-/-}$ mESCs, or *Atxn7l3*$^{-/-}$ MEFs, do not correlate with the modest changes in RNA Polymerase II occupancy observed in the two *Atxn7l3*$^{-/-}$ cellular systems. These

66 observations together indicate that deubiquitylation of histone H2Bub1 does not directly
67 regulate global RNA polymerase II transcription.
68

Introduction

Nucleosomes, composed of a histone $(H3\text{-}H4)_2$ tetramer, flanked by two histone H2A-H2B dimers, wrapped by 147 base pairs of DNA, play a key role in chromatin compaction [1, 2]. The globular domains and the histone tails, which extend from the nucleosome, are the substrates for a vast variety of enzymes carrying out diverse post-translational modifications (PTMs), such as acetylation, phosphorylation, methylation and ubiquitylation [3, 4]. These PTMs are viewed as modulators of accessibility and compaction of chromatin fibres, which regulate essential processes such as transcription, DNA damage repair, chromosome compaction and segregation [5, 6]. Enzymes, often incorporated in multiprotein complexes, are responsible for the addition (writers) or removal (erasers) of specific histone modifications [7]. During mouse embryonic development, dynamic modifications of the chromatin are essential, as the loss of chromatin modifying enzymes, both writers and erasers, can lead to embryonic lethality, although with different severity [8].

Histone H2B can be modified by the dynamic addition or removal of a single ubiquitin (ub1) molecule on lysine 123 in yeast, and on lysine 120 in mammals (H2Bub1). The deposition of mono-ubiquitin onto histone H2B is catalysed by Bre1 in yeast, and by the RNF20/RNF40 complex in mammals [9, 10, 11]. The exact cellular function(s) of the H2Bub1 chromatin mark is not yet fully understood, however it was suggested that the deposition of ubiquitin onto H2B weakens DNA-histone interactions and therefore disrupts chromatin compaction [12]. Histone H2Bub1 mark was suggested to play a role in several DNA-related and epigenetically regulated processes, such as transcription, repair, replication, homologous recombination, as well as in mRNA processing and export [13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. Indeed, chromatin immunoprecipitation

coupled to sequencing (ChIP-seq) studies revealed that H2Bub1 is found at gene bodies of expressed genes and absent from non-expressed chromosomal regions, suggesting that H2Bub1 may be involved in transcription elongation [23, 24, 25, 26, 27]. Intriguingly however, when H2Bub1 deposition was disrupted in mammalian cells by knock-down or knock-out depletion of RNF20, or RNF40, respectively, the expression of only a small subset of genes was affected [14, 19, 28]. Histone H2Bub1 deposition has also been implicated in histone cross talk and has been shown to be a prerequisite for trimethylation of histone 3 lysine 4 (H3K4me3) around promoter regions both in yeast and mammalian cells [29, 30, 31, 32, 33, 34].

H2Bub1 is erased by the de-ubiquitylation (DUB) module of the co-activator SAGA (Spt-Ada-Gcn5 acetyltransferase) complex [35, 36, 37, 38]. The DUB module of the mammalian SAGA complex is composed of the ubiquitin-specific protease 22 (USP22) and the ATXN7, ATXN7L3 and ENY2 adaptor proteins, which are all needed for the full activity of USP22 enzyme [39]. In addition, ATXN7L3 is critical for directing the DUB module substrate specificity towards H2Bub1 [40]. In human cells, depletion of either ENY2 or ATXN7L3 adaptor protein resulted in a non-functional USP22 enzyme, and consequently H2Bub1 was not removed from the genome [22, 26, 39, 41]. It has also been described that two additional USP22-related ubiquitin hydrolases (called USP27X or USP51) can interact with ATXN7L3 and ENY2, which can also deubiquitylate H2Bub1 independently of the SAGA complex [41]. Thus, in mammalian cells the cellular abundance of histone H2Bub1 is regulated by the opposing activities of the ubiquitin E3 ligase complex, RNF20/RNF40, and three related DUB modules, each containing one of the homologuous deubiquitylases: USP22, USP27X or USP51 [41, 42]. USP22-, USP27X- and USP51-containing DUB modules also have non-histone substrates, including TRF1

[43], FBP1 [44], SIRT1 [45, 46], HES1 [47], SNAIL1 [48], and ZEB1 [49]. As *Usp22* null mouse embryos have been described to die around embryonic day (E) E14.5 [46, 50, 51], it seems that the alternative USP27X- and/or USP51-containing DUB modules cannot completely fulfil the role of the USP22-containing DUB module, further suggesting that the three related DUB modules may also have specific functions. *Usp22* mouse mutant studies suggest that USP22 is required to regulate apoptosis by deubiquitylating/stabilizing the class III histone deacetylase SIRT1 and by suppressing p53 functions under DNA damage during embryonic development and/or that USP22 is required for regulating multiple key signalling pathways crucial for vasculature formation in the mouse placenta [46, 50]. Note that no significant phenotypes in *Usp27x* knock out (KO) mouse embryos has been described [47].

Many human cancers exhibit dramatically misregulated levels of H2Bub1 [52, 53, 54] and also the factors involved in the deposition and erasure of H2Bub1 are misregulated in many cancers [55, 56, 57, 58, 59] suggesting that H2Bub1 may play an important role in normal cellular homeostasis [60]. Interestingly, during myogenic differentiation an apparent disconnection between the H2Bub1-H3K4me3 crosstalk was described, as in differentiated myotubes H2Bub1 levels were undetectable, but H3K4me3 levels did not globally change [61]. Moreover, it has been reported that optimal mouse embryonic stem cell (mESC) differentiation requires dynamic changes in histone H2B ubiquitylation patterns, which must occur in a timely and well-coordinated manner [62].

To better understand the role of USP22- and/or ATXN7L3-containing DUB modules in a physiological context and during development, we have generated mice that lack either USP22 or ATXN7L3. *Atxn7l3* null mutants show developmental delay as early as E8.5, while *Usp22*$^{-/-}$ mutant embryos are normal at this stage, but die at E14.5

similarly to what was previously published [46, 50]. These results indicate that USP22 and ATXN7L3 are essential for normal embryonic development, however their requirements are not identical during this process. Histone H2Bub1 levels were only slightly affected in *Usp22* null embryos, while in contrast H2Bub1 levels were strongly increased in *Atxn7l3* null mutants and derived cellular systems. The genome-wide increase of H2Bub1 retention in mESCs and mouse embryo fibroblasts (MEFs) lacking ATXN7L3 was investigated and the consequences of *Atxn7l3* mutation on cellular homeostasis, differentiation, and RNA polymerase II (Pol II) transcription were analysed.

Materials and Methods

Generation and maintenance of *Usp22*$^{+/-}$ and *Atxn7l3*$^{+/-}$ mouse lines

Usp22$^{+/-}$ and *Atxn7l3*$^{+/-}$ mouse lines were generated at the Institut Clinique de la Souris (ICS, Illkirch, France) using mESCs containing the targeting constructs ordered from the International Knockout Mouse Consortium (IKMC), including the Knockout Mouse Programme (KOMP) repository (UC, Davis). In the *Usp22* targeting construct (*Usp22*$^{tm1a(KOMP)Wtsi}$) a *LacZ* and *Neo* cassette were located in intron 1, flanked by *FRT* sequences, and *loxP* sequences were flanking exon 2 (Supplementary Figure 1A). In the *Atxn7l3* targeting construct (*Atxn7l3*$^{tm1.1(KOMP)Wtsi}$) a *LacZ* and *Neo* cassette were located in intron 2, flanked by *FRT* sequences, and the *loxP* sequences were flanking exon 2 to exon 12 (Supplementary Figure 1C). Chimeras were generated by injecting the C57BL/6 mESCs containing the targeting constructs into BALB/C blastocysts. Mice heterozygous for the targeting allele were crossed to a Cre-recombinase deleter strain, in order to generate the null alleles *Usp22*$^{-}$ and *Atxn7l3*$^{-}$, then mice heterozygous for the null allele (*Usp22*$^{+/-}$ or Atxn7l3$^{+/-}$) were intercrossed to generate homozygous mutant

165 embryos (*Usp22*$^{-/-}$ or Atxn7l3$^{-/-}$) as shown in Supplementary Figure 1A and 1C.
166 Genotyping primers are shown in Supplementary Table 1, and example genotyping gels
167 are shown in Supplementary Figure 1B and 1D. The *Atxn7l3*$^{+/-}$ mice were maintained on
168 a mixed B6D2 background. Animal experimentation was carried out according to animal
169 welfare regulations and guidelines of the French Ministry of Agriculture and French
170 Ministry of Higher Education, Research and Innovation.

171 **Generation and maintenance of *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEFs**

172 To generate *Usp22*$^{-/-}$, *Atxn7l3*$^{-/-}$ and control mESCs, timed matings between
173 heterozygous mice were conducted, then at E3.5, pregnant females were sacrificed,
174 uteri were flushed with M2 medium (Sigma-Aldrich), and individual blastocysts were
175 transferred to wells of a 96-well plates pre-coated with 0.1% gelatin. Blastocysts were
176 cultured and expanded in regular mESCs medium (DMEM (4.5 g/l glucose) with 2 mM
177 Glutamax-I, 15% ESQ FBS (Gibco), penicillin, streptomycin, 0.1 mM non-essential
178 amino acids, 0.1% ß-mercaptoethanol, 1500 U/mL LIF and two inhibitors (2i; 3 µM
179 CHIR99021 and 1 µM PD0325901, Axon MedChem). After expansion, mESCs were
180 genotyped and frozen.

181 To generate *Atxn7l3*$^{-/-}$ and control mouse embryonic fibroblasts (MEFs), timed
182 matings between heterozygous mice were conducted, then at E10.5, pregnant females
183 were sacrificed, and embryos were collected. The embryo yolk sacs were collected for
184 genotyping, and the head and gastrointestinal tract were carefully dissected away from
185 embryos. The remaining carcasses were transferred to individual 1.5 ml Eppendorf
186 tubes, and 50 µl of trypsin (0.25% in EDTA, Gibco) was added and gently triturated 5
187 times to dissociate the embryos. The dissociated embryos were incubated in trypsin for
188 5 min at room temperature, then the trypsin was quenched with 500 µl of FCS. Cells

were transferred to individual wells of a 6-well plate pre-coated with 0.1% gelatin and cultured in MEF medium (DMEM, 10% FCS, penicillin and streptomycin). Cells were visualized with a EVOS XL Core Cell Imaging System (#AMEX-1100, Thermo Fisher Scientific) using a LPlan PH2 10x / 0.25 objective.

Protein extraction and Western blot assays

To extract histone proteins, embryos dissected at the indicated embryonic days, or about 5 x10^6 cells were lysed with 100 µl acidic extraction buffer (10 mM Hepes, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT and 0.2 M HCl) freshly complemented with 1× Proteinase Inhibitor Cocktail (Roche) and 10 mM N-ethylmaleimide (Sigma-Aldrich) and incubated on an end-to-end rotator for 2 hours at 4°C. Following the incubation, cell extract was centrifuged at 20 800 x g for 10 min at 4°C, to pellet the acid insoluble material. Ten µl of the supernatant, containing histone proteins, were run on 4–12% gels (Bis-tris NuPAGE Novex, Life Technologies), then proteins were transferred and western blot assays were carried out by using standard methods. The following antibodies were used: anti-H3 (Abcam, #ab1791) anti-H4 (Invitrogen, 3HH4-4G8), anti-H2Bub1 (Cell Signaling Technology, #5546), anti-H3K4me3 (Abcam #8580) and anti-H3K9ac (Merck-Millipore #07-352) were used. Protein levels were quantified by ImageJ.

Actin labelling

Cells were washed twice with 1x PBS, fixed with 4% PFA (Electron Microscopy Science) for 10 min at RT. After fixation, cells were washed three times with 1x PBS, permeabilized with sterile 0.1% Triton X-100 in PBS for 20 min at RT, then washed three times in 1x PSB. Cells were incubated either with phalloidin conjugated to Alexa 488 dye (Phalloidin-iFluor 488, Abcam, as described in the manufacturer's protocol) to label F-actin filaments, or with an anti-β-actin mouse monoclonal antibody (Sigma

Aldrich, A5441) at a dilution of 1:1000 in 1x PBS with 10% FCS, overnight at 4°C. The following day, cells were washed three times with 1x PBS, then β-actin labelled cells were further incubated with secondary goat anti-mouse Alexa 488 antibody (Invitrogen) at a dilution of 1:2000 in 1x PBS with 10% FCS for 1 hr at RT. The cells were washed three times with 1x PBS, then incubated in 20 mM Hoechst 3342 (Thermo Scientific) for 10 min at RT, before being washed three times with 1x PBS, then cells were covered with a coverslip coated in ProLong Gold mounting medium (Invitrogen). Pictures were taken using a Leica DM 4000 B upright microscope equipped with a Photometrics CoolSnap CF Color camera with a HCX PL S-APO 20x/0.50 objective.

Colony formation assay and alkaline phosphatase staining

Three thousand mESCs were seeded on gelatin-coated 6-well plates in regular mESC medium (see above) to form colonies at low density. The medium was exchanged every two days for 6 days. mESC alkaline phosphatase (AP) activity test was performed using Red Substrate Kit, Alkaline Phosphatase (Vector Laboratories) according to the manufacturer's instructions. mESC clones were washed with 1x cold PBS and fixed with 4% PFA for 10 min at RT. After fixation, cells were washed twice with 1x PBS and incubated in 1 ml AP detection system (as recommended by the manufacturer's protocol) for 30 min at RT in the dark. Then cells were washed twice with cold 1x PBS, and visualized with a EVOS XL Core Cell Imaging System (#AMEX-1100, Thermo Fisher Scientific) using a LPlan PH2 4x / 0.13 objective.

Cell proliferation analysis

To determine cell proliferation, a total of $1x10^5$ mESCs per 6-well plate were seeded in regular mESC medium and $3x10^4$ passage two MEF cells per 24-well plate were seeded in MEF medium. The medium was exchanged every two days. Cell

numbers were counted with Countess cell counting chambers (Invitrogen). Statistical analyses were determined by the Mann-Whitney test (ns $p>0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.05$).

Cell cycle analysis

Hundred thousand mESCs were fixed in 70% EtOH overnight at -4°C. After fixation, cells were treated with RNase A (100 µg/ml) (Thermo Fisher Scientific, #EN0531) and stained with propidium iodide (40 µg/ml) (Sigma Aldrich, #P-4170) for 30 min at 37 °C. The acquisition of the DNA content was analysed on FACS CALIBUR (BD Sciences) flow cytometer. Quantitative results were analyzed by FlowJo software (BD Sciences).

Apoptosis analysis using annexin-V staining

At the indicated incubation time, floating cells were collected in culture supernatants and adherent cells were harvested by trypsinization. After collection, cells were washed twice with cold 1X PBS, and about $2x10^5$ cells were resuspend in 100 µl binding buffer (FITC Annexin V Apoptosis Detection Kit, Biolegend). Subsequently, 5 µl FITC Annexin V (FITC Annexin V Apoptosis Detection Kit, Biolegend) and 10 µl propidium iodide was added to the cell suspension. Cells were gently vortexed and incubated in the dark for 15 min at RT. Thereafter, another 400 µl Annexin V binding buffer was added to each tube. Cells were analysed using a FACS CALIBUR (BD Sciences) flow cytometer. Dot plots were generated using the FlowJo software.

RNA-seq and ChIP-seq analyses

For RNA-seq, total RNA was extracted from mESCs and MEFs using the NucleoSpin RNA isolation kit (Macherey-Nagel), according to manufacturer's

instructions. Libraries were generated from the purified RNA using TruSeq Stranded mRNA (Illumina) protocol. After checking the quality of the libraries with the Bioanalyser (Agilent), libraries were sequenced on the Illumina HiSeq 4000 at the GenomEast sequencing platform of IGBMC. The raw sequencing data generated reads were preprocessed in order to remove adapter, polyA and low-quality sequences (Phred quality score below 20), then were mapped to the mouse mm10 genome using STAR [63]. Differential gene expression was measured using the DESeq2 package [64]. For the analysis, only the transcripts expressed more than 100 normalized reads (DESeq2 reads divided by the median of the transcript length in kb) were considered. Using these criteria 11 172 transcripts were expressed in mESCs, and 11 113 transcripts were expressed in MEFs.

ChIP-seq experiments were performed using the protocol described in [65], with some minor modifications, including the use of 10 mM N-ethylmaleimide (Sigma-Aldrich) in all buffers and the use of either the anti-H2Bub1 antibody (MediMabs, NRO3) or the anti-RPB1 CTD Pol II antibody (1PB 7G5; [66]). Briefly, mESCs or MEFs were fixed in 1% PFA for 10 min at room temperature (RT), then the PFA was quenched with glycine at a final concentration of 125 mM for 5 min at RT. Cells were washed two times in 1× cold PBS, scraped, and pelleted. Nuclei were isolated by incubating cells with nuclear isolation buffer (50 mM Tris–HCl pH 8.0, 2 mM EDTA pH 8.0, 0.5% Nonidet P-40, 10% glycerol, 1× protease inhibitors and 10 mM NEM) for 10 min at 4°C with gentle agitation, followed by centrifugation at max speed to pellet the nuclei. Nuclei were resuspended in sonication buffer (0.1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0, 1× protease inhibitors and 10 mM NEM) then chromatin was sheared with the E220 sonicator (Covaris) and chromatin concentration was measured with the Qubit 3.0

(Thermo Fischer Scientific). Approximately of 50 µg of chromatin was used for each IP, which was diluted in ChIP dilution buffer (0.5% Nonidet P-40, 16.7 mM Tris–HCl pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 1× protease inhibitor cocktail and 10mM NEM). Antibodies used for the ChIP included anti-RPB1 CTD (1PB 7G5; [66]) anti-H2Bub1 (MediMab, NRO3), and mouse IgG (Jackson Laboratories) which were incubated with the chromatin overnight with gentle agitation at 4°C. The next day, Dynabeads protein G magnetic beads (Invitrogen) were added for 1 hour, then were isolated and washed for 5 min at 4°C, once with low salt wash buffer (0.1% SDS, 0.5% Nonidet P-40, 2 mM EDTA, 150 mM NaCl, 20 mM and Tris-HCl pH 8.0), once with high salt wash buffer (0.1% SDS, 0.5% Nonidet P-40, 2 mM EDTA, 500 mM NaCl, 20 mM and Tris–HCl pH 8.0), and once with LiCl wash buffer (0.2 M LiCl, 0.5% Nonidet P-40, 0.5% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH8.0), then washed twice with TE buffer, then the beads were incubated in elution buffer (1% SDS, 0.1 M $NaHCO_3$) at 65°C with shaking to elute complexes. Crosslinks were reversed with by adding NaCl at a final concentration of 0.2 M overnight as well as 50 µg/ml RNase A at 65°C and the following day the samples were treated with 20 µg Proteinase K, 26.6 µl of 1 M Tris–HCl pH 7.9, and 13.3 µl of 0.5 M EDTA, and DNA was phenol/chloroform purified and precipitated. The precipitated DNA was used to generate libraries with the MicroPlex Library Preparation kit v2 (Diagenode) for ChIP-seq according to the manufacturer's instructions. The samples were then sequenced on HiSeq 4000 with read lengths of 1 × 50 bp, reads were mapped to the mouse mm10 genome. Samples were normalized and peak calling was performed using the MACS2 software.

Bioinformatics tools and data-analysis methods

Definition of the coordinates of mouse intergenic regions and normalization between ChIP datasets

3115 intergenic regions far away from genes and larger than 100 kb were selected as described previously [26], among them 2755 intergenic regions in mESC containing at least 1 read, and 2738 intergenic regions in MEF cells containing at least 1 read (Supplementary Table 4). The total reads present at these intergenic regions were used for normalization. We calculated the size factor of these intergenic regions for each sample using DESeq2 (version 1.16) [67]. These size factors were used to normalize the data.

Calculation of density values

Density values were defined as follows: density = [(number of aligned reads in a region of interest) / (length of the region of interest in bp)] / (size factor x 10^{-8}). For H2Bub1 datasets, we considered only the gene bodies of expressed genes containing at least 1 ChIP-seq read. Out of 11 172 expressed genes in mESCs, 11 010 contain at least 1 ChIP-seq read. Out of 11 113 expressed genes in MEF cells, 10 946 contain at least 1 ChIP-seq read (Supplementary Table 4).

Generation of average profiles and heat maps

Average profiles and k-means clustering were generated with the seqMINER program [68]. The end of each aligned read was extended to 200 bp in the direction of the read. For the analyses around promoters, the tag density was extracted in a 2 kb window centred on each TSS. For average gene profiles, each gene body was divided into 160 equal bins (the absolute size depending on the gene length). Moreover 20 equally sized bins (250 bp / bin) were created upstream and downstream of genes. Densities were collected for each dataset in each bin.

Data availability

All the datasets generated during the current study are available together in Gene Expression Omnibus (GEO) database under the accession number GSE153587. Individual RNA-seq data can be accessed at GSE153578 and ChIP-seq data at GSE153584.

Results

Loss of the DUB adaptor ATXN7L3 results in a more severe phenotype than the loss of the SAGA DUB enzyme USP22 *in vivo*

Homozygous inactivation of *Usp22* leads to embryonic lethality associated with placentation defects [46, 50]. In order to compare the deubiquitylation requirement for USP22 and ATXN7L3 *in vivo*, we generated *Usp22*$^{+/-}$ mice and *Atxn7l3*$^{+/-}$ mice from mouse mESCs generated by the Knockout Mouse Project (KOMP) Consortium (Figure S1A-D). Concerning the *Usp22*$^{+/-}$ mice we used the same mESC clone as was used by Kosinsky *et al.* (2015) [51], but further deleted *LacZ* and exon 2 of *Usp22*$^{tm1a(KOMP)Wtsi}$ by using FLP and Cre recombinases (see Figure S1A). Next *Usp22*$^{+/-}$*, or Atxn7l3*$^{+/-}$ mice were intercrossed to obtain *Atxn7l3*$^{-/-}$ and *Usp22*$^{-/-}$ homozygous mutants. *Usp22*$^{-/-}$ homozygous mutants started to resorb at E13.5 (Figure 1Ah) and could not be observed after E14.5, similarly to what has been previously published [46, 50] (Table 1, Figure 1A). Similarly, no *Atxn7l3*$^{-/-}$ mutant pups could be retrieved at weaning (Table 2), however, analysis of *Atxn7l3*$^{+/-}$ x *Atxn7l3*$^{+/-}$ litters collected at different stage of development revealed a more severe phenotype when compared to *Usp22*$^{-/-}$ mutants. A growth delay was already observed as early as E7.5 (Figure 1B). At E9.5, *Atxn7l3*$^{-/-}$ mutant embryos

did not turn (Figure 1Bf). From E10.5 onwards, two classes of phenotype were observed; severe and a mild, corresponding to 2/3 and 1/3 of the *Atxn7l3*$^{-/-}$ mutant embryos, respectively. No *Atxn7l3*$^{-/-}$ mutant embryos could be retrieved after E11.5 (Table 2). The mild class embryos were more similar to control embryos but were growth delayed (Figure 1Bi and 1Bl). In some instances, blood pooling could be observed (Figure 1Bh-Bl). The severe class embryos were smaller, failed to turn and displayed shortened trunk, abnormal head development, blood in the heart and enlarged pericardium (Figure 1Bh and 1Bk). Altogether, our data demonstrate that loss of the DUB adaptor protein ATXN7L3 has a more severe effect on embryonic development that the loss of the DUB enzyme, USP22, *in vivo*. This suggests that inactivation of the three related DUB modules in *Atxn7l3* null embryos results in a more severe phenotype than only the knockout of the USP22-containing DUB module.

In contrast to *Usp22*$^{-/-}$ embryos, *Atxn7l3*$^{-/-}$ embryos show strong increase in global H2Bub1 levels

Previous studies in HeLa and 293T cells have shown that depletion of the adaptor protein ATXN7L3 has a more severe effect on the H2Bub1 deubiquitylation activity than the depletion of the DUB enzyme UPS22 [39, 41]. To investigate the importance of USP22 and ATXN7L3 on H2Bub1 deubiquitylation *in vivo*, we analysed global H2Bub1 levels in acidic extracts from E10.5 or, E11.5 *Usp22*$^{-/-}$ and control embryos (Figure 1C), as well as E9.5 or E10.5 *Atxn7l3*$^{-/-}$ and control embryos (Figure 1D). While only minor changes (about 1.2 fold) were observed between controls and *Usp22*$^{-/-}$ mutant embryos lysates (Figure 1C and 1E), an about 4-5 fold increase in global H2Bub1 levels was observed in *Atxn7l3*$^{-/-}$ mutant embryo extracts (Figure 1D and 1F), confirming similar observations in human cells [39, 41]. Interestingly other chromatin marks, such as histone H3K4

trimethylation (H3K4me3; associated with active transcription and the deposition of H2Bub1) or H3K9 acetylation [H3K9ac; deposited by the histone acetyl transferase (HAT) module of SAGA complex] were not affected in *Usp22*$^{-/-}$ or *Atxn7l3*$^{-/-}$ or embryos (Figure 1C and 1D). These results suggest that ATXN7L3 is required for the full activity of the three related DUB modules to regulate global H2Bub1 levels, whereas USP22-containing DUB module is less involved in genome-wide deubiquitylation of H2Bub1. Alternatively, the SAGA deubiquitylation activity on H2Bub1 may be redundant and can be compensated by the two other related DUBs. Furthermore, these observations also suggest that H2Bub1 deubiquitylation and H3K4me3 deposition are not linked, and that the two enzymatic activities of SAGA are not interdependent, as described earlier [26].

Primary *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEF-like cells show abnormal proliferation and phenotypes

As *Usp22*$^{-/-}$ mouse phenotypes have been already described [46, 50] and as the *in vivo* H2Bub1 levels were only weakly affected in *Usp22*$^{-/-}$ mouse embryos, we concentrated our further analyses on *Atxn7l3*$^{-/-}$ mutants. To determine the mechanistic outcome of perturbed DUB function(s), we turned to defined and uniform cell types, such as pluripotent mouse mESCs and mouse embryonic fibroblast (MEF)-like cells, derived from *Atxn7l3*$^{-/-}$ embryos. To this end, from *Atxn7l3* heterozygous intercrosses, mESCs and primary MEFs were generated from E3.5 blastocysts, and E10.5 embryos, respectively. As expected from the *in vivo* data, in both of these *Atxn7l3*$^{-/-}$ cellular systems, global H2Bub1 levels were significantly upregulated, by almost 5-fold in mESCs and about 7.5-fold in MEFs (Figure 2A, and 2B).

When *Atxn7l3*$^{-/-}$ mESCs were analysed, we found that their alkaline phosphatase staining and the expression of pluripotency markers, such as *Pou5f1, Sox2, Klf4, Esrrb*

and *Tfcp2l1* [69], were similar to that in control mESCs (Figure 2C, and Figure S2A), indicating that the pluripotency potential of these cells was not significantly affected by the inactivation of *Atxn7l3*. Similarly, when apoptotic cell death and cell cycle phase distribution were measured, no significant differences were detected when comparing WT and *Atxn7l3*$^{-/-}$ mESCs (Supplementary Figure 2B and 2D). However, we observed that *Atxn7l3*$^{-/-}$ mESCs colonies were more irregular (Figure 2C) and proliferated slower (Figure 2D) as compared to control mESCs.

To study the role of ATXN7L3 in a more differentiated cellular environment, we analysed the phenotype of the MEF-like cells derived from E10.5 embryos. Although the *Atxn7l3*$^{-/-}$ embryos were developmentally delayed, the fact that MEF-like cells could be obtained at this stage suggests that embryonic fibroblasts or their progenitors exist in the *Atxn7l3*$^{-/-}$ embryos. Interestingly, in *Atxn7l3*$^{-/-}$ MEF cultures, many cells had an abnormal round morphology (Figure 2E, right panel). These round *Atxn7l3*$^{-/-}$ cells originated from clusters of cells that proliferated faster than elongated *Atxn7l3*$^{+/+}$ MEFs (Figure 2E). The round *Atxn7l3*$^{-/-}$ cells were present in all MEF cultures generated from E10.5 *Atxn7l3*$^{-/-}$ embryos (n>12 *Atxn7l3*$^{-/-}$ embryos), regardless of the severity of the *Atxn7l3*$^{-/-}$ embryo phenotype. However, the proportion of round cells relative to elongated cells appeared to correlate with the phenotype. Indeed, cultures generated from *Atxn7l3*$^{-/-}$ embryos with the “severe” phenotype, had a greater starting proportion of the round cells compared to MEF cultures generated from *Atxn7l3*$^{-/-}$ embryos with a mild phenotype (data not shown). When cell cycle phase distribution and apoptotic cell death were measured, no significant differences were detected when comparing WT and *Atxn7l3*$^{-/-}$ MEFs (Supplementary Figure 2C and 2E). However, we observed that *Atxn7l3*$^{-/-}$ MEF-s from passage 2 had a tendency to proliferate somewhat slower for the

first three days when compared to WT MEFs, but then started to grow quicker than control MEFs (Figure 2F).

These analyses suggest that the ablation of the ATXN7L3-linked DUB activity, and the resulting increased H2Bub1 levels do not result in severe phenotypic changes in *Atxn7l3*$^{-/-}$ mESCs, but to profound morphological changes and proliferation alterations in *Atxn7l3*$^{-/-}$ MEF-like cells.

Atxn7l3$^{-/-}$ mESCs and MEF-like cells show significant alteration of Pol II transcription, with deregulation of gene expression being more severe in _Atxn7l3_$^{-/-}$ MEF-like cells

To characterize the mESC and MEF transcriptomes and their dependence on ATXN7L3-dependent DUB activity, we measured changes in steady state mRNA levels between *Atxn7l3*$^{+/+}$ and *Atxn7l3*$^{-/-}$ mESCs, as well as between *Atxn7l3*$^{+/+}$ and *Atxn7l3*$^{-/-}$ MEFs by carrying out RNA-seq of polyA+ mRNA. Principal component analysis showed that RNA-seq data obtained from *Atxn7l3*$^{+/+}$ or *Atxn7l3*$^{-/-}$ mESCs, and *Atxn7l3*$^{+/+}$ or *Atxn7l3*$^{-/-}$ MEF-like cells clustered in individual groups, indicating that the main explanation for the variance is the genotype (Supplementary Figure 3A and 3B). As the *Atxn7l3*$^{-/-}$ MEF-like cells display major morphological changes when compared to wild-type MEFs, we first verified whether these cells still belong to the MEF lineage in spite of their unusual morphology. To this end, we investigated whether these cells still maintained a "MEF signature" by comparing the RNA-seq results from three individual *Atxn7l3*$^{-/-}$ and control MEF samples with 921 RNA-seq data from 272 distinct mouse cell types or tissues [70]. This clustering analysis indicated that the *Atxn7l3*$^{-/-}$ MEF-like cells grouped together with *Atxn7l3*$^{+/+}$ MEFs or fibroblasts (Supplementary Figure 3C), suggesting that the mutant cells belong to the fibroblast lineage.

Comparison of RNA level fold changes (log2-fold with a p value cut-off <0.05) between *Atxn7l3*$^{-/-}$ and WT mESCs, or *Atxn7l3*$^{-/-}$ MEF-like cells and WT MEFs, showed that in both *Atxn7l3*$^{-/-}$ samples there are significant numbers of genes which were differentially expressed, both up- and down-regulated (Figure 3A and 3B, and Supplementary Figure 4A and 4B). When compared to control cells, 1163 up-regulated and 1210 down-regulated genes were identified in *Atxn7l3*$^{-/-}$ mESCs, while 1314 up-regulated and 2219 down-regulated transcripts were found in the *Atxn7l3*$^{-/-}$ MEFs (Figure 3A and 3B). These observations suggest that out of approximately 11 000 Pol II genes transcribed above background in mESCs, or in MEFs, ATXN7L3-linked DUB function regulates the transcription of only a subset of the. In both cellular systems, down-regulated, upregulated and unchanged gene sets were validated using RT-qPCR (Supplementary Figure 2A, and Supplementary Figure 3D and 3E). The fold change in the deregulated gene set was much more pronounced in *Atxn7l3*$^{-/-}$ MEF-like cells than in *Atxn7l3*$^{-/-}$ mESCs (Figure 3A and 3B), as in *Atxn7l3*$^{-/-}$ MEFs about 350 transcripts changed their expression 32-fold or more (up and down), while in *Atxn7l3*$^{-/-}$ mESCs only one gene changed its expression 32-fold (Figure 3C). These differences may be in line with the observation that *Atxn7l3*$^{-/-}$ mESCs have a mild cellular phenotype, while *Atxn7l3*$^{-/-}$ MEF-like cells undergo severe morphological changes. In addition, when comparing the down-regulated or up-regulated genes between *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEF-like cells, only very few transcripts were found to be similarly affected in the two cellular systems (Figure 3D and 3E), suggesting that ATXN7L3-linked DUB activity regulates mainly different subset of genes in the two cellular environments.

Next, we used DAVID to determine the gene ontology (GO) of the differentially expressed genes both in *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEFs. GO analyses of

biological process for genes downregulated in *Atxn7l3*$^{-/-}$ mESCs revealed enrichment of GO categories such as “Regulation of transcription, DNA templated”, “Transcription, DNA-templated”, “Negative regulation of transcription from RNA Pol II promoter” and “Cell differentiation” (Figure 4A), while in the upregulated genes the GO categories involving “Metabolic processes” and “Cell adhesion” were enriched (Figure 4B). Similar GO analyses of *Atxn7l3*$^{-/-}$ MEF-like cells indicated that many genes involved in “Multicellular organism development”, “Nervous system development”, “Cell adhesion” and “Cell differentiation” were significantly down-regulated (Figure 4C), while genes belonging mainly in “Metabolic process” and “Immune system processes” were upregulated (Figure 4D). These results together suggest that the ATXN7L3-related DUB activities regulate only a subset of genes in both cellular systems, but these genes and the extent of their deregulation are different in mESCs and MEFs.

Cell adhesion and extracellular matrix genes are downregulated In *Atxn7l3*$^{-/-}$ MEFs

Next, we further investigated the expression changes observed in the cell adhesion GO category, since these sets of downregulated genes could account for the unusual shape of the *Atxn7l3*$^{-/-}$ MEF-like cells. RNA-seq analyses, indicated that a large set of genes coding for proteins belonging to the cell adhesion GO category: such as cadherins, catenins, collagens, laminins, integrins, and other cell adhesion molecules were massively down-regulated in *Atxn7l3*$^{-/-}$ MEF-like cells compared to control MEFs (Figure 5A). The deregulation of several of these genes was confirmed by RT-qPCR analyses (Supplementary Figure 3E).

Cell adhesions proteins form discrete macromolecular complexes and establish a link between the actin cytoskeleton and the extracellular matrix, or adjacent cells. The

organization of the actin cytoskeleton at adhesion sites is tightly regulated and driven by adhesion proteins that are physically linked to the actin cytoskeleton [71, 72]. To determine if the down-regulation of "adhesion" mRNAs, and thus presumably the down-regulation of adhesion proteins, could be responsible for the morphology of the *Atxn7l3*$^{-/-}$ MEF-like cells, we next analysed actin cytoskeletal proteins by fluorescence imaging. Using phalloidin staining, labelling F-actin filaments, and anti-β actin immunofluorescence, we observed a massively reduced abundance of F-actin filaments and β-actin staining in *Atxn7l3*$^{-/-}$ MEFs compared to control MEFs (Figure 5B), suggesting that loss of ATXN7L3 results in a down-regulation of cell adhesion complexes which in turn disrupt the actin cytoskeleton network in MEFs.

Histone H2Bub1 levels increase strongly in the gene bodies of both *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEF-like cells

To gain insights into the changes in the genome-wide distribution of H2Bub1 in *Atxn7l3*$^{-/-}$ mESCs, or *Atxn7l3*$^{-/-}$ MEFs versus WT controls, chromatin immunoprecipitation coupled to high throughput sequencing (ChIP-seq) was performed using an anti-H2Bub1 antibody that recognizes monoubiquitylated H2B. The genomic distribution of H2Bub1 on a couple of housekeeping genes, expressed in both cellular systems, located on different chromosomes was analyzed using Integrative Genomics Viewer (IGV). H2Bub1 levels in both WT cell lines are relatively low, but highly increase in coding regions of both *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEFs, often showing a H2Bub1 enrichment peak downstream of the transcription start site (TSS) (Figure 6A and 6B).

Histone H2Bub1 is deposited on gene bodies by the RNF20/RNF40 complex which is associated through the PAF complex with elongating Pol II [13, 14, 61]. In order to

analyze quantitatively how the loss of the ATXN7L3-linked deubiquitylation activity changes H2Bub1 levels genome-wide, the presence of H2Bub1 over coding sequences of all annotated genes was normalized to intergenic regions and calculated (Materials and Methods). These analyses indicated that in *Atxn7l3*$^{-/-}$ mESCs and in *Atxn7l3*$^{-/-}$ MEFs, the levels of H2Bub1 increase significantly over the gene body regions of annotated genes (Supplementary Figure 5A and 5B), or significantly expressed genes (Figure 6C and 6D). In gene transcribed regions, we observed an about 1.8-fold increase in *Atxn7l3*$^{-/-}$ mESCs when compared to WT controls, and the same type of comparison resulted an about 6.5-fold increase in *Atxn7l3*$^{-/-}$ MEFs (Figure 6C and 6D, Supplementary Figure 5C and 5D).

To further examine the H2Bub1 distribution and retention changes over the bodies of all expressed genes, composite profile of H2Bub1 spanning the entire transcribed region and extending 5 kb upstream from TSSs and 5 kb downstream of the transcription end site (TES) was generated in *Atxn7l3*$^{-/-}$ mESCs versus WT mESCs, or in *Atxn7l3*$^{-/-}$ MEFs versus WT MEFs. (Figure 6E and 6F). The H2Bub1 distribution in this metagene profile obtained from WT mESCs was detectable, over the whole coding regions, with a H2Bub1 enrichment downstream of the TSS region (Figure 6E). In contrast, in *Atxn7l3*$^{-/-}$ mESCs we observed a global increase over the whole transcribed region with an important enrichment in the downstream region from the TSS. Similar results were obtained when we compared WT and *Atxn7l3*$^{-/-}$ MEFs, however with a much stronger increase in H2Bub1 levels on the gene-body regions of *Atxn7l3*$^{-/-}$ MEFs than in *Atxn7l3*$^{-/-}$ mESCs (compare Figure 6E and 6F). These results together show that ATN7L3-linked DUB activity is responsible for the genome-wide deubiquitylation over the coding regions of expressed genes in mouse mESCs and MEFs.

Modest changes in genome-wide Pol II occupancy do not correlate with the strong H2Bub1 increases observed in *Atxn7l3*$^{-/-}$ cells

Next, we wanted to know whether the strong genome-wide H2Bub1 increases over the coding regions observed in the *Atxn7l3*$^{-/-}$ cells would influence Pol II occupancy at promoters and/or in gene bodies. To test this possibility, *Atxn7l3*$^{-/-}$ mESCs and MEFs as well as control cells were subjected to ChIP-seq, using a mouse monoclonal antibody recognizing the C-terminal domain (CTD) of the largest subunit of Pol II (RPB1). Surprisingly, in *Atxn7l3*$^{-/-}$ mESCs or MEFs, Pol II occupancy at selected representative Pol II transcribed genes (Figure 7A and 7B), analysed genome-wide by k-means clustering (Figure 7C and 7D) or by meta-gene plots (Figure 7E and 7F), did not change dramatically when compared with the corresponding WT cells. These analyses showed that Pol II occupancy was almost not affected at the TSS regions and slightly decreased in the gene body regions in *Atxn7l3*$^{-/-}$ mESCs compared to WT cells. In *Atxn7l3*$^{-/-}$ MEFs compared to control cells, Pol II occupancy was weakly decreased at TSSs and very weakly affected on gene body regions (Figure 7E-7F). In contrast, at most of Pol II occupied regions, the levels of H2Bub1 were highly increased in *Atxn7l3*$^{-/-}$ mESCs and MEFs, when compared to control cells. Thus, a global increase in H2Bub1 levels did not induce an important global change in Pol II occupancy across all transcribed genes. In agreement, the RNA-seq data also indicated that only a subset of genes was either down- or up-regulated in both cell types, but no global transcription effects were observed (Figure 3). Nevertheless, when a few selected genes were visualized, we observed a complete loss of Pol II occupancy on down-regulated genes, or a strong increase in Pol II occupancy on up-regulated genes, in *Atxn7l3*$^{-/-}$ MEFs when compared to control cells, but these totally opposite Pol II occupancy changes were often

accompanied by a strong increase in H2Bub1 levels at the gene transcribed regions (Supplementary Figure 6). These results together suggest that a strong global H2Bub1 increase in *Atxn7l3*$^{-/-}$ cells do not majorly deregulate RNA polymerase II levels at transcribed genes, and thus Pol II transcription and H2Bub1 deubiquitylation are not directly coupled.

The promoter proximal paused Pol II and the prominent H2Bub1 peaks upstream of the TSSs do not overlap

It has been suggested that promoter proximal pausing of engaged Pol II is leading to the accumulation of stable transcriptionally competent polymerases about +60 bp downstream of the TSS ([73] and refs therein). Subsequently it was found that a large fraction of engaged, but stopped Pol II around the +60 bp region of promoters does not enter in elongation, but is most probably lost through premature termination [74, 75]. Next, we wanted to analyze whether promoter proximal Pol II peaks observed at transcribed genes around the +60 bp region would overlap with the H2Bub1 peak observed downstream of the TSSs both in WT and *Atxn7l3*$^{-/-}$ cells (mESCs and MEFs), which in return could suggest a link between Pol II escape from promoter proximal pausing and histone H2B ubiquitylation/deubiquitylation mechanisms. As expected meta-gene analyses around the TSSs showed that in both mESCs and MEFs (WT and *Atxn7l3*$^{-/-}$) Pol II peaks gave the highest signal at around the +60 region (Figure 7G and 7H). In contrast, similar meta-gene analyses of the H2Bub1 signal indicated that in WT and *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEFs the H2Bub1 density is low in the +60 regions and reaches its maximum more downstream, in the +300 bp region (Figure 7G and 7H). These observations suggest that the histone H2B ubiquitylation by RNF20/40 or its deubiquitylation by the ATXN7L3-dependent DUB module(s) may not regulate promoter

595 proximal pausing of Pol II, Pol II turnover at promoters and/or the engagement of Pol II
596 into productive transcription.
597
598 **Discussion**
599 **Loss of the DUB adaptor ATXNL3 results in a more severe phenotype than**
600 **the loss of the DUB enzyme of SAGA, USP22**
601 ATXN7L3 is an adaptor protein essential for the function of at least three DUB
602 complexes in mammals, containing either of the ubiquitin-specific proteases: USP22,
603 USP27X or USP51 [41]. The relative abundance and function of the various DUB
604 complexes, their redundant activities and or compensatory mechanisms, in different cell
605 types, at various stages of mouse embryonic development has not been explored.
606 However, work from Koutelou et al. [50] revealed that USP22 is essential for placental
607 development, as was also reported for the deletion of *Supt3*, encoding another SAGA
608 subunit [76]. Consistent with our findings, *Usp22* mutant embryos developed normally up
609 to E12.5, but then die around E13.5-E14.5. It has been reported that *Usp22* is
610 expressed ubiquitously in the embryo and homozygous hypomorphic $Usp22^{lacZ/lacZ}$ mice
611 have a reduced body size and weight [51]. Moreover, in these hypomorphic mice, the
612 proper cell differentiation in the intestinal epithelium and cerebral cortex was perturbed,
613 suggesting that USP22 is involved in the control of cellular differentiation [51]. However,
614 the absence of a strong morphological phenotype in the $Usp22^{-/-}$ null mutant embryos
615 before E13.5 suggests that many key early developmental processes do not require
616 USP22, or that the function of USP22 can be compensated by USP27X, USP51, or
617 another USP. It is however remarkable that placental development in $Usp22^{-/-}$ mutant

618 embryos cannot be compensated by other USPs, suggesting a possible direct
619 requirement of the SAGA complex in placental development.

620 On the other hand, no compensation is expected in *Atxn7l3*$^{-/-}$ mutant embryos as
621 the absence of ATXN7L3 is supposed to inactivate all three SAGA-related DUB
622 complexes [41]. Indeed, *Atxn7l3* loss of function results in a more severe phenotype than
623 that of *Usp22*$^{-/-}$, occurring as early as E7.5. Although at present it is not known whether
624 the deubiquitylation of the epigenetic mark, histone H2Bub1, is linked to the phenotypes
625 of the *Usp22*$^{-/-}$ or *Atxn7l3*$^{-/-}$ mutant embryos, it is interesting to note that there is a
626 parallel between the severity of the embryo phenotypes and the changes in H2Bub1
627 levels. *Usp22*$^{-/-}$ embryos are normal at E10.5 and their genome-wide histone H2Bub1
628 levels do not increase (Figure 1C and 1E), while in contrast E10.5 *Atxn7l3*$^{-/-}$ mutant
629 embryos are seriously affected and their H2Bub1 levels increase 4-5-fold (Figure 1D
630 and 1F).

631 Interestingly, we observed two categories of *Atxn7l3*$^{-/-}$ mutants. The most severely
632 affected *Atxn7l3*$^{-/-}$ embryos (2/3rd of the mutant embryos) are growth retarded, fail to
633 turn and display shortened trunk and abnormal head development. The remaining third
634 of the *Atxn7l3*$^{-/-}$ mutant embryos do turn and only display mild growth delay. It is
635 conceivable that ATXN7L3 is involved in embryo patterning as for example, Nodal
636 signalling mutant embryos, which are defective in the specification of the midline, also
637 fail to turn [77]. Nevertheless, the fact that one third of the *Atxn7l3*$^{-/-}$ mutant embryos
638 escape the severe phenotype suggest that ATXN7L3 could be involved in a
639 developmental checkpoint control at the time of embryo turning. More molecular
640 analyses would be required to study these hypotheses. Remarkably, all *Atxn7l3*$^{-/-}$
641 mutant embryos die around E11.5. As the lethality is much earlier in *Atxn7l3*$^{-/-}$ mutants

than in *Usp22*$^{-/-}$ embryos, in addition to placental defects, defects in the cardiovascular system could also be involved, as enlarged pericardium and blood pooling in the heart are observed in the severely affected mutant *Atxn7l3*$^{-/-}$ embryos. Thus, the comparison of the *Usp22*$^{-/-}$ and *Atxn7l3*$^{-/-}$ embryo phenotypes suggest that the defects observed in *Usp22*$^{-/-}$ embryos could be compensated until E13.5 in the absence of USP22 by the activity of USP27X- and/or USP51-containing DUBs, which would require ATXN7L3 and ENY2 cofactors. Such compensation would not happen in *Atxn7l3*$^{-/-}$ embryo, as in the absence of ATXN7L3 all three related DUBs would be inactive.

The underlying cause of the developmental delay in the *Atxn7l3*$^{-/-}$ embryos could be an impairment in cellular differentiation, as suggested by RNA-seq data comparing *Atxn7l3*$^{-/-}$ mESCs and MEFs, as the category of genes related to “Cellular Differentiation” was massively down-regulated (Figure 4A and 4C). In the absence of ATXN7L3, MEFs were phenotypically different from controls, appearing rounder and smaller (Figure 2E). Since we were unable to generate MEFs from E9.5 WT embryos, we ruled out the possibility that these E10.5 *Atxn7l3*$^{-/-}$ MEFs were simply a primitive MEF cell type occurring in a developmentally delayed embryo. Furthermore, when comparing their transcriptome to that of 272 distinct mouse cell types, the *Atxn7l3*$^{-/-}$ MEFs clustered most closely to fibroblasts (Supplementary Figure 3C), confirming that they are indeed MEFs, despite their strikingly unique phenotype. The round cell phenotype observed in *Atxn7l3*$^{-/-}$ MEFs is similar to the phenotype observed in the triple retinoic acid receptor (RAR) α, β, γ knockout MEFs [78]. While there was no significant reduction in the expression of *Rar* genes in the *Atxn7l3*$^{-/-}$ MEFs compared to control MEFs, the mRNA levels of cellular retinoic acid binding protein (*Crabp1*) gene was reduced by 240-fold in *Atxn7l3*$^{-/-}$ compared to control MEFs (see Supplementary Table

2), potentially resulting in impaired retinoic acid signalling. In the *Rar* triple KO MEFs, many “cellular adhesion” genes were also down-regulated, and the authors of this study concluded that the round *Rar* triple KO MEF phenotype is caused by the misregulation in “cellular adhesion” genes. As many of the “cellular adhesion” genes are also significantly down-regulated in the *Atxn7l3*$^{-/-}$ MEFs (Figure 5A) and CRABP1 levels are seriously reduced, it is conceivable that the “cellular adhesion” genes and the round cellular phenotype are controlled indirectly through retinoic acid and/or RAR-linked signalling.

In conclusion, our results showing that *Usp22* KO embryo phenotypes are less severe agree with the biochemical findings suggesting that in *Usp22* KO cells the activity of only one of the three related DUB modules, the one that can incorporate in the SAGA complex, is eliminated. In contrast, in the *Atxn7l3* KO embryos the activities of all the three related DUB modules are eliminated, thus, causing a more severe phenotype. The fact that *Atxn7l3* KO embryos survive until E7.5, suggests that none of three related DUBs would play an essential role before this embryonic stage, and that also histone H2Bub1 deubiquitylation is not essential for Pol II transcription before this developmental stage.

Histone H2Bub1 deubiquitylation is not linked to global RNA polymerase II transcription

Although histone H2B monoubiquitylation has been linked to increased transcription, transcription elongation, DNA replication, mitosis, and meiosis [79], how this histone modification and the erasing of this mark function is not well understood. Several roles of H2Bub1 in transcription have been proposed. It has been suggested that H2Bub1 stimulates FACT-mediated displacement of an H2A/H2B dimer from the

690 core nucleosome and by that would enhance the passage of Pol II through the
691 nucleosome [23]. Other studies described that H2Bub1 is required for efficient reassembly
692 of nucleosomes behind the elongating Pol II [80, 81]. It was also reported that the effect of
693 H2Bub1 on nucleosome stability is relatively modest [82] and that H2Bub1 can disrupt the
694 higher-order chromatin architecture and lead to an open, accessible conformation fiber,
695 which may favorize gene expression [12].

696 Contrary to H2B ubiquitylation, it is much less well understood whether H2Bub1
697 deubiquitylation would be a process significantly impacting transcription. Previously, by
698 using an *ATXN7L3* knock-down strategy in human HeLa cells we showed that the
699 ATXN7L3-related DUB activities are directed toward the transcribed region of almost all
700 expressed genes, but are only poorly correlated with gene expression [26]. Our present
701 results indicate that impairment of H2Bub1 deubiquitylation does not directly impact
702 transcription, because while we observe a massive H2Bub1 retention at almost every
703 expressed gene in both *Atxn7l3*$^{-/-}$ mESCs and MEFs, Pol II occupancy was only slightly
704 impacted and only limited subsets of genes changed expression in both cellular
705 systems (Figure 3, 6 and 7). Nevertheless, in both cellular systems the lack of
706 correlation between global H2Bub1 increase and consequent genome-wide inhibition of
707 global transcription suggests that the deubiquitylation of H2Bub1 does not directly
708 regulate Pol II transcription. In agreement, the H3K4me3 chromatin mark present at the
709 TSSs of active genes in eukaryotes, of which the levels reflect the amount of
710 transcription and is linked with H2Bub1 deposition [83], did not change either in *Usp22*$^{-/-}$ or
711 in *Atxn7l3*$^{-/-}$ embryos, in spite of the fact that in *Atxn7l3*$^{-/-}$ embryos, the H2Bub1levels
712 were increased by 4-5-fold (Figure 1C and 1D). Similarly, global H3K9ac levels do not
713 change in *Usp22*$^{-/-}$ or in *Atxn7l3*$^{-/-}$ embryos (Figure 1C and 1D). Thus, our study

corroborates other recent studies demonstrating catalytic-independent functions of chromatin modifying complexes in mouse ES cells [84, 85, 86].

In addition, our results also suggest that the dynamic erasure of the H2Bub1 epigenetic mark does not seem to influence global Pol II recruitment and consequent pre-initiation complex formation at promoters and/or the promoter proximal pausing of Pol II, as the high H2Bub1 increase seen in the *Atxn7l3*$^{-/-}$ cells occurs more downstream (+ 300 bp) than the mentioned promoter associated Pol II-dependent events. Whether the observed embryo and cellular phenotypes in the *Atxn7l3*$^{-/-}$ embryos can be directly linked to increased H2Bub1 levels in specific transcribed regions having special chromatin architecture, and/or would be rather linked to deubiquitylation failures of other ubiquitylated protein targets, will need to be further investigated in the future.

Acknowledgements

We thank all members of the Tora lab for protocols, thoughtful discussions and suggestions throughout the course of the work, C. Rochette-Egly for discussion concerning the MEF-like phenotype. We thank V. Hisler for help with mice dissection, C. Hérouard and M. Jung from the GenomEast platform [a member of the 'France Génomique' consortium (ANR-10-INBS-0009], for library preparation, sequencing and preliminary analyses; C. Ebel and M. Philipps for help with FACS analyses, the IGBMC histology platform, the IGBMC cell culture facility and S. Falcone, M. Poirot and F. Memedov of the IGBMC animal facility for animal care taking. This study was supported by grants from European Research Council (ERC) (ERC-2013-Advanced grant 340551, Birtoaction), Agence Nationale de la Recherche (ANR) PICen-19-CE11-0003-02 and EpiCAST-19-CE12-0029-01 grants, NIH 1R01GM131626-01 grant (to LT) and ANR-18-CE12-0026 grant (to DD), fellowships by the IGBMC International PhD program (to FW); and by the IdEx-University of Strasbourg

international PhD program and by the 'Fondation pour la Recherche Médicale' (FRM) association (FDT201904008368) (to VF), and an ANR-10-LABX-0030-INRT grant, a French State fund managed by the ANR under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02.

References

1. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 1997, **389:** 251-260.

2. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 1999, **98:** 285-294.

3. Shilatifard A. Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. *Annu Rev Biochem* 2006.

4. Kouzarides T. Chromatin modifications and their function. *Cell* 2007, **128**(4)**:** 693-705.

5. Tolsma TO, Hansen JC. Post-translational modifications and chromatin dynamics. *Essays Biochem* 2019, **63**(1)**:** 89-96.

6. Cosgrove MS, Boeke JD, Wolberger C. Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* 2004, **11**(11)**:** 1037-1043.

7. Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta* 2014, **1839**(8)**:** 627-643.

8. Festuccia N, Gonzalez I, Navarro P. The Epigenetic Paradox of Pluripotent ES Cells. *J Mol Biol* 2017, **429**(10)**:** 1476-1503.

9. Osley MA. Regulation of histone H2A and H2B ubiquitylation. *Brief Funct Genomic Proteomic* 2006, **5**(3)**:** 179-189.

10. Zhu B, Zheng Y, Pham AD, Mandal SS, Erdjument-Bromage H, Tempst P*, et al.* Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Mol Cell* 2005, **20**(4)**:** 601-611.

11. Shiloh Y, Shema E, Moyal L, Oren M. RNF20-RNF40: A ubiquitin-driven link between gene expression and the DNA damage response. *FEBS Lett* 2011, **585**(18)**:** 2795-2802.

12. Fierz B, Chatterjee C, McGinty RK, Bar-Dagan M, Raleigh DP, Muir TW. Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat Chem Biol* 2011, **7**(2)**:** 113-119.

13. Minsky N, Shema E, Field Y, Schuster M, Segal E, Oren M. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nat Cell Biol* 2008, **10**(4)**:** 483-488.

14. Shema E, Tirosh I, Aylon Y, Huang J, Ye C, Moskovits N*, et al.* The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev* 2008, **22**(19)**:** 2664-2676.

15. Trujillo KM, Osley MA. A role for H2B ubiquitylation in DNA replication. *Mol Cell* 2012, **48**(5)**:** 734-746.

16. Kari V, Shchebet A, Neumann H, Johnsen SA. The H2B ubiquitin ligase RNF40 cooperates with SUPT16H to induce dynamic changes in chromatin structure during DNA double-strand break repair. *Cell Cycle* 2011, **10**(20)**:** 3495-3504.

17. Moyal L, Lerenthal Y, Gana-Weisz M, Mass G, So S, Wang SY*, et al.* Requirement of ATM-dependent monoubiquitylation of histone H2B for timely repair of DNA double-strand breaks. *Mol Cell* 2011, **41**(5)**:** 529-542.

18. Nakamura K, Kato A, Kobayashi J, Yanagihara H, Sakamoto S, Oliveira DV*, et al.* Regulation of homologous recombination by RNF20-dependent H2B ubiquitination. *Mol Cell* 2011, **41**(5)**:** 515-528.

19. Xie W, Nagarajan S, Baumgart SJ, Kosinsky RL, Najafova Z, Kari V*, et al.* RNF40 regulates gene expression in an epigenetic context-dependent manner. *Genome Biol* 2017, **18**(1)**:** 32.

20. Vitaliano-Prunier A, Babour A, Herissant L, Apponi L, Margaritis T, Holstege FC*, et al.* H2B ubiquitylation controls the formation of export-competent mRNP. *Mol Cell* 2012, **45**(1)**:** 132-139.

21. Pirngruber J, Shchebet A, Schreiber L, Shema E, Minsky N, Chapman RD*, et al.* CDK9 directs H2B monoubiquitination and controls replication-dependent histone mRNA 3'-end processing. *EMBO Rep* 2009, **10**(8)**:** 894-900.

22. Evangelista FM, Maglott-Roth A, Stierle M, Brino L, Soutoglou E, Tora L. Transcription and mRNA export machineries SAGA and TREX-2 maintain

monoubiquitinated H2B balance required for DNA repair. *J Cell Biol* 2018, **217**(10)**:** 3382-3397.

23. Pavri R, Zhu B, Li G, Trojer P, Mandal S, Shilatifard A*, et al.* Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* 2006, **125**(4)**:** 703-717.

24. Chandrasekharan MB, Huang F, Sun ZW. Histone H2B ubiquitination and beyond: Regulation of nucleosome stability, chromatin dynamics and the trans-histone H3 methylation. *Epigenetics* 2010, **5**(6)**:** 460-468.

25. Bonnet J, Devys D, Tora L. Histone H2B ubiquitination: signaling not scrapping. *Drug Discov Today Technol* 2014, **12:** e19-27.

26. 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. *Gene Dev* 2014, **28**(18)**:** 1999-2012.

27. Fuchs G, Hollander D, Voichek Y, Ast G, Oren M. Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. *Genome research* 2014, **24**(10)**:** 1572-1583.

28. Jung I, Kim SK, Kim M, Han YM, Kim YS, Kim D*, et al.* H2B monoubiquitylation is a 5'-enriched active transcription mark and correlates with exon-intron structure in human cells. *Genome Res* 2012, **22**(6)**:** 1026-1035.

29. Briggs SD, Xiao T, Sun ZW, Caldwell JA, Shabanowitz J, Hunt DF*, et al.* Gene silencing: trans-histone regulatory pathway in chromatin. *Nature* 2002, **418**(6897)**:** 498.

30. Dover J, Schneider J, Tawiah-Boateng MA, Wood A, Dean K, Johnston M*, et al.* Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *J Biol Chem* 2002, **277**(32)**:** 28368-28371.

31. Ng HH, Xu RM, Zhang Y, Struhl K. Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. *J Biol Chem* 2002, **277**(38)**:** 34655-34657.

32. Sun ZW, Allis CD. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* 2002, **418**(6893)**:** 104-108.

33. Lee JS, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L*, et al.* Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* 2007, **131**(6)**:** 1084-1096.

34. Kim J, Kim JA, McGinty RK, Nguyen UT, Muir TW, Allis CD*, et al.* The n-SET domain of Set1 regulates H2B ubiquitylation-dependent H3K4 methylation. *Mol Cell* 2013, **49**(6)**:** 1121-1133.

35. Daniel JA, Torok MS, Sun ZW, Schieltz D, Allis CD, Yates JR, 3rd*, et al.* Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. *J Biol Chem* 2004, **279**(3)**:** 1867-1871.

36. Henry KW, Wyce A, Lo WS, Duggan LJ, Emre NC, Kao CF*, et al.* Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* 2003, **17**(21)**:** 2648-2663.

37. Zhao Y, Lang G, Ito S, Bonnet J, Metzger E, Sawatsubashi S*, et al.* A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Mol Cell* 2008, **29**(1)**:** 92-101.

38. Zhang XY, Varthi M, Sykes SM, Phillips C, Warzecha C, Zhu W*, et al.* The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression. *Mol Cell* 2008, **29**(1)**:** 102-111.

39. Lang G, Bonnet J, Umlauf D, Karmodiya K, Koffler J, Stierle M*, et al.* The tightly controlled deubiquitination activity of the human SAGA complex differentially modifies distinct gene regulatory elements. *Molecular and cellular biology* 2011, **31**(18)**:** 3734-3744.

40. Morgan MT, Wolberger C. Recognition of ubiquitinated nucleosomes. *Curr Opin Struct Biol* 2017, **42:** 75-82.

41. Atanassov BS, Mohan RD, Lan X, Kuang X, Lu Y, Lin K*, et al.* ATXN7L3 and ENY2 Coordinate Activity of Multiple H2B Deubiquitinases Important for Cellular Proliferation and Tumor Growth. *Mol Cell* 2016, **62**(4)**:** 558-571.

42. Bonnet J, Romier C, Tora L, Devys D. Zinc-finger UBPs: regulators of deubiquitylation. *Trends Biochem Sci* 2008, **33**(8)**:** 369-375.

43. Atanassov BS, Evrard YA, Multani AS, Zhang Z, Tora L, Devys D*, et al.* Gcn5 and SAGA regulate shelterin protein turnover and telomere maintenance. *Mol Cell* 2009, **35**(3)**:** 352-364.

44. Atanassov BS, Dent SY. USP22 regulates cell proliferation by deubiquitinating the transcriptional regulator FBP1. *EMBO Rep* 2011, **12**(9)**:** 924-930.

45. Armour SM, Bennett EJ, Braun CR, Zhang XY, McMahon SB, Gygi SP*, et al.* A high-confidence interaction map identifies SIRT1 as a mediator of acetylation of

914 USP22 and the SAGA coactivator complex. *Mol Cell Biol* 2013, **33**(8)**:** 1487-
915 1502.
916
917 46. Lin Z, Yang H, Kong Q, Li J, Lee SM, Gao B*, et al.* USP22 antagonizes p53
918 transcriptional activation by deubiquitinating Sirt1 to suppress cell apoptosis and
919 is required for mouse embryonic development. *Mol Cell* 2012, **46**(4)**:** 484-494.
920
921 47. Kobayashi T, Iwamoto Y, Takashima K, Isomura A, Kosodo Y, Kawakami K*, et*
922 *al.* Deubiquitinating enzymes regulate Hes1 stability and neuronal differentiation.
923 *FEBS J* 2015, **282**(13)**:** 2411-2423.
924
925 48. Lambies G, Miceli M, Martinez-Guillamon C, Olivera-Salguero R, Pena R, Frias
926 CP*, et al.* TGFbeta-Activated USP27X Deubiquitinase Regulates Cell Migration
927 and Chemoresistance via Stabilization of Snail1. *Cancer Res* 2019, **79**(1)**:** 33-46.
928
929 49. Zhou Z, Zhang P, Hu X, Kim J, Yao F, Xiao Z*, et al.* USP51 promotes
930 deubiquitination and stabilization of ZEB1. *Am J Cancer Res* 2017, **7**(10)**:** 2020-
931 2031.
932
933 50. Koutelou E, Wang L, Schibler AC, Chao HP, Kuang X, Lin K*, et al.* USP22
934 controls multiple signaling pathways that are essential for vasculature formation
935 in the mouse placenta. *Development* 2019, **146**(4).
936
937 51. Kosinsky RL, Wegwitz F, Hellbach N, Dobbelstein M, Mansouri A, Vogel T*, et al.*
938 Usp22 deficiency impairs intestinal epithelial lineage specification in vivo.
939 *Oncotarget* 2015, **6**(35)**:** 37906-37918.
940
941 52. Braunstein M, Liao L, Lyttle N, Lobo N, Taylor KJ, Krzyzanowski PM*, et al.*
942 Downregulation of histone H2A and H2B pathways is associated with
943 anthracycline sensitivity in breast cancer. *Breast Cancer Res* 2016, **18**(1)**:** 16.
944
945 53. Wang Z, Zhu L, Guo T, Wang Y, Yang J. Decreased H2B monoubiquitination and
946 overexpression of ubiquitin-specific protease enzyme 22 in malignant colon
947 carcinoma. *Hum Pathol* 2015, **46**(7)**:** 1006-1014.
948
949 54. Zhang K, Wang J, Tong TR, Wu X, Nelson R, Yuan YC*, et al.* Loss of H2B
950 monoubiquitination is associated with poor-differentiation and enhanced
951 malignancy of lung adenocarcinoma. *Int J Cancer* 2017, **141**(4)**:** 766-777.
952
953 55. Tang B, Liang X, Tang F, Zhang J, Zeng S, Jin S*, et al.* Expression of USP22
954 and Survivin is an indicator of malignant behavior in hepatocellular carcinoma. *Int*
955 *J Oncol* 2015, **47**(6)**:** 2208-2216.
956
957 56. Tang B, Tang F, Li B, Yuan S, Xu Q, Tomlinson S*, et al.* High USP22 expression
958 indicates poor prognosis in hepatocellular carcinoma. *Oncotarget* 2015, **6**(14)**:**
959 12654-12667.
960

961 57. Hu J, Yang D, Zhang H, Liu W, Zhao Y, Lu H*, et al.* USP22 promotes tumor
962 progression and induces epithelial-mesenchymal transition in lung
963 adenocarcinoma. *Lung Cancer* 2015, **88**(3)**:** 239-245.
964
965 58. Duan Y, Huo D, Gao J, Wu H, Ye Z, Liu Z*, et al.* Ubiquitin ligase RNF20/40
966 facilitates spindle assembly and promotes breast carcinogenesis through
967 stabilizing motor protein Eg5. *Nat Commun* 2016, **7:** 12648.
968
969 59. Tarcic O, Pateras IS, Cooks T, Shema E, Kanterman J, Ashkenazi H*, et al.*
970 RNF20 Links Histone H2B Ubiquitylation with Inflammation and Inflammation-
971 Associated Cancer. *Cell Rep* 2016, **14**(6)**:** 1462-1476.
972
973 60. Johnsen SA. The enigmatic role of H2Bub1 in cancer. *FEBS Lett* 2012, **586**(11)**:**
974 1592-1601.
975
976 61. Vethantham V, Yang Y, Bowman C, Asp P, Lee JH, Skalnik DG*, et al.* Dynamic
977 loss of H2B ubiquitylation without corresponding changes in H3K4 trimethylation
978 during myogenic differentiation. *Mol Cell Biol* 2012, **32**(6)**:** 1044-1055.
979
980 62. Fuchs G, Shema E, Vesterman R, Kotler E, Wolchinsky Z, Wilder S*, et al.* RNF20
981 and USP44 regulate stem cell differentiation by modulating H2B
982 monoubiquitylation. *Mol Cell* 2012, **46**(5)**:** 662-673.
983
984 63. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S*, et al.* STAR:
985 ultrafast universal RNA-seq aligner. *Bioinformatics* 2013, **29**(1)**:** 15-21.
986
987 64. Love MI, Huber W, Anders S. Moderated estimation of fold change and
988 dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014, **15**(12)**:** 550.
989
990 65. El-Saafin F, Curry C, Ye T, Garnier JM, Kolb-Cheynel I, Stierle M*, et al.*
991 Homozygous TAF8 mutation in a patient with intellectual disability results in
992 undetectable TAF8 protein, but preserved RNA polymerase II transcription. *Hum*
993 *Mol Genet* 2018, **27**(12)**:** 2171-2186.
994
995 66. Gyenis A, Umlauf D, Ujfaludi Z, Boros I, Ye T, Tora L. UVB induces a genome-
996 wide acting negative regulatory mechanism that operates at the level of
997 transcription initiation in human cells. *PLoS genetics* 2014, **10**(7)**:** e1004483.
998
999 67. Anders S, Huber W. Differential expression analysis for sequence count data.
1000 *Genome Biol* 2010, **11**(10)**:** R106.
1001
1002 68. Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I*, et al.*
1003 seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic acids*
1004 *research* 2011, **39**(6)**:** e35.
1005
1006 69. Martello G, Smith A. The nature of embryonic stem cells. *Annu Rev Cell Dev Biol*
1007 2014, **30:** 647-675.

70. Hutchins AP, Yang Z, Li Y, He F, Fu X, Wang X*, et al.* Models of global gene expression define major domains of cell type and tissue identity. *Nucleic Acids Res* 2017, **45**(5)**:** 2354-2367.

71. Schwarz US, Gardel ML. United we stand: integrating the actin cytoskeleton and cell-matrix adhesions in cellular mechanotransduction. *J Cell Sci* 2012, **125**(Pt 13)**:** 3051-3060.

72. Wehrle-Haller B. Structure and function of focal adhesions. *Curr Opin Cell Biol* 2012, **24**(1)**:** 116-124.

73. Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews Genetics* 2012, **13**(10)**:** 720-731.

74. Krebs AR, Imanci D, Hoerner L, Gaidatzis D, Burger L, Schubeler D. Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol Cell* 2017, **67**(3)**:** 411-422 e414.

75. Erickson B, Sheridan RM, Cortazar M, Bentley DL. Dynamic turnover of paused Pol II complexes at human promoters. *Genes Dev* 2018, **32**(17-18)**:** 1215-1225.

76. Perez-Garcia V, Fineberg E, Wilson R, Murray A, Mazzeo CI, Tudor C*, et al.* Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature* 2018, **555**(7697)**:** 463-468.

77. Vincent SD, Dunn NR, Hayashi S, Norris DP, Robertson EJ. Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes Dev* 2003, **17**(13)**:** 1646-1662.

78. Al Tanoury Z, Piskunov A, Andriamoratsiresy D, Gaouar S, Lutzing R, Ye T*, et al.* Genes involved in cell adhesion and signaling: a new repertoire of retinoic acid receptor target genes in mouse embryonic fibroblasts. *J Cell Sci* 2014, **127**(Pt 3)**:** 521-533.

79. Laribee RN, Fuchs SM, Strahl BD. H2B ubiquitylation in transcriptional control: a FACT-finding mission. *Genes Dev* 2007, **21**(7)**:** 737-743.

80. Fleming AB, Kao CF, Hillyer C, Pikaart M, Osley MA. H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. *Mol Cell* 2008, **31**(1)**:** 57-66.

81. Batta K, Zhang Z, Yen K, Goffman DB, Pugh BF. Genome-wide function of H2B ubiquitylation in promoter and genic regions. *Genes Dev* 2011, **25**(21)**:** 2254-2265.

82. Fierz B, Kilic S, Hieb AR, Luger K, Muir TW. Stability of nucleosomes containing homogenously ubiquitylated H2A and H2B prepared using semisynthesis. *J Am Chem Soc* 2012, **134**(48)**:** 19548-19551.

83. Howe FS, Fischl H, Murray SC, Mellor J. Is H3K4me3 instructive for transcription activation? *Bioessays* 2017, **39**(1)**:** 1-12.

84. Acharya D, Hainer SJ, Yoon Y, Wang F, Bach I, Rivera-Perez JA*, et al.* KAT-Independent Gene Regulation by Tip60 Promotes ESC Self-Renewal but Not Pluripotency. *Cell Rep* 2017, **19**(4)**:** 671-679.

85. Dorighi KM, Swigut T, Henriques T, Bhanu NV, Scruggs BS, Nady N*, et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell* 2017, **66**(4)**:** 568-576 e564.

86. Rickels R, Herz HM, Sze CC, Cao K, Morgan MA, Collings CK*, et al.* Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat Genet* 2017, **49**(11)**:** 1647-1653.

Figure legends

Figure 1: Loss of the SAGA DUB adaptor ATXN7L3 results in a more severe phenotype than loss of the DUB enzyme USP22.

A. Comparison of *Usp22*$^{+/+}$ and *Usp22*$^{-/-}$ littermates from E9.5 to E14.5. **B.** Comparison of *Atxn7l3*$^{+/+}$ and *Atxn7l3*$^{-/-}$ littermates from E7.5 to E11.5. From E10.5 onwards, the *Atxn7l3*$^{-/-}$ embryos can be categorized in 2 phenotypic classes; severe (h, k) and mild (i, l). **C-D.** Western blot analyses of E10.5 and E11.5 *Usp22*$^{+/+}$ and *Usp22*$^{-/-}$ (C), as well as E9.5 and E10.5 *Atxn7l3*$^{+/+}$ and *Atxn7l3*$^{-/-}$ (D) whole embryo lysates using anti-H2Bub1, anti-H3K4me3 and anti-H4 (C) or anti-H2Bub1, anti-H3K4me3 and anti-H3 (D) antibodies. A Ponceau staining view is displayed at the bottom of each panel. M: molecular weight marker (in kDa). The dotted line in (D) indicates where the blot was cut to show comparable number of embryos from each genotype. **E-F**. Western blot

analyses shown in (C-D) were scanned and analysed densitometrically with ImageJ and the Ponceau normalized results are represented for each genotype.

Figure 2: Primary *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEF-like cells show strong increase in H2Bub1 levels, abnormal proliferation and phenotypes.

A. Western blot analysis of H2Bub1 levels in acidic histone extracts obtained from *Atxn7l3*$^{+/+}$ or *Atxn7l3*$^{-/-}$ mESC and *Atxn7l3*$^{+/+}$ or *Atxn7l3*$^{-/-}$ MEF cells. Histone H3 western blot and ponceau stained membranes are shown as loading controls. **B.** Quantification of H2Bub1 levels from (A) by using ImageJ. The y axis represents the fold change compared with WT cells. Histone H2Bub1 quantification was carried out with H3 normalization. Error bars indicate ±SD based on two biological replicates (represented by grey dots). **C.** *Atxn7l3*$^{+/+}$ or *Atxn7l3*$^{-/-}$ mESCs cultured in serum/LIF plus 2i medium for 6 days were either observed by phase contrast microscopy (left panels) or visualized by alkaline phosphatase staining (right panels). Scale bar, 200 μm. **D.** *Atxn7l3*$^{+/+}$ or *Atxn7l3*$^{-/-}$ mESCs cell proliferation was determined by cell counting at the indicated time points. Error bars indicate ±SD based on two biological samples with three technical replicates for each. Statistical significance was calculated using the Mann–Whitney test (ns, $p > 0.05$; *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$). **E.** Morphology of *Atxn7l3* $^{+/+}$ and *Atxn7l3*$^{-/-}$ MEFs derived from E10.5 embryos. Scale bar, 200 μm. **F.** MEF cell number was determined by cell counting at the indicated time points. Error bars indicate ±SD based on two biological samples with three technical replicates for each. Statistical significance was calculated using the Mann–Whitney test (ns, $p > 0.05$; *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$).

Figure 3: *Atxn7l3*$^{-/-}$ mESCs and MEF-like cells show significant deregulation of transcription

A-B. MA-plots of RNA-seq data carried out on polyA$^+$ RNA isolated from *Atxn7l3* $^{+/+}$ and *Atxn7l3*$^{-/-}$ mESCs (A), or from *Atxn7l3* $^{+/+}$ and *Atxn7l3*$^{-/-}$ MEFs (B). Log2 fold changes are shown versus Log2 mean expression signal. Differentially expressed genes were selected using the following thresholds: adjusted *p*-value ≤ 0.05 and absolute value of log2 fold change ≥ 1. Red dots indicate up-regulated genes and blue dots indicates downregulated genes. **C.** The number of significantly affected genes for *Atxn7l3*$^{-/-}$ (KO)/*Atxn7l3*$^{+/+}$ (WT) are represented for either mESCs or MEFs: adjusted *p*-value ≤ 0.05 and absolute value of fold change ≥ 2, 4, 8, 32, 64, 128, 256, separately. **D-E.** Venn diagrams indicate the overlap of down-regulated (**E**) and up-regulated (**F**) genes between mESCs and MEFs.

Figure 4: Gene ontology (GO) analyses of differentially expressed genes both in *Atxn7l3*$^{-/-}$ mESCs, and *Atxn7l3*$^{-/-}$ MEFs versus WT controls

A-D. Results of gene ontology analyses carried out using DAVID bioinformatics resources 6.8 to identify differential gene-function categories (as indicated). Significantly enriched GO terms (-Log_{10} adjusted *p* value<0.05) in biological processes are shown. The number of genes enriched in the given GO terms is also indicated.

Figure 5: Cell adhesion genes are down regulated in *Atxn7l3*$^{-/-}$ MEFs

A. Heat map showing transcript levels belonging to the cell adhesion GO category from the three biological replicates of *Atxn7l3*$^{+/+}$ and *Atxn7l3*$^{-/-}$ MEFs for transcripts that are differentially expressed. Log_2 of normalized expression is shown on the vertical column on the left. **B.** DAPI and immunofluorescence (IF) images of *Atxn7l3*$^{+/+}$ and *Atxn7l3*$^{-/-}$ MEFs stained with anti-β-Actin antibody (left) and phalloidin (right) in MEF cells. The merge of DAPI and IF images is also shown. Scale bar: 100 μm.

Figure 6: Histone H2Bub1 levels increase strongly in the gene bodies of both *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEFs

A-B. IGV genomic snapshots of H2Bub1 binding profiles at three selected genes (*Pgk1*, *Klhl11* and *Acly*). Direction of the transcription is indicated by arrows. Group scaled tag densities on each gene either in mESCs, or in MEFs, are indicated on the left. **C-D.** Boxplots showing the $\log_{10}$(H2Bub1 density) on the gene bodies of expressed transcripts or intergenic regions. Wilcoxon rank sum test with continuity correction (***: *p*-value < 2.2e-16). **E-F.** Average metagene profiles showing H2Bub1 distribution on the bodies of expressed genes. 11 172 expressed genes in mESCs (E) and 11 113 expressed genes in MEFs cells (F) were chosen. TSS: transcription start site. TES: transcription end site. -5000 bp region upstream of the TSS and +5000 bp region downstream of the TES were also included in the average profile analyses.

Figure 7: The modest genome-wide Pol II occupancy changes do not correlate with the strong H2Bub1 increases observed in the *Atxn7l3*$^{-/-}$ mESCs or MEFs

A-B. IGV genomic snapshots of H2Bub1 and Pol II binding profiles at four selected genes (*Zpr1*, *Bud13, Gan* and *Cmip*). Direction of the transcription is indicated by arrows. Group scaled tag densities on each gene either in mESCs, or in MEFs, are indicated on the left. **C-D.** K-means clustering showing the distribution of Pol II and H2Bub1 on expressed genes (from -5000 upstream from the TSS to + 5000 downstream of the TES) in control and *Atxn7l3*$^{-/-}$ mESC (C) and MEF (D). **E-F** Average metagene profiles showing Pol II distribution on bodies of expressed genes (from -5000 upstream from the TSS to + 5000 downstream of the TES) in control and *Atxn7l3*$^{-/-}$ mESCs (E) and MEFs (F). **G-H.** Average profiles depicting Pol II and H2Bub distribution

around the TSS (TSS -1 kb / +1 kb) of expressed genes in control and *Atxn7l3*$^{-/-}$ mESCs (G) and MEFs (F).

Table 1: Offsprings from *Usp22*$^{+/-}$ intercrosses

Stage	*Usp22*$^{+/+}$	*Usp22*$^{+/-}$	*Usp22*$^{-/-}$	Total	Number of litters
E9.5	3 (16.7%)	9 (50%)	6 (33.3%)	18	2
E10.5	5 (23.8%)	11 (52.4%)	5 (23.8%)	21	3
E12.5	8 (19.5%)	21 (51.2%)	12 (29.3%)	41	5
E13.5	4 (28.6%)	7 (50%)	3 (21.4%)	14	2
E14.5	6 (27.3%)	10 (45.4%)	6* (27.3%)	22	3
weaning	93 (37.6%)	154 (62.4%)	0 (0%)	247	37

* dead embryo (no beating heart)

Table 2: Offsprings from *Atxn7l3*$^{+/-}$ intercrosses

Stage	*Atxn7l3*$^{+/+}$	*Atxn7l3*$^{+/-}$	*Atxn7l3*$^{-/-}$	Total	Number of litters
E7.5	10 (47.6%)	5 (23.8%)	6 (28.6%)	21	2
E8.5	20 (31.2%)	35 (54.7%)	9 (14.1%)	64	7
E9.5	13 (25.5%)	26 (51%)	12 (23.5%)	51	6
E10.5	53 (28.8%)	83 (45.1%)	48 (26.1%)	184	21
E11.5	7 (28%)	12 (48%)	6 (24%)	25	3
E12.5	9 (47.4%)	10 (52.6%)	0 (0%)	19	3
weaning	138 (44.7%)	171 (55.3%)	0 (0%)	309	47

Figure 1:

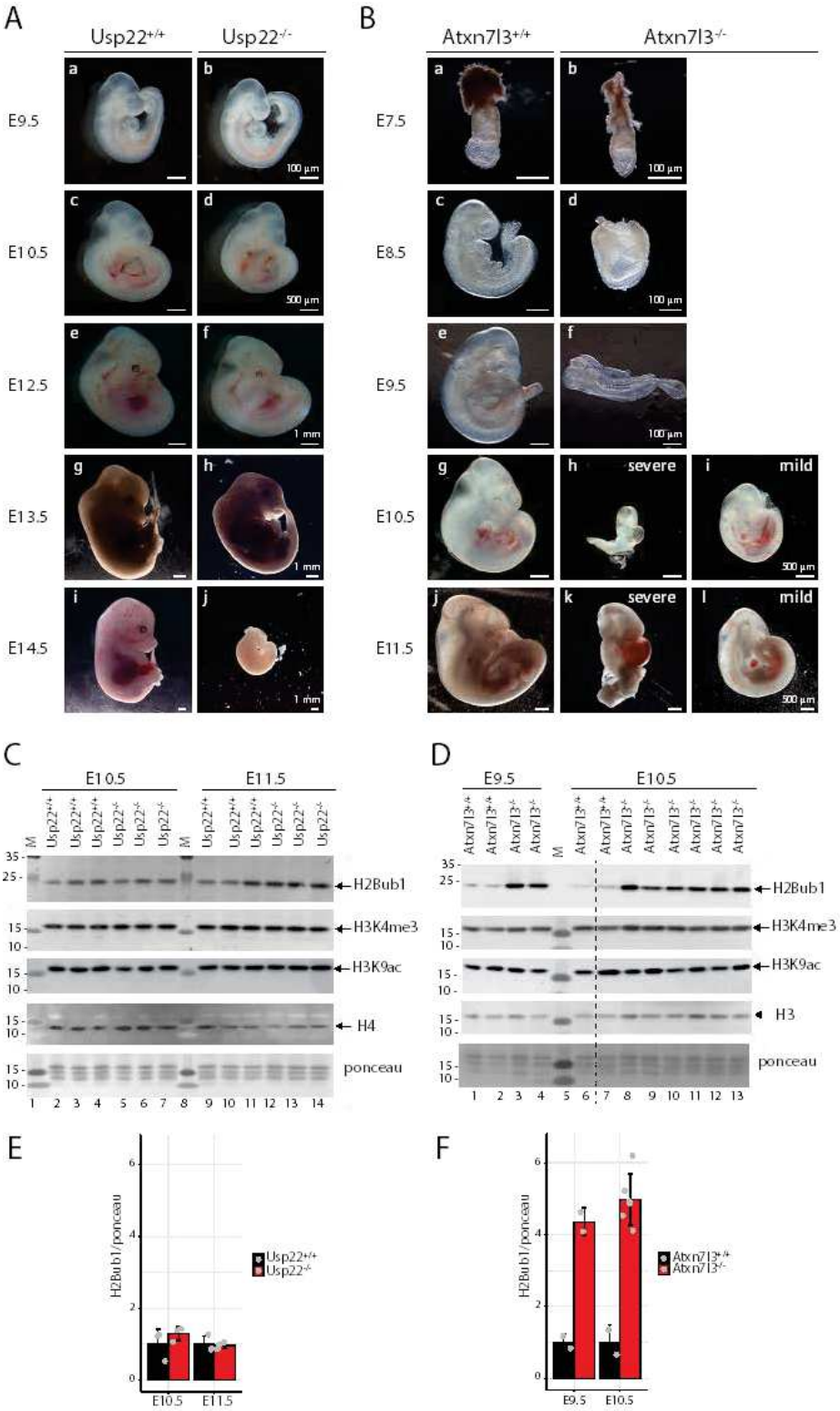

A
Usp22+/+
Usp22-/-
E9.5
E10.5
E12.5
E13.5
E14.5
B
Atxn7l3+/+
Atxn7l3-/-
E7.5
E8.5
E9.5
E10.5
E11.5
severe
mild
C
E10.5
E11.5
H2Bub1
H3K4me3
H3K9ac
H4
ponceau
D
E9.5
E10.5
H2Bub1
H3K4me3
H3K9ac
H3
ponceau
E
H2Bub1/ponceau
Usp22+/+
Usp22-/-
E10.5
E11.5
F
H2Bub1/ponceau
Atxn7l3+/+
Atxn7l3-/-
E9.5
E10.5

Figure 2:

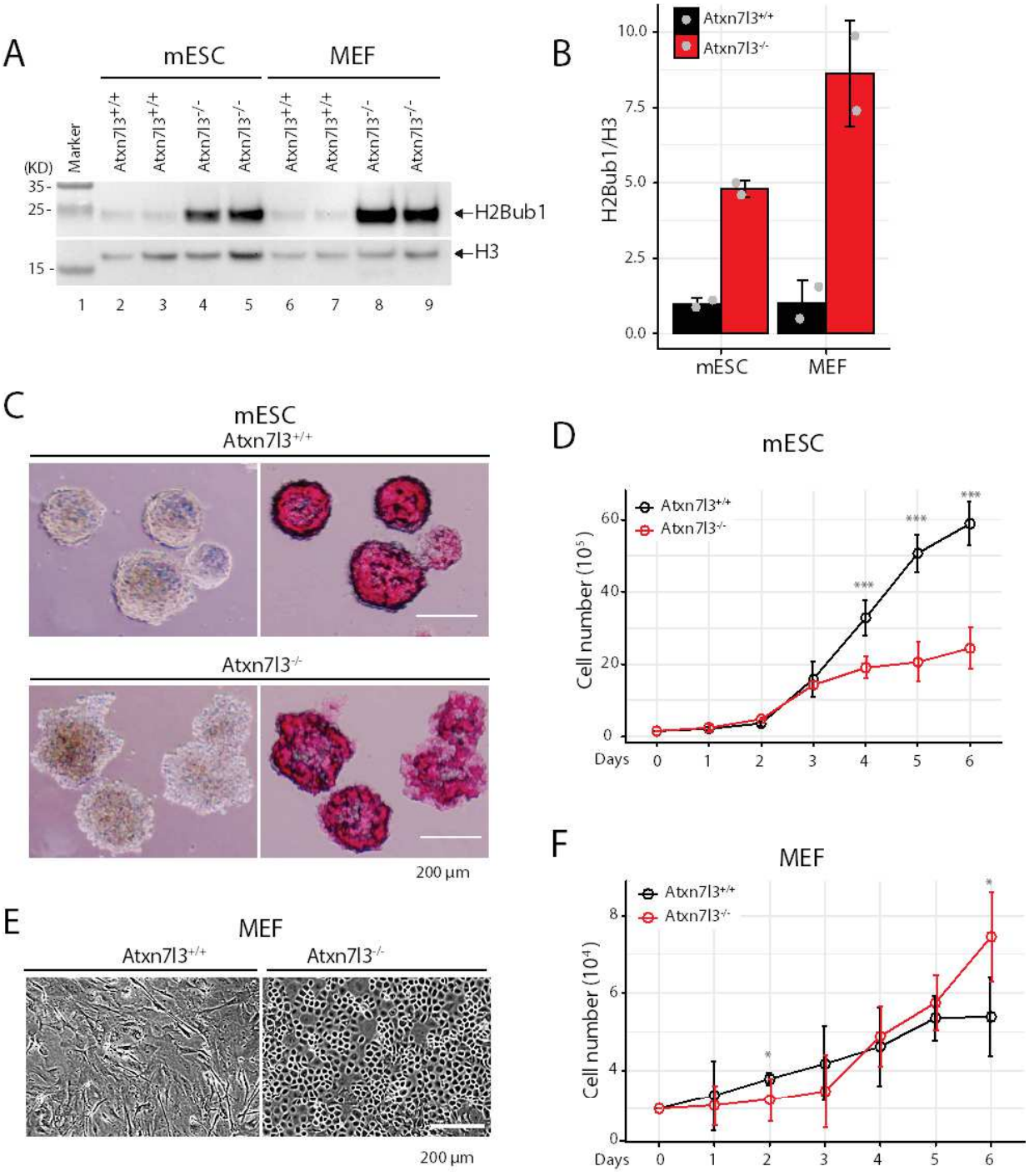
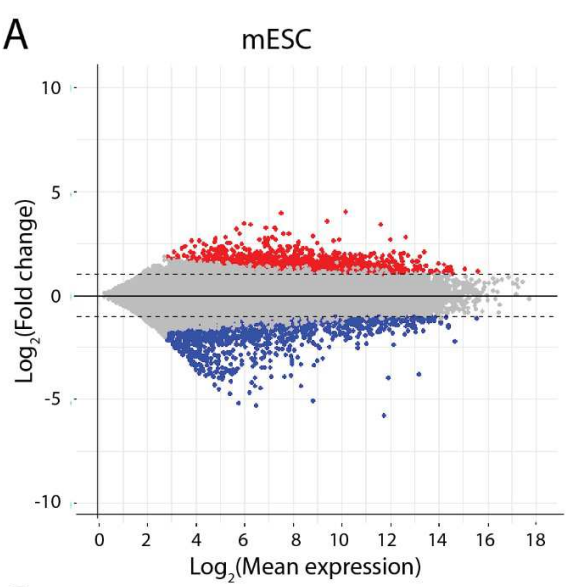
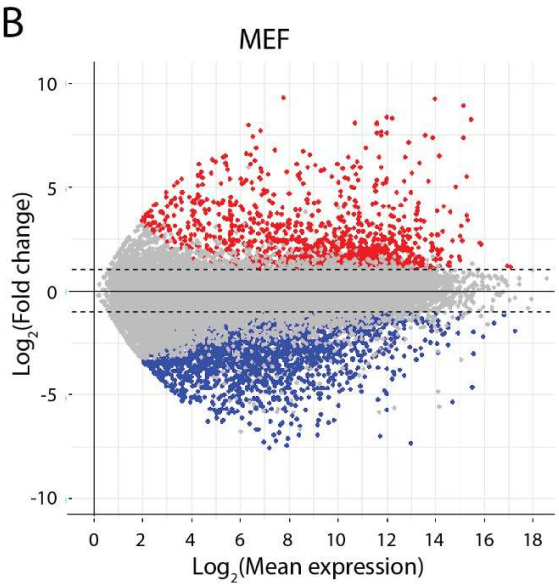

A
mESC
MEF
Marker
Atxn7l3+/+
Atxn7l3+/+
Atxn7l3-/-
Atxn7l3-/-
Atxn7l3+/+
Atxn7l3+/+
Atxn7l3-/-
Atxn7l3-/-
(KD)
35-
25-
15-
←H2Bub1
←H3
1 2 3 4 5 6 7 8 9
B
Atxn7l3+/+
Atxn7l3-/-
H2Bub1/H3
10.0
7.5
5.0
2.5
0.0
mESC
MEF
C
mESC
Atxn7l3+/+
Atxn7l3-/-
200 µm
D
mESC
Atxn7l3+/+
Atxn7l3-/-
Cell number (10^5)
60
40
20
0
Days 0 1 2 3 4 5 6
***
***
***
E
MEF
Atxn7l3+/+
Atxn7l3-/-
200 µm
F
MEF
Atxn7l3+/+
Atxn7l3-/-
Cell number (10^4)
8
6
4
0
Days 0 1 2 3 4 5 6
*
*

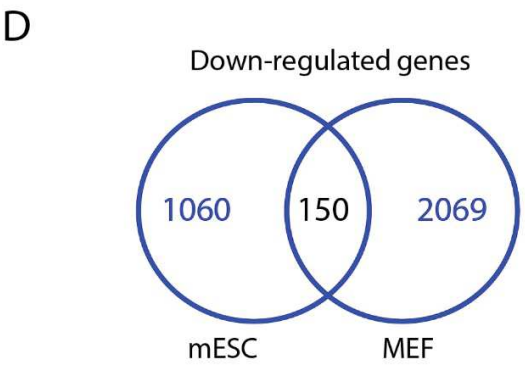
Figure 3:





C

	mESC		MEF	
Fold change	Differentially expressed genes		Differentially expressed genes	
KO vs WT	Up-regulated	Down-regulated	Up-regulated	Down-regulated
2	1163	1210	1314	2219
4	110	157	603	1568
8	10	25	328	915
16	0	2	173	354
32	0	1	98	118
64	0	0	45	30
128	0	0	20	23
256	0	0	9	7



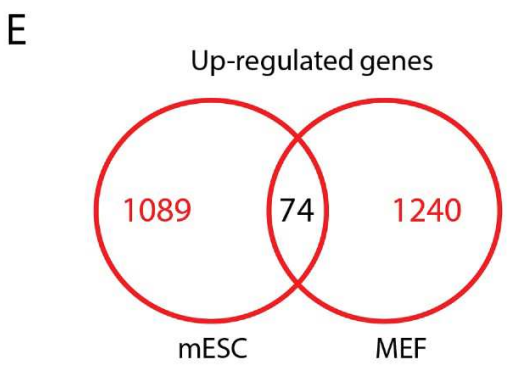


Figure 4:

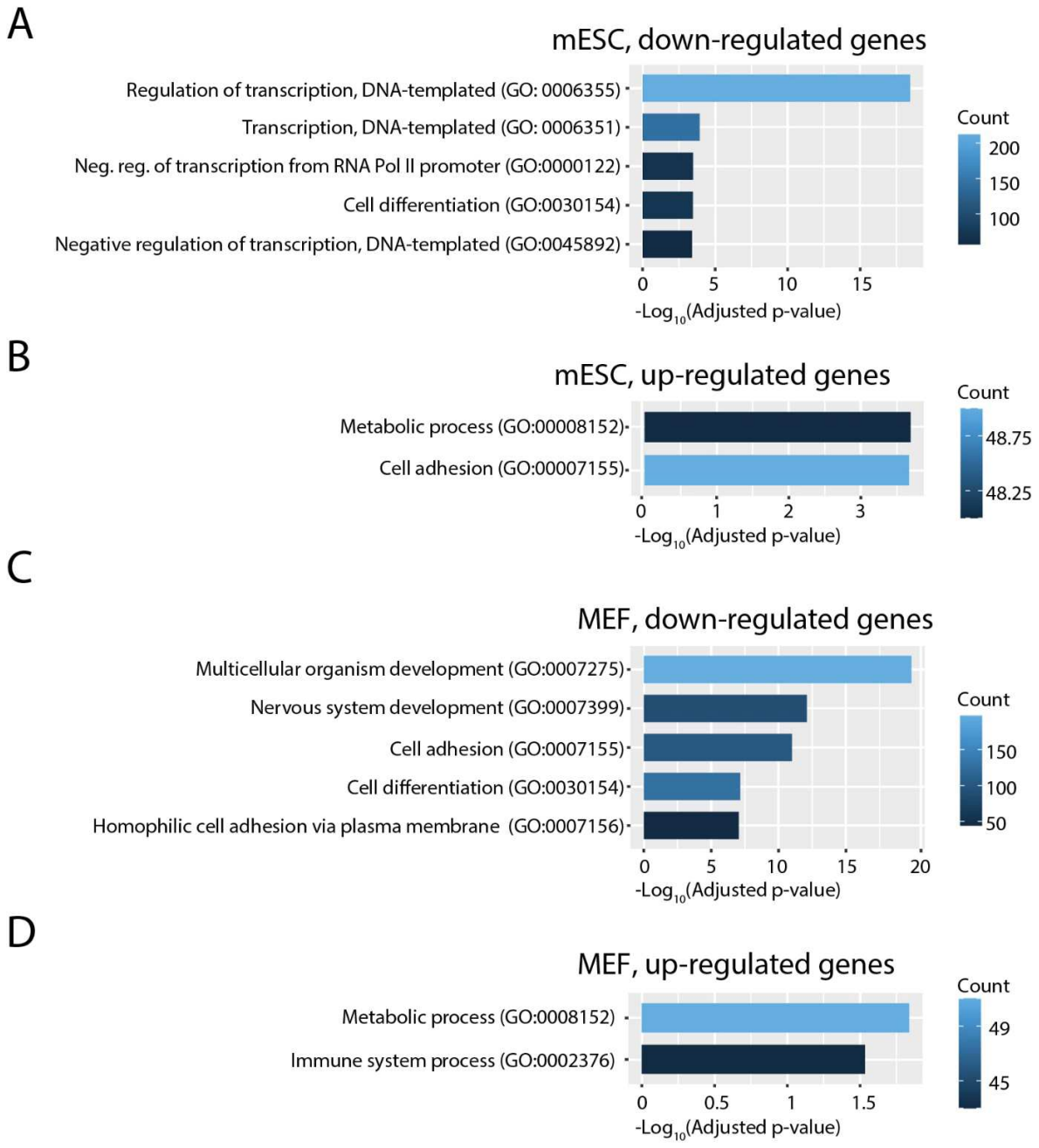

A
mESC, down-regulated genes
Regulation of transcription, DNA-templated (GO: 0006355)
Transcription, DNA-templated (GO: 0006351)
Neg. reg. of transcription from RNA Pol II promoter (GO:0000122)
Cell differentiation (GO:0030154)
Negative regulation of transcription, DNA-templated (GO:0045892)
0 5 10 15
-Log10(Adjusted p-value)
Count
200
150
100
B
mESC, up-regulated genes
Metabolic process (GO:00008152)
Cell adhesion (GO:00007155)
0 1 2 3
-Log10(Adjusted p-value)
Count
48.75
48.25
C
MEF, down-regulated genes
Multicellular organism development (GO:0007275)
Nervous system development (GO:0007399)
Cell adhesion (GO:0007155)
Cell differentiation (GO:0030154)
Homophilic cell adhesion via plasma membrane (GO:0007156)
0 5 10 15 20
-Log10(Adjusted p-value)
Count
150
100
50
D
MEF, up-regulated genes
Metabolic process (GO:0008152)
Immune system process (GO:0002376)
0 0.5 1 1.5
-Log10(Adjusted p-value)
Count
49
45

Figure 5:

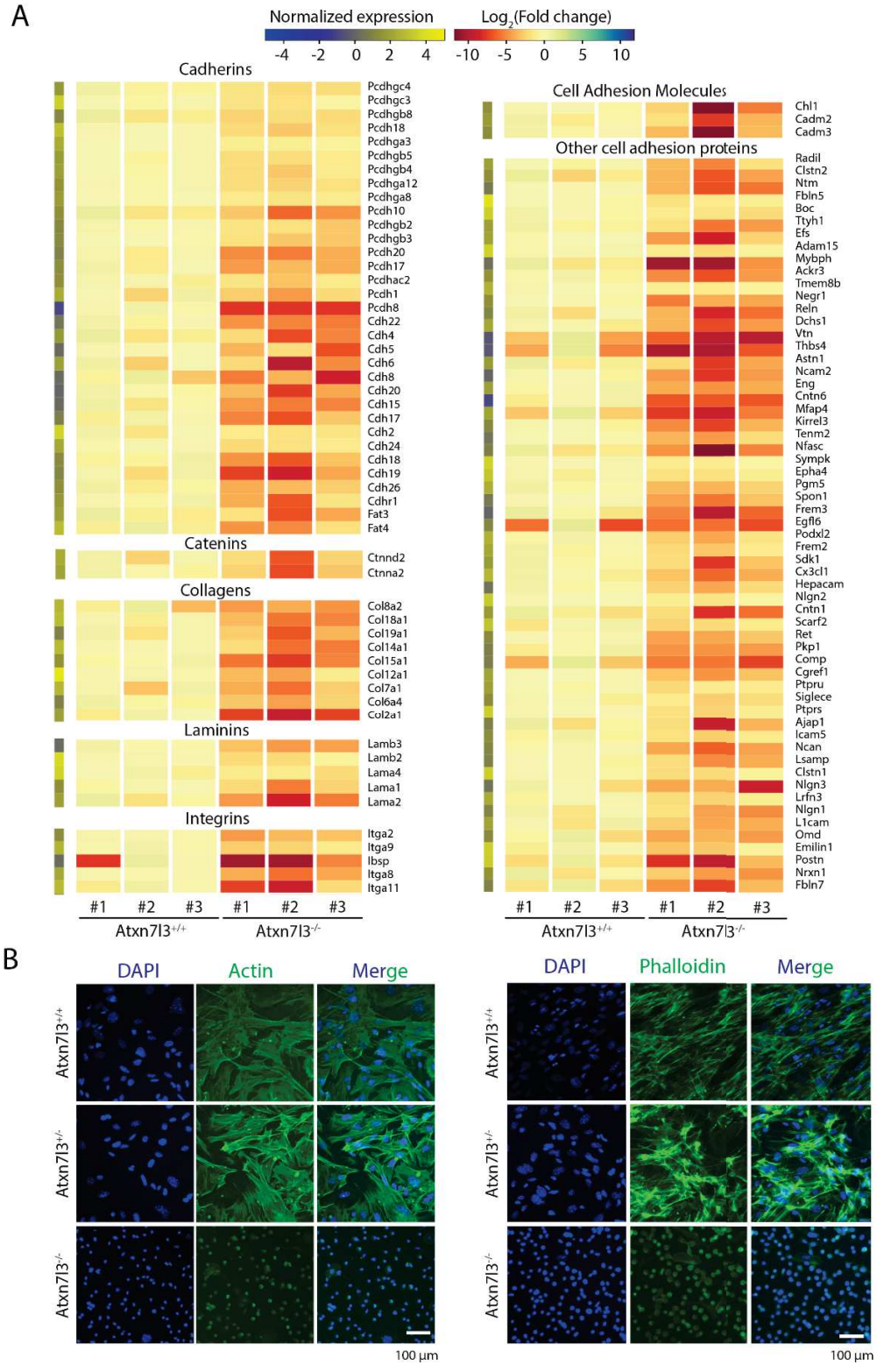

A
Normalized expression
-4 -2 0 2 4
Log2(Fold change)
-10 -5 0 5 10
Cadherins
Pcdhgc4
Pcdhgc3
Pcdhgb8
Pcdh18
Pcdhga3
Pcdhgb5
Pcdhgb4
Pcdhga12
Pcdhga8
Pcdh10
Pcdhgb2
Pcdhgb3
Pcdh20
Pcdh17
Pcdhac2
Pcdh1
Pcdh8
Cdh22
Cdh4
Cdh5
Cdh6
Cdh8
Cdh20
Cdh15
Cdh17
Cdh2
Cdh24
Cdh18
Cdh19
Cdh26
Cdhr1
Fat3
Fat4
Catenins
Ctnnd2
Ctnna2
Collagens
Col8a2
Col18a1
Col19a1
Col14a1
Col15a1
Col12a1
Col7a1
Col6a4
Col2a1
Laminins
Lamb3
Lamb2
Lama4
Lama1
Lama2
Integrins
Itga2
Itga9
Ibsp
Itga8
Itga11
#1 #2 #3 #1 #2 #3
Atxn7l3+/+
Atxn7l3-/-
Cell Adhesion Molecules
Chl1
Cadm2
Cadm3
Other cell adhesion proteins
Radil
Clstn2
Ntm
Fbln5
Boc
Ttyh1
Efs
Adam15
Mybph
Ackr3
Tmem8b
Negr1
Reln
Dchs1
Vtn
Thbs4
Astn1
Ncam2
Eng
Cntn6
Mfap4
Kirrel3
Tenm2
Nfasc
Sympk
Epha4
Pgm5
Spon1
Frem3
Egfl6
Podxl2
Frem2
Sdk1
Cx3cl1
Hepacam
Nlgn2
Cntn1
Scarf2
Ret
Pkp1
Comp
Cgref1
Ptpru
Siglece
Ptprs
Ajap1
Icam5
Ncan
Lsamp
Clstn1
Nlgn3
Lrfn3
Nlgn1
L1cam
Omd
Emilin1
Postn
Nrxn1
Fbln7
#1 #2 #3 #1 #2 #3
Atxn7l3+/+
Atxn7l3-/-
B
DAPI
Actin
Merge
Atxn7l3+/+
Atxn7l3+/-
Atxn7l3-/-
100 µm
DAPI
Phalloidin
Merge
Atxn7l3+/+
Atxn7l3+/-
Atxn7l3-/-
100 µm

Figure 6:

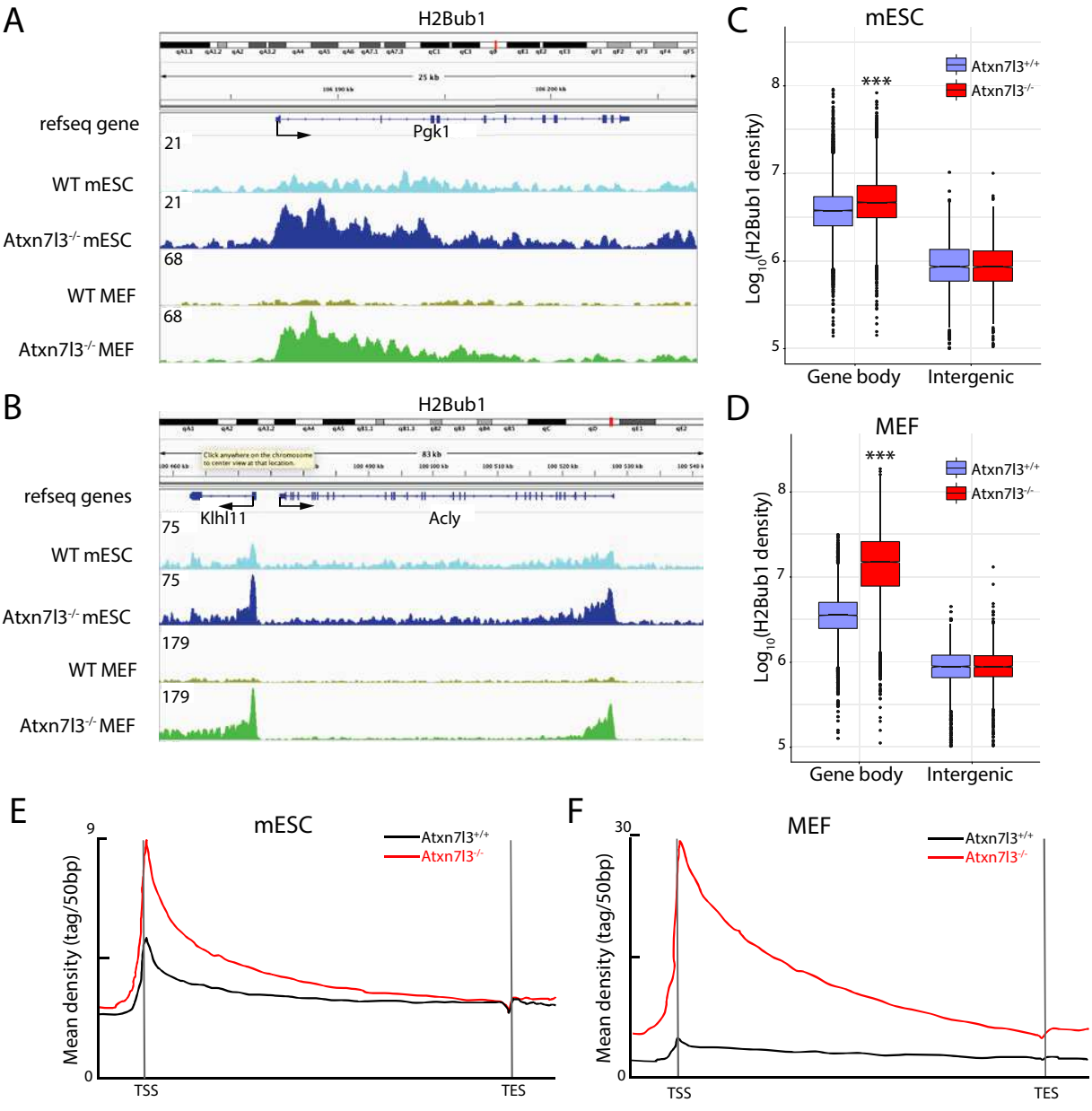
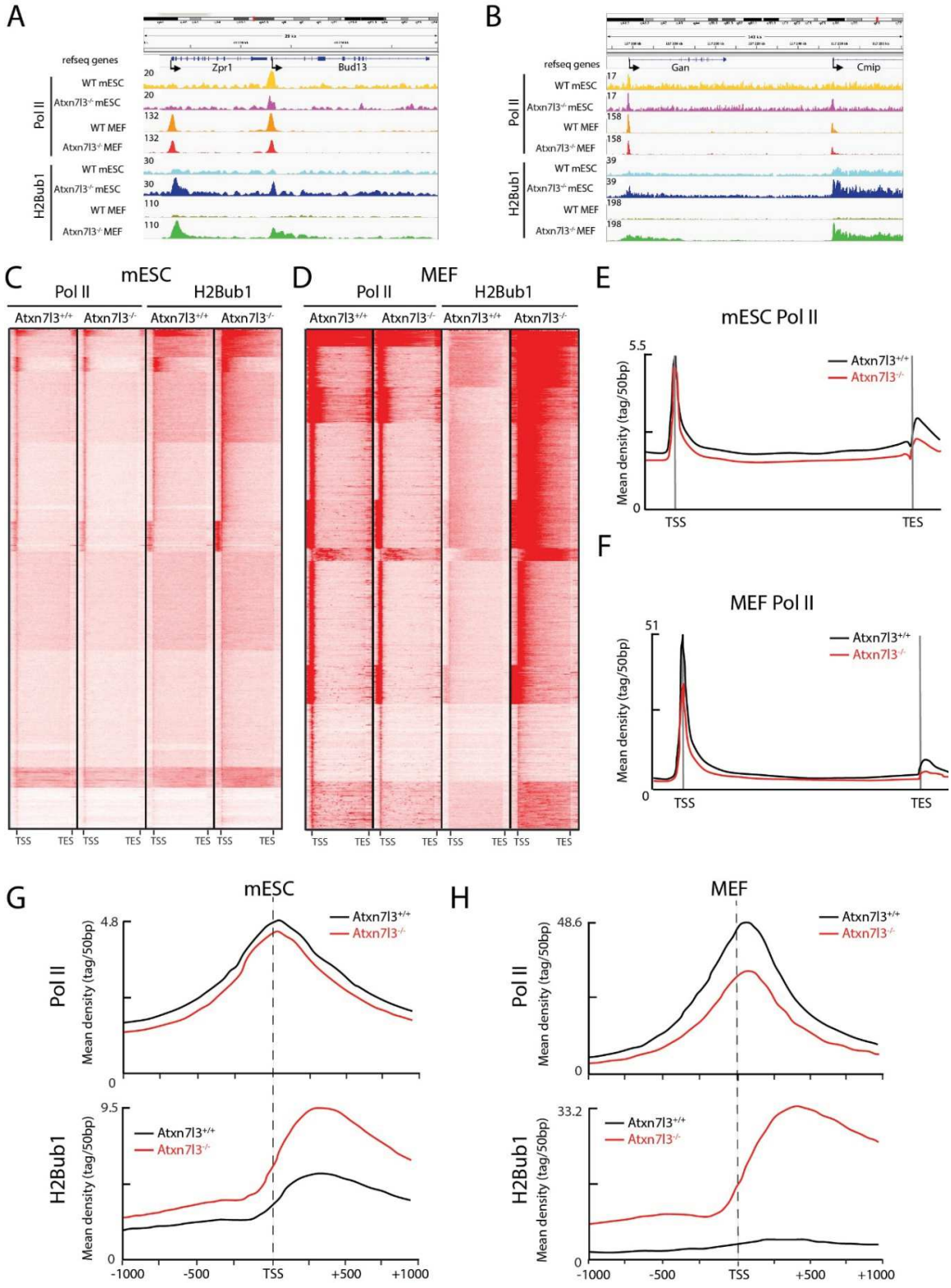

A
H2Bub1
refseq gene
Pgk1
WT mESC
Atxn7l3-/- mESC
WT MEF
Atxn7l3-/- MEF
21
21
68
68
B
H2Bub1
refseq genes
Klhl11
Acly
WT mESC
Atxn7l3-/- mESC
WT MEF
Atxn7l3-/- MEF
75
75
179
179
C
mESC
***
Atxn7l3+/+
Atxn7l3-/-
Log10(H2Bub1 density)
Gene body
Intergenic
D
MEF
***
Atxn7l3+/+
Atxn7l3-/-
Log10(H2Bub1 density)
Gene body
Intergenic
E
mESC
Atxn7l3+/+
Atxn7l3-/-
Mean density (tag/50bp)
TSS
TES
F
MEF
Atxn7l3+/+
Atxn7l3-/-
Mean density (tag/50bp)
TSS
TES

Figure 7:


A
refseq genes
Zpr1
Bud13
Pol II
H2Bub1
WT mESC
Atxn7l3-/- mESC
WT MEF
Atxn7l3-/- MEF
B
refseq genes
Gan
Cmip
C
mESC
Pol II
H2Bub1
Atxn7l3+/+
Atxn7l3-/-
TSS
TES
D
MEF
E
mESC Pol II
Mean density (tag/50bp)
F
MEF Pol II
G
mESC
H
MEF
-1000
-500
TSS
+500
+1000

Supplementary figure legends

Figure S1: Deletion of the *Usp22* and *Atxn7l3* genes in mouse

A. Generation of the *Usp22*$^{tm1d(KOMP)Wtsi}$ allele (*Usp22*$^{-}$) after FLP and CRE recombination of the *Usp22*$^{tm1a(KOMP)Wtsi}$ initial allele. The primers used for genotyping are indicated on the maps. **B.** PCR analysis of DNA samples using the Ef and Er primers from *Usp22*$^{+/+}$ and *Usp22*$^{-/-}$ mice. The 200 and 500 bp bands correspond to the WT and null alleles, respectively. **C.** Generation of the *Atxn7l3*$^{m1.1(KOMP)Wtsi}$ allele (*Atxn7l3*$^{-}$) after FLP recombination of the *Atxn7l3*$^{m1(KOMP)Wtsi}$ initial allele. The primers used for genotyping are indicated on the maps. **D.** PCR analysis of DNA samples using the Ef and Er primers from *Atxn7l3*$^{+/+}$, *Atxn7l3*$^{+/-}$ and *Atxn7l3*$^{-/-}$ mice. The 200 and 500 bp bands correspond to the WT and null alleles, respectively.

Figure S2:

A. RT-qPCR analysis of genes associated with pluripotency in *Atxn7l3*$^{+/+}$ (black) and *Atxn7l3*$^{-/-}$ (red) ESCs. Y axis indicates the relative mRNA expression to the *Pgk1* housekeeping gene. *Atxn7l3*$^{+/+}$ RNA expression level is normalized to 1. Error bars represent ±SD from three biological samples with three technical replicates (represented by grey dots) for each. **(B)** Apoptosis measured by Annexin V and Propidium iodide (PI) staining and quantified by flow cytometry in mESC and MEF cells. **(C)** Quantification of cell cycle phase distribution by flow cytometry from Propidium iodide (PI) treated *Atxn7l3*$^{+/+}$ (black) and *Atxn7l3*$^{-/-}$ (red) mESC and MEF cells. Error bars indicate ±SD based on three biological replicates (represented by grey dots).

Figure S3

A-B. Principal component analysis of control (black) and *Atxn7l3*$^{-/-}$ (red) RNA-seq data in mESCs (A) and MEFs (B).

C. Hierarchical clustering of *Atxn7l3*$^{-/-}$ and control MEF RNA-seq data with 921 RNA-seq data from 272 distinct mouse cell types or tissues. **D.** RT-qPCR analysis of up-regulated, down-regulated and unchanged genes from RNA-seq in mESC. Y axis indicates the relative mRNA expression to the *Pgk1* housekeeping gene in *Atxn7l3*$^{-/-}$ mESC compared to WT controls. WT gene expression is normalized to 1. Error bars represent ±SD from two biological and three technical replicates (represented by grey dots). **(E)** RT-qPCR analysis of up-regulated, down-regulated and unchanged genes from RNA-seq in *Atxn7l3*$^{-/-}$ MEFs compared to WT control. Y axis indicates the relative mRNA expression to the *Pgk1 and Hsp90ab1* housekeeping genes. Error bars represent ±SD from two biological and three technical replicates (represented by grey dots)

Figure S4

A-B. Volcano plots comparing gene expression between *Atxn7l3*$^{-/-}$ and WT control mESC (A) and MEF(B). Blue dots correspond to significantly differentially expressed genes with adjusted *p*-values ≤ 0.05 and absolute $\log_2$(Fold change) ≥ 1. Green dots indicate genes with adjusted *p*-values < 0.05 and absolute $\log_2$(Fold change) < 1. Red dots indicate genes with adjusted *p*-values > 0.05 and absolute $\log_2$(Fold change) > 1. Grey dots indicate adjusted *p*-values > 0.05 and absolute $\log_2$(Fold change) < 1.

Figure S5

A-B. Heat maps showing the distribution of H2Bub1 on bodies of 31277 annotated transcripts in mESCs(A) and 29450 annotated transcripts in MEFs(B). **C-D.** Scatter plots representing H2Bub1 densities in control cells relative to *Atxn7l3*$^{-/-}$ mESC (C) and MEF (D). From 11172 expressed transcripts in mESCs, 11010 expressed transcripts containing at least 1 read were selected (blue dot). From 11113 expressed transcripts in MEF cells, 10946 expressed transcripts containing at least 1 read were selected (blue dot).

Figure S6

A-E H2Bub1 and Pol II binding profiles are shown at four selected genes in MEFs using the IGV genome browser. Direction of the transcription is indicated by arrows. Scaled tag densities for each gene are indicated on the left.

Figure S1:

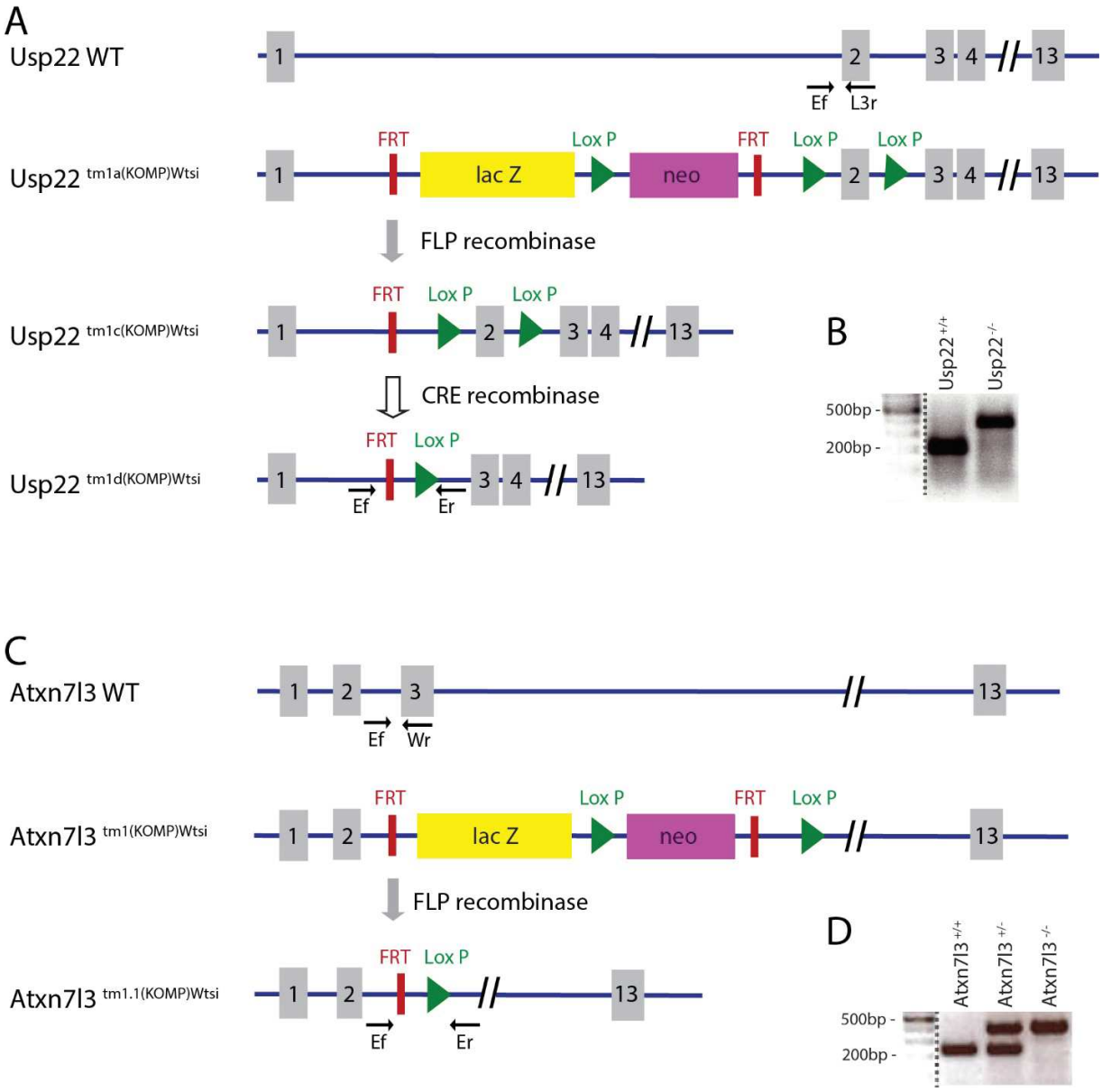

A
Usp22 WT
1
2
3
4
13
Ef
L3r
Usp22 tm1a(KOMP)Wtsi
FRT
lac Z
Lox P
neo
FRT
Lox P
Lox P
FLP recombinase
Usp22 tm1c(KOMP)Wtsi
CRE recombinase
Usp22 tm1d(KOMP)Wtsi
Er
B
Usp22+/+
Usp22-/-
500bp -
200bp -
C
Atxn7l3 WT
Wr
Atxn7l3 tm1(KOMP)Wtsi
Atxn7l3 tm1.1(KOMP)Wtsi
D
Atxn7l3+/+
Atxn7l3+/-
Atxn7l3-/-

Figure S2:

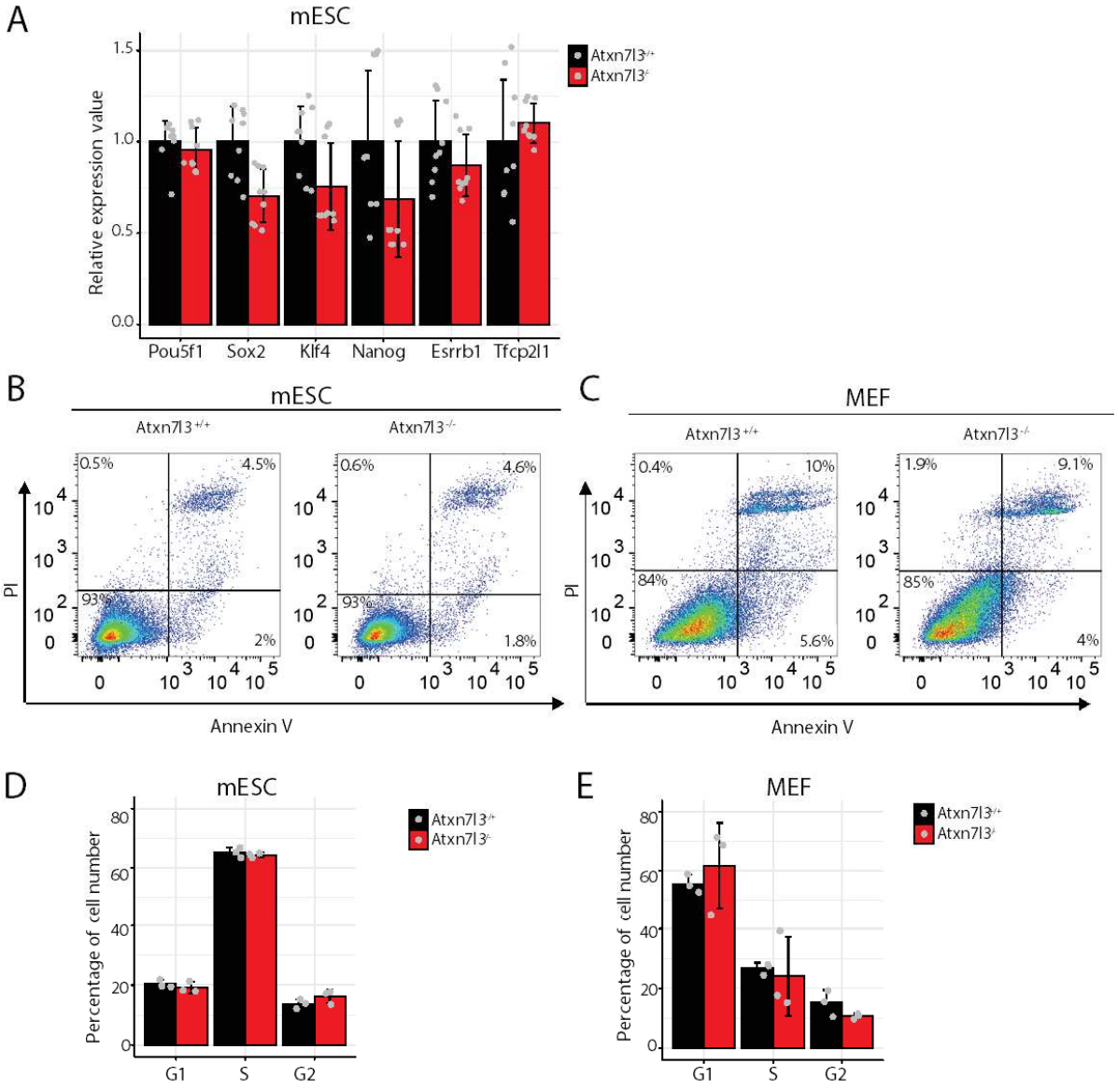

A
mESC
Relative expression value
Atxn7l3+/+
Atxn7l3-/-
Pou5f1
Sox2
Klf4
Nanog
Esrrb1
Tfcp2l1
B
mESC
C
MEF
PI
Annexin V
0.5%
4.5%
93%
2%
0.6%
4.6%
1.8%
0.4%
10%
84%
5.6%
1.9%
9.1%
85%
4%
D
mESC
E
MEF
Percentage of cell number
G1
S
G2

Figure S3:

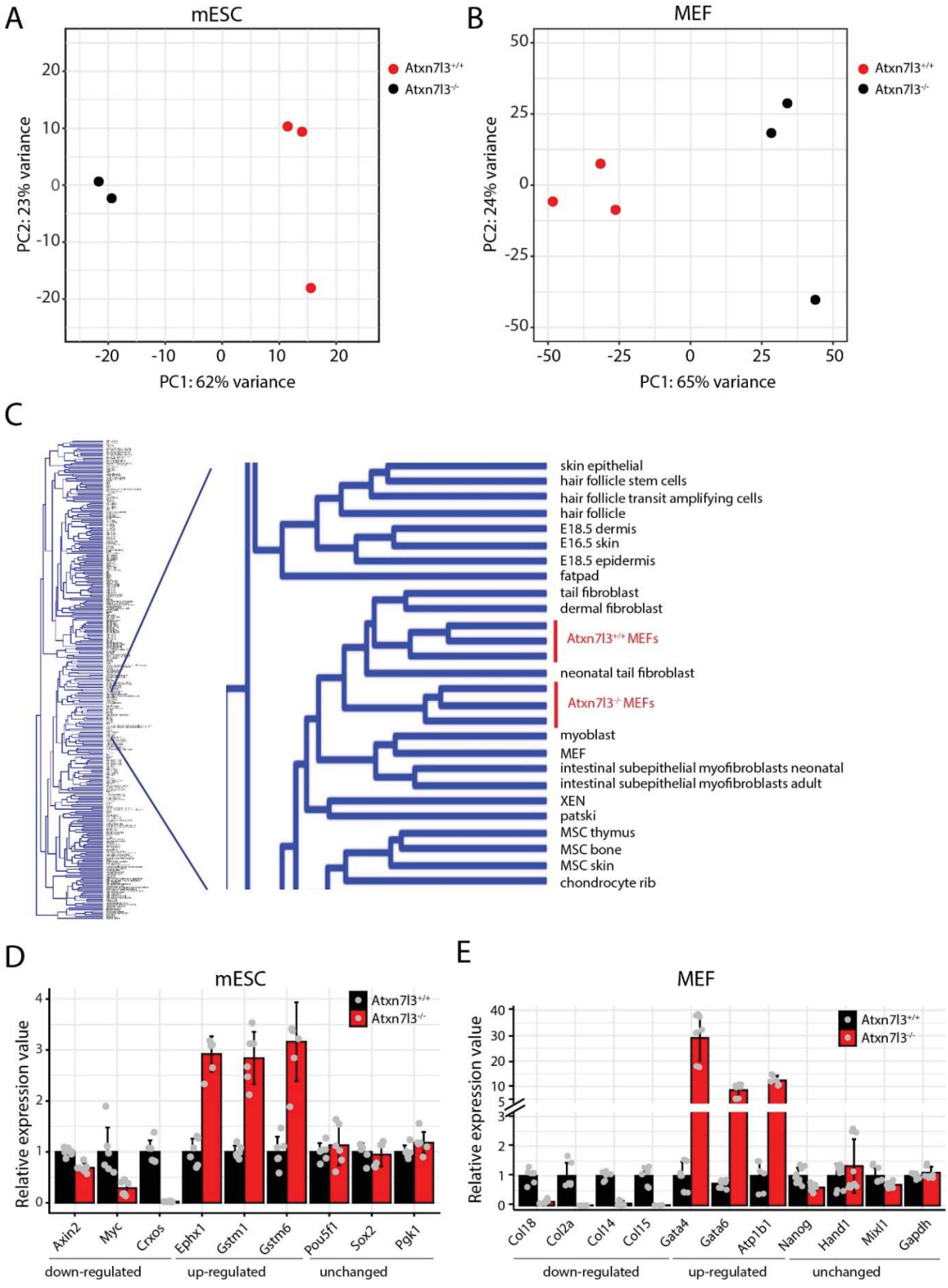

A
mESC
PC2: 23% variance
PC1: 62% variance
Atxn7l3+/+
Atxn7l3-/-
B
MEF
PC2: 24% variance
PC1: 65% variance
C
skin epithelial
hair follicle stem cells
hair follicle transit amplifying cells
hair follicle
E18.5 dermis
E16.5 skin
E18.5 epidermis
fatpad
tail fibroblast
dermal fibroblast
Atxn7l3+/+ MEFs
neonatal tail fibroblast
Atxn7l3-/- MEFs
myoblast
MEF
intestinal subepithelial myofibroblasts neonatal
intestinal subepithelial myofibroblasts adult
XEN
patski
MSC thymus
MSC bone
MSC skin
chondrocyte rib
D
mESC
Relative expression value
Axin2
Myc
Cxos
Ephx1
Gstm1
Gstm6
Pou5f1
Sox2
Pgk1
down-regulated
up-regulated
unchanged
E
MEF
Relative expression value
Col18
Col2a
Col14
Col15
Gata4
Gata6
Atp1b1
Nanog
Hand1
Mixl1
Gapdh
down-regulated
up-regulated
unchanged

Figure S4:

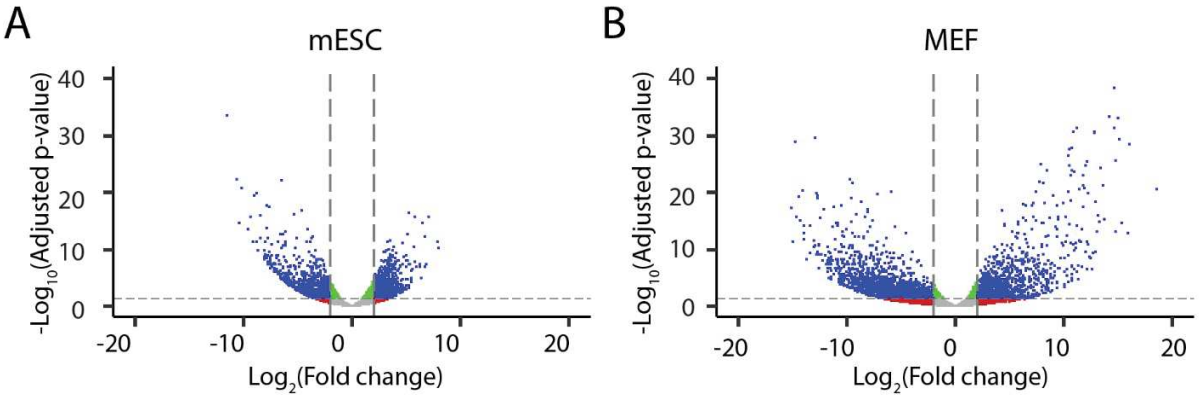

A
mESC
B
MEF
-Log10(Adjusted p-value)
Log2(Fold change)
40
30
20
10
0
-20
-10
0
10
20

Figure S5:

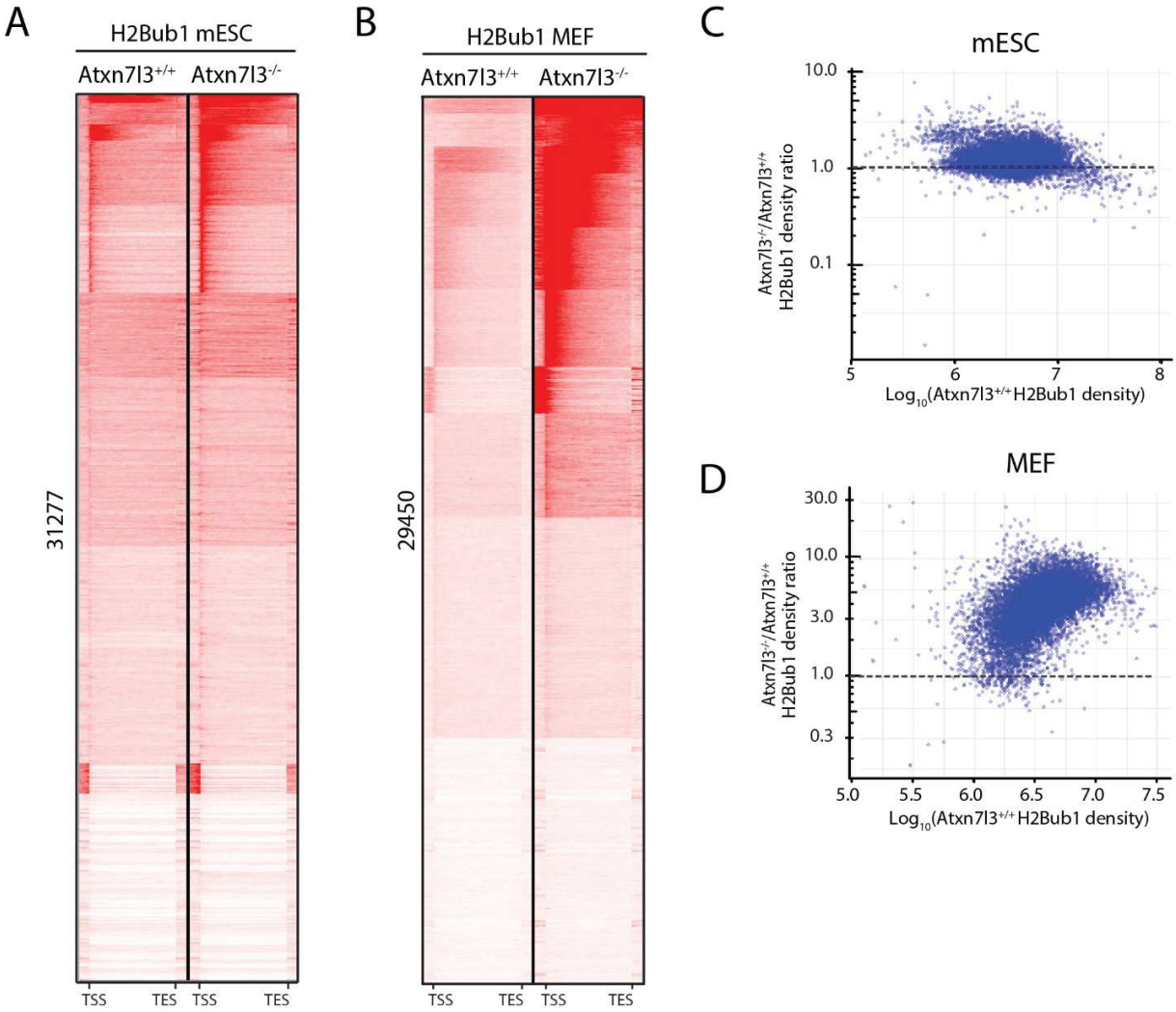
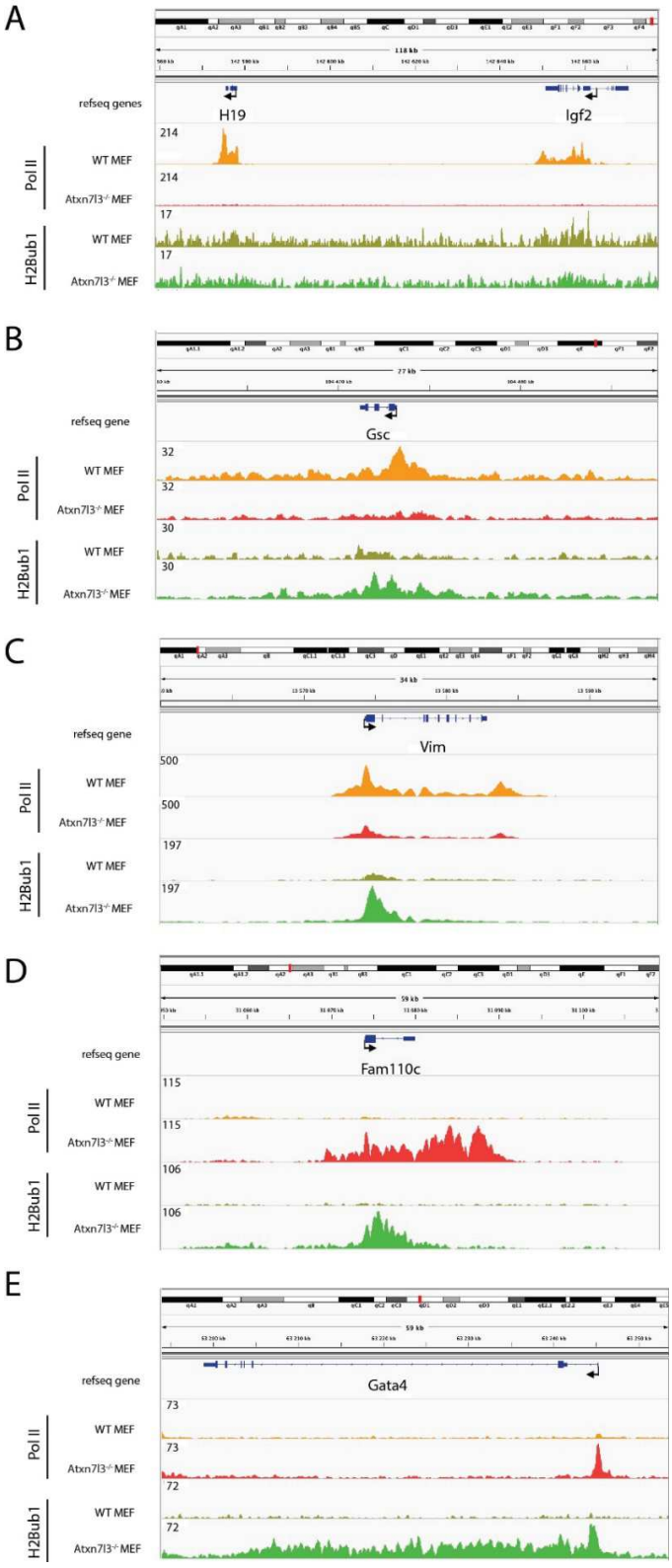

A
H2Bub1 mESC
Atxn7l3+/+ Atxn7l3-/-
31277
TSS TES TSS TES
B
H2Bub1 MEF
Atxn7l3+/+ Atxn7l3-/-
29450
TSS TES TSS TES
C
mESC
Atxn7l3-/-/Atxn7l3+/+ H2Bub1 density ratio
10.0
1.0
0.1
5 6 7 8
Log10(Atxn7l3+/+ H2Bub1 density)
D
MEF
Atxn7l3-/-/Atxn7l3+/+ H2Bub1 density ratio
30.0
10.0
3.0
1.0
0.3
5.0 5.5 6.0 6.5 7.0 7.5
Log10(Atxn7l3+/+ H2Bub1 density)

Figure S6:


A
refseq genes
H19
Igf2
Pol II
WT MEF
Atxn7l3-/- MEF
H2Bub1
WT MEF
Atxn7l3-/- MEF
214
214
17
17
B
refseq gene
Gsc
32
32
30
30
C
refseq gene
Vim
500
500
197
197
D
refseq gene
Fam110c
115
115
106
106
E
refseq gene
Gata4
73
73
72
72

3.2 Unpublished results

3.2.1 Testing the role of the DUBm in mESC

As showed in the first part, *Atxn7l3* knock-out (KO) mice die around E11.5 indicating that the three ATXN7L3-related DUBs activities have an essential function during mouse development. However, the molecular analysis of embryonic signalling is often limited by the small size and heterogeneity of embryonic tissues. The generation of ESCs and the development of ESC differentiation technics that mimic embryonic cell differentiation allow to further study our questions in an *in vitro* system system. To better understand the mechanistic basis of the *Atxn7l3*-dependent DUBm-s, we generated two ESC cell lines in which the *Atxn7l3* gene was inactivated: the first one was generated from the blastocysts of *Atxn7l3*$^{-/-}$ mice (described in our submitted publication), and a second ESC line where a homozygous *Atxn7l3* inactivation was obtained in E14 mESCs by using the CRISP-Cas9 technique (Figure 6-1) (obtained by Veronique Fischer, PhD candidate in the lab). Unlike blastocysts derived mESCs, the E14 mESCs can be cultured without feeder cells and its phenotype is more uniform. Therefore, the E14 mESCs are more easily to handle compared with our blastocysts derived mESCs. Considering the *Atxn7l3*$^{-/-}$ mice are development delayed as early as E7.5 which is a stage undergoing gastrulation. E14 mESCs are used to test how *Atxn7l3* affects the expression of gastrulation-related genes by *in vitro* differentiation experiment.

	Screened	n° HET	n° HOM	% of HET	% of HOM
***Atxn7l3* KO**	13	5	4	30,7	38,5

Figure 6-1. *Atxn7l3* KO E14 mESCs were obtained by using the CRISP-Cas9 technique. HET: heterozygous; HOM: homozygote.

3.2.1.1 Loss of *Atxn7l3* affects mESCs self-renewal

To understand the function of *Atxn7l3*-dependent DUBm-s in mESC self-renewal, we did colony-formation assay in medium containing serum with LIF and 2i (GSK3 and MEK inhibitors). We found that depletion of *Atxn7l3* the mESCs were still alive but inhibited the efficiency of colony-formation (Figure 6-2A, B). Consistently, *Atxn7l3* deletion also caused a decrease in cell number after 3 days of culture (Figure 6-2D). Moreover, our cell cycle analysis showed that *Atxn7l3* loss

of function led to a strong delay of G1 to S phase transition and a moderate delay from G2/M to G1 phase transition (Figure 6-2E). Together, these results indicate that *Atxn7l3* is essential for mESC self-renewal.

To test whether loss of *Atxn7l3* affects mESC pluripotency, the expression of several pluripotency marker genes was assessed by RT-qPCR. Our results showed that the expression of the tested pluripotency marker genes, like *Klf4, Nanog*, *Sox2* and *Pou5f1* (Martello and Smith, 2014), wassimilar between WT and *Atxn7l3* KO mESCs (Figure 6-2F). Besides, *Atxn7l3* KO mESCs still maintain pluripotency, as they were Alkaline Phosphatase (AP) positive (Figure 6-2C). Together, these results suggest that ATXN7L3-related DUBm-s facilitate mESC self-renewal, but have no obvious effect on mESC pluripotency maintenance, which are similar to the blastocysts derived *Atxn7l3*$^{-/-}$ mESCs

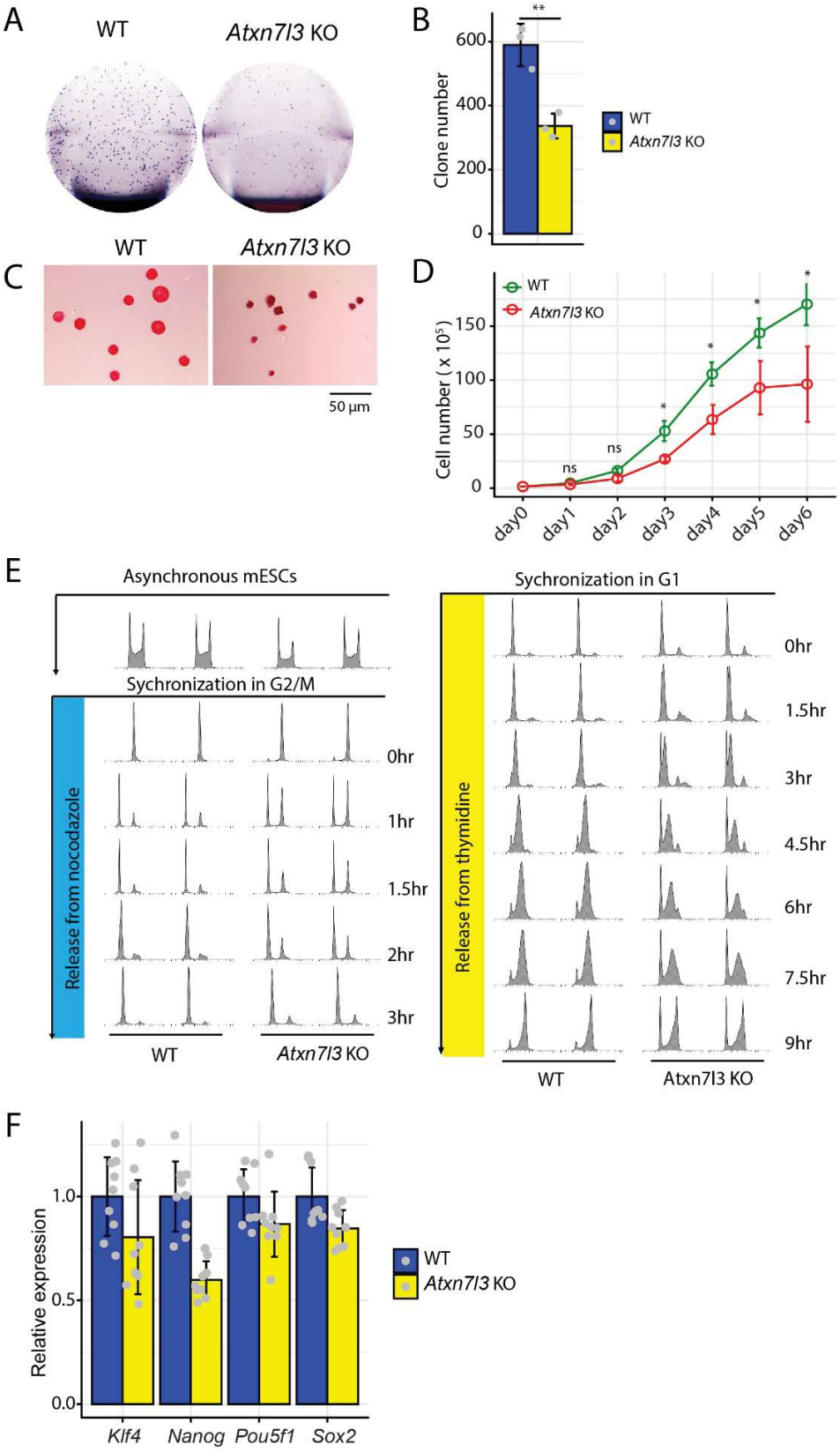

A
WT
Atxn7l3 KO
B
Clone number
600
400
200
0
**
WT
Atxn7l3 KO
C
WT
Atxn7l3 KO
50 μm
D
WT
Atxn7l3 KO
Cell number (x 10⁵)
150
100
50
0
ns
ns
*
*
*
*
day0
day1
day2
day3
day4
day5
day6
E
Asynchronous mESCs
Sychronization in G2/M
Release from nocodazole
0hr
1hr
1.5hr
2hr
3hr
WT
Atxn7l3 KO
Sychronization in G1
Release from thymidine
0hr
1.5hr
3hr
4.5hr
6hr
7.5hr
9hr
WT
Atxn7l3 KO
F
Relative expression
1.0
0.5
0.0
Klf4
Nanog
Pou5f1
Sox2
WT
Atxn7l3 KO

Figure 6-2. *Atxn7l3* promotes mESCs self-renewal by facilitating cell cycle transition. (A) Morphology of WT and *Atxn7l3* KO mESC colonies cultured in serum/LIF plus 2i medium for 6 days. (B) Quantification mESC colonies from Figure 1A. Error bars indicate ±SD based on three biological samples (represented by grey dots). Statistical significance was calculated using the Mann–Whitney test (ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$). (C) Alkaline phosphatase staining of colonies arising from *Atxn7l3* KO and WT mESCs. Scale bar, 50 μm. (D) The number of mESCs was determined by the cell counting chamber slide at the indicated time points. Error bars indicate ±SD based on three biological samples with three technical replicates, respectively. Statistical significance was calculated using the Mann–Whitney test (ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$). (E) FACS analysis of cell cycle progression after synchronization in G1 or G2 phases by thymidine or nocodazole (as indicated) in *Atxn7l3* KO and WT mESCs. Two biological samples were tested. (F) RT-qPCR results show the expression of pluripotency related genes. Y axis indicates the mRNA expression relative to the housekeeping gene (*Pgk1*). Error bars represent ±SD from three biological samples with three technical replicates (represented by grey dots), respectively.

3.2.1.2 *Atxn7l3* deletion affects the expression of gastrulation-related genes

To understand why *Atxn7l3*$^{-/-}$ embryos are early embryonic lethality, we used *in vitro* mESC differentiation assays to define the roles played by ATXN7L3 during embryonic development. Activin A cooperating with FGF2 induces mESC differentiation into epiblast-like cell (EpiLC) (Chen et al., 2016). During this process, mESCs undergo morphological transformation and express many lineage commitment markers with down-regulation of some pluripotency genes (Brons et al., 2007; Chen et al., 2016). To test whether *Atxn7l3* affects the ability of mESCs to differentiate into EpiLCs, we treated WT or *Atxn7l3* KO E14 mESCs with FGF2 and Activin A as depicted in Figure 6-3A. We found that *Atxn7l3* KO did not influence the EpiLCs morphology (Figure 6-3B). Besides, our RT-qPCR data showed that the pluripotency genes (*Sox2, Klf4, Esrrb)* were down-regulated, whereas the expression of the post implantation epiblast primed genes (*Fgf5, Dnmt3b* and *Otx2*) were up-regulated upon mESCs differentiated into EpiLCs in both WT and *Atxn7l3* KO cells (Figure 6-3C). Thus, ATXN7L3 has no obvious effect on the transition from ESCs to EpiLCs.

Epiblast cells make fate decisions towards mesoderm, endoderm, or ectoderm (Guo et al., 2009). In this process, signaling activator Activin A mimicking Nodal signaling combined with CHIR

mimicking Wnt signaling can stimulate EpiLCs differentiation into primitive streak cells. We found that the expression of EpiLC genes were decreased, whereas the expression of primitive streak genes (*Foxa2, Brachury*, *Wnt3* and *Cdh2)* was increased upon stimulating gastrulation in WT cells. On the contrary, *Atxn7l3* KO EpiLCs failed to induce the expression of primitive streak genes, including *Foxa2, Brachury, and Cdh2*. Additionally, there is a switch of cadherin types from CDH2/N-cadherin to CDH1/E-cadherin during gastrulation (Gheldof and Berx, 2013). However, the *Atxn7l3* KO primitive streak cells expressed higher *Cdh1* but lower *Cdh2* than WT primitive streak cells (Figure 6-3D). In conclusion, these data suggest that ATXN7L3 may be required to facilitate gastrulation transition by promoting proper gene expression levels.

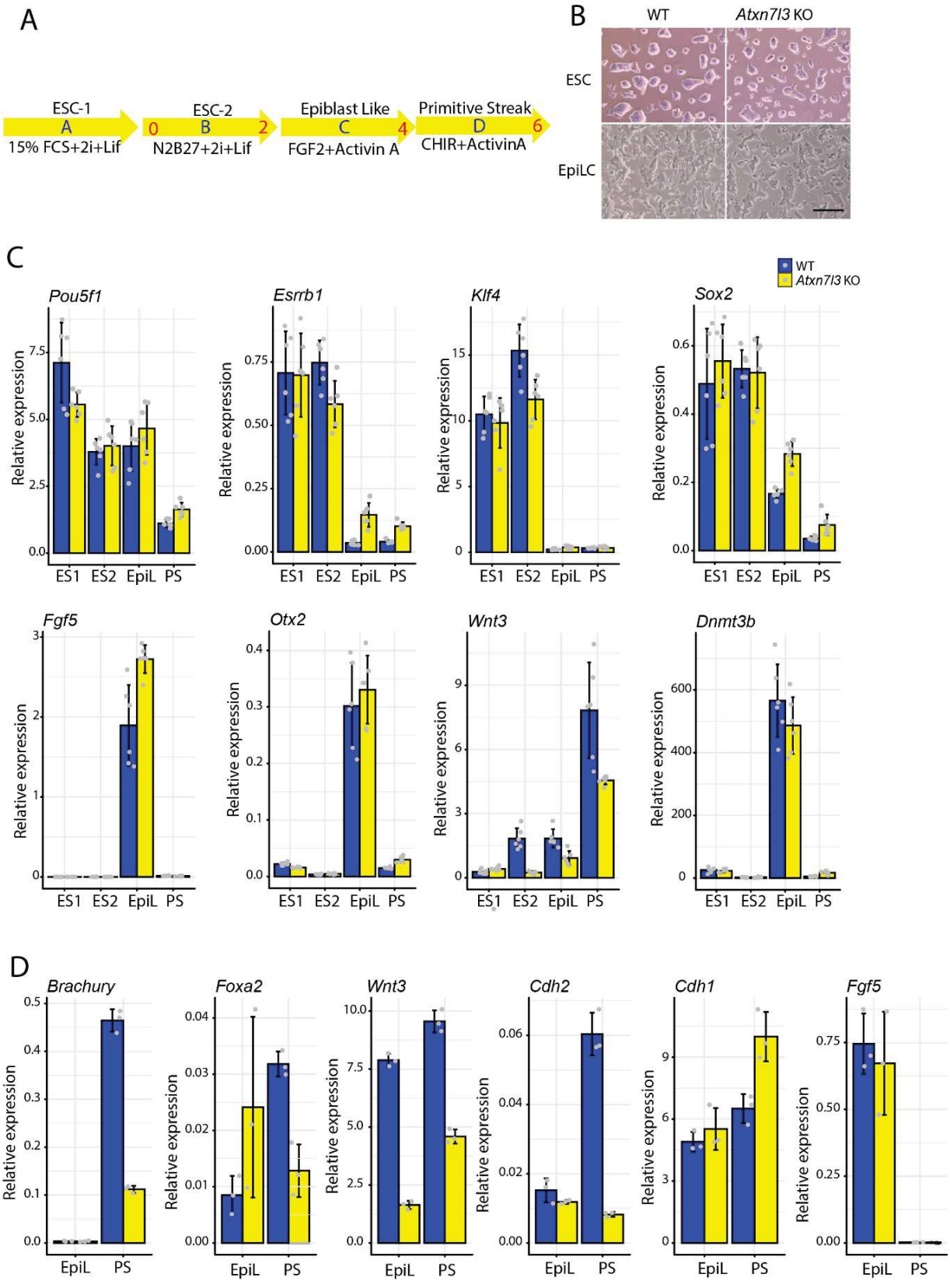

A
ESC-1
A
15% FCS+2i+Lif
0
ESC-2
B
N2B27+2i+Lif
2
Epiblast Like
C
FGF2+Activin A
4
Primitive Streak
D
CHIR+ActivinA
6
B
WT
Atxn7l3 KO
ESC
EpiLC
C
WT
Atxn7l3 KO
Pou5f1
Esrrb1
Klf4
Sox2
Fgf5
Otx2
Wnt3
Dnmt3b
Relative expression
ES1 ES2 EpiL PS
D
Brachury
Foxa2
Wnt3
Cdh2
Cdh1
Fgf5
EpiL PS

Figure 6-3. *Atxn7l3* facilitates gastrulation transition. (A) Flow diagram to generate primitive streak cells from mESCs. (B) The morphology of WT and *Atxn7l3* KO mESC colonies in serum/LIF plus 2i medium for 3 days (up); The morphology of WT and *Atxn7l3* KO EpiLCs in serum-free medium with FGF2 and Activin A for 2 days (down). Scale bar: 400 μm. (C) RT-qPCR analysis of the expression of pluripotent genes. Y axis indicates the mRNA expression relative to the housekeeping gene (*Pgk1*). Error bars represent ±SD from three biological samples with three technical replicates (represented by grey dots), respectively. (D) RT-qPCR analysis of the expression of gastrulation genes. Y axis indicates the mRNA expression relative to the housekeeping gene (*Pgk1*). Error bars represent ±SD from two biological samples with three technical replicates (represented by grey dots), respectively.

3.2.1.3 *Atxn7l3* promotes embryoid body growth

ESCs tend to aggregate when cultured in suspension without inhibitors of differentiation (2i and LIF) and undergo stepwise morphological change to form embryoid bodies (EBs). This process recapitulate the transition from ESC to embryonic gastrulation (Li et al., 2003), thereby providing opportunities to define molecular events *in vitro*. During ESC to EB differentiation the mRNA expression of *Atxn7l3* peaked in WT ESCs and gradually reduced during differentiation (Figure 6-4A). Consistently, we found that ATXN7L3 protein levels were also high in WT mESCs and at day 2 WT EBs, but then it decreased gradually during EB differentiation. Importantly, ATXN7L3 could not be detected in KO ESCs or EBs (Figure 6-4B). These results indicate that *Atxn7l3* might be required during early EB differentiation stage. Next we measured the diameters of the EBs at indicated time point (Figure 6-4C). The average size of EBs was smaller in *Atxn7l3* KO condition compared with its control condition (Figure 6-4D). Interestingly, we also found that *Atxn7l3* deletion did not influence the stability of SAGA core-module subunit SUPT7L, but affected the stability of its partner protein ENY2 (Figure 6-4B). Together these results indicate that ATXN7L3-related DUBm-s may facilitate EB growth during differentiation process.

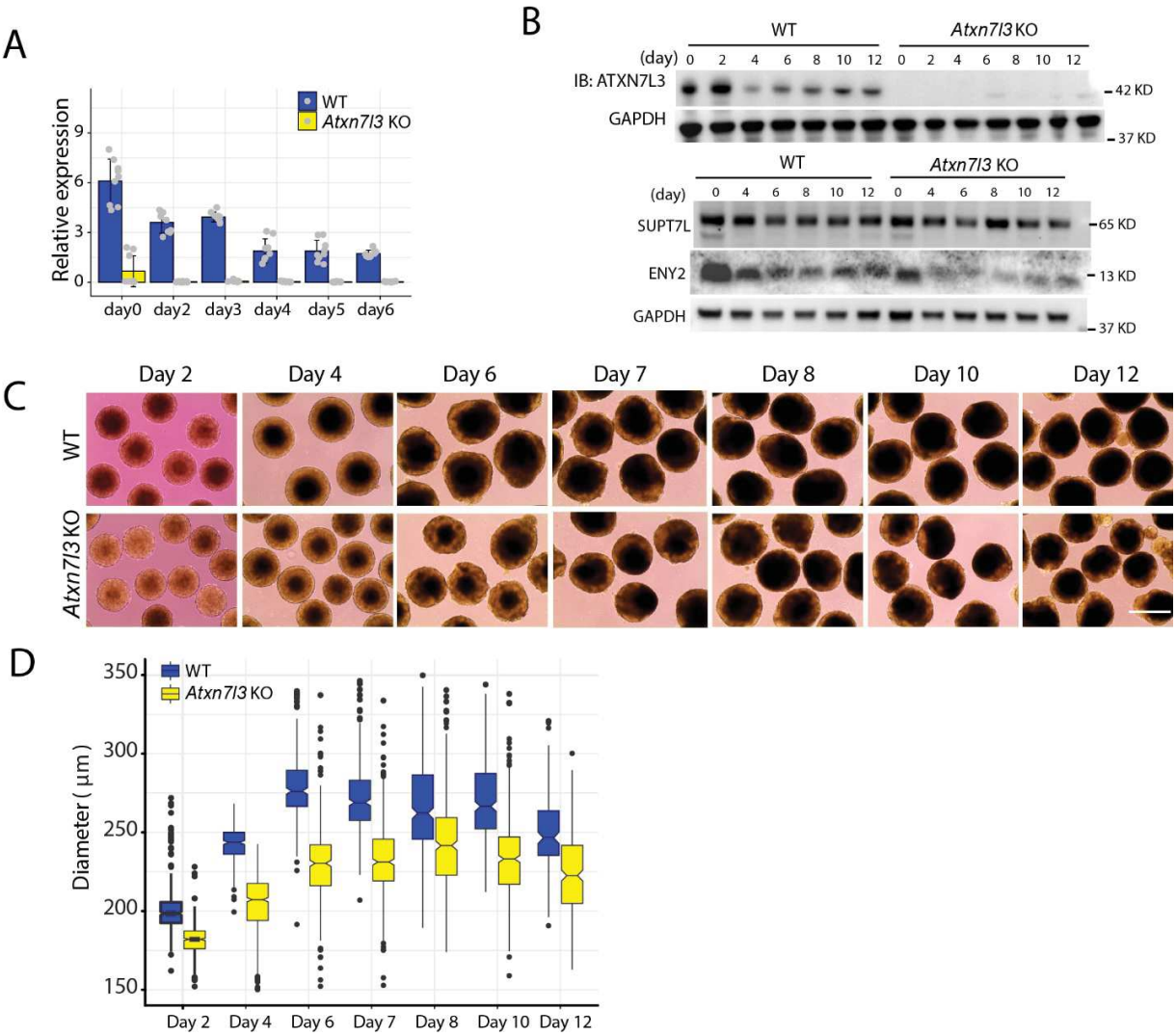


Figure 6-4. The size of *Atxn7l3* KO EBs is smaller compared with WT condition EBs. (A) Western blot analysis of ATXN7L3, SPT7 and ENY2 in whole cell extracts obtained from WT and *Atxn7l3* KO EBs. GAPDH is shown as loading control. (B) RT-qPCR analysis of the expression of *Atxn7l3*. Y axis indicates the mRNA expression relative to the housekeeping gene (*Pgk1*). Error bars represent ±SD from three biological samples with three technical replicates (represented by grey dots), respectively. (C) Morphology of WT and *Atxn7l3* KO EBs cultured in 20% serum medium until 12 days. Scale bar: 500µm. (D) Quantification of the diameters of EBs in *Atxn7l3* KO and WT mESCs.

3.2.1.4 Compromised differentiation potential of *Atxn7l3* null EBs

To understand how *Atxn7l3* affects the embryonic differentiation process, firstly we tested the expression of the pluripotent gene *Pou5f1*. We observed that expression of *Pou5f1* was down-regulated upon induction of differentiation in both WT and *Atxn7l3* KO cells. As the epiblast gives

rise to all three germ layers during gastrulation (Rivera-Perez and Hadjantonakis, 2014), we wonder whether *Atxn7l3* loss affects lineage formation. Therefore, we examined marker genes that expressed in the embryonic epiblast, including *Flk1, Gsc, Foxa2, Mesp1, Brachury, Mixl1, Otx2 and Fgf5* (Kamiya et al., 2011; Kurokawa et al., 2004; Sumi et al., 2013; Yamanaka et al., 2010). Consistent with the results from monolayer differentiation protocol (Figure 6-4), we observed no obvious difference in the expression of *Otx2* and *Fgf5* between WT and the *Atxn7l3* KO at day 4 (Figure 6-5A). However, the expression of mesoderm-specific genes (*Brachury*, *Flk1*, and *Mesp1*) and endoderm-specific gene (*FoxA2*) were lower in the *Atxn7l3* KO EBs than WT EBs at day 4 (Figure 6-5A).

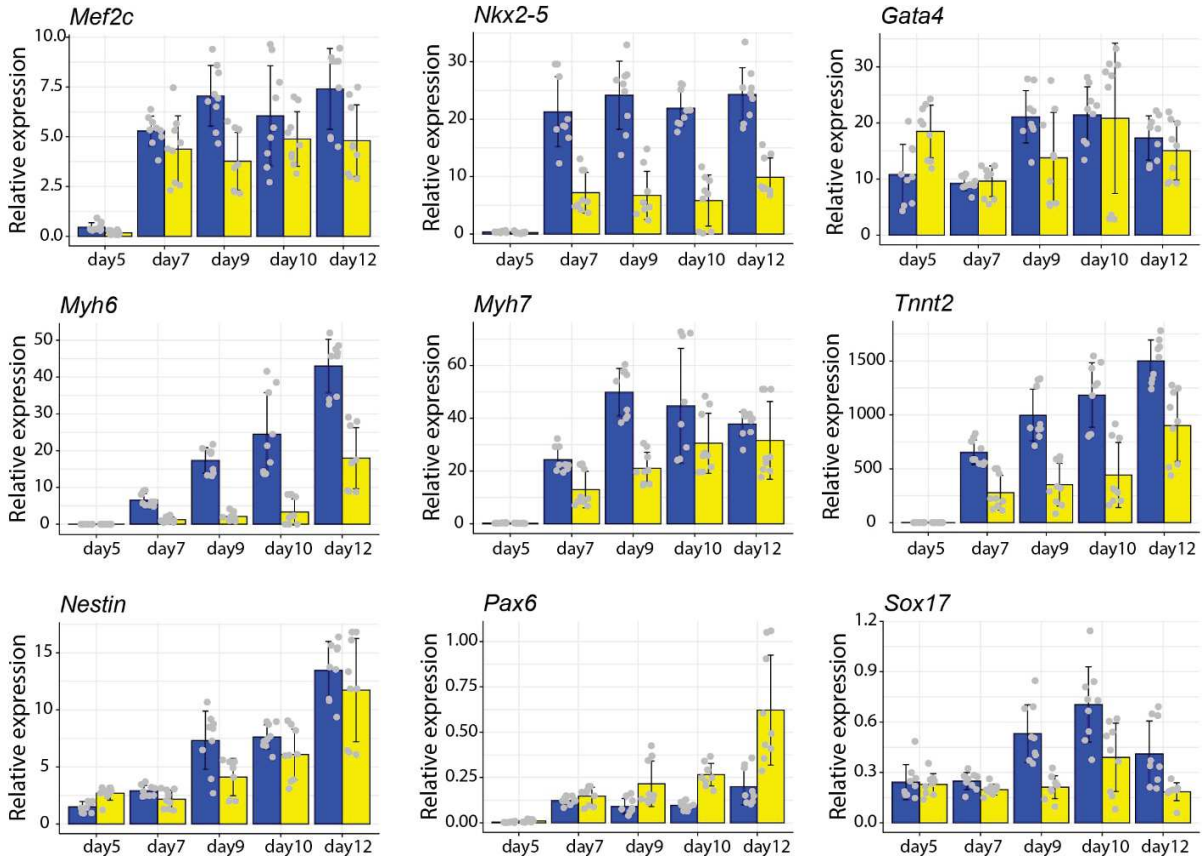
To further identify the effects of *Atxn7l3* deletion on lineage formation, we tested mRNA expression of lineage genes at late EB stage. Specifically, we observed significantly lower expression of mesoderm-specific genes (*Mef2c, Nkx2-5, Myh6, Myh7 and Tnnt2*) *in Atxn7l3* KO cells compared with WT cells (Figure 6-5B), whereas, expression of endoderm-specific genes was unaffected (*Gata4*) or downregulated (*Sox17*) in *Atxn7l3* KO cells (Figure 6-5B). Moreover, expression of ectoderm-specific genes (*Nestin* and *Pax6*) were almost unchanged between WT and *Atxn7l3* KO (Figure 6-5B). These data suggested that the differentiation potential of *Atxn7l3* KO EBs is compromised.

In support of this hypothesis, *Atxn7l3* mutant ESCs displayed abnormal cardiomyocyte differentiation. For example, *Atxn7l3* KO EBs has lower expression of cardiac muscle troponin T (cTnT) protein than WT EBs by immunofluorescence (Figure 6-5C). Besides, cTnT started to express at day 6 and reached the highest level at day 12 in WT EBs. However, the expression of cTnT was dramatically delayed in *Atxn7l3* KO condition, as cTnT was just slightly detected at day 12 in *Atxn7l3* KO EBs (Figure 6-5C).

Notably, *Atxn7l3* did not influence the expression of neuron-specific class III β-tubulin (β-Tubulin III) at day 12 EBs (Figure 6-5E). To further characterize the effects of *Atxn7l3* loss on neural ectoderm formation, we utilized a new protocol to specifically direct mESCs toward defined neuronal lineage (Bibel et al., 2004). We tested the ectoderm-specific protein (PAX6 and NESTIN) at day 8 neuronal precursor cells by immunofluorescence (Figure 6-5F, G) and found that they are

similarly expressed in both WT and *Atxn7l3* KO conditions (Figure 6-5H). Consistently, *Atxn7l3* did not influence the expression of β-Tubulin III at day 10 neurons (Figure 6-5I). To sum up these results suggest that *Atxn7l3* is important for the differentiation of mesoderm but not for neural ectoderm.

B


Mef2c
Nkx2-5
Gata4
Myh6
Myh7
Tnnt2
Nestin
Pax6
Sox17
Relative expression
day5
day7
day9
day10
day12

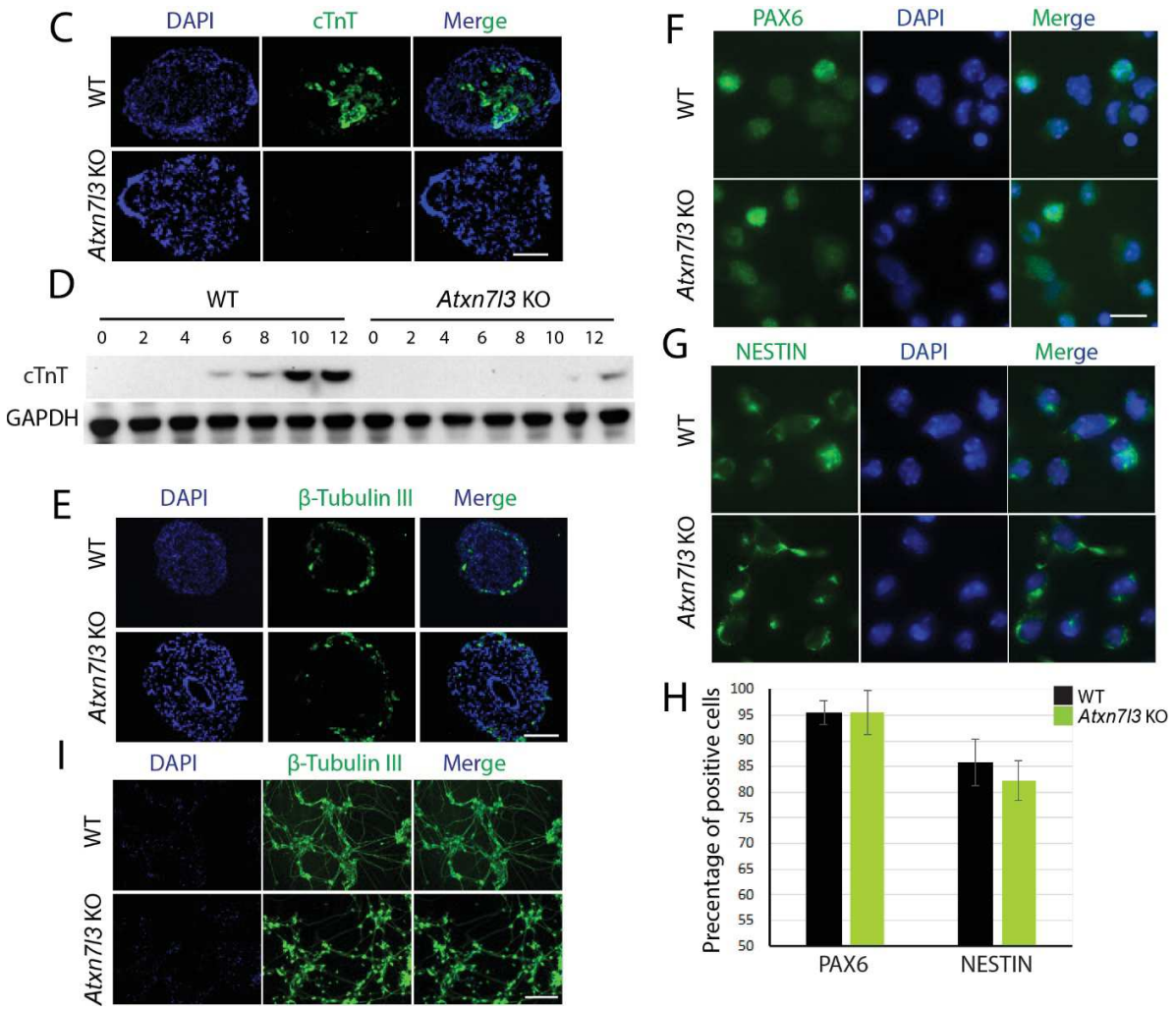


Figure 6-5. *Atxn7l3* affects the differentiation of cardiomyocyte, but has no effects on neural differentiation. (A, B) RT-qPCR analyze the expression of genes during EB differentiation. Y axis indicates the expression of tested genes relative to the housekeeping gene (*Pgk1*). Error bars represent ±SD from three biological samples with three technical replicates (represented by grey dots), respectively. (C) DAPI and immunofluorescence images of WT and *Atxn7l3* KO EB cryosections stained with anti-cTnT. Scale bar: 100 μm. (D) Western blot analysis of cTnT protein in whole cell extracts obtained from WT or *Atxn7l3* KO EBs at indicated time point. GAPDH is shown as a loading control (E) DAPI and IF images of WT and *Atxn7l3* KO EB cryosections stained with β-tubulin III. Scale bar: 100 μm. (F, G) DAPI and IF images of WT and *Atxn7l3* KO neural precursor cells stained with PAX6 (F) and NESTIN (G). Scale bar: 100 μm. (H) Quantify the percentage of PAX6 and NESTIN positive cells in *Atxn7l3* KO and WT conditions. Error bars indicate ±SD based on three biological samples. (I) DAPI and IF images of WT and *Atxn7l3* KO neural cells stained with β-tubulin III. Scale bar: 100 μm.

3.2.1.5 ATXN7L3-dependent DUBm-s deubiquitinate H2Bub1 during EB differentiation

In mammalian cells, depletion of either ENY2 or ATXN7L3 adaptor protein resulted in a non-functional USP22 enzyme, therefore cells fail to remove H2Bub1 (Atanassov et al., 2016; Lang et al., 2011). To better understand the molecular basis underlying the defects caused by *Atxn7l3* loss, we tested the dynamic change of H2Bub1 during EB differentiation. In WT condition, H2Bub1 was gradually increased for about 1.5 times at days 4, then it kept stable until days 12. However, H2Bub1 was huge increased around 3 times all the time in *Atxn7l3* cells compared with WT cells (Figure 6-6A, B). These observations indicate that the consistently increased H2Bub1 may cause the compromised differentiation potential of *Atxn7l3* KO EBs. Additionally, Histone H2Bub1 was reported to play a critical role in regulating autophagy (Chen et al., 2017). To identify whether H2B deubiquitination affects EBs differentiation through affecting autophagy, we detected the autophagic flux by measuring LC3-II turnover. Our results show that H2B deubiquitination has no effect on the change of LC3B I and LC3B II (Figure 6-6C). Therefore, the constantly high H2Bub1 may compromise EB differentiation without affecting autophagy.

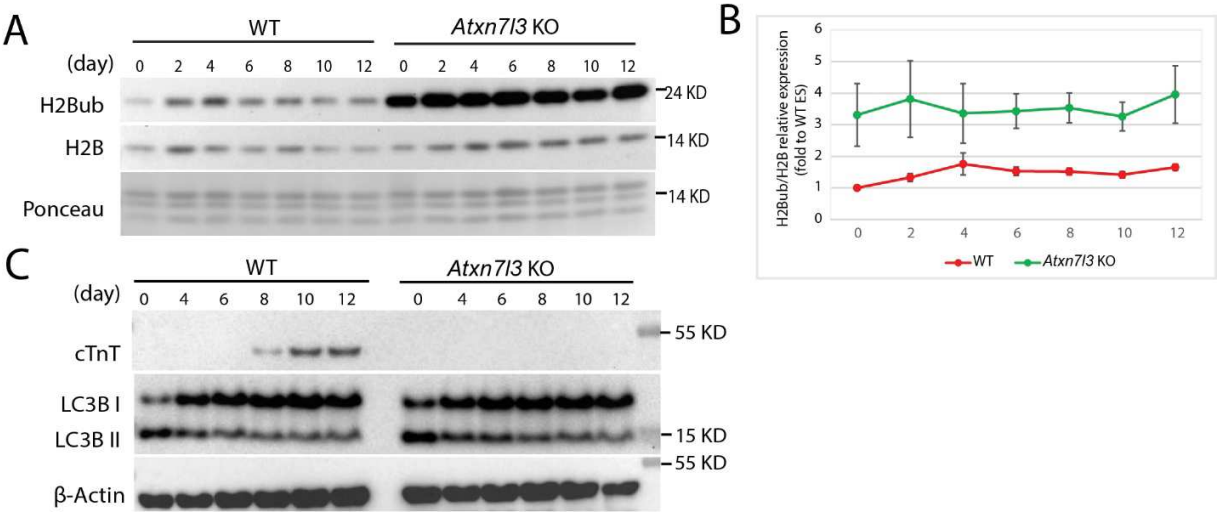


Figure 6-6. H2Bub1 is deubiquitinated by ATXN7L3-dependent DUBm-s. (A) Western blot analysis of H2Bub1 in acidic histone extracts obtained from WT and *Atxn7l3* KO EBs. Histone H2B western blot and ponceau stained membranes are shown as loading controls. (B) Western blot analyses shown in Figure 5A were scanned and analyzed densitometrically with ImageJ and the H2B normalized results are represented for each

genotype. (C) Western blot analysis of cTnT, LC3B I and LC3B II levels in whole cell extracts obtained from WT or *Atxn7l3* KO EBs. β-Actin is shown as loading control.

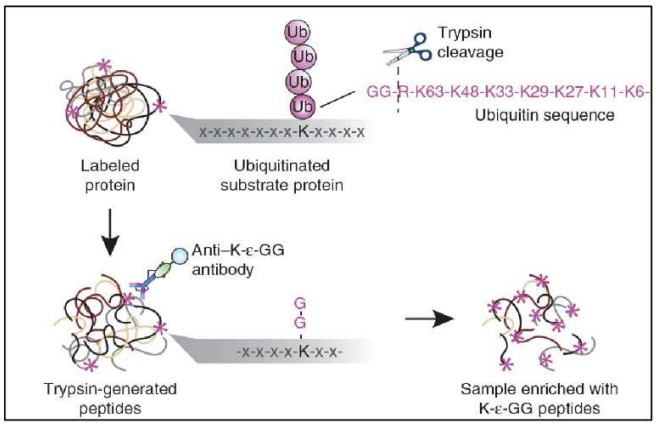
3.2.2 Proteomic screening of the DUBm-targeted proteins

Our RNA sequencing data showed that *Atxn7l3* deletion resulted in deregulation of about two thousand genes in mESCs out of how many expressed. Although H2Bub1 levels significantly increased, the genome-wide occupancy of Pol II was only modestly changed in *Atxn7l3*$^{-/-}$ ESCs. Therefore, H2Bub1 deubiquitination did not seem to regulate directly global RNA polymerase II transcription. It is thus conceivable that embryonic death of the *Atxn7l3*$^{-/-}$ embryo could be a consequence of the activity of the DUBm-s on other proteins that remained to be identified.

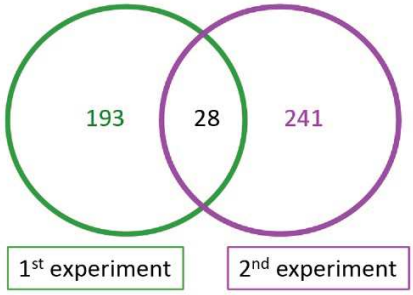
To fully understand the role of DUBm, we carried out a screening for the ATXN7L3-related DUBm-dependent accumulation of ubiquitylated proteins in *Atxn7l3* KO mESCs. This analysis relies on affinity capture of ubiquitylated peptides using an antibody specific for the di-glycine tag that is linked to ubiquitylated lysine residues (Kim et al., 2011) (Figure 6-7A). With the help of IGBMC proteomic facility, enriched ubiquitylated peptides from WT and *Atxn7l3*$^{-/-}$ ESCs were subjected to LC-MS/MS analysis for quantitative Ub profiling. For mass spectrometry experiment, we did two biological samples with three technical replicates respectively.

In one of the biological replicates, we identified 3014 di-glycine containing peptides (ub-peptides). We took the Extracted Ion Chromatogram (XIC) value to indicate the abundance of the peptide. Among these peptides, the abundance of 220 ub-peptides were up-regulated for above 2 times in *Atxn7l3* KO mESCs. Similarly, in the second biological replicate, 2859 ub-peptides were identified and 268 ub-peptides had more than 2 times increased XIC values in *Atxn7l3* KO mESCs compared with WT mESC. Interestingly, 28 ub-peptides were commonly up-regulated in the two biological samples (Figure 6-7B), including two H2BubK120 peptides that were dramatically increased about 5 times in *Atxn7l3* KO mESCs (Figure 6-7C). These results suggest that in addition to H2Bub1, there may be additional targets of ATXN7L3-dependent DUBm-s. Further *in vivo* and *in vitro* experiments will be needed to validate and functionally characterize the deubiquitinylation of these potential target proteins ATXN7L3-dependent DUBm-s.

A



B



C

Gene	Protein	Modification	Peptide sequence	XIC-Ratio (KO/WT)	
				Replicate1	Replicate2
H2bc12	Histone H2B type 1-K	[K120]	[K].AVTKYTSAK.[-]	5.170871	5.170871
Hist1h2bb	Histone H2B type 1-B	[K120]	[K].AVTKYTSSK.[-]	5.223704	17.23939
H2az1	Histone H2A.Z	[K15]	[K].TKAVSR.[S]	3.225117	3.781668
Hist2h2aa1	Histone H2A.2	[K119]]	[K].KTESHHK.[A]	2.821705	3.641841
Ass1	Argininosuccinate synthase	[K121]	[K].YVSHGATGKGNDQVR.[F]	4.621175	2.342173
Gfpt1	GFPT1	[K48]	[R].GYDSAGVGLDGGNDKDWEANACK.[I]	7.567482	2.491019
Gja1	Gap junction alpha-1 protein	[K303]	[R].NYNKQASEQNWANYSAEQNR.[M]	2.084184	3.283049
Glrx	Glutaredoxin-1	[K28]	[R].KTQEILSQLPFK.[Q]	2.80753	2.120307
Glul	Glutamine synthetase	[K95]	[K].DPNKLVLCEVFK.[Y]	3.220777	2.032001
Gprc5a	Retinoic acid-induced protein 3	[K287]	[R].SPTDYPVEDAFCKPQLMK.[Q]	4.880742	2.602568
Impact	IMPACT	[K291]	[R].NILVEKNFTNTPDESTK.[N]	4.707008	3.808589
Impact	IMPACT	[K228]	[R].IFCEDKQTFLQDCEDDGETAAGGR.[L]	7.658704	3.664426
Itm2b	Integral membrane protein 2B	[K13]	[K].VTFNSALAQKEAK.[K]	3.274426	2.209967
Itm2c	Integral membrane protein 2C	[K14]	[K].ISFQPAVAGIKADK.[A]	3.934877	2.254886
Morc3	MORC family CW-type zinc finger prote	[K809]	[K].SECSQASCTESKSEVDEMAVQLDDVFR.[Q]	2.110574	7.276584
Morf4l2	Mortality factor 4-like protein 2	[K175]]	[K].SQGNVDNKEYAVNEVVGGIK.[E]	3.556765	4.328213
Ndfip2	NEDD4 family-interacting protein 2	[K168]	[K].AKAAALAAAAADAPQR.[N]	2.100936	2.409053
Neu3	Sialidase-3	[K154]]	[K].DLTEEVIGSEVKR.[W]	3.439615	4.339878
Peg10	Retrotransposon-derived protein PEG1	[K813]	[R].NVKDGLMTPTVAPNGAQVLQVK.[R]	20.64694	2.019716
Polr2a	DNA-directed RNA polymerase II subur	[K1268]	[R].IMNSDENKMQEEEEVVDK.[M]	10	2.020821
Rab7a	Ras-related protein Rab-7a	[K175]	[R].NALKQETEVELYNEFPEPIK.[L]	4.381428	2.983415
Set	SET	[K166]	[K].EFHLNESGDPSSKSTEIK.[W]	2.094103	2.158809
Stt3a	STT3A	[K444]	[K].KQQDSTYPIK.[N]	2.514924	2.174666
Stx3	Syntaxin-3	[K100)]	[K].SMEKHIEEDEVR.[S]	3.355341	5.401167
Tfrc	Transferrin receptor protein 1	[K39]	[R].QVDGDNSHVEMKLAADEEENADNNMK.[A]	2.009193	2.474409
Tmem59	Transmembrane protein 59	K15]	[R].SQTEEHEEAGPLPTK.[V]	8.171958	4.261056
Tmem59	Transmembrane protein 59	[K315]	[R].SQTEEHEEAGPLPTKVNLAHSEI.[-]	2.734929	2.964126
Uhrf1	E3 ubiquitin-protein ligase UHRF1	[K196]	[K].YDDYPEHGVDIVKAK.[N]	7.380629	15.49221

Figure 6-7. Proteomic analysis (UbiScan) identifies potential candidate substrates of ATXN7L3-dependent DUBms. (A) Strategy for proteome-wide screen to find *Atxn7l3*-dependent deubiquitylation (Udeshi et al., 2013). (B) Venn diagram showed the number of ub-peptides, of which the XIC value is up-regulated for more than 2 times in *Atxn7l3* KO mESCs compared with WT mESC. (C) This table showed the information of the 28 ub-peptides that were commonly up-regulated in the two biological samples.

3.2.3 Analyze the role of the DUBm in DNA damage process

In eukaryotic cells, unrepaired DNA lesions are barriers for elongating Pol II. Arrested polymerase not only blocks the passage of subsequent RNA polymerase, but it also prevents the exposure of damaged site to DNA repair factors (Lavigne et al., 2017; Nakazawa et al., 2020). It has been reported that UV damage-induced Pol II stalling stimulated H2Bub1 deubiquitylation and that H2Bub1 in *ubp8Δubp10Δ* mutant strains increased the UV-induced Pol II degradation (Mao et al., 2014). These observations suggest that cells respond to Pol II arrest by deubiquitylating H2Bub1 to coordinate DNA repair and Pol II degradation. However, the mechanism for this is still not fully understand.

Interestingly, in the above identified 28 common potential ATXN7L3-related DUBm targets we have identified the lysine 1268 of the largest subunit (RPB1) of Pol II (see Figure 6-8A). Thus, to test whether the ATXN7L3-related DUBm-s regulate the ubiquitination state of Pol II, we enriched ubiquitinated proteins from ESC extracts prepared either from WT cells or from KO cells. To enrich ubiquinited proteins we carried out a GST-DSK2 pull-down assay. GST-DSK2 can bind ubiquitinated proteins with its ubiquitin-associated (UBA) domain originating from the ubiquitin-binding protein Dsk2 (Tufegdzic Vidakovic et al., 2019). In response to UV irradiation, the RPB1-pSer2 (S2-phosphorylated) was changed to polyubiquitinated in both WT and *Atxn7l3* KO ESCs. Compared to WT, the ubiquitinated RPB1-pSer2 was lower in *Axtn7l3* KO ESCs, whereas the ubiquitinated total RPB1 was almost unchanged (Figure 6-8B). These preliminary results may suggest that Atxn7l3 could specifically facilitate/regulate the ubiquitination of elongating Pol II.

To elaborate whether the reduced RPB1-pSer2 phosphorylation affects DNA damage repair process, we tested the efficiency of γH2AX of deposition, which represents a mean of visualizing individual DSBs (Rogakou et al., 1999; Rogakou et al., 1998). Interestingly, we found that γH2AX recruitment was delayed in *Atxn7l3* KO mESCs compared with WT (Figure 6-8C) suggesting that polyubiquitinated RPB1-pSer2 facilitated DNA damage repair might be linked to DNA repair. Nevertheless, many further experiments will be required to better understand this potential regulatory link.

A

Annotated Sequence	Modifications in Master Proteins	# PSMs WT	# PSMs KO	Ratio(KO/WT)
[R].IMNSDENKMQEEEEVVDK.[M]	RPB1 1xGG [K1268]	1	5	5
[K].NICEGGEEMDNKFGVEQPEGDEDLTK.[E]	RPB1 1xGG [K163]	3	7	2.333333333
[K].FGVEQPEGDEDLTKEK.[G]	RPB1 1xGG [K177]	6	8	1.333333333
[R].VLSEKDVDPVR.[T]	RPB1 1xGG [K1350]	3	2	0.666666667
[K].TGSSAQKSLSEYNNFK.[S]	RPB1 1xGG [K758)]	4	2	0.5
[R].VIFPTGDSKVVLPCNLLR.[M]	RPB1 1xGG [K976]	4	N/A	0

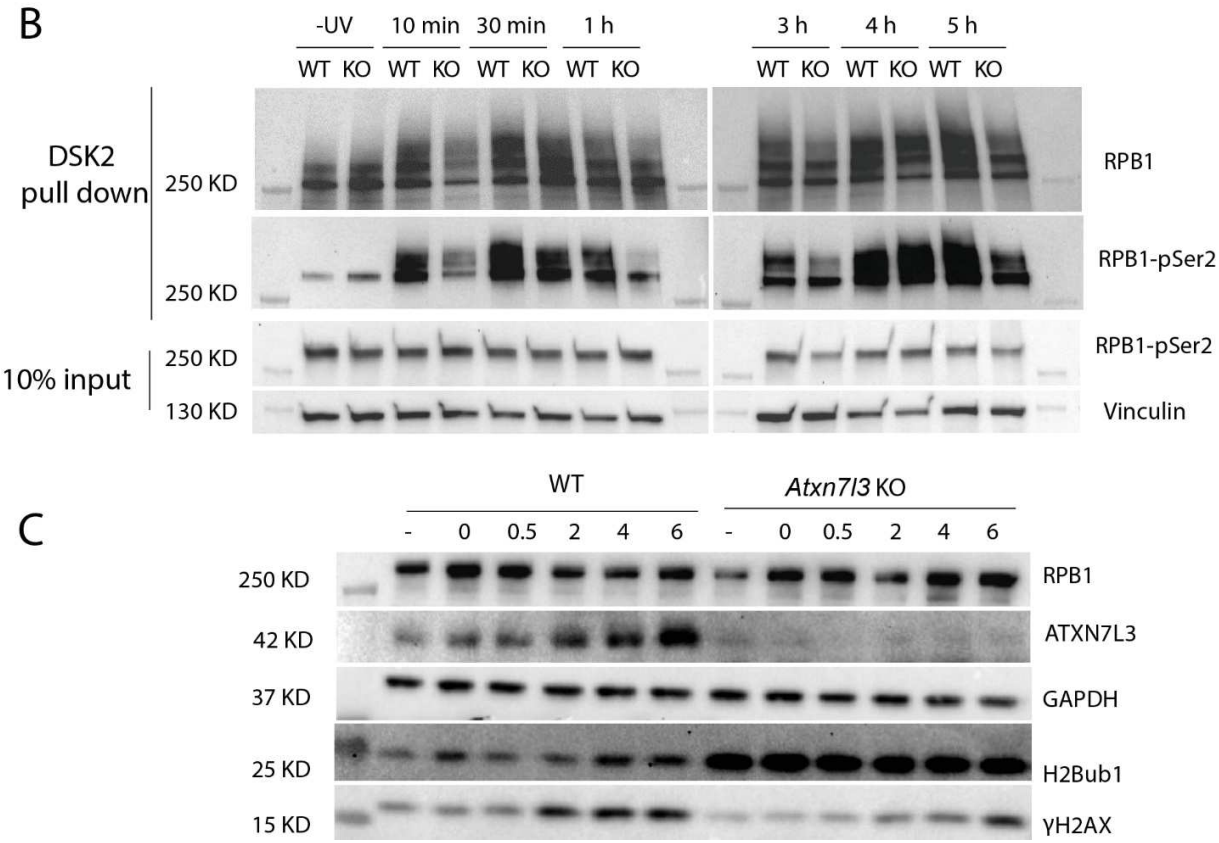
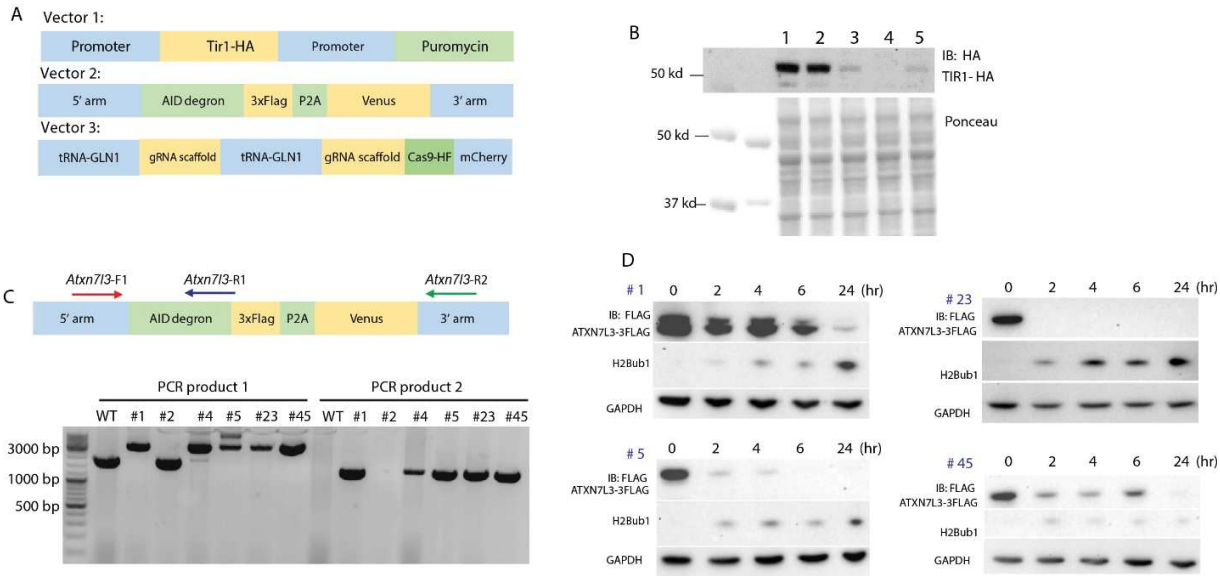


Figure 6-8. *Atxn7l3* affects DNA damage repair. (A) RPB1 ubiquitinated sites identified in UbiScan. (B) Dsk2 pull down following western blot to analysis RPB1-pSer2 and RPB1 in WT and *Atxn7l3* deletion mESC, before and after UV irradiation (20 J/m^2) treatment. Vinculin western blot is shown as loading control. (C) Western blot analysis of RPB1, ATXN7L3, H2Bub1 and γH2AX in whole cell extracts obtained from WT and *Atxn7l3* KO ESCs. GAPDH western blot is shown as loading control.

3.2.4 Other results

3.2.4.1 Strategy to generate ATXN7L3 conditional deletion mESC

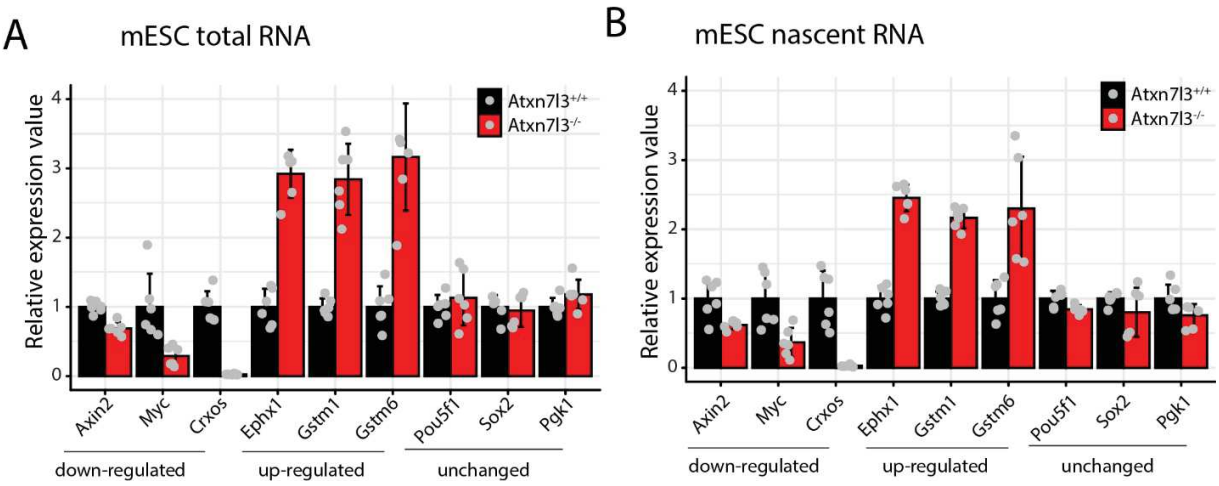
To decide which stage was affected by *Atxn7l3* deletion during gastrulation process, we generated conditional deletion *Atxn7l3*-AID-3xFlag mESC lines by CRISP-Cas9 technique. It has been reported that auxin promoted the interaction between the AID degron tagged protein and the E3 ubiquitin ligase SCF (TIR1) in plant, thus the former proteins are rapidly polyubiquitylated for degradation by the proteasome (Nishimura and Kanemaki, 2014). Based on this mechanism, firstly we transfected Vector 1 with Tir1 genes into mESCs to obtain TIR1 overexpression cell line (clone 1 and clone2) (Sup. 1A, B). Then Vector 2 and 3 was transfected into TIR1-HA transfected cell line (clone 2), and cultured for 2 days before FACS selection. After PCR validation, we got five clones, including #1, #4, #5, #23, #45, are *Atxn7l3*-AID-3xFlag positive (Sup. 1C). Upon auxin treatment, ATXN7L3 was degraded and H2Bub1 was increased within 24 hours (Sup. 1D).



Supplementary figure 1. Strategy to generate ATXN7L3 conditional deletion mESC. (A) Model showing the structures of the three vectors used for generating *Tir1*, *Atxn7l3*-AID ESC lines. (B) Western blot analysis of TIR1-HA protein. (C) PCR validation for *Atxn7l3*-AID-3xFlag clones. *Atxn7l3*-F1 and *Atxn7l3*-R1 primers were used to amplify the DNA containing the 5' arm of *Atxn7l3* and the whole AID-3xFlag DNA sequence; *Atxn7l3*-F1 and *Atxn7l3*-R2 primers are used to amplify the DNA between 5' arm and part of the AID-3xFlag DNA sequence. (C) Western blot analysis of ATXN7L3-3xFlag and H2Bub1 in whole cell extracts obtained from AID tagged clones untreated or treated with auxin for 24 hours. GAPDH western blot is shown as loading control.

3.2.4.2 ATXN7L3 has no global effect on nascent RNA transcription

Eukaryotes can modulate RNA levels by altering RNA synthesis, processing or decay. Therefore, the same RNA steady-state levels can be caused via multiple pathways (Duffy et al., 2019). To test whether *Atxn7l3* influences the global synthesis of nascent RNA, we used 4-Thiouridine (s^4U) to metabolic labeling nascent RNA (Duffy et al., 2019), followed by RT-qPCT. We found that the nascent RNA showed down-regulated, up-regulated or unchanged genes, which is similar to steady state total RNA. These results suggested that *Atxn7l3* has no effect on both the global transcription of nascent RNA and the processing or decay of these nascent RNA into mature RNA.



Supplementary figure 2. (A) RT-qPCR analyze the total RNA levels of genes that are down-regulated, up-regulated or unchanged according to RNA sequencing results in mESCs. (B) RT-qPCR analyze the nascent RNA levels of the same genes in panel A. Y axis (in panel A and panel B) indicate the expression of tested genes relative to the spiking in genes (*Gapdh*) from *drosophila*. Error bars represent ±SD from two biological samples with three technical replicates (represented by grey dots), respectively.

4. Discussion

4.1 Roles of *Atxn7l3* in self-renewal and differentiation of mESCs

Firstly, we found that ATXN7L3 facilitated the self-renewal of mESC by enhancing the transition from G1 to S phase. Consistently, USP22 has been described to be critical for progression through the G1 phase of the cell cycle by deubiquitylating G1 phase cyclin CCND1 (Gennaro et al., 2018). In addition, the DUBm of SAGA has also been implicated in DNA repair (Evangelista et al., 2018; Mao et al., 2014). Thus, DNA repair defects may also contribute to the growth defects of *Atxn7l3* KO mESCs.

Secondly, we found that *Atxn7l3* deletion had no obvious effect on the pluripotency of mESCs, as the normal expression of pluripotency-related genes upon *Atxn7l3* deletion. However, further studies found that *Atxn7l3* affected the processes of mESCs differentiation. These observations raise the possibility that H2Bub1 deubiquitination mainly regulates inducible but not constitutive transcription. Indeed, H2Bub1 has been associated with the regulation of inducible genes, such as HOX genes that involved in cell differentiation and relatively long genes that induced by retinoic acid (Fuchs et al., 2012; Karpiuk et al., 2012; Materne et al., 2016; Zhu et al., 2005). Therefore, although there is a general correlation between H2Bub1 and gene expression, the dynamic of H2Bub1 might only fine-turning the transcriptional activity in mESCs (Fuchs et al., 2014; Minsky et al., 2008).

Thirdly, our results indicated that *Atxn7l3* affected the expression gastrulation-related genes. During gastrulation process, epiblast cells migrate out the primitive streak to form the mesoderm and the endoderm. In contrast, cells that do not pass through the primitive streak give rise to the surface ectoderm and the neural tissues (Murry and Keller, 2008; Tam and Loebel, 2007). Interestingly, we found that ATXN7L3 was required for mesodermal derived cardiomyocyte differentiation, but not for ectoderm neural precursor development *in vitro*. In line with this, our *in vivo* result showed that *Atxn7l3* KO caused a developmental delay as early as E7.5 when the embryos went through gastrulation. These observations suggest that ATXN7L3 and its related DUBm-s may promote the gastrulation process. Given that gastrulation is controlled by the coordinated activation and regional inhibition of the Wnt, Nodal, and BMP-signaling pathways (Conlon et al., 1994; Gadue et al., 2005; Hogan, 1996; Schier, 2003; Yamaguchi, 2001), ATXN7L3 and its related DUBm-s may affect part of these pathways by modulating the

recruitment of relevant TFs. Yet further *in vivo* and *in vitro* experiments need to validate these hypotheses.

4.2 Roles of H2Bub1 deubiquitination in transcriptional regulation

4.2.1 H2Bub1 deubiquitination does not regulate global transcription elongation.

Histone H2Bub1 has been linked to increased transcription, transcription elongation and DNA replication (Laribee et al., 2007). In yeast, histone H2BK123ub1 is deposited by the E3 ubiquitin ligase Bre1 (an orthologue of RNF20/RNF40 proteins in human cells), together with the E2 ubiquitin-conjugating enzyme Rad6 and the E1 ubiquitin-activating enzyme Uba1 (Hwang et al., 2003; Kim et al., 2009). Generally, the transcription elongation factor PAF complex recruits Rad6 and the elongating Pol II, which stimulates the deposition of H2Bub1 at actively transcribed regions (Wood et al., 2003a; Xiao et al., 2005). Genome wide approaches revealed a nonrandom distribution of H2Bub1 within active gene bodies. Moreover, in mammalian cells it seems that H2Bub1 is significantly reduced following the first internal exon (Huff et al., 2010). Thus, it has been suggested that H2B ubiquitylation was coupled with the elongation rate of RNA polymerase II (Fuchs et al., 2014; Minsky et al., 2008). Compared with other histone modifications, such as H3K36me3 and H3K79me2, H2Bub1 is highly dynamic during the transcription process (Fuchs et al., 2014). It has been reported that H2Bub1 was quickly erased by the DUBm within 10 min (Bonnet et al., 2014). However, contrary to H2B ubiquitylation, it is much less well understood whether H2Bub1 deubiquitylation would be a process significantly impacting transcription in mammalian cells.

Previously, by using an *ATXN7L3* knock-down strategy in human HeLa cells, our laboratory showed that the ATXN7L3-related DUB activities are directed toward the transcribed region of almost all expressed genes, but are only poorly correlated with gene expression (Bonnet et al., 2014). Our present results indicate that impairment of H2Bub1 deubiquitylation does not directly impact transcription. For example, we found a massive H2Bub1 retention at almost every expressed gene in both *Atxn7l3*$^{-/-}$ mESCs and MEFs. Nevertheless, in both cellular systems, the Pol II occupancy was only slightly impacted. Therefore, the lack of correlation between global

H2Bub1 increase and consequent genome-wide inhibition of global transcription suggests that the deubiquitylation of H2Bub1 does not directly regulate Pol II transcription.

4.2.2 Potential roles of H2Bub1 deubiquitination at promoter and enhancer regions.

In addition to regulate transcriptional elongation, H2Bub1 was reported to have a repressive function at the promoter and enhancer regions. For example, in yeast, H2Bub1 inhibited the occupancy of Pol II at normally quiescent promoters by assisting nucleosome reassembly (Batta et al., 2011). Moreover, H2Bub1 mediated nucleosome reassembly was suggested to suppress cryptic transcriptional initiation at certain genes by blocking access of the transcription machinery at promoters (Chandrasekharan et al., 2009; Fleming et al., 2008). Consistently, biochemical analyses found that the nucleosome stability was enhanced when H2Bub1 levels increase (Chandrasekharan et al., 2009). This feature of H2Bub1 was also suggested to affect enhancer activity. As H2Bub1 inhibited the activity of inducible enhancer by impairing the chromatin access to INO80 which promoted histone H2A.Z eviction (Segala et al., 2016). Therefore, whether the observed embryo and cellular phenotypes in the *Atxn7l3*$^{-/-}$ embryos can be directly linked to increased H2Bub1 levels in specific enhancer or promoter regions having special chromatin architecture will need to be further investigated in the future.

4.3 Multiple complexes regulate the DUBm of SAGA

The DUBm of SAGA is composed of the ubiquitin-specific protease 22 (USP22) and three adaptor proteins (ATXN7, ATXN7L3 and ENY2). The N-terminal ZnF domain of ATXN7L3 can dock these small DUB complexes to the H2A/H2B acidic patch (Morgan et al., 2016). Interestingly, the adaptor protein, ENY2, was also part of TREX-2 complex which played an essential role in mRNA export (Fischer et al., 2004; Fischer et al., 2002; Gonzalez-Aguilera et al., 2008; Rodriguez-Navarro et al., 2004; Wilmes et al., 2008). We found that *Atxn7l3* deletion affected the stability of its partner protein ENY2. This observation suggests that the DUBm of SAGA might influence TREX-2-mediated mRNA export process by modulating the stability of ENY2. Besides, in humans, ENY2 also interacted with ATXN7L3B in the cytoplasm, which limited the ENY2-ATXN7L3 interaction in the nucleus (Li et al., 2016a). Moreover, ENY2 and ATXN7L3 also

comprised two independent DUB module variants, containing either USP27X or USP51, are catalytically active on mono-ubiquitinated H2B (Atanassov et al., 2016). We found that *Atxn7l3* loss of function results in a more severe phenotype than that of *Usp22*$^{-/-}$ embryos. Consistently, the inactivation of ATXN7L3 resulted in increased H2Bub1 levels, whereas USP22 almost has no effect on H2Bub1 levels. Therefore, the activity of USP27X- and/or USP51-containing DUBms might compensate the function of USP22. However, it is still unknown how SAGA and these two DUBm variants contribute to global or specific genomic locations. Together, the mandatory incorporation of DUBm subunit within SAGA, TREX-2, ATXN7L3B or the DUBm variants might regulate the activity of the DUBm by sequestering the limited ENY2 and/or ATXN7L3 subunits. However, further experiment was needed to validate these hypotheses.

5. Perspectives

5.1 Potential role of H2Bub1 deubiquitination in nucleosome dynamics

The nucleosome is quite dynamic, which undergoes assembly and disassembly cycles during transcription process. *In vitro* experiments showed that nucleosomal assembly was initiated by the occupancy of H3–H4 tetramer, and then two H2A–H2B dimers wrapped the remaining DNA (Arents et al., 1991; Smith and Stillman, 1991). The disassembly of nucleosomes is thought to occur through a reversal process. During nucleosome disassembly, firstly, the interface between the H2A−H2B dimers and the $(H3-H4)_2$ tetramer was opened up, followed by the removal of either one or both of the H2A−H2B dimers (Gansen et al., 2007; Gansen et al., 2009; Li et al., 2005). FACT (facilitates chromatin transcription) is an essential histone chaperone that plays an important role in regulating chromatin structure (Bondarenko et al., 2015). Several theories and models exist to explain how Pol II deals with nucleosomes during the transcription process, i.e. bypassing nucleosomes (Kassabov et al., 2003; Owen-Hughes et al., 1996; Xu et al., 2020), partially disassembling nucleosomes (Kulaeva et al., 2007) or completely evicting nucleosomes from the transcribed DNA template (Dion et al., 2007; Jamai et al., 2007; Kimura and Cook, 2001). According to the partial disassembly model, to allow the passage of Pol II, FACT would displace the H2A/H2B dimer from the core nucleosomes (Belotserkovskaya et al., 2003; Kireeva et al., 2002). Moreover, FACT was also shown to reassemble the nucleosome in the wake of elongating Pol II (Belotserkovskaya et al., 2003; Stevens et al., 2011). Therefore, FACT may also be involved in maintaining nucleosome integrity after Pol II passage.

H2Bub1 has been shown to facilitate displacement of H2A-H2B dimer by interacting with FACT (Pavri et al., 2006). In turn, FACT also promotes H2Bub1 deubiquitination via stimulating the enzyme activity of Ubp10 in yeast (Nune et al., 2019). However, in some chromatin contexts, H2Bub1 promote nucleosome reassembly after Pol II passage via stabilizing the association of FACT with chromatin in *S. cerevisiae* (Batta et al., 2011; Fleming et al., 2008). Moreover, FACT and H2Bub1 globally repress antisense transcripts near the 5’ end of genes and inside gene bodies, respectively (Murawska et al., 2020). These observations revealed unexpected interplay between H2Bub1 and FACT to regulate nucleosome dynamics. Given, H2Bub1 is highly dynamic during transcription (Fuchs et al., 2014), timely deubiquitinatilation of H2Bub1 might also participate in the nucleosome remodelling process, of which the exact molecular mechanism need future work.

Secondly, H2Bub1 might affect the access to the 'acidic patch' domain within the nucleosomes. This nucleosome surface possesses a cluster of eight acidic residues to form a negatively charged 'acidic patch' domain (Kalashnikova et al., 2013). This 'acidic patch' domain was reported to act as an interface for many nucleosome binding proteins (McGinty and Tan, 2016), including FACT, PRC1, the SAGA DUB module, ATAC complex, and remodelers of CHD and SWI/SNF family (Dann et al., 2017; Hodges et al., 2017; McGinty et al., 2014; Morgan et al., 2016; Skrajna et al., 2020). Interestingly, H2Bub1 is located adjacent to 'acidic patch' domain. The change in the position of ubiquitin also has the potential to indirectly affect the way in which other factors interact with ubiquitinylated nucleosomes. For example, ubiquitin is positioned on the wrapped side of the nucleosome to occlude the access to the acidic patch. Whereas on the unwrapped side of the nucleosome, ubiquitin positions away from the acidic patch, which permits other factors accessing to the acidic patch. In this way, H2Bub1 may provide means for regulating access to the acidic patch (Sundaramoorthy et al., 2018). Thus, H2Bub1 deubiquitination might indirectly regulate the affinity of acidic patch associated proteins.

Thirdly, the DUBm of SAGA might affect the condensation of chromatin by coordinating with the potential state of H2A.Z. H2A.Z is enriched at the nucleosome-depleted region of active transcriptional start sites (TSS) (Nekrasov et al., 2012). Therefore, H2A.Z is suggested to be necessary for the binding of the transcriptional machinery by facilitating the establishment of NDR. On the other hand, the capacity of H2A.Z to regulate chromatin dynamic is also dependent on H2A.Z posttranscriptional modifications. Generally, the acetylation of H2A.Z destabilizes the nucleosome and in turn the NDR becomes more competent to recruit the transcriptional machinery (Bruce et al., 2005; Ishibashi et al., 2009). In contrast to acetylation modification, the ubiquitination of H2A.Z is associated with the transcriptionally silent heterochromatin (Ku et al., 2012; Sarcinella et al., 2007). Interestingly, our proteomic analysis (UbiScan) results showed that the abundance of H2A.ZK15ub was increased for about 3 times in *Atxn7l3* KO mESCs, suggesting the DUBm of SAGA might regulate nucleosome dynamic by deubiquitinating H2A.Z. Moreover, biochemical study reveals that the deposition of H2Bub1 is highly sensitive to H2A.Z and H2A modifications, which might contribute to the spatial organization of H2Bub1 on gene bodies (Wojcik et al., 2018). Therefore, the relationship between the SAGA DUBm and H2A.Z will need to be further investigated.

5.2 Potential role of H2Bub1 deubiquitination in histone crosstalk

In addition to influence nucleosome dynamics, the monoubiqutylation of histone H2B also facilitates di- and tri-methylation of H3K4 and H3K79 through the recruitment of relevant enzymes (Lee et al., 2007). Each of these histone modifications has been widely linked to actively transcribed genes by direct recruitment of various chromatin-modifying factors (Ruthenburg et al., 2007). Therefore, it has been suggested that H2Bub1 promotes efficient transcription elongation by recruiting transcriptional elongation factors and a "crosstalk" with other histone modifications. Unexpectedly, we found that the global H3K4me3 levels were unaffected, in spite of the fact that the H2Bub1 levels were increased in *Atxn7l3*$^{-/-}$ embryos. This result suggests that the function of H2B ubiquitination and deubiquitination might not be reversible processes in regard to the histone crosstalk process .Interestingly, our GSEA analysis showed that most of the down-regulated genes in *Atxn7l3*$^{-/-}$ MEF cells contain a high-CpG-density promoter (HCP) bearing H3K27me3 modification. Further investigation would be needed to clarify how H2Bub1 deubiquitinating affects H3K27me3 regulated genes during embryonic development.

5.3 Whether the SAGA DUBm functions corporately with its HATm?

The core structural module of SAGA can deliver TBP to gene promoters and regulates global Pol II transcription in yeast (Baptista et al., 2017; Papai et al., 2020; Warfield et al., 2017). Whereas the other modules of SAGA only regulate a subset of genes specifically in response to environment cues like DNA damage or developmental signals(Lang et al., 2011; Nagy et al., 2009; Sellam et al., 2009). The mechanism for this is still unknown. The DUBm of SAGA is responsible for the removal of mono-ubiquitin from histone H2B (Bonnet et al., 2014). The HATm of SAGA is responsible for acetylating histone H3 (Bonnet et al., 2014; Hassan et al., 2002). Besides, both the DUB and HAT modules of SAGA were suggested to function without the whole SAGA complex (Atanassov et al., 2016; Nagy et al., 2009; Nagy and Tora, 2007). Recently, crystal structure study suggested that the nucleosome binding of the SAGA complex can displace the HATm and DUB modules from the core module in yeast (Wang et al., 2020a). In this case, these two catalytic modules can move around or downstream of the TSS, meanwhile the core module recruits TBP at the promoter (Wang et al., 2020a). However, due to the highly dynamic interaction between the SAGA complex and the chromatin, the binding sites of the SAGA complex on chromatin are still

unclear in mammalian cells. Thereby whether the occupancy of the DUBm (or HATm) on chromatin needs the SAGA complex and whether the DUBm and HATm cooperatively or separately regulate gene expression through affecting histone modifications are still unknown.

5.4 Are there potential non-histone targets of the DUBm?

The enzymatic activity of USP22, USP27X and USP51 are dependent on the adaptor protein ATXN7L3 in mammals (Atanassov et al., 2016). In addition to H2Bub1, a multitude of substrates have been identified as targets of these ubiquitin proteases. For example, USP22 can deubiquitinate H2Aub1, TRF1, CCNB1, CCND1 and FBP1 (Atanassov and Dent, 2011; Atanassov et al., 2009; Gennaro et al., 2018; Lang et al., 2011; Lin et al., 2015; Schones et al., 2008). USP27X stabilizes Snai1, BH3-only protein Bim and Hes1 (Kobayashi et al., 2015; Lambies et al., 2019; Weber et al., 2016). USP51 promotes deubiquitination of ZEB1 and H2AK13, 15ub (Zhang et al., 2020; Zhou et al., 2017). Thereby, the phenotype of embryonic development in *Atxn7l3*$^{-/-}$ embryos might be caused by the failure of deubiquitinating these proteins or some other unknown novel target proteins.

6. Conclusion

My Ph.D. research focused on mechanisms of epigenetics in transcriptional regulation, with a special interest in the role of H2Bub1 deubiquitination in mouse embryonic development and mESC differentiation. We have in place CRISPR-mediated mutagenesis and genome-wide molecular approaches (RNAseq, ChIPseq and Proteomic) to address these questions. The deubiquitinase module (DUBm) of SAGA contains ubiquitin-specific protease 22 (USP22) and three adaptor proteins, ATXN7, ATXN7L3 and ENY2, which are responsible for the removal of mono-ubiquitin from histone H2B. We found that *Atxn7l3*$^{-/-}$ embryos were developmental delayed as early as E8.5 and died around E12.5. To get better insight into ATXN7L3, we carried out *in vitro* mESC differentiation assays. Surprisingly, we found that ATXN7L3 promoted the differentiation of cardiomyocyte cells, but not ectoderm neural precursor. Thereby, ATXN7L3 might function in a tissue-specific manner. To understand the molecular mechanisms underlying these phenotypes, we performed transcriptomic and ChIP-seq analyses from *Atxn7l3*$^{-/-}$ mESC. Unexpectedly, although H2Bub1 levels significantly increased in the gene body of every expressed gene, the genome-wide occupancy of Pol II was only modestly changed in *Atxn7l3*$^{-/-}$ ESCs. Thus, H2Bub1 deubiquitination did not directly regulate global Pol II transcription and the embryonic phenotypes of the *Atxn7l3*$^{-/-}$ embryo could be a consequence of the activity of the DUBm on other proteins that of which the identification has been started during my thesis, but awaits for further validation experiments.

7. Materials and methods

7.1 Generation *Usp22*$^{+/-}$ and *Atxn7l3*$^{+/-}$ mouse lines

Usp22$^{+/-}$ and *Atxn7l3*$^{+/-}$ mouse lines were generated at the Institut Clinique de la Souris (ICS, Illkirch, France) using mESCs containing the targeting constructs ordered from the International Knockout Mouse Consortium (IKMC), including the Knockout Mouse Programme (KOMP) repository (UC, Davis). In the *Usp22* targeting construct (*Usp22*$^{tm1a(KOMP)Wtsi}$) a *LacZ* and *Neo* cassette were located in intron 1, flanked by *FRT* sequences, and *loxP* sequences were flanking exon 2. In the *Atxn7l3* targeting construct (*Atxn7l3*$^{tm1.1(KOMP)Wtsi}$) a *LacZ* and *Neo* cassette were located in intron 2, flanked by *FRT* sequences, and the *loxP* sequences were flanking exon 2 to exon 12. Chimeras were generated by injecting the C57BL/6 mESCs containing the targeting constructs into BALB/C blastocysts. Mice heterozygous for the targeting allele were crossed to a Cre-recombinase deleter strain, in order to generate the null alleles *Usp22*$^{-}$ and *Atxn7l3*$^{-}$, then mice heterozygous for the null allele (*Usp22*$^{+/-}$ or *Atxn7l3*$^{+/-}$) were intercrossed to generate homozygous mutant embryos (*Usp22*$^{-/-}$ or *Atxn7l3*$^{-/-}$). The *Atxn7l3*$^{+/-}$ mice were maintained on a mixed B6D2 background.

7.2 Generation *Atxn7l3*$^{-/-}$ mESCs and *Atxn7l3*$^{-/-}$ MEFs

To generate *Usp22*$^{-/-}$, *Atxn7l3*$^{-/-}$ and control mESCs, timed matings between heterozygous mice were conducted, then at E3.5, pregnant females were sacrificed, uteri were flushed with M2 medium (Sigma-Aldrich), and individual blastocysts were transferred to wells of a 96-well plates pre-coated with 0.1% gelatin. Blastocysts were cultured and expanded in regular mESCs medium (DMEM (4.5 g/l glucose) with 2 mM Glutamax-I, 15% ESQ FBS (Gibco), penicillin, streptomycin, 0.1 mM non-essential amino acids, 0.1% ß-mercaptoethanol, 1500 U/mL LIF and two inhibitors (2i; 3 µM CHIR99021 and 1 µM PD0325901, Axon MedChem)). After expansion, mESCs were genotyped and frozen.

To generate *Atxn7l3*$^{-/-}$ and control mouse embryonic fibroblasts (MEFs), timed matings between heterozygous mice were conducted, then at E10.5, pregnant females were sacrificed, and embryos were collected. The embryo yolk sacs were collected for genotyping, and the head and gastrointestinal tract were carefully dissected away from embryos. The remaining carcasses were transferred to individual 1.5 ml Eppendorf tubes, and 50 µl of trypsin (0.25% in EDTA, Gibco)

was added and gently triturated 5 times to dissociate the embryos. The dissociated embryos were incubated in trypsin for 5 min at room temperature, then the trypsin was quenched with 500 µl of FCS. Cells were transferred to individual wells of a 6-well plate pre-coated with 0.1% gelatin and cultured in MEF medium (DMEM, 10% FCS, penicillin and streptomycin). Cells were visualized with an EVOS XL Core Cell Imaging System (#AMEX-1100, Thermo Fisher 192 Scientific) using an LPlan PH2 10x / 0.25 objective.

7.3 Protein extraction and western blot assays

To extract histone proteins, embryos dissected at the indicated embryonic days, or about 5 x10^6 cells were lysed with 100 µl acidic extraction buffer (10 mM Hepes, pH 7.9, 1.5 mM MgCl2, 10mM KCl, 0.5 mM DTT and 0.2 M HCl) freshly complemented with 1× Proteinase Inhibitor Cocktail (PIC) and 10 mM N-ethylmaleimide (Sigma-Aldrich) and incubated on an end-to-end rotator for 2 hours at 4°C. Following the incubation, cell extract was centrifuged at 20 800 x g for 10 min at 4°C, to pellet the acid insoluble material. Ten µl of the supernatant, containing histone proteins, were run on 4–12% gels (Bis-tris NuPAGE Novex, Life Technologies), then proteins were transferred and western blot assays were carried out by using standard methods. Protein levels were quantified by ImageJ.

7.4 Immunofluorescence

Cells were washed twice with 1x PBS, fixed with 4% PFA (Electron Microscopy Science) for 10 min at RT. After fixation, cells were washed three times with 1x PBS, permeabilized with sterile 0.1% Triton X-100 in PBS for 20 min at RT, then washed three times in 1x PSB. Cells were incubated either with phalloidin conjugated to Alexa dye (Phalloidin-iFluor 488, Abcam, as described in the manufacturer's protocol) to label F-actin filaments, or with an anti-β-actin mouse monoclonal antibody (Sigma Aldrich, A5441) at a dilution of 1:1000 in 1x PBS with 10% FCS, overnight at 4°C. The following day, cells were washed three times with 1x PBS, then β-actin labelled cells were further incubated with secondary goat anti-mouse Alexa 488 antibody (Invitrogen) at a dilution of 1:2000 in 1x PBS with 10% FCS for 1 hr at RT. The cells were washed three times with 1x PBS, then incubated in 20 mM Hoechst 3342 (Thermo Scientific) for 10 min at RT, before being washed three times with 1x PBS, then cells were covered with a coverslip

coated in ProLong Gold mounting medium (Invitrogen). Pictures were taken using a Leica DM 4000 B upright microscope equipped with a Photometrics CoolSnap CF Color camera with a HCX PL S-APO 20x/0.50 objective.

7.5 Colony formation assay and alkaline phosphatase staining

Three thousand mESCs were seeded on gelatin-coated 6-well plates in regular mESC medium (see above) to form colonies at low density. The medium was exchanged every two days for 6 days. mESC alkaline phosphatase (AP) activity test was performed using Red Substrate Kit, Alkaline Phosphatase (Vector Laboratories) according to the manufacturer's instructions. mESC clones were washed with 1x cold PBS and fixed with 4% PFA for 10 min at RT. After fixation, cells were washed twice with 1x PBS and incubated in 1 ml AP detection system (as recommended by the manufacturer's protocol) for 30 min at RT in the dark. Then cells were washed twice with cold 1x PBS, and visualized with a EVOS XL Core Cell Imaging System (#AMEX- 1100, Thermo Fisher Scientific) using a LPlan PH2 4x / 0.13 objective.

7.6 Cell proliferation analysis

To determine cell proliferation, a total of $1x10^5$ mESCs per 6-well plate were seeded in regular mESC medium and 3x104 passage two MEF cells per 24-well plate were seeded in MEF medium. The medium was exchanged every two days. Cell numbers were counted with Countess cell counting chambers (Invitrogen). Statistical analyses were determined by the Mann-Whitney test (ns $p>0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; 239 *** $p \leq 0.05$).

7.7 Cell cycle analysis

Hundred thousand mESCs were fixed in 70% EtOH overnight at -4°C. After fixation, cells were treated with RNase A (100 µg/ml) (Thermo Fisher Scientific, #EN0531) and stained with propidium iodide (40 µg/ml) (Sigma Aldrich, #P-4170) for 30 min at 37 °C. The acquisition of the DNA content was analysed on FACS CALIBUR (BD Sciences) flow cytometer. Quantitative results were analyzed by FlowJo software (BD Sciences).

7.8 Apoptosis analysis using annexin-V staining

At the indicated incubation time, floating cells were collected in culture supernatants and adherent cells were harvested by trypsinization. After collection, cells were washed twice with cold 1X PBS, and about $2x10^5$ cells were resuspend in 100 µl binding buffer (FITC Annexin V Apoptosis Detection Kit, Biolegend). Subsequently, 5 µl FITC Annexin V (FITC Annexin V Apoptosis Detection Kit, Biolegend) and 10 µl propidium iodide was added to the cell suspension. Cells were gently vortexed and incubated in the dark for 15 min at RT. Thereafter, another 400 µl Annexin V binding buffer was added to each tube. Cells were analysed using a FACS CALIBUR (BD Sciences) flow cytometer. Dot plots were generated using the FlowJo software.

7.9 Nascent RNA extraction

3 x 15 cm plates of E14 mESCs were used. The growing cells were supplemented with a final concentration of 500 µM 4sU for labelling up to 15 min. Upon harvesting, cells were washed once with ice cold 1xPBS, scraped into 1 ml Trizol reagent per 15 cm plate, collected the lysates of all three plates in one 15 ml tube and further homogenized by using a syringe with the smallest aperture possible, then aliquot in 1,5 ml tube for RNA extraction. For spike-in, the cell number ratio we used is 10:1 of mESC: *Drosophila* S2 cells. S2 cells are labelled in the same way as mESC cells. Total RNA was isolated with Trizol extraction. The Trizol reagent treated cells were incubated for 5 min at RT to permit complete dissociation of the nucleoprotein complex, added 0.2 ml of chloroform per 1ml of Trizol Reagent. Centrifuged the sample at 12,000 x g for 15 minutes at 4°C and removed the aqueous phase into a new 1.5 ml tube. 0.5 ml of 100% isopropanol was added to the aqueous phase to precipitation RNA pellet. After washing with 1 ml of 75% ethanol, the RNA was treated with DNase using TURBO DNA-free Kit to remove DNA. Isolated RNA was dissolved in sterile nuclease-free water (Bio-lab) and quantified using Nano-drop. Thereafter, 250 µg total RNA in 100 µl DEPC-treated water and sonication on Covaris E220 to fragment RNA. Fragmented RNA was in a range between 10 kb and 200 bp (average of >1.5 kb). Equal amounts of RNA were then biotinylated as follows: to purify nascent RNA, fragmented total RNA was incubated at 60°C for 10 min and immediately chilled on ice for 2 minutes. Biotinylation labeling buffer (100 mM Tris-HCl, pH 7.5 and 10 mM EDTA) was added together with 400 µl DMSO, 200 µl biotin-HPDP (of 1 mg/ml stock) and added DEPC-treated RNase-free water to

have a total volume of 1 ml. The reaction proceeded at room temperature for 3 hours in darkness. Biotinylated RNA was then isolated in two sequential rounds of phase separation (adding equal volumes of chloroform, centrifugation at 4 °C 12000 x g for 10 min and isolating the upper fraction). RNA was precipitated by over-night incubation at -80°C with equal volumes of isopropanol and 1:10 (v/v) of 5 M NaCl. Clean RNA was reconstituted in 100 ml of nuclease-free water. To capture biotinylated RNAs, magnetic streptavidin beads (μMACS Streptavidin beads and kit, Miltenyi) was used according to the manufacturer's instructions. Eluted RNA was then isolated using RNA Clean & Concentrator – 25 kit (Zymo Research) and reconstituted in 15 μl DEPC-treated water. Aliquots were taken for further reverse-transcription and qRT-PCR analyses.

7.10 Purification of GST-Dsk2 protein

Transform One Shot BL21 (DE3) bacterial cells with pGEX3-Dsk2 plasmid, and plate cells on ampicillin selection plates. Keep at 37 °C overnight. Pick a single colony and inoculate into 20 ml of LB containing 100 μg/ml ampicillin (LBamp) and shake at 37 °C at 200 rpm overnight for the pre-inoculum culture. Inoculate 300 ml of LBamp with 5 ml of the pre-inoculum in 2L Erlenmeyer flask. Shake at 37 °C at 200 rpm. When the OD600 reaches 0.8, induce the culture with 1mM (final concentration) of IPTG. Shake at 30 °C at 200 rpm for 4 h. Aliquot the culture into 50 ml falcon tubes and centrifuge at 4,500 rpm for 10 min to pellet bacteria. To each cell pellets, add 15 ml of cold PBSA containing protease inhibitors and resuspend the pellet completely by careful pipetting, or vortexing. Avoid denaturing proteins, often signified by bubbles in the mixture. Combine all 6 samples (90 ml total) into one 200 ml glass beaker. Sonicate with a tip probe sonicator (Branson Digital Sonifier 250) at 33% output, with 15 s ON, 30 s OFF pulses, for a total ON pulse duration of 10 min. Keep the sample on ice at all times. Add Triton-X100 to a final concentration of 0.5%, mix gently. Incubate on ice for 30 min. Transfer the sonicated lysates to appropriate vessels and centrifuge at 4,500 rpm for 30 min in falcon tubes. Resuspend the glutathione sepharose beads well and take 3 ml of suspension into a fresh tube. Spin at 500 g for 5 min at 4 °C and remove supernatant carefully. Wash once with cold PBSA, and then resuspend in 3.3 ml cold PBSA. For binding of GST-Dsk2 to the beads, add 1 ml of well-resuspended glutathione sepharose bead solution from the previous step to each 30 ml of cleared lysate. Also add DTT to a final concentration of 2 mM. Rotate gently in the cold room for at least 4 h or overnight. Wash the beads twice with ice-cold PBSA, 0.1% Triton X-100, containing protease inhibitors. Wash once more

with PBSA without Triton X-100, but containing protease inhibitors. Add 30 ml of PBSA containing protease inhibitors and 0.02% sodium azide to the prepared Dsk2 beads and store at 4 °C.

7.11 Dsk2 pulldown

Take 0.5 ml of GST-Dsk2 bead suspension (equivalent to 25 µl packed beads), to deplete/enrich ubiquitylated proteins from 1 mg of whole cell protein extract. Keep both beads and protein samples on ice at all times. For each cell lysate sample, pipet 1 mg of total protein into 2 ml safe-lock Eppendorf tube and slowly adjust all samples to the same volume with TENT buffer containing protease inhibitors, phosphatase inhibitors and 2mM freshly made NEM. Typically, the final sample volume should be between 700 µl and 1 ml. Prewash the beads in bulk. Spin beads at 500 g for 5 min at 4 °C, remove supernatant and wash once with TENT buffer containing protease inhibitors, phosphatase inhibitors and 2mM NEM. Gently resuspend beads in a smaller volume of TENT buffer containing protease inhibitors, phosphatase inhibitors and 2mM NEM (typically 220 µl per sample). Avoid making bubbles. Aliquot the same volume (typically 200 µl) of well-resuspended Dsk2 bead slurry to each sample. Rotate on a turning wheel/rotator (low to moderate speed) in the cold room for several hours to overnight. Spin the samples at 500 g for 5 min at 4 °C, remove supernatant and save as “unbound” fraction. Wash the beads carefully twice with 1 ml of TENT buffer containing protease inhibitors, phosphatase inhibitors and 2mM NEM. Wash the beads carefully once with 1 ml of PBS containing protease inhibitors, phosphatase inhibitors and 2mM NEM. Spin at 500 g for 5 min at 4 °C and remove as much supernatant as possible. Re-spin a few times if necessary. At this point, any remaining liquid may also be removed with a fine pipet-tip. To each bead sample, add 40 µl of Laemmli buffer containing DTT, mix by brief vortexing, and boil at 96–98 °C for 5 min. Spin the samples and save supernatant which now contains the enriched, ubiquitylated proteins.

7.12 RNA-seq and ChIP-seq analyses

For RNA-seq, total RNA was extracted from mESCs and MEFs using the NucleoSpin RNA isolation kit (Macherey-Nagel), according to manufacturer's instructions. Libraries were generated from the purified RNA using TruSeq Stranded mRNA (Illumina) protocol. After checking the

quality of the libraries with the Bioanalyser (Agilent), libraries were sequenced on the Illumina HiSeq 4000 at the GenomEast sequencing platform of IGBMC. The raw sequencing data generated reads were preprocessed in order to remove adapter, polyA and low-quality sequences (Phred quality score below 20), then were mapped to the mouse mm10 genome using STAR 63. Differential gene expression was measured using the DESeq2 package. For the analysis, only the transcripts expressed more than 100 normalized reads (DESeq2 reads divided by the median of the transcript length in kb) were considered. Using these criteria 11 172 transcripts were expressed in mESCs, and 11 113 transcripts were expressed in MEFs.

In the ChIP-seq experiments, we added 10 mM N-ethylmaleimide (Sigma-Aldrich) into all buffers and the use of either the anti-H2Bub1 antibody (MediMabs, NRO3) or the anti-RPB1 CTD Pol II antibody (1PB 7G5). Briefly, mESCs or MEFs were fixed in 1% PFA for 10 min at room temperature (RT), then the PFA was quenched with glycine at a final concentration of 125 mM for 5 min at RT. Cells were washed two times in 1× cold PBS, scraped, and pelleted. Nuclei were isolated by incubating cells with nuclear isolation buffer (50 mM Tris–HCl pH 8.0, 2 mM EDTA pH 8.0, 0.5% Nonidet P-40, 10% glycerol, 1× protease inhibitors and 10 mM NEM) for 10 min at 4°C with gentle agitation, followed by centrifugation at max speed to pellet the nuclei. Nuclei were resuspended in sonication buffer (0.1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0, 1× protease inhibitors and 10 mM NEM) then chromatin was sheared with the E220 sonicator (Covaris) and chromatin concentration was measured with the Qubit 3.0 (Thermo Fischer Scientific). Approximately of 50 µg of chromatin was used for each IP, which was diluted in ChIP dilution buffer (0.5% Nonidet P-40, 16.7 mM Tris–HCl pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 1× protease inhibitor cocktail and 10mM NEM). Antibodies used for the ChIP included anti-RPB1 CTD (1PB 7G5; 66) anti-H2Bub1 (MediMab, NRO3), and mouse IgG (Jackson Laboratories) which were incubated with the chromatin overnight with gentle agitation at 4°C. The next day, Dynabeads protein G magnetic beads (Invitrogen) were added for 1 hour, then were isolated and washed for 5 min at 4°C, once with low salt wash buffer (0.1% SDS, 0.5% Nonidet P-40, 2 mM EDTA, 150 mM NaCl, 20 mM and Tris-HCl pH 8.0), once with high salt wash buffer (0.1% SDS, 0.5% Nonidet P-40, 2 mM EDTA, 500 mM NaCl, 20 mM and Tris–HCl pH 8.0), and once with LiCl wash buffer (0.2 M LiCl, 0.5% Nonidet P-40, 0.5% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH8.0), then washed twice with TE buffer, then the beads were incubated in elution

buffer (1% SDS, 0.1 M NaHCO3) at 65°C with shaking to elute complexes. Crosslinks were reversed with by adding NaCl at a final concentration of 0.2 M overnight as well as 50 µg/ml RNase A at 65°C and the following day the samples were treated with 20 µg Proteinase K, 26.6 µl of 1 M Tris–HCl pH 7.9 and 13.3 µl of 0.5 M EDTA, and DNA was phenol/chloroform purified and precipitated. The precipitated DNA was used to generate libraries with the MicroPlex Library Preparation kit v2 (Diagenode) for ChIP-seq according to the manufacturer's instructions. The samples were then sequenced on HiSeq 4000 with read lengths of 1 × 50 bp, reads were mapped to the mouse mm10 genome. Samples were normalized and peak calling was performed using the MACS2 software.

7.13 Embryoid body (EB) formation

For EB differentiation, 1.5×10^6 cells were plated in nonadherent bacterial 10 cm plates (CA 39 QUA-01) in differentiation medium. Media were replaced every other day. The differentiation medium contained high-glucose DMEM with 20% fetal serum, 100 mM β-mercaptoethanol, 2 mM non-essential amino acids.

For hanging drop method, 3000 cells per 20ul medium in one drop was plated in nonadherent bacterial 10 cm plates. About 70 drops were set in one 10 cm plate, then invert cultured these drops with 5 ml water in the corresponding cover to avoid the drops drying. After 48 hours, 15ml differentiation media was added and the EBs were cultured for 12 days. Media were replaced every other day.

7.14 Neuronal differentiation

For EB formation, 3×10^6 ES cells were plated onto nonadherent bacterial dishes (Greiner) in differentiation medium (DMEM with 10% fetal serum, 100 mM β-mercaptoethanol, 2 mM non-essential amino acids) and incubated for 8 days. Medium was changed every other day and 5 µM retinoic acid was added after 4 days. EBs were dissociated into single cells and put in 24-well plate pretreated PORN- and laminin-coated. The 24-well plate was coated with a solution of 10 µg/ml PORN solution in borate buffer (150 mM, pH 8.4) and placed overnight in the incubator. After washing the plate three times with PBS (H2O in the case of polyornithine), laminin (~0.5 µg/cm^2)

was added directly to the PBS solution and the plate returned to the incubators for at least 2 h. EBs were washed twice with PBS and trypsinized by incubating them 5 min in a water bath at 37 °C in a 0.05% trypsin solution in 0.04% EDTA/PBS. EBs were then gently but thoroughly resuspended in 10 ml EB medium, centrifuged for 5 min at 1,000 r.p.m at room temperature. The pellet was resuspended in N2 medium [125 ml DMEM (4.5 g/l glucose), 125 ml F-12, 25 ug/ml insulin (bovine or porcine), 50 ug/ml transferrin, 30 nM sodium selenite, 50 ug/ml BSA, 1% penicillin/streptomycin. pH 7.0-7.8] and the cell suspension were filtered through a 40-µm nylon cell strainer. Dissociated cells can be frozen at this stage if needed. After removal of laminin from the plates, the cell suspension was immediately added at a density of 1.5×10^5 cells per cm^2. The N2 medium was changed after 2 h and again after 2 days for IF experiment.

7.15 Epiblast like cell differentiation

For mESCs expansion, mESCs were first cultured in mESC medium (DMEM (4.5 g/l glucose) with 2 mM Glutamax-I, 15% ESQ FBS (Gibco), penicillin, streptomycin, 0.1 mM non-essential amino acids, 0.1% ß-mercaptoethanol, 1500 U/mL LIF and two inhibitors (2i; 3 µM CHIR99021 and 1 µM PD0325901, Axon MedChem)) cultured for two days, then the serum-free N2B27-based medium was used for 2 days. For differentiation of mouse ESCs into EpiLCs, cells were washed with PBS, trypsinized, and strained. A total of 200,000–300,000 cells per 10 cm^2 were plated on tissue culture dishes pretreated with 5 mg/ml Fibronectin (Millipore) in N2B27-based medium supplemented with 1% KSR (Invitrogen), 12 mg/ml bFGF (R&D Scientific) and 20 ng/ml Activin A (R&D Scientific) and cultured for 2 days. Subsequently, EpiLCs were treated with CHIR and Activin A for the last 2 days to induce gastrulation.

7.16 List of primers

Primers for q-PCR			
Gene	**Forward 5' - 3'**	**Reverse 5 '- 3'**	**Reference**
Atp1b1	GGAGGAAGGCAGCTGGAAG	GATGGTCCCGATGAAGATGC	
Atxn7l3	CTTCTCTGAGCCATAGGACCA	CCCCCACCTGGAGAAGTG	
Brachury	GTATTCCCAATGGGGGTGGCT	CCTTAGAGCTGGGTACCTCTC	(Shilu et al., 2016)
Chd1	CACCCGAGCTCAGTGTTTG	CAAAGCCATGAGGAGACCTG	(Carla et al., 2013)
Cdh2	AGCACACCTTCACCCAACAT	TGACATCTGTCACCGTGATG	(Carla et al., 2013)
Col14a1	TGGGAGTACCTGGACCTCAG	TCAGGGGCAGGAGCTTAGTA	
Col18a1	CAGCTGCCTCCCTTCCAG	AGGGTCATCGATTTGTGAGA	
Col2a1	GTGGCAGAGATGGAGAACCT	CCTTGCATGACTCCCATCTG	
Crxos	GCCCTGGATGGTACCTCTTC	TGTGCTTACAGCTGGTCGAG	
Dnmt3b	CTCGCAAGGTGTGGGCTTTTGTAAC	CTGGGCATCTGTCATCTTTGCACC	
Eomes	GGCCTACCAAAACACGGATATC	TTTCTGAAGCCGTGTACATGGA	(Hidetoshi et al., 2009)
Ephx1	ATGACTGGGAAGGAACCAGG	GACATCCGCAAGTTCGTGTC	
Esrrb1	GAGGACTCCGCCATCAAAT	TAGTGGTAGCCAGAGGCAATGT	
Flk-1	CCAAGCTCAGCACACAGAAA	CCTGGGAATGGTGAGTGTTT	(Carla et al., 2013)
Foxa2	CGAGTTAAAGTATGCTGGGAG	TATGTGTTCATGCCATTCATCC	(Antonio et al., 2017)
Gapdh	TTCACCACCATGGAGAAGGC	CCCTTTTGGCTCCACCCT	
Gata4	CAGCAGCAGCAGTGAAGAGATG	ACCAGGCTGTTCCAAGAGTCC	(Yeh et al., 2014)
Gata6	TCTACACAAGCGACCACCTCAG	GCCAGAGCACACCAAGAATCC	(Yeh et al., 2014)
Gsc	GAGAACCTCTTCCAGGAGAC	TTCTTAAACCAGACCTCCACC	(Antonio et al., 2017)
Gstm1	CTAGTGAGTGCCCGTGTAGC	TGCCTACATGAAGAGTAGCCG	
Gstm6	CCAACACCGGCACTCCAT	ATATGAAGACCAGCCGCTTCC	
Hand1	CACCACCTACCACCGCAGTA	CCTTCTTGGGTCCTGAGCCTTT	(Yeh et al., 2014)
Hsp90ab1	ACCTGGGAACCATTGCTAAG	AGAATCCGACACCAAACTGC	
Klf4	GTGGGTTAGCGAGTTGGAAA	GTGCAGCTTGCAGCAGTAAC	
Fgf5	CAGATCTCCCGGATGGCAAAG	GCGGACGATAGGTATTATAGCT	
mCol15a1	GAGGTGGCTGCTCTCCATC	AAAGCTGTAAGCCGGGAAAC	
Mef2c	CTGAGCGTGCTGTGCGACTGT	GCTCTCGTGCGGCTCGTTGTA	(Qin et al., 2017)
Mesp1	GTCACTCGGTCCTGGTTTAAGC	TGCGTACTGGAACGATGGGT	(Qin et al., 2017)
Mixl1	TCCTCCATTGCCCTGCTCCT	ACGCCTCCTCCAGTCATGCT	(Yeh et al., 2014)
Myc	CTGACAGAACTGATGCGCTG	GGCTGAAGCTTACAGTCCCAA	
Myh6	GCCCAGTACCTCCGAAAGTC	GCCTTAACATACTCCTTGTC	(Peter et al., 2016)

Myh7	ACAACCCCTACGATTATGCGT	ACGTCAAAGGCACTATCCGTG	(Qin et al., 2017)
Nanog	GAAATCCCTTCCCTCGCCATC	CTCAGTAGCAGACCCTTGTAAGC	
Nestin	ACCAAAGCCTCTTAGAAATGACC	CTCCATACCTCCTTCATTCAGTG	
Nkx2.5	CAAGTGCTCTCCTGCTTTCC	GGCTTTGTCCAGCTCCACT	(Qin et al., 2017)
Notch1	ACAACAACGAGTGTGAGTCC	ACACGTGGCTCCTGTATATG	(Kensuke et al., 2011)
Otx2	CTTCATGAGGGAAGAGGTGG	GGCCTCACTTTGTTCTGACC	(Li et al., 2018)
PAX6	CAGTCAGACCTCCTCATACTCGT	ACTGTTCATGTGTGTTTGCATGT	
Pgk1	TACCTGCTGGCTGGATGG	CACAGCCTCGGCATATTTCT	
Pou5f1	CTAGCATTGAGAACCGTGTGAG	GATTGGCGATGTGAGTGATCT	
Sox17	CTCCAGAAACTGCAGACCAGA	TGGAGGTGCTGCTCATTGTAT	(Carla et al., 2013)
Sox2	CCAGCGCATGGACAGCTA	GCTGCTCCTGCATCATGCT	
Tfcp2l1	ACTACAACCAGCACAACTCTGG	CCCATTCTCAGGAGATAGCTG	
TnnT2	GGCAGAACCGCCTGGCTGAA	CTGCCACAGCTCCTTGGCCT	(Qin et al., 2017)
Wnt3	CAAGCACAACAATGAAGCAGGC	TCGGGACTCACGGTGTTTCTC	(Nicholas et al., 2017)
Zscan4d	ATGATTGGCGAAAGCGACGG	TTCAGCCACAAGACCAACCTG	

Primers for genotyping		
Usp22	Ef	CCTCTTCATCTTTCTGTACCTGACCCA
	Er	ACATCTCTTGGGCACTGAGCGC
	L3r	ACCTACAATGCCAGAACTGGGGTG
Atxn7l3	Ef	CAAAGAAAGCAGCATGCTTGGTCAGG
	Er	CCTGCAGAGGAAAGAGGCACAGAG
	Wr	CAGGAAGAAGTAGCCACACTTAACAGC

7.17 List of antibodies

Name		Species	Company	Catalog
H2Bub1	Western blot	Rabbit	Cell signaling technology	5546
H3K4me3	Western blot	Rabbit	Abcam	8580
H3K9ac	Western blot	Rabbit	Merck-Millipore	07-352
ATXN7L3	Western blot	Rabbit	“in house”	2325
ENY2	Western blot	Rabbit	abcom	ab183622
SUPT7L	Western blot	Rabbit	Bethyl	A302-803A
LC3B	Western blot	Rabbit	Abcam	Ab51520
PRB1(Ser2)	Western blot	Rat	“in house”	3E10
γH2AX	Western blot	Mouse	Abcam	Ab22551
H3	Western blot	Rabbit	Abcam	ab1791
H4	Western blot	Mouse	Invitrogen	MA3-050
GAPDH	Western blot	Mouse	Sigma-Aldrich	MAB374
VINCULIN	Western blot	Mouse	sigma	V9131
H2Bub1	ChIP	Mouse	MediMabs	MM-0029-P
Pol II (RPB1)	ChIP	Mouse	“in house”	PB-7G5
β-Tublin III	Immunofluorescence	Mouse	BioLegend	MMS-435P
β-actin	Western blot	Mouse	Sigma	A5441
cTnT	Immunofluorescence/ Western blot	Mouse	ThermoFisher	MA5-12960
PAX6	Immunofluorescence	Mouse	DSHB	PAX6-S
NESTIN	Immunofluorescence	Mouse	DSHB	rat-401

8. References

Adam, S., Dabin, J., and Polo, S.E. (2015). Chromatin plasticity in response to DNA damage: The shape of things to come. DNA Repair (Amst) *32*, 120-126.
Adamczewski, J.P., Rossignol, M., Tassan, J.P., Nigg, E.A., Moncollin, V., and Egly, J.M. (1996). MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH. EMBO J *15*, 1877-1884.
Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720-731.
Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191-1200.
Akalin, A., Fredman, D., Arner, E., Dong, X., Bryne, J.C., Suzuki, H., Daub, C.O., Hayashizaki, Y., and Lenhard, B. (2009). Transcriptional features of genomic regulatory blocks. Genome Biol *10*, R38.
Alabert, C., and Groth, A. (2012). Chromatin replication and epigenome maintenance. Nat Rev Mol Cell Biol *13*, 153-167.
Alipour, E., and Marko, J.F. (2012). Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res *40*, 11202-11212.
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.
Anandapadamanaban, M., Andresen, C., Helander, S., Ohyama, Y., Siponen, M.I., Lundstrom, P., Kokubo, T., Ikura, M., Moche, M., and Sunnerhagen, M. (2013). High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. Nat Struct Mol Biol *20*, 1008-1014.
Anderson, C.J., Baird, M.R., Hsu, A., Barbour, E.H., Koyama, Y., Borgnia, M.J., and McGinty, R.K. (2019). Structural Basis for Recognition of Ubiquitylated Nucleosome by Dot1L Methyltransferase. Cell Rep *26*, 1681-1690 e1685.
Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B.*, et al.* (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183-197.
Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci U S A *88*, 10148-10152.
Armache, A., Yang, S., Martinez de Paz, A., Robbins, L.E., Durmaz, C., Cheong, J.Q., Ravishankar, A., Daman, A.W., Ahimovic, D.J., Klevorn, T.*, et al.* (2020). Histone H3.3 phosphorylation amplifies stimulation-induced transcription. Nature *583*, 852-857.
Armache, K.J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proc Natl Acad Sci U S A *100*, 6964-6968.
Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Biol *10*, 91-103.
Asahina, K., Tsai, S.Y., Li, P., Ishii, M., Maxson, R.E., Jr., Sucov, H.M., and Tsukamoto, H. (2009). Mesenchymal origin of hepatic stellate cells, submesothelial cells, and perivascular mesenchymal cells during mouse liver development. Hepatology *49*, 998-1011.
Atanassov, B.S., and Dent, S.Y. (2011). USP22 regulates cell proliferation by deubiquitinating the transcriptional regulator FBP1. EMBO Rep *12*, 924-930.

Atanassov, B.S., Evrard, Y.A., Multani, A.S., Zhang, Z., Tora, L., Devys, D., Chang, S., and Dent, S.Y. (2009). Gcn5 and SAGA regulate shelterin protein turnover and telomere maintenance. Mol Cell *35*, 352-364.

Atanassov, B.S., Mohan, R.D., Lan, X., Kuang, X., Lu, Y., Lin, K., McIvor, E., Li, W., Zhang, Y., Florens, L.*, et al.* (2016). ATXN7L3 and ENY2 Coordinate Activity of Multiple H2B Deubiquitinases Important for Cellular Proliferation and Tumor Growth. Mol Cell *62*, 558-571.

Atlasi, Y., and Stunnenberg, H.G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. Nat Rev Genet *18*, 643-658.

Auernhammer, C.J., Bousquet, C., Chesnokova, V., and Melmed, S. (2000). SOCS proteins: modulators of neuroimmunoendocrine functions. Impact on corticotroph LIF signaling. Ann N Y Acad Sci *917*, 658-664.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.

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.

Balasubramanian, R., Pray-Grant, M.G., Selleck, W., Grant, P.A., and Tan, S. (2002). Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation. J Biol Chem *277*, 7989-7995.

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.

Baniahmad, A., Muller, M., Steiner, C., and Renkawitz, R. (1987). Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. EMBO J *6*, 2297-2303.

Bannister, A.J., and Kouzarides, T. (2005). Reversing histone methylation. Nature *436*, 1103-1106.

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.

Bartholomew, B. (2014). Regulating the chromatin landscape: structural and mechanistic perspectives. Annu Rev Biochem *83*, 671-696.

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

Beck, S., Le Good, J.A., Guzman, M., Ben Haim, N., Roy, K., Beermann, F., and Constam, D.B. (2002). Extraembryonic proteases regulate Nodal signalling during gastrulation. Nat Cell Biol *4*, 981-985.

Becker, P.B., and Workman, J.L. (2013). Nucleosome remodeling and epigenetics. Cold Spring Harb Perspect Biol *5*.

Beddington, R.S. (1983). Histogenetic and neoplastic potential of different regions of the mouse embryonic egg cylinder. J Embryol Exp Morphol *75*, 189-204.

Beddington, R.S., and Robertson, E.J. (1998). Anterior patterning in mouse. Trends Genet *14*, 277-284.

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.

Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E.J., and Constam, D.B. (2006). The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. Dev Cell *11*, 313-323.

Bergeron, K.F., Cardinal, T., Toure, A.M., Beland, M., Raiwet, D.L., Silversides, D.W., and Pilon, N. (2015). Male-biased aganglionic megacolon in the TashT mouse line due to perturbation of silencer elements in a large gene desert of chromosome 10. PLoS Genet *11*, e1005093.

Bergink, S., Salomons, F.A., Hoogstraten, D., Groothuis, T.A., de Waard, H., Wu, J., Yuan, L., Citterio, E., Houtsmuller, A.B., Neefjes, J.*, et al.* (2006). DNA damage triggers nucleotide excision repair-dependent monoubiquitylation of histone H2A. Genes Dev *20*, 1343-1352.
Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature *529*, 551-554.
Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. Genome Biol *5*, R62.
Bertero, A., Madrigal, P., Galli, A., Hubner, N.C., Moreno, I., Burks, D., Brown, S., Pedersen, R.A., Gaffney, D., Mendjan, S.*, et al.* (2015). Activin/nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. Genes Dev *29*, 702-717.
Bhaumik, S.R., and Green, M.R. (2002). Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters in vivo. Mol Cell Biol *22*, 7365-7371.
Bian, C., Xu, C., Ruan, J., Lee, K.K., Burke, T.L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B.O.*, et al.* (2011). Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. EMBO J *30*, 2829-2842.
Bibel, M., Richter, J., Schrenk, K., Tucker, K.L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nat Neurosci *7*, 1003-1009.
Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. Cell *152*, 1270-1284.
Biterge, B., and Schneider, R. (2014). Histone variants: key players of chromatin. Cell Tissue Res *356*, 457-466.
Blau, J., Xiao, H., McCracken, S., O'Hare, P., Greenblatt, J., and Bentley, D. (1996). Three functional classes of transcriptional activation domain. Mol Cell Biol *16*, 2044-2055.
Bleichenbacher, M., Tan, S., and Richmond, T.J. (2003). Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. J Mol Biol *332*, 783-793.
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.
Boeuf, H., Hauss, C., Graeve, F.D., Baran, N., and Kedinger, C. (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. J Cell Biol *138*, 1207-1217.
Bogdanovic, O., Fernandez-Minan, A., Tena, J.J., de la Calle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S.J., Veenstra, G.J., and Gomez-Skarmeta, J.L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res *22*, 2043-2053.
Bohm, V., Hieb, A.R., Andrews, A.J., Gansen, A., Rocker, A., Toth, K., Luger, K., and Langowski, J. (2011). Nucleosome accessibility governed by the dimer/tetramer interface. Nucleic Acids Res *39*, 3093-3102.
Boiani, M., and Scholer, H.R. (2005). Regulatory networks in embryo-derived pluripotent stem cells. Nat Rev Mol Cell Biol *6*, 872-884.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., *et al.* (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855 e1816.

Bondarenko, M.T., Maluchenko, N.V., Valieva, M.E., Gerasimova, N.S., Kulaeva, O.I., Georgiev, P.G., and Studitsky, V.M. (2015). [Structure and function of histone chaperone FACT]. Mol Biol (Mosk) *49*, 891-904.

Bondue, A., and Blanpain, C. (2010). Mesp1: a key regulator of cardiovascular lineage commitment. Circ Res *107*, 1414-1427.

Bondue, A., Lapouge, G., Paulissen, C., Semeraro, C., Iacovino, M., Kyba, M., and Blanpain, C. (2008). Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. Cell Stem Cell *3*, 69-84.

Bonenfant, D., Coulot, M., Towbin, H., Schindler, P., and van Oostrum, J. (2006). Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry. Mol Cell Proteomics *5*, 541-552.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A., *et al.* (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557-572 e524.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczynski, B., Riddell, A., and Furlong, E.E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet *44*, 148-156.

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

Bonnet, J., Wang, Y.H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W., Timmers, H.T., Tora, L., Devys, D., *et al.* (2010). The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. EMBO Rep *11*, 612-618.

Bossard, P., and Zaret, K.S. (2000). Repressive and restrictive mesodermal interactions with gut endoderm: possible relation to Meckel's Diverticulum. Development *127*, 4915-4923.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Branco, M.R., and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol *4*, e138.

Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R., and Nasmyth, K. (1985). Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. Cell *41*, 41-48.

Brand, M., Leurent, C., Mallouh, V., Tora, L., and Schultz, P. (1999). Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. Science *286*, 2151-2153.

Brand, M., Moggs, J.G., Oulad-Abdelghani, M., Lejeune, F., Dilworth, F.J., Stevenin, J., Almouzni, G., and Tora, L. (2001). UV-damaged DNA-binding protein in the TFTC complex links DNA damage recognition to nucleosome acetylation. EMBO J *20*, 3187-3196.

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K., *et al.* (2004). Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. Nat Biotechnol *22*, 707-716.

Brannan, K., and Bentley, D.L. (2012). Control of Transcriptional Elongation by RNA Polymerase II: A Retrospective. Genet Res Int *2012*, 170173.
Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J.*, et al.* (2012). mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. Mol Cell *46*, 311-324.
Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A.*, et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.
Brown, C.E., Howe, L., Sousa, K., Alley, S.C., Carrozza, M.J., Tan, S., and Workman, J.L. (2001). Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. Science *292*, 2333-2337.
Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.
Bruce, K., Myers, F.A., Mantouvalou, E., Lefevre, P., Greaves, I., Bonifer, C., Tremethick, D.J., Thorne, A.W., and Crane-Robinson, C. (2005). The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. Nucleic Acids Res *33*, 5633-5639.
Buckingham, M., Meilhac, S., and Zaffran, S. (2005). Building the mammalian heart from two sources of myocardial cells. Nat Rev Genet *6*, 826-835.
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.
Bunch, H., Lawney, B.P., Burkholder, A., Ma, D., Zheng, X., Motola, S., Fargo, D.C., Levine, S.S., Wang, Y.E., and Hu, G. (2016). RNA polymerase II promoter-proximal pausing in mammalian long non-coding genes. Genomics *108*, 64-77.
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.
Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev Biol *210*, 30-43.
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.
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.
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.
Butler, J.E., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev *16*, 2583-2592.

Butryn, A., Schuller, J.M., Stoehr, G., Runge-Wollmann, P., Forster, F., Auble, D.T., and Hopfner, K.P. (2015). Structural basis for recognition and remodeling of the TBP:DNA:NC2 complex by Mot1. Elife *4*.
Cadigan, K.M., and Nusse, R. (1997). Wnt signaling: a common theme in animal development. Genes Dev *11*, 3286-3305.
Cai, G., Imasaki, T., Yamada, K., Cardelli, F., Takagi, Y., and Asturias, F.J. (2010). Mediator head module structure and functional interactions. Nat Struct Mol Biol *17*, 273-279.
Cai, Z., Zhang, M.X., Tang, Z., Zhang, Q., Ye, J., Xiong, T.C., Zhang, Z.D., and Zhong, B. (2020). USP22 promotes IRF3 nuclear translocation and antiviral responses by deubiquitinating the importin protein KPNA2. J Exp Med *217*.
Cairns, B.R. (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. Nat Struct Mol Biol *14*, 989-996.
Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.
Camus, A., Perea-Gomez, A., Moreau, A., and Collignon, J. (2006). Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. Dev Biol *295*, 743-755.
Cano, A., Perez-Moreno, M.A., Rodrigo, I., Locascio, A., Blanco, M.J., del Barrio, M.G., Portillo, F., and Nieto, M.A. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat Cell Biol *2*, 76-83.
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.
Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P.*, et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.
Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development *132*, 885-896.
Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. Nat Struct Mol Biol *20*, 290-299.
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.
Cevher, M.A., Shi, Y., Li, D., Chait, B.T., Malik, S., and Roeder, R.G. (2014). Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. Nat Struct Mol Biol *21*, 1028-1034.
Chakravarthy, S., Gundimella, S.K., Caron, C., Perche, P.Y., Pehrson, J.R., Khochbin, S., and Luger, K. (2005). Structural characterization of the histone variant macroH2A. Mol Cell Biol *25*, 7616-7624.
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.
Chambeyron, S., and Bickmore, W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev *18*, 1119-1130.
Chandrasekharan, M.B., Huang, F., and Sun, Z.W. (2009). Ubiquitination of histone H2B regulates chromatin dynamics by enhancing nucleosome stability. Proc Natl Acad Sci U S A *106*, 16686-16691.
Chandy, M., Gutierrez, J.L., Prochasson, P., and Workman, J.L. (2006). SWI/SNF displaces SAGA-acetylated nucleosomes. Eukaryot Cell *5*, 1738-1747.

Chapman, R.D., Heidemann, M., Hintermair, C., and Eick, D. (2008). Molecular evolution of the RNA polymerase II CTD. Trends Genet *24*, 289-296.
Chasman, D.I., Flaherty, K.M., Sharp, P.A., and Kornberg, R.D. (1993). Crystal structure of yeast TATA-binding protein and model for interaction with DNA. Proc Natl Acad Sci U S A *90*, 8174-8178.
Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev Cell *10*, 615-624.
Chen, C., and Shen, M.M. (2004). Two modes by which Lefty proteins inhibit nodal signaling. Curr Biol *14*, 618-624.
Chen, F.X., Smith, E.R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *19*, 464-478.
Chen, G., Schell, J.P., Benitez, J.A., Petropoulos, S., Yilmaz, M., Reinius, B., Alekseenko, Z., Shi, L., Hedlund, E., Lanner, F.*, et al.* (2016). Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. Genome Res *26*, 1342-1354.
Chen, S., Jing, Y., Kang, X., Yang, L., Wang, D.L., Zhang, W., Zhang, L., Chen, P., Chang, J.F., Yang, X.M.*, et al.* (2017). Histone H2B monoubiquitination is a critical epigenetic switch for the regulation of autophagy. Nucleic Acids Res *45*, 1144-1158.
Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.C., Nilges, M.*, et al.* (2010). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J *29*, 717-726.
Cheng, B., and Price, D.H. (2007). Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J Biol Chem *282*, 21901-21912.
Cheung, A.C., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249-253.
Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev *15*, 3319-3329.
Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, II (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412-415.
Choi, J.K., and Howe, L.J. (2009). Histone acetylation: truth of consequences? Biochem Cell Biol *87*, 139-150.
Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X.*, et al.* (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *361*.
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.
Ciruna, B., and Rossant, J. (2001). FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. Dev Cell *1*, 37-49.
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.
Clements, A., Poux, A.N., Lo, W.S., Pillus, L., Berger, S.L., and Marmorstein, R. (2003). Structural basis for histone and phosphohistone binding by the GCN5 histone acetyltransferase. Mol Cell *12*, 461-473.

Coin, F., Oksenych, V., and Egly, J.M. (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. Mol Cell *26*, 245-256.
Coin, F., Proietti De Santis, L., Nardo, T., Zlobinskaya, O., Stefanini, M., and Egly, J.M. (2006). p8/TTD-A as a repair-specific TFIIH subunit. Mol Cell *21*, 215-226.
Colbert, T., and Hahn, S. (1992). A yeast TFIIB-related factor involved in RNA polymerase III transcription. Genes Dev *6*, 1940-1949.
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. (2016). Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond. Annu Rev Biochem *85*, 265-290.
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.
Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B., and Robertson, E.J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development *120*, 1919-1928.
Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.
Corden, J.L., Cadena, D.L., Ahearn, J.M., Jr., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci U S A *82*, 7934-7938.
Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev *33*, 960-982.
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. (2019). Organization and regulation of gene transcription. Nature *573*, 45-54.
Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640-649.
Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863-1876.
Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet *2*, 292-301.
Cremer, T., Kurz, A., Zirbel, R., Dietzel, S., Rinke, B., Schrock, E., Speicher, M.R., Mathieu, U., Jauch, A., Emmerich, P.*, et al.* (1993). Role of chromosome territories in the functional compartmentalization of the cell nucleus. Cold Spring Harb Symp Quant Biol *58*, 777-792.
Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.
Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res *19*, 24-32.

Cunningham, T.J., Zhao, X., and Duester, G. (2011). Uncoupling of retinoic acid signaling from tailbud development before termination of body axis extension. Genesis *49*, 776-783.
D'Arcy, S., Martin, K.W., Panchenko, T., Chen, X., Bergeron, S., Stargell, L.A., Black, B.E., and Luger, K. (2013). Chaperone Nap1 shields histone surfaces used in a nucleosome and can put H2A-H2B in an unconventional tetrameric form. Mol Cell *51*, 662-677.
Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L., Lis, J.T., Siepel, A., and Kraus, W.L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol Cell *50*, 212-222.
Dey, A., Chitsaz, F., Abbasi, A., Misteli, T., and Ozato, K. (2003). The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. Proc Natl Acad Sci U S A 100, 8758-8763.
Dann, G.P., Liszczak, G.P., Bagert, J.D., Muller, M.M., Nguyen, U.T.T., Wojcik, F., Brown, Z.Z., Bos, J., Panchenko, T., Pihl, R.*, et al.* (2017). ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. Nature *548*, 607-611.
Darnell, J.E., Jr. (1996). Reflections on STAT3, STAT5, and STAT6 as fat STATs. Proc Natl Acad Sci U S A *93*, 6221-6224.
Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. Nature *468*, 911-920.
De Craene, B., and Berx, G. (2013). Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer *13*, 97-110.
de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev *26*, 11-24.
Dehaan, R.L. (1963). Migration patterns of the precardiac mesoderm in the early chick embrvo. Exp Cell Res *29*, 544-560.
Dekker, J. (2008). Gene regulation in the third dimension. Science *319*, 1793-1794.
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.
Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev *30*, 1357-1382.
Devault, A., Martinez, A.M., Fesquet, D., Labbe, J.C., Morin, N., Tassan, J.P., Nigg, E.A., Cavadore, J.C., and Doree, M. (1995). MAT1 ('menage a trois') a new RING finger protein subunit stabilizing cyclin H-cdk7 complexes in starfish and Xenopus CAK. EMBO J *14*, 5027-5036.
Dhillon, N., Raab, J., Guzzo, J., Szyjka, S.J., Gangadharan, S., Aparicio, O.M., Andrews, B., and Kamakaka, R.T. (2009). DNA polymerase epsilon, acetylases and remodellers cooperate to form a specialized chromatin structure at a tRNA insulator. EMBO J *28*, 2583-2600.
Di-Gregorio, A., Sancho, M., Stuckey, D.W., Crompton, L.A., Godwin, J., Mishina, Y., and Rodriguez, T.A. (2007). BMP signalling inhibits premature neural differentiation in the mouse embryo. Development *134*, 3359-3369.
Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of replication-independent histone turnover in budding yeast. Science *315*, 1405-1408.
Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W.*, et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331-336.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Doamekpor, S.K., Sanchez, A.M., Schwer, B., Shuman, S., and Lima, C.D. (2014). How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. Genes Dev *28*, 1323-1336.

Doble, B.W., and Woodgett, J.R. (2003). GSK-3: tricks of the trade for a multi-tasking kinase. J Cell Sci *116*, 1175-1186.

Dominski, Z., and Marzluff, W.F. (1999). Formation of the 3' end of histone mRNA. Gene *239*, 1-14.

Donda, A., Schulz, M., Burki, K., De Libero, G., and Uematsu, Y. (1996). Identification and characterization of a human CD4 silencer. Eur J Immunol *26*, 493-500.

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002). Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem *277*, 28368-28371.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schujiers, J., Lee, T.I., Zhao, K.*, et al.* (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374-387.

Dubaele, S., Proietti De Santis, L., Bienstock, R.J., Keriel, A., Stefanini, M., Van Houten, B., and Egly, J.M. (2003). Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. Mol Cell *11*, 1635-1646.

Duffy, E.E., Schofield, J.A., and Simon, M.D. (2019). Gaining insight into transcriptome-wide RNA population dynamics through the chemistry of 4-thiouridine. Wiley Interdiscip Rev RNA *10*, e1513.

Eichner, J., Chen, H.T., Warfield, L., and Hahn, S. (2010). Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. EMBO J *29*, 706-716.

Ellisdon, A.M., Jani, D., Kohler, A., Hurt, E., and Stewart, M. (2010). Structural basis for the interaction between yeast Spt-Ada-Gcn5 acetyltransferase (SAGA) complex components Sgf11 and Sus1. J Biol Chem *285*, 3850-3856.

Ema, M., Takahashi, S., and Rossant, J. (2006). Deletion of the selection cassette, but not cis-acting elements, in targeted Flk1-lacZ allele reveals Flk1 expression in multipotent mesodermal progenitors. Blood *107*, 111-117.

Engel, C., Gubbey, T., Neyer, S., Sainsbury, S., Oberthuer, C., Baejen, C., Bernecky, C., and Cramer, P. (2017). Structural Basis of RNA Polymerase I Transcription Initiation. Cell *169*, 120-131 e122.

Erdel, F., Krug, J., Langst, G., and Rippe, K. (2011). Targeting chromatin remodelers: signals and search mechanisms. Biochim Biophys Acta *1809*, 497-508.

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.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A.*, et al.* (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol Cell *38*, 452-464.

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., and Werner, M. (2008). Mediator-dependent recruitment of TFIIH modules in preinitiation complex. Mol Cell *31*, 337-346.

Evangelista, F.M., Maglott-Roth, A., Stierle, M., Brino, L., Soutoglou, E., and Tora, L. (2018). Transcription and mRNA export machineries SAGA and TREX-2 maintain monoubiquitinated H2B balance required for DNA repair. J Cell Biol *217*, 3382-3397.

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

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell *36*, 593-608.

Fan, L., Arvai, A.S., Cooper, P.K., Iwai, S., Hanaoka, F., and Tainer, J.A. (2006). Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. Mol Cell *22*, 27-37.

Fang, J., Chen, T., Chadwick, B., Li, E., and Zhang, Y. (2004). Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. J Biol Chem *279*, 52812-52815.

Feaver, W.J., Svejstrup, J.Q., Henry, N.L., and Kornberg, R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell *79*, 1103-1109.

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.

Fischer, T., Rodriguez-Navarro, S., Pereira, G., Racz, A., Schiebel, E., and Hurt, E. (2004). Yeast centrin Cdc31 is linked to the nuclear mRNA export machinery. Nat Cell Biol *6*, 840-848.

Fischer, T., Strasser, K., Racz, A., Rodriguez-Navarro, S., Oppizzi, M., Ihrig, P., Lechner, J., and Hurt, E. (2002). The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. EMBO J *21*, 5843-5852.

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.

Fleming, A.B., Kao, C.F., Hillyer, C., Pikaart, M., and Osley, M.A. (2008). H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. Mol Cell *31*, 57-66.

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.

Fraser, P., and Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. Nature *447*, 413-417.

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, G., and Oren, M. (2014). Writing and reading H2B monoubiquitylation. Biochim Biophys Acta *1839*, 694-701.

Fuchs, G., Shema, E., Vesterman, R., Kotler, E., Wolchinsky, Z., Wilder, S., Golomb, L., Pribluda, A., Zhang, F., Haj-Yahya, M.*, et al.* (2012). RNF20 and USP44 regulate stem cell differentiation by modulating H2B monoubiquitylation. Mol Cell *46*, 662-673.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep *15*, 2038-2049.
Funa, N.S., Schachter, K.A., Lerdrup, M., Ekberg, J., Hess, K., Dietrich, N., Honore, C., Hansen, K., and Semb, H. (2015). beta-Catenin Regulates Primitive Streak Induction through Collaborative Interactions with SMAD2/SMAD3 and OCT4. Cell Stem Cell *16*, 639-652.
Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science *361*, 1341-1345.
Gadue, P., Huber, T.L., Nostro, M.C., Kattman, S., and Keller, G.M. (2005). Germ layer induction from embryonic stem cells. Exp Hematol *33*, 955-964.
Gadue, P., Huber, T.L., Paddison, P.J., and Keller, G.M. (2006). Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. Proc Natl Acad Sci U S A *103*, 16806-16811.
Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A.*, et al.* (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282-286.
Gaiser, F., Tan, S., and Richmond, T.J. (2000). Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. J Mol Biol *302*, 1119-1127.
Galonska, C., Ziller, M.J., Karnik, R., and Meissner, A. (2015). Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. Cell Stem Cell *17*, 462-470.
Gangloff, Y.G., Romier, C., Thuault, S., Werten, S., and Davidson, I. (2001). The histone fold is a key structural motif of transcription factor TFIID. Trends Biochem Sci *26*, 250-257.
Gansen, A., Hauger, F., Toth, K., and Langowski, J. (2007). Single-pair fluorescence resonance energy transfer of nucleosomes in free diffusion: optimizing stability and resolution of subpopulations. Anal Biochem *368*, 193-204.
Gansen, A., Valeri, A., Hauger, F., Felekyan, S., Kalinin, S., Toth, K., Langowski, J., and Seidel, C.A. (2009). Nucleosome disassembly intermediates characterized by single-molecule FRET. Proc Natl Acad Sci U S A *106*, 15308-15313.
Garcia-Martinez, V., and Schoenwolf, G.C. (1993). Primitive-streak origin of the cardiovascular system in avian embryos. Dev Biol *159*, 706-719.
Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet *7*, 703-713.
Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. Science *272*, 830-836.
Gennaro, V.J., Stanek, T.J., Peck, A.R., Sun, Y., Wang, F., Qie, S., Knudsen, K.E., Rui, H., Butt, T., Diehl, J.A.*, et al.* (2018). Control of CCND1 ubiquitylation by the catalytic SAGA subunit USP22 is essential for cell cycle progression through G1 in cancer cells. Proc Natl Acad Sci U S A *115*, E9298-E9307.
Gheldof, A., and Berx, G. (2013). Cadherins and epithelial-to-mesenchymal transition. Prog Mol Biol Transl Sci *116*, 317-336.
Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L.*, et al.* (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity *32*, 317-328.

Gibbons, B.J., Brignole, E.J., Azubel, M., Murakami, K., Voss, N.R., Bushnell, D.A., Asturias, F.J., and Kornberg, R.D. (2012). Subunit architecture of general transcription factor TFIIH. Proc Natl Acad Sci U S A *109*, 1949-1954.
Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Mol Cell *49*, 773-782.
Gilbert, W., and Muller-Hill, B. (1966). Isolation of the lac repressor. Proc Natl Acad Sci U S A *56*, 1891-1898.
Gilfillan, S., Stelzer, G., Piaia, E., Hofmann, M.G., and Meisterernst, M. (2005). Efficient binding of NC2.TATA-binding protein to DNA in the absence of TATA. J Biol Chem *280*, 6222-6230.
Gilmour, D.S., and Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol Cell Biol *6*, 3984-3989.
Glatt, S., Alfieri, C., and Muller, C.W. (2011). Recognizing and remodeling the nucleosome. Curr Opin Struct Biol *21*, 335-341.
Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Mol Cell Biol *29*, 5455-5464.
Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. Science *300*, 980-984.
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.
Gonzalez-Aguilera, C., Tous, C., Gomez-Gonzalez, B., Huertas, P., Luna, R., and Aguilera, A. (2008). The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. Mol Biol Cell *19*, 4310-4318.
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.
Gout, J.F., Li, W., Fritsch, C., Li, A., Haroon, S., Singh, L., Hua, D., Fazelinia, H., Smith, Z., Seeholzer, S.*, et al.* (2017). The landscape of transcription errors in eukaryotic cells. Sci Adv *3*, e1701484.
Grant, P.A., Duggan, L., Cote, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F.*, et al.* (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev *11*, 1640-1650.
Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R.H., Michaelis, J., and Werner, F. (2011). The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. Mol Cell *43*, 263-274.
Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu Rev Biochem *57*, 159-197.
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.
Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W.*, et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948-951.
Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063-1069.
Guo, S., Yamaguchi, Y., Schilbach, S., Wada, T., Lee, J., Goddard, A., French, D., Handa, H., and Rosenthal, A. (2000). A regulator of transcriptional elongation controls vertebrate neuronal development. Nature *408*, 366-369.
Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yanez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B.*, et al.* (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693-707 e614.
Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol *19*, 621-637.
Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C.*, et al.* (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360-369.
Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol *11*, 394-403.
Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705-736.
Hake, S.B., Garcia, B.A., Kauer, M., Baker, S.P., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2005). Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. Proc Natl Acad Sci U S A *102*, 6344-6349.
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.
Han, Y., Luo, J., Ranish, J., and Hahn, S. (2014). Architecture of the Saccharomyces cerevisiae SAGA transcription coactivator complex. EMBO J *33*, 2534-2546.
Hanna, J.H., Saha, K., and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell *143*, 508-525.
Harper, T.M., and Taatjes, D.J. (2018). The complex structure and function of Mediator. J Biol Chem *293*, 13778-13785.
Harvey, R.P. (2002). Patterning the vertebrate heart. Nat Rev Genet *3*, 544-556.
Hassan, A.H., Awad, S., Al-Natour, Z., Othman, S., Mustafa, F., and Rizvi, T.A. (2007). Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. Biochem J *402*, 125-133.
Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., and Workman, J.L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369-379.
Hatzis, P., and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol Cell *10*, 1467-1477.
He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. Nature *495*, 481-486.
He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., and Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. Nature *533*, 359-365.
Heard, E. (2004). Recent advances in X-chromosome inactivation. Curr Opin Cell Biol *16*, 247-255.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W.*, et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.
Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.
Helmlinger, D., and Tora, L. (2017). Sharing the SAGA. Trends Biochem Sci *42*, 850-861.
Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B., and Ahmad, K. (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res *19*, 460-469.
Henry, K.W., Wyce, A., Lo, W.S., Duggan, L.J., Emre, N.C., Kao, C.F., Pillus, L., Shilatifard, A., Osley, M.A., and Berger, S.L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev *17*, 2648-2663.
Herbig, E., Warfield, L., Fish, L., Fishburn, J., Knutson, B.A., Moorefield, B., Pacheco, D., and Hahn, S. (2010). Mechanism of Mediator recruitment by tandem Gcn4 activation domains and three Gal11 activator-binding domains. Mol Cell Biol *30*, 2376-2390.
Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13-23.
Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. Nature *463*, 474-484.
Hodges, A.J., Gloss, L.M., and Wyrick, J.J. (2017). Residues in the Nucleosome Acidic Patch Regulate Histone Occupancy and Are Important for FACT Binding in Saccharomyces cerevisiae. Genetics *206*, 1339-1348.
Hofmann, M., Schuster-Gossler, K., Watabe-Rudolph, M., Aulehla, A., Herrmann, B.G., and Gossler, A. (2004). WNT signaling, in synergy with T/TBX6, controls Notch signaling by regulating Dll1 expression in the presomitic mesoderm of mouse embryos. Genes Dev *18*, 2712-2717.
Hogan, B.L. (1996). Bone morphogenetic proteins in development. Curr Opin Genet Dev *6*, 432-438.
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.
Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E.T., Nijmeijer, B., Kotthoff, C., Rybin, V., Amlacher, S.*, et al.* (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. Nature *499*, 111-114.
Howe, F.S., Fischl, H., Murray, S.C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? Bioessays *39*, 1-12.
Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev *26*, 2119-2137.
Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C., and Zhao, K. (2013). 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.
Huang, R.Y., Guilford, P., and Thiery, J.P. (2012). Early events in cell adhesion and polarity during epithelial-mesenchymal transition. J Cell Sci *125*, 4417-4422.
Huang, S., Li, X., Yusufzai, T.M., Qiu, Y., and Felsenfeld, G. (2007). USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. Mol Cell Biol *27*, 7991-8002.

Huff, J.T., Plocik, A.M., Guthrie, C., and Yamamoto, K.R. (2010). Reciprocal intronic and exonic histone modification regions in humans. Nat Struct Mol Biol *17*, 1495-1499.
Hwang, W.W., Venkatasubrahmanyam, S., Ianculescu, A.G., Tong, A., Boone, C., and Madhani, H.D. (2003). A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. Mol Cell *11*, 261-266.
Hyman, A.A., Weber, C.A., and Julicher, F. (2014). Liquid-liquid phase separation in biology. Annu Rev Cell Dev Biol *30*, 39-58.
Ishibashi, T., Dryhurst, D., Rose, K.L., Shabanowitz, J., Hunt, D.F., and Ausio, J. (2009). Acetylation of vertebrate H2A.Z and its effect on the structure of the nucleosome. Biochemistry *48*, 5007-5017.
Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R., and Kadonaga, J.T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell *90*, 145-155.
Izban, M.G., and Luse, D.S. (1992). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. Genes Dev *6*, 1342-1356.
Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.
Jamai, A., Imoberdorf, R.M., and Strubin, M. (2007). Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. Mol Cell *25*, 345-355.
Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell *19*, 523-534.
Jiang, H., Shukla, A., Wang, X., Chen, W.Y., Bernstein, B.E., and Roeder, R.G. (2011). Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell *144*, 513-525.
Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev *21*, 1519-1529.
Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.
Johnson, D.G., and Dent, S.Y. (2013). Chromatin: receiver and quarterback for cellular signals. Cell *152*, 685-689.
Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407.
Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 167-177.
Jung, I., Kim, S.K., Kim, M., Han, Y.M., Kim, Y.S., Kim, D., and Lee, D. (2012). H2B monoubiquitylation is a 5'-enriched active transcription mark and correlates with exon-intron structure in human cells. Genome Res *22*, 1026-1035.
Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip Rev Dev Biol *1*, 40-51.
Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kalashnikova, A.A., Porter-Goff, M.E., Muthurajan, U.M., Luger, K., and Hansen, J.C. (2013). The role of the nucleosome acidic patch in modulating higher order chromatin structure. J R Soc Interface *10*, 20121022.

Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R.G., Meisterernst, M., and Burley, S.K. (2001). Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. Cell *106*, 71-81.

Kamenova, I., Mukherjee, P., Conic, S., Mueller, F., El-Saafin, F., Bardot, P., Garnier, J.M., Dembele, D., Capponi, S., Timmers, H.T.M.*, et al.* (2019). Co-translational assembly of mammalian nuclear multisubunit complexes. Nat Commun *10*, 1740.

Kamiya, D., Banno, S., Sasai, N., Ohgushi, M., Inomata, H., Watanabe, K., Kawada, M., Yakura, R., Kiyonari, H., Nakao, K.*, et al.* (2011). Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. Nature *470*, 503-509.

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

Kaplan, C.D., Larsson, K.M., and Kornberg, R.D. (2008). The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. Mol Cell *30*, 547-556.

Karpiuk, O., Najafova, Z., Kramer, F., Hennion, M., Galonska, C., Konig, A., Snaidero, N., Vogel, T., Shchebet, A., Begus-Nahrmann, Y.*, et al.* (2012). The histone H2B monoubiquitination regulatory pathway is required for differentiation of multipotent stem cells. Mol Cell *46*, 705-713.

Kassabov, S.R., Zhang, B., Persinger, J., and Bartholomew, B. (2003). SWI/SNF unwraps, slides, and rewraps the nucleosome. Mol Cell *11*, 391-403.

Kattman, S.J., Adler, E.D., and Keller, G.M. (2007). Specification of multipotential cardiovascular progenitor cells during embryonic stem cell differentiation and embryonic development. Trends Cardiovasc Med *17*, 240-246.

Kattman, S.J., Huber, T.L., and Keller, G.M. (2006). Multipotent flk-1+ cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. Dev Cell *11*, 723-732.

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.

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.

Kettenberger, H., Armache, K.J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell *114*, 347-357.

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

Kim, J.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, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.

Kim, W., Bennett, E.J., Huttlin, E.L., Guo, A., Li, J., Possemato, A., Sowa, M.E., Rad, R., Rush, J., Comb, M.J.*, et al.* (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell *44*, 325-340.

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.

Kimura, H., and Cook, P.R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J Cell Biol *153*, 1341-1353.

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

Kitajima, S., Takagi, A., Inoue, T., and Saga, Y. (2000). MesP1 and MesP2 are essential for the development of cardiac mesoderm. Development *127*, 3215-3226.

Klein, B.J., Bose, D., Baker, K.J., Yusoff, Z.M., Zhang, X., and Murakami, K.S. (2011). RNA polymerase and transcription elongation factor Spt4/5 complex structure. Proc Natl Acad Sci U S A *108*, 546-550.

Kobayashi, T., Iwamoto, Y., Takashima, K., Isomura, A., Kosodo, Y., Kawakami, K., Nishioka, T., Kaibuchi, K., and Kageyama, R. (2015). Deubiquitinating enzymes regulate Hes1 stability and neuronal differentiation. FEBS J *282*, 2411-2423.

Kohler, A., Schneider, M., Cabal, G.G., Nehrbass, U., and Hurt, E. (2008). Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. Nat Cell Biol *10*, 707-715.

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

Kolesnikova, O., Ben-Shem, A., Luo, J., Ranish, J., Schultz, P., and Papai, G. (2018). Molecular structure of promoter-bound yeast TFIID. Nat Commun *9*, 4666.

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.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. Trends Biochem Sci *30*, 235-239.

Kosa, P.F., Ghosh, G., DeDecker, B.S., and Sigler, P.B. (1997). The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. Proc Natl Acad Sci U S A *94*, 6042-6047.

Kostek, S.A., Grob, P., De Carlo, S., Lipscomb, J.S., Garczarek, F., and Nogales, E. (2006). Molecular architecture and conformational flexibility of human RNA polymerase II. Structure *14*, 1691-1700.

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.

Koutelou, E., Wang, L., Schibler, A.C., Chao, H.P., Kuang, X., Lin, K., Lu, Y., Shen, J., Jeter, C.R., Salinger, A.*, et al.* (2019). USP22 controls multiple signaling pathways that are essential for vasculature formation in the mouse placenta. Development *146*.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schubeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Mol Cell *67*, 411-422 e414.

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

Ku, M., Jaffe, J.D., Koche, R.P., Rheinbay, E., Endoh, M., Koseki, H., Carr, S.A., and Bernstein, B.E. (2012). H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. Genome Biol *13*, R85.

Kulaeva, O.I., Gaykalova, D.A., and Studitsky, V.M. (2007). Transcription through chromatin by RNA polymerase II: histone displacement and exchange. Mutat Res *618*, 116-129.

Kurokawa, D., Takasaki, N., Kiyonari, H., Nakayama, R., Kimura-Yoshida, C., Matsuo, I., and Aizawa, S. (2004). Regulation of Otx2 expression and its functions in mouse epiblast and anterior neuroectoderm. Development *131*, 3307-3317.

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., and Lis, J.T. (2013). Control of transcriptional elongation. Annu Rev Genet *47*, 483-508.

Laganiere, J., Deblois, G., Lefebvre, C., Bataille, A.R., Robert, F., and Giguere, V. (2005). From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. Proc Natl Acad Sci U S A *102*, 11651-11656.

Lagrange, T., Kim, T.K., Orphanides, G., Ebright, Y.W., Ebright, R.H., and Reinberg, D. (1996). High-resolution mapping of nucleoprotein complexes by site-specific protein-DNA photocrosslinking: organization of the human TBP-TFIIA-TFIIB-DNA quaternary complex. Proc Natl Acad Sci U S A *93*, 10620-10625.

Lambies, G., Miceli, M., Martinez-Guillamon, C., Olivera-Salguero, R., Pena, R., Frias, C.P., Calderon, I., Atanassov, B.S., Dent, S.Y.R., Arribas, J.*, et al.* (2019). TGFbeta-Activated USP27X Deubiquitinase Regulates Cell Migration and Chemoresistance via Stabilization of Snail1. Cancer Res *79*, 33-46.

Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol *15*, 178-196.

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

Laribee, R.N., Fuchs, S.M., and Strahl, B.D. (2007). H2B ubiquitylation in transcriptional control: a FACT-finding mission. Genes Dev *21*, 737-743.

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

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature *547*, 236-240.

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell *152*, 1021-1036.

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

Lawson, K.A. (1999). Fate mapping the mouse embryo. Int J Dev Biol *43*, 773-775.
Lee, C.S., Friedman, J.R., Fulmer, J.T., and Kaestner, K.H. (2005a). The initiation of liver development is dependent on Foxa transcription factors. Nature *435*, 944-947.
Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005b). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Mol Cell Biol *25*, 9674-9686.
Lee, J.E., Wang, C., Xu, S., Cho, Y.W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W*., et al.* (2013). H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife *2*, e01503.
Lee, J.S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R., and Shilatifard, A. (2007). Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell *131*, 1084-1096.
Lee, K.K., Sardiu, M.E., Swanson, S.K., Gilmore, J.M., Torok, M., Grant, P.A., Florens, L., Workman, J.L., and Washburn, M.P. (2011). Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. Mol Syst Biol *7*, 503.
Lee, M.T., Bonneau, A.R., and Giraldez, A.J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. Annu Rev Cell Dev Biol *30*, 581-613.
Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K*., et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.
Lempiainen, H., and Halazonetis, T.D. (2009). Emerging common themes in regulation of PIKKs and PI3Ks. EMBO J *28*, 3067-3073.
Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet *13*, 233-245.
Lepoivre, C., Belhocine, M., Bergon, A., Griffon, A., Yammine, M., Vanhille, L., Zacarias-Cabeza, J., Garibal, M.A., Koch, F., Maqbool, M.A*., et al.* (2013). Divergent transcription is associated with promoters of transcriptional regulators. BMC Genomics *14*, 914.
Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr Biol *20*, R754-763.
Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. Cell *145*, 502-511.
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, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.
Li, C., Irrazabal, T., So, C.C., Berru, M., Du, L., Lam, E., Ling, A.K., Gommerman, J.L., Pan-Hammarstrom, Q., and Martin, A. (2018). The H2B deubiquitinase Usp22 promotes antibody class switch recombination by facilitating non-homologous end joining. Nat Commun *9*, 1006.
Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005). Rapid spontaneous accessibility of nucleosomal DNA. Nat Struct Mol Biol *12*, 46-53.
Li, S., Edgar, D., Fassler, R., Wadsworth, W., and Yurchenco, P.D. (2003). The role of laminin in embryonic cell polarization and tissue organization. Dev Cell *4*, 613-624.

Li, W., Atanassov, B.S., Lan, X., Mohan, R.D., Swanson, S.K., Farria, A.T., Florens, L., Washburn, M.P., Workman, J.L., and Dent, S.Y.R. (2016a). Cytoplasmic ATXN7L3B Interferes with Nuclear Functions of the SAGA Deubiquitinase Module. Mol Cell Biol *36*, 2855-2866.
Li, W., Notani, D., and Rosenfeld, M.G. (2016b). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet *17*, 207-223.
Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.
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.
Lin, Z., Tan, C., Qiu, Q., Kong, S., Yang, H., Zhao, F., Liu, Z., Li, J., Kong, Q., Gao, B.*, et al.* (2015). Ubiquitin-specific protease 22 is a deubiquitinase of CCNB1. Cell Discov *1*.
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.
Lindsley, R.C., Gill, J.G., Kyba, M., Murphy, T.L., and Murphy, K.M. (2006). Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. Development *133*, 3787-3796.
Lindsley, R.C., Gill, J.G., Murphy, T.L., Langer, E.M., Cai, M., Mashayekhi, M., Wang, W., Niwa, N., Nerbonne, J.M., Kyba, M.*, et al.* (2008). Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. Cell Stem Cell *3*, 55-68.
Liu, P., Kenney, J.M., Stiller, J.W., and Greenleaf, A.L. (2010). Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. Mol Biol Evol *27*, 2628-2641.
Liu, Y., Asakura, M., Inoue, H., Nakamura, T., Sano, M., Niu, Z., Chen, M., Schwartz, R.J., and Schneider, M.D. (2007). Sox17 is essential for the specification of cardiac mesoderm in embryonic stem cells. Proc Natl Acad Sci U S A *104*, 3859-3864.
Louder, R.K., He, Y., Lopez-Blanco, J.R., Fang, J., Chacon, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature *531*, 604-609.
Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., and Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. Nature *558*, 318-323.
Lu, H., Zawel, L., Fisher, L., Egly, J.M., and Reinberg, D. (1992). Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. Nature *358*, 641-645.
Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.
Luk, E., Vu, N.D., Patteson, K., Mizuguchi, G., Wu, W.H., Ranjan, A., Backus, J., Sen, S., Lewis, M., Bai, Y.*, et al.* (2007). Chz1, a nuclear chaperone for histone H2AZ. Mol Cell *25*, 357-368.
Luo, Z., Lin, C., Guest, E., Garrett, A.S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M.P., and Shilatifard, A. (2012). The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. Mol Cell Biol *32*, 2608-2617.
Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet *11*, 761-772.
Mao, P., Meas, R., Dorgan, K.M., and Smerdon, M.J. (2014). UV damage-induced RNA polymerase II stalling stimulates H2B deubiquitylation. Proc Natl Acad Sci U S A *111*, 12811-12816.

Marciano, G., and Huang, D.T. (2016). Structure of the human histone chaperone FACT Spt16 N-terminal domain. Acta Crystallogr F Struct Biol Commun *72*, 121-128.
Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A.*, et al.* (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590-604.
Martello, G., and Smith, A. (2014). The nature of embryonic stem cells. Annu Rev Cell Dev Biol *30*, 647-675.
Martin-Puig, S., Wang, Z., and Chien, K.R. (2008). Lives of a heart cell: tracing the origins of cardiac progenitors. Cell Stem Cell *2*, 320-331.
Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A *78*, 7634-7638.
Martinez, E., Palhan, V.B., Tjernberg, A., Lymar, E.S., Gamper, A.M., Kundu, T.K., Chait, B.T., and Roeder, R.G. (2001). Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. Mol Cell Biol *21*, 6782-6795.
Martire, S., Gogate, A.A., Whitmill, A., Tafessu, A., Nguyen, J., Teng, Y.C., Tastemel, M., and Banaszynski, L.A. (2019). Phosphorylation of histone H3.3 at serine 31 promotes p300 activity and enhancer acetylation. Nat Genet *51*, 941-946.
Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The human and mouse replication-dependent histone genes. Genomics *80*, 487-498.
Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet *7*, 29-59.
Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A.*, et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol *9*, 625-635.
Materne, P., Vazquez, E., Sanchez, M., Yague-Sanz, C., Anandhakumar, J., Migeot, V., Antequera, F., and Hermand, D. (2016). Histone H2B ubiquitylation represses gametogenesis by opposing RSC-dependent chromatin remodeling at the ste11 master regulator locus. Elife *5*.
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.
Mazurkiewicz, J., Kepert, J.F., and Rippe, K. (2006). On the mechanism of nucleosome assembly by histone chaperone NAP1. J Biol Chem *281*, 16462-16472.
McGinty, R.K., Henrici, R.C., and Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. Nature *514*, 591-596.
McGinty, R.K., and Tan, S. (2016). Recognition of the nucleosome by chromatin factors and enzymes. Curr Opin Struct Biol *37*, 54-61.
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.
Meriesh, H.A., Lerner, A.M., Chandrasekharan, M.B., and Strahl, B.D. (2020). The histone H4 basic patch regulates SAGA-mediated H2B deubiquitination and histone acetylation. J Biol Chem *295*, 6561-6569.
Mesnard, D., Guzman-Ayala, M., and Constam, D.B. (2006). Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. Development *133*, 2497-2505.

Milligan, L., Sayou, C., Tuck, A., Auchynnikava, T., Reid, J.E., Alexander, R., Alves, F.L., Allshire, R., Spanos, C., Rappsilber, J.*, et al.* (2017). RNA polymerase II stalling at pre-mRNA splice sites is enforced by ubiquitination of the catalytic subunit. Elife *6*.
Minsky, N., Shema, E., Field, Y., Schuster, M., Segal, E., and Oren, M. (2008). Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nat Cell Biol *10*, 483-488.
Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371-378.
Mohn, F., and Schubeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet *25*, 129-136.
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.
Moretti, A., Caron, L., Nakano, A., Lam, J.T., Bernshausen, A., Chen, Y., Qyang, Y., Bu, L., Sasaki, M., Martin-Puig, S.*, et al.* (2006). Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. Cell *127*, 1151-1165.
Morgan, M.T., Haj-Yahya, M., Ringel, A.E., Bandi, P., Brik, A., and Wolberger, C. (2016). Structural basis for histone H2B deubiquitination by the SAGA DUB module. Science *351*, 725-728.
Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., and Washburn, M.P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Mol Cell *34*, 168-178.
Murakami, K., Gibbons, B.J., Davis, R.E., Nagai, S., Liu, X., Robinson, P.J., Wu, T., Kaplan, C.D., and Kornberg, R.D. (2012). Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation. Proc Natl Acad Sci U S A *109*, 4816-4821.
Murawska, M., Schauer, T., Matsuda, A., Wilson, M.D., Pysik, T., Wojcik, F., Muir, T.W., Hiraoka, Y., Straub, T., and Ladurner, A.G. (2020). The Chaperone FACT and Histone H2B Ubiquitination Maintain S. pombe Genome Architecture through Genic and Subtelomeric Functions. Mol Cell *77*, 501-513 e507.
Murr, R., Vaissiere, T., Sawan, C., Shukla, V., and Herceg, Z. (2007). Orchestration of chromatin-based processes: mind the TRRAP. Oncogene *26*, 5358-5372.
Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell *132*, 661-680.
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.
Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. Proc Natl Acad Sci U S A *90*, 8424-8428.
Nagy, Z., Riss, A., Romier, C., le Guezennec, X., Dongre, A.R., Orpinell, M., Han, J., Stunnenberg, H., and Tora, L. (2009). The human SPT20-containing SAGA complex plays a direct role in the regulation of endoplasmic reticulum stress-induced genes. Mol Cell Biol *29*, 1649-1660.
Nagy, Z., and Tora, L. (2007). Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. Oncogene *26*, 5341-5357.
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.

Naito, A.T., Shiojima, I., Akazawa, H., Hidaka, K., Morisaki, T., Kikuchi, A., and Komuro, I. (2006). Developmental stage-specific biphasic roles of Wnt/beta-catenin signaling in cardiomyogenesis and hematopoiesis. Proc Natl Acad Sci U S A *103*, 19812-19817.
Nakazawa, Y., Hara, Y., Oka, Y., Komine, O., van den Heuvel, D., Guo, C., Daigaku, Y., Isono, M., He, Y., Shimada, M.*, et al.* (2020). Ubiquitination of DNA Damage-Stalled RNAPII Promotes Transcription-Coupled Repair. Cell *180*, 1228-1244 e1224.
Narita, N., Bielinska, M., and Wilson, D.B. (1997). Wild-type endoderm abrogates the ventral developmental defects associated with GATA-4 deficiency in the mouse. Dev Biol *189*, 270-274.
Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D.K., Hasegawa, J., Omori, M., Inukai, N.*, et al.* (2003). Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. Mol Cell Biol *23*, 1863-1873.
Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell *154*, 490-503.
Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annu Rev Genet *46*, 1-19.
Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335-338.
Nekrasov, M., Amrichova, J., Parker, B.J., Soboleva, T.A., Jack, C., Williams, R., Huttley, G.A., and Tremethick, D.J. (2012). Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics. Nat Struct Mol Biol *19*, 1076-1083.
Ng, H.H., Dole, S., and Struhl, K. (2003). The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. J Biol Chem *278*, 33625-33628.
Ng, H.H., Xu, R.M., Zhang, Y., and Struhl, K. (2002). Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. J Biol Chem *277*, 34655-34657.
Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature *414*, 322-325.
Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature *377*, 119-128.
Nishimura, K., and Kanemaki, M.T. (2014). Rapid Depletion of Budding Yeast Proteins via the Fusion of an Auxin-Inducible Degron (AID). Curr Protoc Cell Biol *64*, 20 29 21-16.
Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev *12*, 2048-2060.
Nogales, E., Louder, R.K., and He, Y. (2017). Structural Insights into the Eukaryotic Transcription Initiation Machinery. Annu Rev Biophys *46*, 59-83.
Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944 e922.
Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.*, et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.
Norris, D.P., Brennan, J., Bikoff, E.K., and Robertson, E.J. (2002). The Foxh1-dependent autoregulatory enhancer controls the level of Nodal signals in the mouse embryo. Development *129*, 3455-3468.

Nune, M., Morgan, M.T., Connell, Z., McCullough, L., Jbara, M., Sun, H., Brik, A., Formosa, T., and Wolberger, C. (2019). FACT and Ubp10 collaborate to modulate H2B deubiquitination and nucleosome dynamics. Elife *8*.

Ogbourne, S., and Antalis, T.M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem J *331 (Pt 1)*, 1-14.

Ohkuma, Y., Sumimoto, H., Horikoshi, M., and Roeder, R.G. (1990). Factors involved in specific transcription by mammalian RNA polymerase II: purification and characterization of general transcription factor TFIIE. Proc Natl Acad Sci U S A *87*, 9163-9167.

Okamoto, T., Yamamoto, S., Watanabe, Y., Ohta, T., Hanaoka, F., Roeder, R.G., and Ohkuma, Y. (1998). Analysis of the role of TFIIE in transcriptional regulation through structure-function studies of the TFIIEbeta subunit. J Biol Chem *273*, 19866-19876.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet *15*, 234-246.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev *10*, 2657-2683.

Owen-Hughes, T., Utley, R.T., Cote, J., Peterson, C.L., and Workman, J.L. (1996). Persistent site-specific remodeling of a nucleosome array by transient action of the SWI/SNF complex. Science *273*, 513-516.

Pandey, N.B., Chodchoy, N., Liu, T.J., and Marzluff, W.F. (1990). Introns in histone genes alter the distribution of 3' ends. Nucleic Acids Res *18*, 3161-3170.

Pankotai, T., Komonyi, O., Bodai, L., Ujfaludi, Z., Muratoglu, S., Ciurciu, A., Tora, L., Szabad, J., and Boros, I. (2005). The homologous Drosophila transcriptional adaptors ADA2a and ADA2b are both required for normal development but have different functions. Mol Cell Biol *25*, 8215-8227.

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

Parua, P.K., Booth, G.T., Sanso, M., Benjamin, B., Tanny, J.C., Lis, J.T., and Fisher, R.P. (2018). A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. Nature *558*, 460-464.

Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N.*, et al.* (2010). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Res *38*, 4958-4969.

Patel, A.B., Louder, R.K., Greber, B.J., Grunberg, S., Luo, J., Fang, J., Liu, Y., Ranish, J., Hahn, S., and Nogales, E. (2018). Structure of human TFIID and mechanism of TBP loading onto promoter DNA. Science *362*.

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.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell *125*, 703-717.

Pei, Y., and Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. J Biol Chem *277*, 19639-19648.

Peinado, H., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat Rev Cancer *7*, 415-428.

Perry, P., Sauer, S., Billon, N., Richardson, W.D., Spivakov, M., Warnes, G., Livesey, F.J., Merkenschlager, M., Fisher, A.G., and Azuara, V. (2004). A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction. Cell Cycle *3*, 1645-1650.
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.
Piunti, A., and Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. Science *352*, aad9780.
Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. Nature *533*, 353-358.
Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F.*, et al.* (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature *518*, 376-380.
Poleshko, A., Shah, P.P., Gupta, M., Babu, A., Morley, M.P., Manderfield, L.J., Ifkovits, J.L., Calderon, D., Aghajanian, H., Sierra-Pagan, J.E.*, et al.* (2017). Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. Cell *171*, 573-587 e514.
Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. Crit Rev Biochem Mol Biol *48*, 575-608.
Prokocimer, M., Davidovich, M., Nissim-Rafinia, M., Wiesel-Motiuk, N., Bar, D.Z., Barkan, R., Meshorer, E., and Gruenbaum, Y. (2009). Nuclear lamins: key regulators of nuclear structure and activities. J Cell Mol Med *13*, 1059-1085.
Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.
Ptashne, M. (1967). Specific binding of the lambda phage repressor to lambda DNA. Nature *214*, 232-234.
Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.
Qi, D., Larsson, J., and Mannervik, M. (2004). Drosophila Ada2b is required for viability and normal histone H3 acetylation. Mol Cell Biol *24*, 8080-8089.
Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y.*, et al.* (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell *174*, 744-757 e724.
Rach, E.A., Winter, D.R., Benjamin, A.M., Corcoran, D.L., Ni, T., Zhu, J., and Ohler, U. (2011). Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. PLoS Genet *7*, e1001274.
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.
Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.
Ramachandran, S., Haddad, D., Li, C., Le, M.X., Ling, A.K., So, C.C., Nepal, R.M., Gommerman, J.L., Yu, K., Ketela, T.*, et al.* (2016). The SAGA Deubiquitination Module Promotes DNA Repair and Class Switch Recombination through ATM and DNAPK-Mediated gammaH2AX Formation. Cell Rep *15*, 1554-1565.
Ramkumar, N., and Anderson, K.V. (2011). SnapShot: mouse primitive streak. Cell *146*, 488-488 e482.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.
Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A *90*, 7923-7927.
Rasmussen, E.B., and Lis, J.T. (1995). Short transcripts of the ternary complex provide insight into RNA polymerase II elongational pausing. J Mol Biol *252*, 522-535.
Rasmussen, T.P., Huang, T., Mastrangelo, M.A., Loring, J., Panning, B., and Jaenisch, R. (1999). Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing. Nucleic Acids Res *27*, 3685-3689.
Ravarani, C.N.J., Flock, T., Chavali, S., Anandapadamanaban, M., Babu, M.M., and Balaji, S. (2020). Molecular determinants underlying functional innovations of TBP and their impact on transcription initiation. Nat Commun *11*, 2384.
Redkar, A., Montgomery, M., and Litvin, J. (2001). Fate map of early avian cardiac progenitor cells. Development *128*, 2269-2279.
Reeves, W.M., and Hahn, S. (2005). Targets of the Gal4 transcription activator in functional transcription complexes. Mol Cell Biol *25*, 9092-9102.
Reinke, H., and Horz, W. (2003). Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. Mol Cell *11*, 1599-1607.
Rivera-Perez, J.A., and Hadjantonakis, A.K. (2014). The Dynamics of Morphogenesis in the Early Mouse Embryo. Cold Spring Harb Perspect Biol *7*.
Robinson, P.J., Trnka, M.J., Pellarin, R., Greenberg, C.H., Bushnell, D.A., Davis, R., Burlingame, A.L., Sali, A., and Kornberg, R.D. (2015). Molecular architecture of the yeast Mediator complex. Elife *4*.
Rodriguez-Navarro, S., Fischer, T., Luo, M.J., Antunez, O., Brettschneider, S., Lechner, J., Perez-Ortin, J.E., Reed, R., and Hurt, E. (2004). Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. Cell *116*, 75-86.
Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci *21*, 327-335.
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.
Rogakou, E.P., Boon, C., Redon, C., and Bonner, W.M. (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. J Cell Biol *146*, 905-916.
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.
Roider, H.G., Lenhard, B., Kanhere, A., Haas, S.A., and Vingron, M. (2009). CpG-depleted promoters harbor tissue-specific transcription factor binding signals--implications for motif overrepresentation analyses. Nucleic Acids Res *37*, 6305-6315.
Rosado-Lugo, J.D., and Hampsey, M. (2014). The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition. J Biol Chem *289*, 33916-33926.
Rossignol, M., Kolb-Cheynel, I., and Egly, J.M. (1997). Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH. EMBO J *16*, 1628-1637.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795-804.
Roy, R., Adamczewski, J.P., Seroz, T., Vermeulen, W., Tassan, J.P., Schaeffer, L., Nigg, E.A., Hoeijmakers, J.H., and Egly, J.M. (1994). The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. Cell *79*, 1093-1101.
Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell *25*, 15-30.
Saga, Y., Hata, N., Kobayashi, S., Magnuson, T., Seldin, M.F., and Taketo, M.M. (1996). MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. Development *122*, 2769-2778.
Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J., and Inoue, T. (1999). MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Development *126*, 3437-3447.
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.
Samara, N.L., Datta, A.B., Berndsen, C.E., Zhang, X., Yao, T., Cohen, R.E., and Wolberger, C. (2010). Structural insights into the assembly and function of the SAGA deubiquitinating module. Science *328*, 1025-1029.
Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J.*, et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A *112*, E6456-6465.
Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109-113.
Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G.P., Heron, M., Williams, H., Soding, J., Stewart, A., and Svejstrup, J.Q. (2014). RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. Cell *157*, 1037-1049.
Sarcinella, E., Zuzarte, P.C., Lau, P.N., Draker, R., and Cheung, P. (2007). Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. Mol Cell Biol *27*, 6457-6468.
Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med *10*, 55-63.
Sato, N., Sanjuan, I.M., Heke, M., Uchida, M., Naef, F., and Brivanlou, A.H. (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. Dev Biol *260*, 404-413.
Sawada, S., Scarborough, J.D., Killeen, N., and Littman, D.R. (1994). A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. Cell *77*, 917-929.
Schier, A.C., and Taatjes, D.J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. Genes Dev *34*, 465-488.
Schier, A.F. (2003). Nodal signaling in vertebrate development. Annu Rev Cell Dev Biol *19*, 589-621.
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.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S.*, et al.* (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet *42*, 53-61.
Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.
Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L.*, et al.* (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51-56.
Segala, G., Bennesch, M.A., Pandey, D.P., Hulo, N., and Picard, D. (2016). Monoubiquitination of Histone H2B Blocks Eviction of Histone Variant H2A.Z from Inducible Enhancers. Mol Cell *64*, 334-346.
Sellam, A., Askew, C., Epp, E., Lavoie, H., Whiteway, M., and Nantel, A. (2009). Genome-wide mapping of the coactivator Ada2p yields insight into the functional roles of SAGA/ADA complex in Candida albicans. Mol Biol Cell *20*, 2389-2400.
Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.
Shema, E., Tirosh, I., Aylon, Y., Huang, J., Ye, C., Moskovits, N., Raver-Shapira, N., Minsky, N., Pirngruber, J., Tarcic, G.*, et al.* (2008). The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. Genes Dev *22*, 2664-2676.
Sheridan, R.M., Fong, N., D'Alessandro, A., and Bentley, D.L. (2019). Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5' Pause Release, Termination, and Transcription Elongation Rate. Mol Cell *73*, 107-118 e104.
Shi, L., Wen, H., and Shi, X. (2017). The Histone Variant H3.3 in Transcriptional Regulation and Human Disease. J Mol Biol *429*, 1934-1945.
Shin, Y., Chang, Y.C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M., and Brangwynne, C.P. (2019). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. Cell *176*, 1518.
Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T.*, et al.* (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A *100*, 15776-15781.
Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.
Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C.*, et al.* (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc Natl Acad Sci U S A *110*, 2876-2881.
Sikorski, R.S., Boguski, M.S., Goebl, M., and Hieter, P. (1990). A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell *60*, 307-317.
Sim, Y.J., Kim, M.S., Nayfeh, A., Yun, Y.J., Kim, S.J., Park, K.T., Kim, C.H., and Kim, K.S. (2017). 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. Stem Cell Reports *8*, 1312-1328.

Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., and Arndt, K.M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO J *22*, 1846-1856.
Sims, R.J., 3rd, Nishioka, K., and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. Trends Genet *19*, 629-639.
Skrajna, A., Goldfarb, D., Kedziora, K.M., Cousins, E.M., Grant, G.D., Spangler, C.J., Barbour, E.H., Yan, X., Hathaway, N.A., Brown, N.G.*, et al.* (2020). Comprehensive nucleosome interactome screen establishes fundamental principles of nucleosome binding. Nucleic Acids Res *48*, 9415-9432.
Smale, S.T., and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell *57*, 103-113.
Smith, E., and Shilatifard, A. (2010). The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. Mol Cell *40*, 689-701.
Smith, S., and Stillman, B. (1991). Stepwise assembly of chromatin during DNA replication in vitro. EMBO J *10*, 971-980.
Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat Rev Genet *14*, 204-220.
Smolle, M., and Workman, J.L. (2013). Transcription-associated histone modifications and cryptic transcription. Biochim Biophys Acta *1829*, 84-97.
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.
Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet *13*, 613-626.
Srivastava, D. (2006). Making or breaking the heart: from lineage determination to morphogenesis. Cell *126*, 1037-1048.
Stadhouders, R., Filion, G.J., and Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. Nature *569*, 345-354.
Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y.*, et al.* (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat Genet *50*, 238-249.
Stainier, D.Y. (2001). Zebrafish genetics and vertebrate heart formation. Nat Rev Genet *2*, 39-48.
Stefansson, O.A., Hermanowicz, S., van der Horst, J., Hilmarsdottir, H., Staszczak, Z., Jonasson, J.G., Tryggvadottir, L., Gudjonsson, T., and Sigurdsson, S. (2017). CpG promoter methylation of the ALKBH3 alkylation repair gene in breast cancer. BMC Cancer *17*, 469.
Stegeman, R., Spreacker, P.J., Swanson, S.K., Stephenson, R., Florens, L., Washburn, M.P., and Weake, V.M. (2016). The Spliceosomal Protein SF3B5 is a Novel Component of Drosophila SAGA that Functions in Gene Expression Independent of Splicing. J Mol Biol *428*, 3632-3649.
Stevens, J.R., O'Donnell, A.F., Perry, T.E., Benjamin, J.J., Barnes, C.A., Johnston, G.C., and Singer, R.A. (2011). FACT, the Bur kinase pathway, and the histone co-repressor HirC have overlapping nucleosome-related roles in yeast transcription elongation. PLoS One *6*, e25644.
Stewart, C.L., Kaspar, P., Brunet, L.J., Bhatt, H., Gadi, I., Kontgen, F., and Abbondanzo, S.J. (1992). Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. Nature *359*, 76-79.

Stoller, J.Z., Huang, L., Tan, C.C., Huang, F., Zhou, D.D., Yang, J., Gelb, B.D., and Epstein, J.A. (2010). Ash2l interacts with Tbx1 and is required during early embryogenesis. Exp Biol Med (Maywood) *235*, 569-576.

Strahl, B.D., and Briggs, S.D. (2020). The SAGA continues: The rise of cis- and trans-histone crosstalk pathways. Biochim Biophys Acta Gene Regul Mech, 194600.

Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development *132*, 2093-2102.

Suganuma, T., and Workman, J.L. (2011). Signals and combinatorial functions of histone modifications. Annu Rev Biochem *80*, 473-499.

Sumi, T., Oki, S., Kitajima, K., and Meno, C. (2013). Epiblast ground state is controlled by canonical Wnt/beta-catenin signaling in the postimplantation mouse embryo and epiblast stem cells. PLoS One *8*, e63378.

Sun, X., Meyers, E.N., Lewandoski, M., and Martin, G.R. (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. Genes Dev *13*, 1834-1846.

Sundaramoorthy, R., Hughes, A.L., El-Mkami, H., Norman, D.G., Ferreira, H., and Owen-Hughes, T. (2018). Structure of the chromatin remodelling enzyme Chd1 bound to a ubiquitinylated nucleosome. Elife *7*.

Suto, R.K., Clarkson, M.J., Tremethick, D.J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat Struct Biol *7*, 1121-1124.

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.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51-61.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P.*, et al.* (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92-104.

Takaoka, K., and Hamada, H. (2012). Cell fate decisions and axis determination in the early mouse embryo. Development *139*, 3-14.

Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D.M., Lovell-Badge, R., Papaioannou, V.E., and Kondoh, H. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. Nature *470*, 394-398.

Tam, P.P., and Loebel, D.A. (2007). Gene function in mouse embryogenesis: get set for gastrulation. Nat Rev Genet *8*, 368-381.

Tan, S., Hunziker, Y., Sargent, D.F., and Richmond, T.J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. Nature *381*, 127-151.

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.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

Thakar, A., Gupta, P., Ishibashi, T., Finn, R., Silva-Moreno, B., Uchiyama, S., Fukui, K., Tomschik, M., Ausio, J., and Zlatanova, J. (2009). H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. Biochemistry *48*, 10852-10857.

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.
Thiery, J.P., and Sleeman, J.P. (2006). Complex networks orchestrate epithelial-mesenchymal transitions. Nat Rev Mol Cell Biol *7*, 131-142.
Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.
Thompson, C.M., and Young, R.A. (1995). General requirement for RNA polymerase II holoenzymes in vivo. Proc Natl Acad Sci U S A *92*, 4587-4590.
Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131-3141.
Tome, J.M., Tippens, N.D., and Lis, J.T. (2018). Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. Nat Genet *50*, 1533-1541.
Torigoe, S.E., Urwin, D.L., Ishii, H., Smith, D.E., and Kadonaga, J.T. (2011). Identification of a rapidly formed nonnucleosomal histone-DNA intermediate that is converted into chromatin by ACF. Mol Cell *43*, 638-648.
Trowitzsch, S., Viola, C., Scheer, E., Conic, S., Chavant, V., Fournier, M., Papai, G., Ebong, I.O., Schaffitzel, C., Zou, J.*, et al.* (2015). Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. Nat Commun *6*, 6011.
Tsai, K.L., Tomomori-Sato, C., Sato, S., Conaway, R.C., Conaway, J.W., and Asturias, F.J. (2014). Subunit Architecture and Functional Modular Rearrangements of the Transcriptional Mediator Complex. Cell *158*, 463.
Tsai, K.L., Yu, X., Gopalan, S., Chao, T.C., Zhang, Y., Florens, L., Washburn, M.P., Murakami, K., Conaway, R.C., Conaway, J.W.*, et al.* (2017). Mediator structure and rearrangements required for holoenzyme formation. Nature *544*, 196-201.
Tufegdzic Vidakovic, A., Harreman, M., Dirac-Svejstrup, A.B., Boeing, S., Roy, A., Encheva, V., Neumann, M., Wilson, M., Snijders, A.P., and Svejstrup, J.Q. (2019). Analysis of RNA polymerase II ubiquitylation and proteasomal degradation. Methods *159-160*, 146-156.
Tufegdzic Vidakovic, A., Mitter, R., Kelly, G.P., Neumann, M., Harreman, M., Rodriguez-Martinez, M., Herlihy, A., Weems, J.C., Boeing, S., Encheva, V.*, et al.* (2020). Regulation of the RNAPII Pool Is Integral to the DNA Damage Response. Cell *180*, 1245-1261 e1221.
Udeshi, N.D., Mertins, P., Svinkina, T., and Carr, S.A. (2013). Large-scale identification of ubiquitination sites by mass spectrometry. Nat Protoc *8*, 1950-1960.
Ueno, S., Weidinger, G., Osugi, T., Kohn, A.D., Golob, J.L., Pabon, L., Reinecke, H., Moon, R.T., and Murry, C.E. (2007). Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. Proc Natl Acad Sci U S A *104*, 9685-9690.
Umlauf, D., Bonnet, J., Waharte, F., Fournier, M., Stierle, M., Fischer, B., Brino, L., Devys, D., and Tora, L. (2013). The human TREX-2 complex is stably associated with the nuclear pore basket. J Cell Sci *126*, 2656-2667.
Uversky, V.N. (2002). Natively unfolded proteins: a point where biology waits for physics. Protein Sci *11*, 739-756.
Uversky, V.N., Gillespie, J.R., and Fink, A.L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins *41*, 415-427.

van Attikum, H., Fritsch, O., and Gasser, S.M. (2007). Distinct roles for SWR1 and INO80 chromatin remodeling complexes at chromosomal double-strand breaks. EMBO J *26*, 4113-4125.
van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T.*, et al.* (2014). Classification of intrinsically disordered regions and proteins. Chem Rev *114*, 6589-6631.
van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell *169*, 780-791.
Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922-926.
Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res *24*, 896-905.
Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol *16*, 178-189.
Verger, A., Monte, D., and Villeret, V. (2019). Twenty years of Mediator complex structural studies. Biochem Soc Trans *47*, 399-410.
Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G.*, et al.* (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967-980.
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.
Vethantham, V., Yang, Y., Bowman, C., Asp, P., Lee, J.H., Skalnik, D.G., and Dynlacht, B.D. (2012). Dynamic loss of H2B ubiquitylation without corresponding changes in H3K4 trimethylation during myogenic differentiation. Mol Cell Biol *32*, 1044-1055.
Vettese-Dadey, M., Grant, P.A., Hebbes, T.R., Crane- Robinson, C., Allis, C.D., and Workman, J.L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. EMBO J *15*, 2508-2518.
Volpi, E.V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R.D., Goldsworthy, M., Powis, S.H., Ragoussis, J.*, et al.* (2000). Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. J Cell Sci *113 (Pt 9)*, 1565-1576.
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, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet *15*, 69-81.
Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O.*, et al.* (2012). Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. Cell *150*, 1147-1157.

Wang, C., Lee, J.E., Lai, B., Macfarlan, T.S., Xu, S., Zhuang, L., Liu, C., Peng, W., and Ge, K. (2016). Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proc Natl Acad Sci U S A *113*, 11871-11876.
Wang, H., Dienemann, C., Stutzer, A., Urlaub, H., Cheung, A.C.M., and Cramer, P. (2020a). Structure of the transcription coactivator SAGA. Nature *577*, 717-720.
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, L., Koutelou, E., Hirsch, C., McCarthy, R., Schibler, A., Lin, K., Lu, Y., Jeter, C., Shen, J., Barton, M.C.*, et al.* (2018). GCN5 Regulates FGF Signaling and Activates Selective MYC Target Genes during Early Embryoid Body Differentiation. Stem Cell Reports *10*, 287-299.
Wang, Y., Huang, Y., Liu, J., Zhang, J., Xu, M., You, Z., Peng, C., Gong, Z., and Liu, W. (2020b). Acetyltransferase GCN5 regulates autophagy and lysosome biogenesis by targeting TFEB. EMBO Rep *21*, e48335.
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., Dyer, J.O., Seidel, C., Box, A., Swanson, S.K., Peak, A., Florens, L., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2011). Post-transcription initiation function of the ubiquitous SAGA complex in tissue-specific gene activation. Genes Dev *25*, 1499-1509.
Weake, V.M., Lee, K.K., Guelman, S., Lin, C.H., Seidel, C., Abmayr, S.M., and Workman, J.L. (2008). SAGA-mediated H2B deubiquitination controls the development of neuronal connectivity in the Drosophila visual system. EMBO J *27*, 394-405.
Weake, V.M., Swanson, S.K., Mushegian, A., Florens, L., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2009). A novel histone fold domain-containing protein that replaces TAF6 in Drosophila SAGA is required for SAGA-dependent gene expression. Genes Dev *23*, 2818-2823.
Weake, V.M., and Workman, J.L. (2010). Inducible gene expression: diverse regulatory mechanisms. Nat Rev Genet *11*, 426-437.
Weber, A., Heinlein, M., Dengjel, J., Alber, C., Singh, P.K., and Hacker, G. (2016). The deubiquitinase Usp27x stabilizes the BH3-only protein Bim and enhances apoptosis. EMBO Rep *17*, 724-738.
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.
Weil, P.A., and Blatti, S.P. (1976). HeLa cell deoxyribonucleic acid dependent RNA polymerases: function and properties of the class III enzymes. Biochemistry *15*, 1500-1509.
Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Biol *17*, 155-169.
Weinmann, R., Raskas, H.J., and Roeder, R.G. (1974). Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. Proc Natl Acad Sci U S A *71*, 3426-3439.
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.
Welstead, G.G., Schorderet, P., and Boyer, L.A. (2008). The reprogramming language of pluripotency. Curr Opin Genet Dev *18*, 123-129.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684-687.

Williams, S.K., Truong, D., and Tyler, J.K. (2008). Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. Proc Natl Acad Sci U S A *105*, 9000-9005.

Wilmes, G.M., Bergkessel, M., Bandyopadhyay, S., Shales, M., Braberg, H., Cagney, G., Collins, S.R., Whitworth, G.B., Kress, T.L., Weissman, J.S.*, et al.* (2008). A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. Mol Cell *32*, 735-746.

Wilson, M.D., Harreman, M., and Svejstrup, J.Q. (2013). Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. Biochim Biophys Acta *1829*, 151-157.

Wojcik, F., Dann, G.P., Beh, L.Y., Debelouchina, G.T., Hofmann, R., and Muir, T.W. (2018). Functional crosstalk between histone H2B ubiquitylation and H2A modifications and variants. Nat Commun *9*, 1394.

Wollmann, P., Cui, S., Viswanathan, R., Berninghausen, O., Wells, M.N., Moldt, M., Witte, G., Butryn, A., Wendler, P., Beckmann, R.*, et al.* (2011). Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP. Nature *475*, 403-407.

Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R., and Choo, K.H. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res *20*, 351-360.

Wong, L.H., Ren, H., Williams, E., McGhie, J., Ahn, S., Sim, M., Tam, A., Earle, E., Anderson, M.A., Mann, J.*, et al.* (2009). Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. Genome Res *19*, 404-414.

Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F.*, et al.* (2003a). Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol Cell *11*, 267-274.

Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003b). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. J Biol Chem *278*, 34739-34742.

Worden, E.J., Hoffmann, N.A., Hicks, C.W., and Wolberger, C. (2019). Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. Cell *176*, 1490-1501 e1412.

Worden, E.J., Zhang, X., and Wolberger, C. (2020). Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome. Elife *9*.

Woychik, N.A., McKune, K., Lane, W.S., and Young, R.A. (1993). Yeast RNA polymerase II subunit RPB11 is related to a subunit shared by RNA polymerase I and III. Gene Expr *3*, 77-82.

Wu, C.Y., Kang, H.Y., Yang, W.L., Wu, J., Jeong, Y.S., Wang, J., Chan, C.H., Lee, S.W., Zhang, X., Lamothe, B.*, et al.* (2011). Critical role of monoubiquitination of histone H2AX protein in histone H2AX phosphorylation and DNA damage response. J Biol Chem *286*, 30806-30815.

Wu, S.M., Fujiwara, Y., Cibulsky, S.M., Clapham, D.E., Lien, C.L., Schultheiss, T.M., and Orkin, S.H. (2006). Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. Cell *127*, 1137-1150.

Wutz, G., Varnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J.*, et al.* (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J *36*, 3573-3599.

Wyce, A., Xiao, T., Whelan, K.A., Kosman, C., Walter, W., Eick, D., Hughes, T.R., Krogan, N.J., Strahl, B.D., and Berger, S.L. (2007). H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. Mol Cell *27*, 275-288.
Xiao, T., Kao, C.F., Krogan, N.J., Sun, Z.W., Greenblatt, J.F., Osley, M.A., and Strahl, B.D. (2005). Histone H2B ubiquitylation is associated with elongating RNA polymerase II. Mol Cell Biol *25*, 637-651.
Xie, W., Nagarajan, S., Baumgart, S.J., Kosinsky, R.L., Najafova, Z., Kari, V., Hennion, M., Indenbirken, D., Bonn, S., Grundhoff, A.*, et al.* (2017). RNF40 regulates gene expression in an epigenetic context-dependent manner. Genome Biol *18*, 32.
Xu, J., Wang, W., Xu, L., Chen, J.Y., Chong, J., Oh, J., Leschziner, A.E., Fu, X.D., and Wang, D. (2020). Cockayne syndrome B protein acts as an ATP-dependent processivity factor that helps RNA polymerase II overcome nucleosome barriers. Proc Natl Acad Sci U S A.
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, W., Edmondson, D.G., Evrard, Y.A., Wakamiya, M., Behringer, R.R., and Roth, S.Y. (2000). Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. Nat Genet *26*, 229-232.
Xue, H., Yao, T., Cao, M., Zhu, G., Li, Y., Yuan, G., Chen, Y., Lei, M., and Huang, J. (2019). Structural basis of nucleosome recognition and modification by MLL methyltransferases. Nature *573*, 445-449.
Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell *21*, 227-237.
Yamaguchi, T.P. (2001). Heads or tails: Wnts and anterior-posterior patterning. Curr Biol *11*, R713-724.
Yamaguchi, T.P., Takada, S., Yoshikawa, Y., Wu, N., and McMahon, A.P. (1999a). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. Genes Dev *13*, 3185-3190.
Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999b). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41-51.
Yamanaka, Y., Lanner, F., and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development *137*, 715-724.
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., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G.*, et al.* (2020). Guidelines and definitions for research on epithelial-mesenchymal transition. Nat Rev Mol Cell Biol *21*, 341-352.
Yang, J., and Corces, V.G. (2011). Chromatin insulators: a role in nuclear organization and gene expression. Adv Cancer Res *110*, 43-76.
Yang, J., and Weinberg, R.A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. Dev Cell *14*, 818-829.
Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol Cell *19*, 535-545.
Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature *414*, 317-322.

Yella, V.R., and Bansal, M. (2017). DNA structural features of eukaryotic TATA-containing and TATA-less promoters. FEBS Open Bio *7*, 324-334.
Yilmaz, M., and Christofori, G. (2009). EMT, the cytoskeleton, and cancer cell invasion. Cancer Metastasis Rev *28*, 15-33.
Yilmaz, M., and Christofori, G. (2010). Mechanisms of motility in metastasizing cells. Mol Cancer Res *8*, 629-642.
Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.
Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.
Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940-954.
Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., and Roeder, R.G. (2015). RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science *350*, 1383-1386.
Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell Res *21*, 564-578.
Zentner, G.E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. Nat Struct Mol Biol *20*, 259-266.
Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res *21*, 1273-1283.
Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications - writers that read. EMBO Rep *16*, 1467-1481.
Zhang, Y. (2006). It takes a PHD to interpret histone methylation. Nat Struct Mol Biol *13*, 572-574.
Zhang, Y., Cao, R., Wang, L., and Jones, R.S. (2004). Mechanism of Polycomb group gene silencing. Cold Spring Harb Symp Quant Biol *69*, 309-317.
Zhang, Z., Li, J., Ou, Y., Yang, G., Deng, K., Wang, Q., Wang, Z., Wang, W., Zhang, Q., Wang, H.*, et al.* (2020). CDK4/6 inhibition blocks cancer metastasis through a USP51-ZEB1-dependent deubiquitination mechanism. Signal Transduct Target Ther *5*, 25.
Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., Suzuki, E., Le Guezennec, X., Stunnenberg, H.G., Krasnov, A.*, et al.* (2008). A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. Mol Cell *29*, 92-101.
Zheng, H., and Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. Nat Rev Mol Cell Biol *20*, 535-550.
Zhou, Q., Li, T., and Price, D.H. (2012). RNA polymerase II elongation control. Annu Rev Biochem *81*, 119-143.
Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol Cell *29*, 69-80.
Zhou, Z., Zhang, P., Hu, X., Kim, J., Yao, F., Xiao, Z., Zeng, L., Chang, L., Sun, Y., and Ma, L. (2017). USP51 promotes deubiquitination and stabilization of ZEB1. Am J Cancer Res *7*, 2020-2031.

Zhu, B., Zheng, Y., Pham, A.D., Mandal, S.S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2005). Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Mol Cell *20*, 601-611.

Zinn, K., DiMaio, D., and Maniatis, T. (1983). Identification of two distinct regulatory regions adjacent to the human beta-interferon gene. Cell *34*, 865-879.

Zuleger, N., Boyle, S., Kelly, D.A., de las Heras, J.I., Lazou, V., Korfali, N., Batrakou, D.G., Randles, K.N., Morris, G.E., Harrison, D.J.*, et al.* (2013). Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. Genome Biol *14*, R14.

Fang WANG

Le rôle du module de deubiquitination SAGA dans la régulation de la transcription

Résumé en Français

Les coactivateurs régulent l'accessibilité de la chromatine en déposant et retirant des modifications post-traductionnelles des histones.Le complexe coactivateur SAGA (Spt-Ada-Gcn5 acetyltransferase) est conservés chez les eucaryotes et est organisé en modules fonctionnels. Le module de déubiquitination (DUBm) de SAGA est composé de la protéase ubiquitine-spécifique 22 (USP22) et de trois protéines dites « adaptrices », ATXN7, ATXN7L3 et ENY2. L'ensemble des protéines du DUBm sont requises pour le clivage de molécules de mono-ubiquitine sur les histones H2B. Ici, nous étudions le rôle du DUBm de SAGA dans la régulation de la transcription. Nous avons démontré que la protéine ATXN7L3 est essentielle pour le développement embryonnaire. Pour avoir un meilleur aperçu de la fonction d'ATXN7L3, nous avons effectué des expériences de différentiation *in vitro* de cellules souches embryonnaires de souris (mESC) en l'absence d'ATXN7L3. Etonnement, nous avons observé qu'ATXN7L3 promeut la différenciation des mESC en cardiomyocytes, mais pas en précurseurs de l'ectoderme neural. De ce fait, ATNX7L3 pourrait fonctionner de manière tissue-spécifique. Afin de comprendre les mécanismes moléculaires expliquant ces phénotypes, nous avons effectué des analyses transcriptionnelles et ChIP-seq de mESC *Atxn7l3*$^{-/-}$. De façon inattendue, les niveaux de H2Bub1 sont significativement plus élevés dans le corps de l'ensemble des gènes transcrits en l'absence d'ATXN7L3. Cependant, l'occupation de l'ARN polymérase II sur l'ensemble de ces gènes ne varie que modestement dans ces cellules *Atxn7l3*$^{-/-}$. Ainsi, la déubiquitination de H2Bub1 ne régule pas directement la transcription par l'ARN polymérase II de l'ensemble du génome et les phénotypes embryonnaires dans des embryons *Atxn7l3*$^{-/-}$ pourraient être la conséquence de l'activité de déubiquitination d'autres protéines.

Mots-clés : régulation transcriptionnelle, développement embryonnaire, ATXN7L3, H2Bub, Pol II

Abstract in English

Coactivator complexes regulate chromatin accessibility by dynamically depositing or removing PTMs on histones. SAGA (Spt-Ada-Gcn5 acetyltransferase) is an evolutionary conserved multi-subunit co-activator complex with a modular organization. The deubiquitylation module (DUBm) of SAGA is composed of the ubiquitin-specific protease 22 (USP22) and three adaptor proteins, ATXN7, ATXN7L3 and ENY2, which are all required for the removal of mono-ubiquitin (ub1) from histone H2B. Here we investigated the role of SAGA deubiquitinase module in transcriptional regulation. We found that *Atxn7l3* is essential for embryonic development. To get better insight into ATXN7L3, we carried out *in vitro* mESC differentiation assays. Surprisingly, we found that ATXN7L3 promoted the differentiation of cardiomyocyte cells, but not ectoderm neural precursor. Thereby, ATXN7L3 might function in a tissue-specific manner. To understand the molecular mechanisms underlying these phenotypes, we performed transcriptomic and ChIP-Seq analyses from *Atxn7l3*$^{-/-}$ mESC. Unexpectedly, although H2Bub1 levels significantly increased in the gene body of every expressed gene, the genome-wide occupancy of Pol II was only modestly changed in *Atxn7l3*$^{-/-}$ ESCs. Thus, H2Bub1 deubiquitination did not directly regulate global Pol II transcription and the embryonic phenotypes of the *Atxn7l3*$^{-/-}$ embryo could be a consequence of the activity of the DUBm on other proteins.

Key words: Transcriptional regulation, embryonic development, ATXN7L3, H2Bub, Pol II