PRC1, PRC2 and BAP1 : Three tightly-linked chromatin modifiers involved in transcriptional regulation

Ming-Kang Lee

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Institut Curie

PRC1, PRC2 and BAP1: Three tightly-linked chromatin modifiers involved in transcriptional regulation

Soutenue par
Ming-Kang LEE
Le 20 Septembre 2021

Ecole doctorale n° 515
Ecole Doctorale Complexité du Vivant

Spécialité
Génétique et Génomique

Composition du jury :

Robert, SCHNEIDER PhD, Institute of Functional Epigenetics	*Président*
Jean-Christophe, ANDRAU PhD, IGMM	*Rapporteur*
Sophie, POSTEL-VINAY MD, PhD, Institut Gustave Roussy	*Rapporteur*
Ines, DRINNENBERG PhD, Institut Curie	*Examinateur*
Raphaël, MARGUERON PhD, Institut Curie	*Directeur de thèse*

institutCurie | PSL

Table of Contents

Acknowledgements

First of all, I would like to express my gratitude to my jury members, Dr. Andrau Dr. Drinnenberg, Dr. Postel-Vinay and Dr. Schneider, who kindly accepted the invitation and dedicate their time and energy on my dissertation. I would also like to thank my thesis committee members Dr. Batsché and Dr. Schoumacher and once again Dr. Schneider, who gave me very constructive advice throughout my PhD thesis.

I would like to thank my supervisor Raphaël, who had faith in me and gave me this chance four years ago when I barely had any knowledge in the field. I would like to thank you for always being patient, kind and honest with me. I still remember the first time that when I was helpless, you approached to my bench teaching me how to dot blots. I appreciate that your door is always open to discussion, that you never turn me down when I needed your help and that you have spent a great deal of time training me into critical thinking and sharpening my mind. I have learned so much from you. Your attitude and integrity towards science have always inspired me to become a better scientist.

I would like to thank all the members in the team.
I cannot express how much I thank you, **Laia**, the queen of bioinfo. You are the best, the kindest and the most talented postdoc! You have helped me so much with my project and have taught me so much with your sharp mind. You were always kind and generous in giving me advice even though you are constantly occupied. I miss the days we were working long hours in the cell culture room, chatting and laughing about all the crazy loads of experiments that we needed to do. You have taught me the right way to do science and you have given me so much strength and support when I was down, lost and stressed. You are the model I looked up to, and I have become who I am today because of you.

Seynabou, mon cœur, mon amour, my wife in the lab that I had no idea when we got married. How can I go through my PhD without you? How crazy was it that you happened to sit next to me when you arrived in the lab as a M2 student, and we became inseparable ever since. We laughed so much together, we cried so hard together. Even though you were so crazy that I demanded divorce so many

times, you are always the one that calms me down and give me strength. You are the person who has the integrity and never bow down to injustice. I am going to miss you so much when it is my time to move on. I wish you all the best and I know you will thrive in your last year of PhD.

Pierre, Pierre, you crazy tough, hard, stony pierre! It will always be a mystery how we have become friends. Even though to date you still called me your colleague, but I know you have treated me like a real friend. It is never easy to go through all this PhD process, but I am happy and lucky to have gone through together with you. You will always be my mentally unstable colleague. You have helped and listened to me so much. We have been through so many things, and I am glad you are the one that I am running to the finish line with. You, me and Seynabou, I will definitely miss the crazy times that we pull this off together.

I would like to thank **Audrey**. You have been taking care of me since the day one when I entered the lab. You have always been nice and kind to me and constantly checking me if I was doing fine.

I would like to thank **Katia**. You gave me so much mental support especially throughout the last few months of my thesis. Your words helped me stay composed and you never hesitated to put your work aside when I was in need.

I would like to thank my students **Suzanne** and **Emma**. Even though sometimes I was completely occupied by my work, you guys are always being supportive and I had enjoying working with you. Thank you for being super considerate and help me on the projects.

I would also like to thank **Tiphaine, Manuela, Camille, Samuel and Aurélien**, it was such a golden time having you guys in the lab, we shared so much joy and laugh. **Daniel and Michel**, you guys are full of knowledge and thank you for teaching me so much.

I would like to thank the Mass Spectrometry platform, especially Florent and Damarys, who had helped me a lot with the project and never hesitated to help me when I needed it. I would like to that Aurélien from the Microscopy platform. You

helped me a lot to acquire the photos. I enjoyed spending hours with you taking photos and chatting about basically everything. You have helped me a lot with imaging and analysis. I would like to thank La Ligue contre le cancer, who granted me the funding for my fourth year.

I would like to thank the IC3I and training unit. If were not for this amazing international PhD program, I would never have had the chance to enjoy such a wonderful experience in Curie. This program not only brought our fellow PhD students together, but it also gave us a lifetime friendship that we would always cherish.

I would like to thank my good friends **Darine** and **Tommaso**. I have no idea why you decided to be friends with me, probably your biggest mistakes in life so far. You guys are always there for me. We shared the good times, the bad times and difficult times. Your words always give me strength and encourage me go forward. You guys are the lighthouse of my life in Paris. We have shared so much wine, cheese, beer together, and I will always love you guys. I would also like to thank other wonderful friends I met through the IC3I program. We grow together, we learn together and we prosper together.

I would like to thank my friends and family in Taiwan. Even though we are 13-hour of (direct) flight apart, you are always there to give me support when I needed you guys. **Anya and Allen**, we have been so close since uni, and here we are, after 10 years, we still share so much with each other. All the chats on the phone for few hours, and still feel like it is not enough. I can easily remember all the insane things we have done together. Every time going back to Taiwan, we have created more fond memories of us.

Special thanks to Regina Filange, Chanandler Bong, Kim Adams, Monana, Mrs. Ross and Mr. Rachel. You have made my life full of joy and laughers. Even though sometimes it feels like I am stuck in the second gear, and it hasn't been my days, my week, my month and even my year, you guys are always there for me.

Abstract (*English*)

In eukaryotes, the maintenance of cell identity entails the precise control of gene expression, which results from the concerted actions of transcription factors and factors controlling chromatin structure. Polycomb repressive complex 1 and 2 (PRC1 and PRC2) are chromatin modifiers that orchestrate transcriptional repression by catalyzing H2Aub and H3K27me3, respectively. By contrast, BRCA1-associated protein 1 (BAP1) promotes transcription by removing H2Aub, acting as an antagonist of PRC1. However, the detailed mechanism of how BAP1 regulates transcription remains largely elusive. The interplay between PRC1 and PRC2 is also far from being fully understood. My PhD study aimed at investigating the underlying mechanisms for these two important questions.

(1) BAP1 is recruited to a subset of active enhancers where it stabilizes BRD4 occupancy.

In these studies, we showed that BAP1 promotes transcription by opposing PRC1 activity, and that BAP1 is mostly inert in its absence. Genome-wide analysis revealed that BAP1 is recruited to a subset of active enhancers. Besides, inactivation of BAP1 led to accumulation of H2Aub and impaired BRD4 recruitment. Consistently, super-resolution microscopy demonstrated reduced condensates of BRD4 and MED1 in BAP1-KO cells. This suggests that BAP1 has a crucial function for the integrity of a subset of enhancers. Importantly, by treating isogenic cells with BET inhibitors, we showed that cells mutant for BAP1 display a more pronounced proliferative response. This result suggests that further perturbation of enhancers function could be a therapeutic strategy for BAP1-null malignancies.

(2) PRC2 represses transcription independently of PRC1

PRC1 and PRC2 are long considered cooperating to maintain gene repression. However, analyzing transcriptomic profiles of PRC1-null, PRC2-null and PRC1/2-null cells, we demonstrated that both PRC1 and PRC2 can autonomously repress transcription. Through both unbiased and candidate-based approaches, we focus on identifying downstream effectors of PRC2-mediated silencing in the absence of PRC1. This includes investigating the roles of previously proposed H3K27me3 readers. While this study is still ongoing, it is likely that it will reveal new actor for PRC2-mediated repression.

Abstract (*French*)

Chez les eucaryotes, la maintenance de l'identité cellulaire implique le contrôle précis de l'expression des gènes. Ceci résulte de l'action concertée des facteurs de transcription et des facteurs contrôlant la structure de la chromatine. Les complexes répresseurs Polycomb (PRC1 et PRC2) sont des modificateurs de la chromatine qui orchestrent la répression transcriptionnelle en catalysant respectivement l'ubiquitinylation de H2A (H2Aub) et la méthylation de H3K27. A l'inverse, BAP1 (BRCA1-Associated Protein 1) favorise la transcription en retirant H2Aub, agissant donc comme un antagoniste de PRC1. Toutefois, les détails du mécanisme par lequel BAP1 régule la transcription restent mal compris. L'interaction entre PRC1 et PRC2 est également un sujet encore débattu. Mon projet de thèse visait à étudier ces deux importantes questions.

(1) La protéine BAP1 est localisée à une fraction des enhancers où elle stabilise le recrutement de BRD4.

Dans ces études, nous avons montré que BAP1 favorise la transcription en s'opposant au complexe PRC1 et que BAP1 est inerte en son absence. Des analyses à l'échelle du génome entier ont révélé que la protéine BAP1 est recrutée à une fraction des enhancers. Par ailleurs, l'inactivation de BAP1 amène à l'accumulation de H2Aub et à l'altération du recrutement de BRD4. En accord avec ces résultats, des expériences de microscopie à super résolution indiquent une réduction des condensées de BRD4 et de MED1 dans les cellules knockout pour BAP1. Cela suggère que BAP1 a un rôle crucial pour l'intégrité de certains enhancers. De façon importante, en traitant des cellules isogèniques avec des inhibiteurs de BET, nous avons montré que les cellules mutantes pour BAP1 montrent une sensibilité particulière à l'inhibition de la prolifération. Ces résultats suggèrent que promouvoir les perturbations des enhancers pourrait constituer une stratégie thérapeutique dans les pathologies où le gène BAP1 est muté.

(2) PRC2 réprime la transcription indépendamment de PRC1

PRC1 et PRC2 ont été considérés depuis longtemps comme agissant de concert pour maintenir la répression. Toutefois, en analysant les profiles transcriptomiques de cellules où soit PRC1, soit PRC2, soit les deux sont inactivés, nous avons démontré que PRC1 et PRC2 peuvent agir de façon autonome pour réprimer la transcription. Au travers d'approches non-biaisées et d'approches

basées sur une sélection de gènes candidats, nous essayons d'identifier les effecteurs de cette répression dépendant exclusivement de PRC2. Cela implique l'étude de protéines préalablement proposées comme interagissant avec H3K27me3. Cette étude est en cours mais il est probable qu'elle va révéler de nouveaux acteurs de la répression dépendant de PRC2.

List of figures

Acronyms and abbreviations

3D: Three dimensional

AEBP2: AE binding protein 2

AML: Acute myeloid leukemia

Ash1/2: Absent, small, or homeotic discs 1/2

Asx: Additional sex combs

ASXL1-3: Additional sex combs-like proteins 1-3

ATAC-seq: Assay for transposase-accessible chromatin using sequencing

AUTS2: Autism susceptibility gene 2 protein

BAF: BRG1/BRM-associated factor

BAHCC1: BAH domain and coiled-coil containing 1

BAHD1/2: Bromo adjacent homology domain containing 1/2

BAP1: BRCA1-associated protein 1

BCOR: BCL6 corepressor

BCORL1: BCL6 corepressor like 1

BET: Bromodomain and extraterminal domain

BRCA1: Breast cancer gene 1

BRD4: Bromodomain-containing protein 4

Brm: Brahma

CAF1: Chromatin assembly factor 1

CAGE-analysis: Cap analysis gene expression analysis

CBP: CREB binding protein

CBX: Chromobox Homologs

CDK7-9: Cyclin dependent kinase 7-9

CDYL: Chromodomain Y Like

CDYL2: Chromodomain Y Like 2

CHD7: Chromodomain helicase DNA binding protein 7

ChIA-PET: Chromatin interaction analysis by paired-end tag sequencing

ChIP: Chromatin immunoprecipitation

CK2: Casein kinase 2

cKO: Conditional KO

COMPASS: Complex of proteins associated with Set 1

cPRC1: canonical PRC1

CPSF: Cleavage and polyadenylation specificity factor

CRISPR: Clustered regularly interspaced short palindromic repeats

Cryo-EM: Cryogenic electron microscopy

CTCF: CCCTC-binding factor

CTD: C-terminal domain

CUT&RUN: Cleavage Under Targets and Release Using Nuclease

DCAF7: DDB1 and CUL4 associated factor 7

DNA: Deoxyribonucleic Acid

DSIF: DRB sensitivity inducing factor

DUB: deubiquitinase

E(z): Enhancer of zeste

E2F6: E2F transcription factor 6

EED: Embryonic ectoderm development

EHMT: Euchromatic histone lysine methyltransferase

EIF1AX: Eukaryotic translation initiation factor 1A X-linked

EMSA: Electrophoretic mobility shift assay

ENCODE: Encyclopedia of DNA Elements

EPOP: Elongin BC and Polycomb repressive complex 2 associated protein

EPR-1: Effector of Polycomb repression 1

eRNAs: Enhancer RNAs

Esc: Extra sex combs

Escl: Extra sex comb like

Ez: Enhancer of zeste

EZH1: Enhancer of zeste homolog 1

EZH2: Enhancer of zeste homolog 2

EZHIP: EZH1/2 inhibitory protein

FAIRE-seq: Formaldehyde-assisted isolation of regulatory elements-suquencing

FBRS: Fibrosin

FBRSL1: Fibrosin like 1

FBXL10: F-box and leucine-rich repeat protein 10

FOXK1/2: Forkhead box K1/2

fs(1)h: female sterile (1) homeotic

GFP: Green fluorescent protein

GNA11: G Protein subunit alpha Q

GNAQ: G Protein subunit alpha 11

G-o-f: Gain-of-function

GRO-seq: Global run-on sequencing

GTFs: General transcription factors
H2Aub: Histone H2A ubiquitination
H3K27: Histone H3 lysine 27
H3K36: Histone H3 lysine 36
H3K4: Histone H3 lysine 4
HCF-1: Host cell factor 1
HDAC: Histone deacetylase
HMTase: Histone methyltransferase
Hox: Homeotic genes
HP1: Heterochromatin protein 1
IDRs: Intrinsically disordered regions
IF: Immunofluorescence
IP3R3: Inositol 1,4,5-trisphosphate receptor type 3
JARID2: Jumonji and AT-rich interaction domain containing 2
KD: Knockdown
Kdm2: Lysine demethylase 2
KDM2B: Lysine demethylase 2B
Kis: Kismet
KLF4/5: Kruppel like factor 4/5
KMT: Lysine methyltransferase
KO: Knock out
L3MBTL2: L3MBTL histone methyl-lysine binding protein 2
LIF: Leukemia inhibitory factor
lncRNA: Long non-coding RNA
L-o-f: Loss-of-function
LSD1: Lysine demethylase 1A
MAX: MYC associated factor X
MBD5/6: Methyl-CpG binding domain protein 5/6
MED1: Mediator complex subunit 1
MEFs: Mouse embryonic fibroblasts
mESCs: mouse embryonic stem cells
MGA: MAX gene-associated protein
MLL1-4: Mixed-lineage leukemia protein 1-4
MM: Multiple myeloma
MPRA: Massively parallel reporter assays

mRNA: Messenger RNA
MTF2: Metal response element binding transcription factor 2
ncRNA: Non-coding RNA
NELF5: Negative elongation factor 5
NGS: Next generation sequencing
NPCs: Neural progenitor cells
NSD-1: Nuclear receptor SET domain-containing protein 1
Nurf55: Nucleosome-remodeling factor subunit 55
OGT: O-linked N-acetylglucosamine (GlcNAc) transferase
PALI1/2: PRC2 associated LCOR isoform 1/2
PBAF: Polybromo-associated BAF
Pc: Polycomb
PcG: Polycomb group
PCGF1-6: Polycomb group ring finger 1-6
PCL1-3: Polycomb-like homologs 1-3
PFA: Posterior fossa type A
Ph: Polyhomeotic
PHC1-3: Polyhomeotic homolog 1-3
PHD: Plant homeodomain
PHF1: PHD finger protein 1
PHF19: PHD finger protein 19
Pho: Pleiohomeotic
PhoRC: Pho repressive complex
PIC: Pre-initiation complex
POL II: RNA polymerase II
PPP2R1A: Protein phosphatase 2 scaffold subunit A alpha
PRC1: Polycomb repressive complex 1
PRC2: Polycomb repressive complex 2
PR-DUB: Polycomb repressive deubiquitinase
PRE: Polycomb repressive element
PRO-seq: Precision nuclear run-on sequencing
Psc: Posterior sex combs
P-TEFb: Positive transcription elongation factor b
PTM: Post-translational modification
RbAp46/48: Rb-associated protein 46/48

REST: RE1 silencing transcription factor
RING1A/B: Really interesting new gene 1 protein A/B
RNA: Ribonucleic acid
RSF1: Remodeling and spacing factor 1
RUVBL2: RuvB like AAA ATPase 2
RYBP: RING1 And YY1 binding protein
SAM: Sterile alpha motif
Sce: Sex comb extra
Scm: Sex comb on midleg
SCM1-2: Scm homolog 1-2
Ses: Super-enhancers
SET: Su(var)3-9, Enhancer-of-zeste and Trithorax
Set1: SET domain containing 1
SF3B1: Splicing factor 3b subunit 1
Sfmbt: Scm like with four Mbt domains
shRNA: Short-hairpin RNA
siRNA: Short interfering RNA
SKP1: S-phase kinase associated protein 1
SMARCA2/4: SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 2/4
STARR-seq: Short for self-transcribing active regulatory region sequencing
STAT3: Signal transducer and activator of transcription 3
Su(z)12: Suppressor of zeste 12
SV40: Simian vacuolating virus 40
SWI2/SNF2: Switch [swi] genes and Sucrose nonfermenting [snf] genes
TADs: Topologically associated domains
T-ALL: T-cell acute lymphoblastic leukemia
TFs: Transcription factor
Trr: Trithorax-related
Trx: Trithorax
TrxG: Trithorax group
TSSs: Transcription starting sites
UCH: Ubiquitin carboxyl-terminal
UM: Uveal melanoma
USP: Ubiquitin specific peptidase

vPRC1: variant PRC1

WDR5: WD repeat domain 5

XCI: X chromosome inactivation

YAF2: YY1 associated factor 2

YY1: Yin and Yang 1

Introduction

Preface: Chromatin organization and transcriptional regulation

One of the key questions in biology that has remained puzzling is how a single (fertilized) egg develops into an individual with highly specialized cell types while bearing identical genetic information. Presumably, extrinsic and intrinsic signals that establish and maintain distinct combinations of genes shape cellular plasticity and identity. This expression pattern reckons on deployment of an array of transcription factors (TFs), DNA elements, molecular signatures of chromatin, which in concert, articulate particular chromatin features that drive cell versatility.

In eukaryotes, DNA is wrapped around histones to form chromatin. The structural unit of chromatin is termed nucleosome, which consists of 147 bp of DNA and an octamer of 2 repeats of the 4 histones: H2A, H2B, H3 and H4. The fifth histone, H1, is a linker histone that binds to the entry and exits sites of DNA on the nucleosomes. The C-terminus of histones protrudes outwards from nucleosomes where the residues can undergo post-translational modifications (PTMs)[1]. The organization of genome is not stochastic but rather compartmentalized in a sophisticated fashion that contribute to transcriptional regulation[2]. In early days, euchromatin and heterochromatin were first observed based on light microscopy and chromatin dyes. With the advancements of methodology and innovation of novel techniques, the nuclear architecture has been deciphered at a rapid pace. It is shown that individual chromosomes occupy distinct regions in the nucleus termed chromosome territories[3]. As the saying goes, birds of a feather flock together. At megabase scale, genomic regions that share similar features tend to interact with each other. That is, actively transcribed regions frequently establish segregation that seclude inactive clusters, and *vice versa*. These active compartments are referred as "compartment A", and inactive ones as "compartment B"[4]. Within the compartment, chromatin is further partitioned into topologically associated domains (TADs) that show strong self-interacting activities[5]. Chromatin loop extrusion established by cohesin complex facilitates chromatin folding within the TADs[6]. Such higher-order organizations grant long-range interactions that hurdles the linearly spatial limitations, which enables communications between regulatory elements (e.g. enhancer and promoters) and encompasses the establishment, maintenance and propagation of transcriptional and/or epigenetic cues. Indeed, dynamics of 3D genome architecture participates

to transcriptional regulation throughout development, lineage-specification and other processes **(Figure 1)**[2,7].

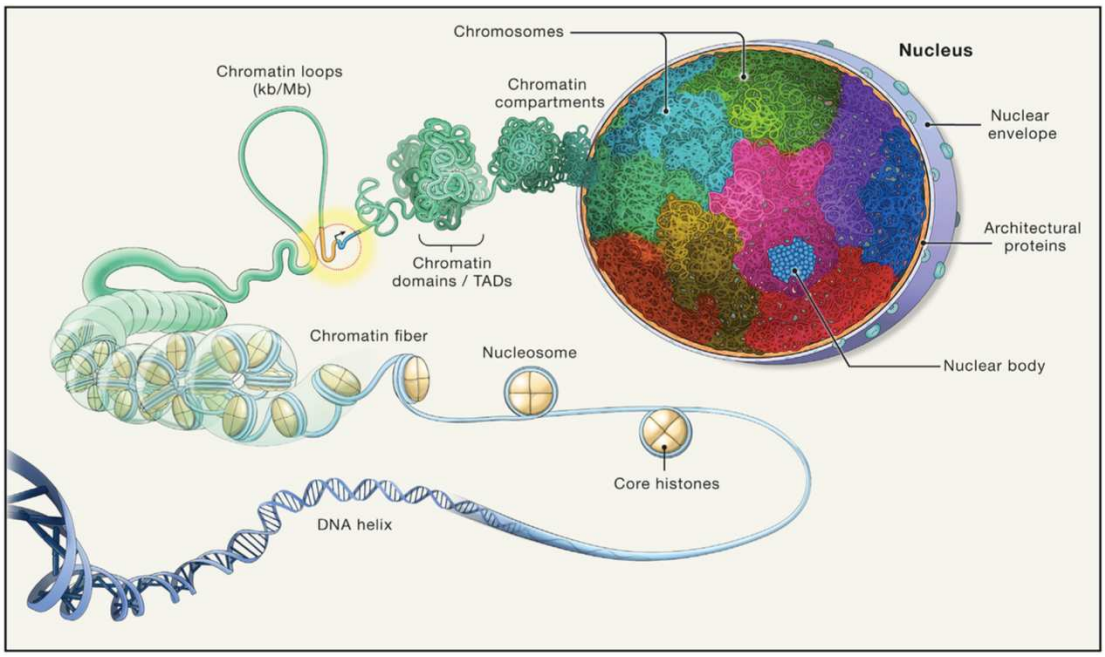

Figure 1. Chromatin organization (Misteli 2020)

Transcription is an intricate process that is tightly regulated by a variety of factors in a multi-step fashion. Transcription of messenger RNA (mRNA) and some non-coding RNA (ncRNA) are carried out by RNA polymerase II (POL II), which catalyzes DNA-dependent synthesis of RNA. To initiate transcription, POL II recognizes the promoter region of the genes, opens the DNA duplex and embark on RNA synthesis. When RNA grows to a critical length, the polymerase escapes from the promoter and enters the elongation phase. The subsequent elongation complex governs the extension of RNA chain until reaching the termination signal where it releases DNA and RNA[8] (**Figure 2**).

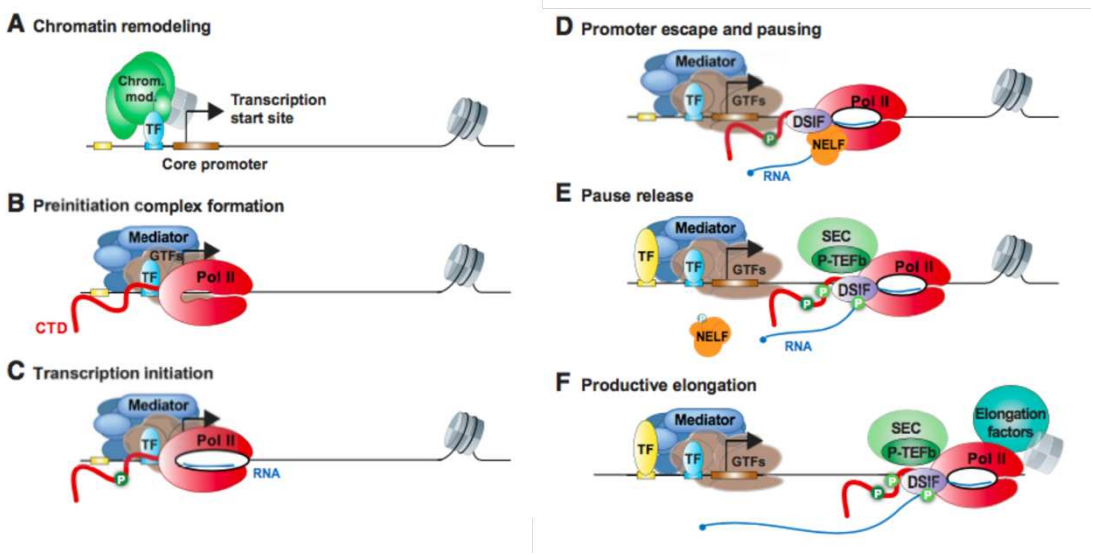

Figure 2. The processes of transcription (Core and Adelman 2020)

By default, the structure of chromatin hampers the access of the polymerase, and the enzyme itself cannot recognize promoter elements[9]. Instead, loading of transcription machinery ascribes to TFs binding to sequence-specific motifs on promoters and enhancers that recruits chromatin co-activators and Mediators to facilitate chromatin accessibility and recruitment of POL II[10,11]. To take effect, the RNA polymerase II, together with general transcription factors (GTFs)—TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH—assembles into the pre-initiation complex (PIC) and make extensive contact with DNA[12]. The C-terminal domain (CTD) of the POL II RPB1 subunit is comprised of numerous hepta-amino acid repeats of $Y_1S_2P_3T_4S_5P_6S_7$, which are subjected to PTMs such as phosphorylation on serine residues (Ser2, Ser5 and Ser7) across transcription cycle. These modifications stimulate stage-appropriate interactions with transcription initiation, elongation, and RNA processing factors[13]. Collectively, the GTFs contributes to transcriptional initiation through various mechanisms. For example, TFIIB directly recruits POL II to promoter and functions in promoter opening[14]. TFIIH subunit CDK7 kinase phosphorylates Ser5 and Ser7 on the CTD domain to facilitate promoter escape to enter elongation state[15].

As the transcription initiates, the polymerase catalyzes synthesis of nascent RNA that is complementary to template DNA. The newly synthesized RNA is tightly

associated with DNA template, forming a RNA-DNA hybrid within POL II active sites.

Once the RNA synthesis reaches to a critical point, an elongation complex assembles, which extends the newly synthesized RNA strand in a processed manner. However, approximately 50 bp downstream of transcription starting sites (TSSs), the machinery is paused, and awaits further signals prior to entering productive transcription[16]. RNA POL II pausing is a ubiquitous event in all transcribed genes. During which, the complex and the nascent transcript remain stable throughout this stage[17]. The fate of the proximally-paused early elongation complex is absolutely decisive for gene output, as successful pause-release guarantees the production of full-length RNA, and that the paused complex itself inhibit new series of transcriptional initiation[18]. Therefore, pause-release regulates transcription frequency per unit of time. Paused elongation complex is stabilized by DSIF and NELF5 which restrict the mobility of POL II. The release of paused POL II into gene body requires the CDK9 kinase, a subunit of positive transcription elongation factor b (P-TEFb)[16]. P-TEFb is recruited to promoters through interactions with TFs, Mediator, and coactivators[19].The complex phosphorylates DSIF, NELF, and Ser2 on the POL II CTD. DSIF phosphorylation introduces allosteric conformational change and dissociates NELF from POL II. P-TEFb also triggers formation of an activated complex of which SPT6 binds the phosphorylated linker to the CTD and the PAF complex competes the binding with NELF[20]. Phosphorylation of Ser2 on the CTD domain help recruit co-activators that maximize elongation rate throughout the course of time. As P-TEFb plays as a gatekeeper in control of pause-release, inhibition of P-TEFb activity traps Pol II in early elongation at nearly all mRNA promoters[17,21]. Once the process reaches downstream of a polyadenylation signal, the nascent transcript is cleaved by CPSF complex. The transcription termination occurs as both of the polymerase and RNA sequence is released from the DNA template[22].

Transcriptional events of given genes can also be regulated by epigenetic signals that establish, maintain and reverse certain transcriptional states. The epigenetic information resides in self-propagating and self-sustainable molecular signatures that drive transcriptional response, which can be established *de novo* from environmental or developmental cues; or are inherited from previously

experienced stimuli. These signals must perpetuate throughout cell division to confer the "memory" of transcriptional states, but are subject to alternations if under further alternative cues[23]. Taken as an example, a fertilized egg is able to self-renew and develop into various types of progenitor cells with distinct programs guided by extrinsic and intrinsic cues. These progenitors amplify and give rise to different classes of terminal differentiated cells, of which the transcriptional output and cellular identity are maintained. Intricate epigenetic states involves orchestration of converging and diverging signals, including TFs, DNA methylation[24], PTMs on histones[7,25] and non-coding RNAs[26]. Epigenetic inheritance can be achieved by trans-acting and/or cis-acting mechanisms. Trans-acting mechanism can maintain the transcriptional states by self-propagation through positive feedback loops and networks of TFs. After each cell division, the inherited TFs in the cellular content continuously function on DNA element to confer the states. In contrast, cis-acting factors transmit the molecular signatures on chromatin where they impose an effect. These include DNA methylation and certain types of histone modifications. The establishment, reinforcement and transmission of the epigenetic landscape are three distinct but highly connected processes. Considering its widespread activities on chromatin feature, epigenetic mechanisms are likely the crux of many physiological processes in organisms.

The following chapters will describe functional regulations by chromatin factors in mammalian system, notably the antagonistic Polycomb (PcG) and Trithorax (TrxG) groups of proteins, the relatively newly identified histone de-ubiquitinase BAP1 complex, and how these processes incorporate with cis-regulatory elements (i.e. enhancers) to modulate transcriptional landscapes.

Transcriptional regulation by Polycomb repressive complexes

In Drosophila, the antagonistic mechanism– Polycomb and Trithorax complexes– regulates a plethora of genes, where mutations of genes within the complexes lead to embryonic lethality and developmental abnormalities. Subsequent works identify them as chromatin modifiers that are crucial for gene regulation. This chapter will center on the current knowledge of Polycomb machinery, covering from the discovery, diversity to its functionality in transcriptional control.

History of Polycomb group of genes

Almost 80 years ago, Eleanor Slifer reported a recessive mutation rendering a phenotype with ectopic sex combs on midlegs and hindlegs in adult male flies. Based on the mutant phenotype, it was named extra sex combs (*esc*). Several years later, a dominant mutation with similar phenotype was identified by Pam Lewis, and was given the name Polycomb (*Pc*)[27]. It was later discovered that *Pc*-mutated larvae showed posterior transformation, implicating a more general role of Polycomb in repressing homeotic genes (*Hox*), a cluster important for somite segmentation and patterning[28]. Ever since, a collection of genes whose mutant phenotypes resembled those of ectopic sex combs or misexpression of *Hox* genes were enlisted as Polycomb group of genes. It is estimated that 15 genes encode PcG proteins in Drosophila.

Polycomb repressive complexes coordinate transcriptional repression

Aside from the paradigmatic HOX genes, PcG proteins repress transcriptional activities of a plethora of genes that are important for development, pluripotency and differentiation. Biochemical studies showed that PcG proteins assemble mainly into two evolutionarily conserved multimeric complexes: Polycomb repressive complex 1 and 2 (PRC1 and PRC2)[29,30]. The two complexes are chromatin modifiers that repress transcription largely through their histone modifying activities. PRC1 is an E3 ubiquitin ligase that catalyzes mono-ubiquitination of histone H2A at lysine 119 in mammals (H2A lysine 118 in flies; herein referred to as H2Aub)[31,32], whilst PRC2 is a methyltransferase (KMT) accounting for mono-, di- and tri- methylation on histone H3 at lysine 27

(H3K27me1/2/3) in a step-wise fashion[33,34].The PcG genes, notably PRC2 complex, are functionally conserved throughout evolution to control critical biological processes. Homologous or analogous complexes have been identified widely in eukaryotes, including plants, fungi, nematodes, drosophila and vertebrates[29]. For example, PcG system controls vernalization in plants[35]. They are also required to repress transposable elements in paramecium[36]. The core complexes are conserved between flies and mammals. However, emergence of vertebrate-specific homologs and co-factors have added layers of complexity in the regulatory mechanisms. To name a few: functional redundancy of RING1A and RING1B in H2A ubiquitination[37,38], stage-dependent expression of versatile CBX paralogs[39], and functional divergence of EZH1 and EZH2[40,41].

During early embryonic development in mouse, deletion of either PRC1 and PRC2 component lead to fatal defects at implantation or early gastrulation stages[42,43]. In mouse embryonic stem cells (mESCs), PRC1 and PRC2 coordinate gene repression of a cohort of developmental genes to maintain pluripotency. Genome-wide mapping suggests the binding sites of PRC1 and PRC2 are mostly at loci enriched in CpG and GA motifs (CpG island) around promoters, and that both complexes and their marks largely colocalize[44-48]. Loss of either complex results in large-scale de-repression in mESC[37,45]. It is clear that loss of PRC1 activity halts self-renew and de-represses lineage-specific genes[37,49]. However, the effects of PRC2 loss on pluripotency remain debatable. Complete abrogation of PRC2 activity by EED knock out (KO)[45] or EZH2/1 dKO[50] de-represses lineage-specific genes but does not compromise cell proliferation. Of note, in a supposedly similar setting (SUZ12 knock out), others did not observe much effect of abrogating PRC2 activity as long as the cells remain undifferentiated[51]. PRC2-null cells are able to form teratomas but with reported smaller sizes than WT, and skewed development toward endodermal layers. Hence, while PRC1 is required for pluripotency in mESCs, loss of PRC2 displays milder consequences on pluripotency[52]. Interestingly, simultaneous ablation of PRC1 and PRC2 introduces larger magnitude of gene de-repression, suggesting functional redundancy in repression[52]. Of note, most H3K27me3-enriched promoters are also marked by H3K4me3, a modification deposited by TrxG complex and associated with gene activation in mESCs. The co-existence of both active and repressive marks is termed “bivalent domain”. Bivalent genes encode key transcription factors important for morphogenesis and

cell-fate commitment (as expected for H3K27me3-enriched genes), where the transcriptional activity is poised and awaits to be switched on or off according to addition clues[45,53,54]. Early studies dissecting roles of PcG system have proposed a PRC2-initiated hierarchical recruitment whereby PRC2 is recruited to chromatin, depositing H3K27me3, and the mark is further recognized by PRC1 subunits chromodomain proteins CBXs, thereby promoting PRC1 recruitment[33,55,56]. An alternative model has been proposed considering that JARID2 containing-PRC2 complex can recognize H2A ubiquitination. Therefore, it is proposed that PRC2 recruitment could be stabilized by PRC1[57,58]. Detailed mechanisms for PRC1/PRC2 recruitment will be discussed further below. Nevertheless, this mutually reinforcing interactions potentiate fidelity and propagation of the PcG system (**Figure 3**).

The importance of PRC1 and PRC2 has also been documented in differentiation of various somatic lineages[54], such as haematopoietic lineage, mesenchymal stem cells, epidermal homeostasis[59-62].

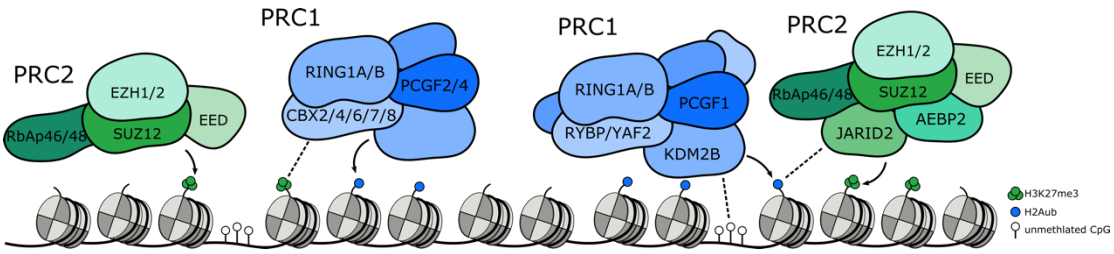

Figure 3. Polycomb repressive complexes coordinate transcriptional silencing.

Polycomb repressive complex 1

Genomic mapping in mESCs revealed RING1B broadly co-localized with PRC2 at promoter regions[45,47], but it can also occupy regions that are not bound by PRC2[63,64]. Indeed, a handful of studies both in fly and in mammals have demonstrated that recruitment of PRC1 can be established by binding to pre-existing H3K27me3 or by *de novo* tethering to DNA mediated by variant PRC1 subunits[25]. The diverse mechanisms driving PRC1 recruitment potentially reflect the heterogeneity of PRC1 complex, of which, distinct subunits exploiting different approaches to target chromatin. Depletion of both RING1A and RING1B in mESCs

completely abrogate H2Aub in concert with genome-wide reduction of PRC2 occupancy and H3K27me3 enrichment at Polycomb target sites. This introduces global de-repression and differentiation, and finally halts cell proliferation in mESCs[37,49,63], suggesting that PRC1 controls at least in part PRC2 occupancy.

How PRC1 conducts repression awaits to be unraveled. Potentially, the enzymatic product H2Aub can recruit PRC2 component and confers transcriptional repression (the question therefore shifts to how PRC2 controls transcription). However, since PRC1 localizes on a subset of promoters devoid of PRC2, this implies that RING1B/ H2Aub are capable of silencing target genes independently of PRC2. A recent study reports that the remodeling and spacing factor RSF1 is a novel H2Aub reader that contributes to H2Aub-mediated silencing[65]. However, there is only minor overlap of de-repressed genes resulted from respective RING1B-KO and RSF1-KO in HeLa cells, insinuating the contribution of RSF-1 is not major. Alternatively, H2Aub *per se* might repress transcription as H2Aub inhibits the deposition of H3K4 di- and tri- methylation (H3K4me2/3) *in vitro*, which consequently compromises transcriptional initiation but not elongation[66]. H2Aub may also plays a role in sequestration of POL II loading onto the promoters in mESCs, but the link between H2Aub and POL II configuration requires further studies[67]. Aside from its catalysis, PRC1 could initiate chromatin compaction to generate less accessible environment for transcription machinery[68,69].

Versatility of PRC1 complex

In *Drosophila*, PRC1 complex contains the catalytic subunit Sex comb extra (Sce)/dRing, Pc, Posterior sex combs (Psc), Polyhomeotic (Ph) and Sex comb on midleg (Scm). Kdm2, homolog of mammalian demethylase KDM2B, forms a variant PRC1 complex. In mammals, the system is diversified by versatile homologs and accessory co-factors, which give rise to functionally distinct PRC1 complexes in humans[27].

Systematic proteomic analysis identified mammalian core PRC1 consists of E3 ligase RING1A or RING1B, and one of the six PCGF subunits (Psc homologs)[70]. The complexes can be further classified into canonical PRC1 (cPRC1) or variant PRC1

(vPRC1) complexes, regarding on the presence or absence of the CBX subunit (Pc homolog). Canonical PRC1 contains one of the CBX proteins (CBX2, CBX4, CBX6-8), PHC1-3 (Ph homolog), SCM1-2 (Scm homolog) and PCGF2/MEL-18 or PCGF4/BMI-1. In variant PRC1 complex, CBX is replaced by RYBP or YAF2, and assembles with either one of the six PCGF proteins. Each PCGF-RING1A/B incorporates distinct set of associated polypeptides, emcompassing the functional versatility of PRC1 **(Figure 4)**[7,25,71].

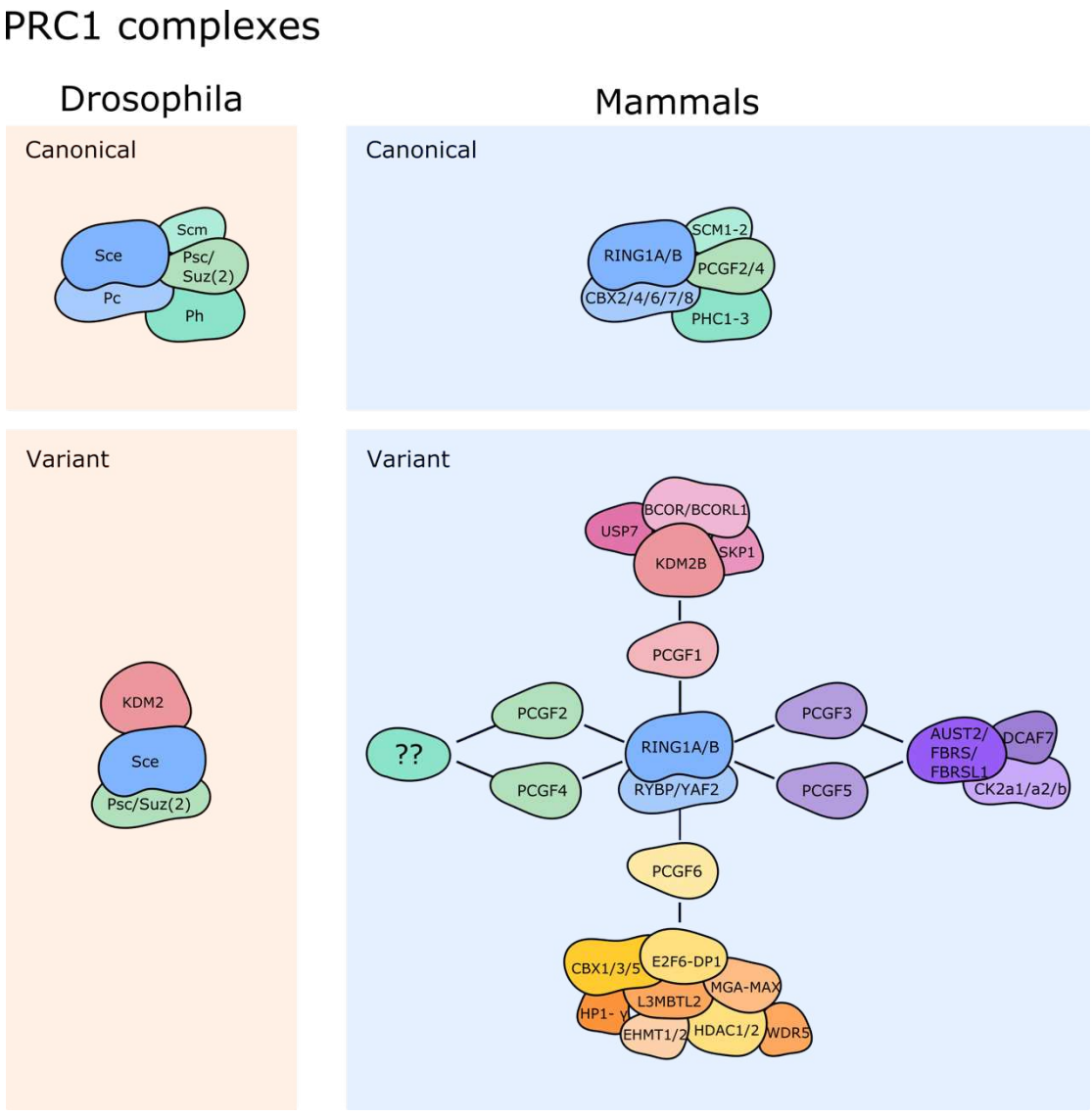

Figure 4. Polycomb repressive complex 1

Canonical PRC1

Canonical PRC1 contains CBX subunit that binds to H3K27me3. Cells with mutated PRC2 shows eviction of PRC1 from PRC2 targets, suggesting a causal link of PRC2 mediating PRC1 recruitment[33]. However, deletion of PRC2 shows only reduction, but not complete loss of PRC1, and has little effects on global H2Aub, suggesting the existence of other mechanisms that drive PRC1 recruitment[72]. The canonical PRC1 compacts chromatin and restricts the access of trans-acting factors to DNA to prevent transcription. For example, loss of RING1B decompacts chromatin at Hoxb locus and abrogates silencing in mESCs[73]. The chromatin compaction can be mediated by the subunits PHC2 and CBX2. The sterile alpha motif (SAM) of PHC2 plays an essential role in nucleosome oligomerization and PRC1 clustering[74]. Inhibition of polymerization abolishes PRC1 binding and leads to gene de-repression[74]. The intrinsically disordered regions (IDRs) of CBX2 were linked to chromatin compaction[75], and were described to be crucial for formation of phase separated droplets termed "Polycomb body"[76,77]. Single-molecule tracking of RING1B shows that PRC1 binds more stably and with higher intensity within the Polycomb bodies[78]. These suggest that cPRC1 modulates the structural configuration of chromatin, which helps self-propagation and contributes to gene silencing.

Variant PRC1

Individual PCGF homologs assemble with distinct chromatin factors and thereby confering the heterogeneity of vPRC1 complexes. Variant PRC1.1 contains H3K36-specific demethylase KDM2B, USP7, BCOR/BCORL1 and SKP1. As for vPRC1.3 & 1.5, PCGF3/5 interact with WD40 domain protein DCAF7, AUTS2/FBRS/FBRSL1 and CK2a1/a2/b. PCGF6 defines the largest vPRC1, which is composed of WDR5, heterochromatin factors HP1-γ and CBX1/3/5, and transcription factor E2F6 complex (E2F6-DP1, L3MBTL2, MGA-MAX, HDAC1/2 and EHMT1/2). The interactors of RYBP-PCGF2/4-PRC1 remain to be determined **(Figure 4)**[70,79].

PRC1 recognizes its own H2Aub mark through the RYBP subunit, suggesting a likelihood of feed-forward mechanism that could be involved in spreading along the chromatin[80]. It was demonstrated *in vitro* that both RYBP and YAF2 stimulates the ubiquitinating activity of PRC1[81]. Using de novo tethering assay, Blackledge and colleagues reported that vPRC1 is the main effector in catalyzing H2A

ubiquitination in mESCs. On the contrary, the canonical CBX-PCGF2/4-PRC1 complex has little E3 ligase activity. Variant PRC1-dependent H2Aub controls genome-wide PRC2 localization and H3K27me3 deposition, providing a "PRC1-driving-PRC2" recruitment model[64]. Lastly, different PRC1 complexes can occupy distinct genomic loci, implying a cooperative system in transcriptional silencing[70].

Functional divergence of variant PRC1 complexes

PRC1.1

Sequence analysis of Polycomb nucleation sites shows enrichment for CpG islands in the vicinity of promoters[82]. Studies in mESCs showed that KDM2B/FBXL10, a H3K36 specific demethylase, displayed strong affinity towards unmethylated CpG island through its CxxC-ZF domain. KDM2B preferentially binds to unmethylated DNA, as demonstrated by electrophoretic mobility shift assay (EMSA). Indeed, KDM2B was displaced from methylated promoter during differentiation in mESCs. Artificial tethering of KDM2B to a heterologous promoter recruits RING1B and PCGF1, and subsequently leads to H2AK119 ubiquitination. KDM2B is found to co-localized with RING1B, whose depletion leads to decreased RING1B binding, loss of H2Aub, and consequent gene de-repression as well as early exit of pluripotency. These observations establish a direct link between targeted recognition of CpG island and PRC1 recruitment[83-85].

PRC1.3 and PRC1.5

In mESCs, PCGF3 and PCGF5 was reported to stably interaction with Xist RNA, suggesting a direct role in Xist-RNA mediated X chromosome inactivation (XCI). Despite single ablation of either PCGF3 or PCGF5 does not affect H2Aub coating on inactivated X chromosome, deletion of both homologous proteins strongly reduces Xist-dependent H2Aub and H3K27me3 deposition, and is associated to failed silencing of X-linked genes. Consistently, recruitment of PCGF3/5 can occur in the absence of PRC2. PCGF3/5 double deletion is embryonic lethal, but female embryos display more profound defects and embryonic lethality is reported as early as E9.5. These observations illustrate functional redundancy between PCGF3 and PCGF5 and, more importantly, highlight a direct link between X inactivation and PRC1.3/5-mediated H2A ubiquitination[86].

PRC1.6

PCGF6-PRC1 binds at promoters that are specifically enriched in germline related genes such as the ones regulating spermatogenesis and meiosis. Deletion of PCGF6 leads to H2Aub reduction, upregulation of the target genes, and hence perturbs cell identity and self-renew. Mechanistically, PCGF6 recruitment requires the TF dimer MAX-MGA within the complex to bind to the E-box motif. MAX or MGA knockdown (KD) reduces occupancy of PCGF6, and gives rise to subsequent de-repression on PRC1.6 target genes. In vivo study showed partial penetrance of embryonic lethality in PCGF6$^{-/-}$ progeny, ranging from blastocyst to post-implantation stage[87].

PRC1-mediated gene silencing requires the combinatory actions of vPRC1s.

Genomic profiling of individual PCGF suggests that they bind to distinct genomic region[70], alluding to a model whereby diverse PRC1 complexes contribute concurrently to gene silencing. In mouse ESCs, while PRC1 deletion causes massive gene activation, removal of respective vPRC1 complexes does not fully recapitulate the PRC1-null phenotype[49,88]. For example, removal of canonical PRC1, which acts downstream of PRC2, merely contributes to gene de-repression, in accordance with the study proposing that PRC2 is dispensable for gene silencing in mES cells[51]. Loss of PCGF1 perturbs H2Aub enrichment and PRC2 recruitment in concert with mild de-repression. PCGF3/5 double KO causes little gene upregulation but results in major reduction in H2Aub in bulk and across the genome, pointing out their roles in maintaining pervasive H2Aub. In contrast, simultaneous deletion of PCGF1/3/5/6 resembles those of PRC1-null context. Hence, the additive action of the different vPRC1 complexes together defines PRC1-mediated repression, and that individual variant contributes to distinct pools of H2Aub[49].

Enzymatic activity is central to PRC1-mediated repression

As abovementioned, PRC1 is proposed to contribute to repression via both histone modifying activity and chromatin compaction. However, several studies have challenged this view by showing that catalytically inactive PRC1 complex sustain repressive states. In a fly clonal assay, embryos carrying zygotic and maternal point mutation form of Sce (I48A) renders the enzyme to be catalytically dead. These mutant embryos show no H2Aub accumulation with reported decreased H3K27me3 level, but no derepression of the canonical *Hox* genes in late embryonic stage. Nonetheless, these embryos cease development at the end of embryogenesis. In parallel, a form of mutant H2A that could not be ubiquitylated in culture cells model also suggested lack of Hox loci de-repression similar to those of catalytic dead embroys[89]. Besides, in mouse ESCs, re-expression of WT or catalytic dead RING1B could both rescue the de-repression of Hox genes[73], and mutated RING1B did not introduce genome-wide de-repression[69]. These indicate that H2Aub is dispensable for stable Hox silencing. However, the study in flies is concluded solely based on the observation at the paradigmatic Hox loci, and overlooked that the ubiquitinating activity of PRC1 is essential for development. Also, the studies in mESCs overlooked the contribution of RING1A.

Hypothetically, if H2Aub is dispensable for PcG-mediated repression, then it is hard to reconcile with the facts that: (1) vPRC1-mediated H2Aub accumulation controls PRC2 recruitment[64]. (2) PRC2 alone is not sufficient to encompass Polycomb-mediated transcriptional repression, as more recent work reported mild de-repression in PRC2-null mESCs[50,51]. (3) Removal of cPRC1 shows little impact in transcription[49]. Importantly, part of these contradictory data might reflect the fact that the catalytic dead I53A RING1B mutant turns out to be hypomorphic, that is, it displays residual E3 ligase activity[90].

Recent work employed a conditional double mutant RING1B (I53A-D56K) that rendered complete abrogation of PRC1 catalysis in mESCs. Transcriptomic analysis of RING1B$^{I53A\text{-}D56K}$ mutant recapitulated those of conditional PRC1-KO. While recruitment of PRC1 is independent of enzymatic function, components of PRC2 were evicted in concert with substantial loss of H3K27me3, thereby, cPRC1

binding could not take place, and Polycomb domain formation was compromised. Together, these implicate that catalytic activity of PRC1 is central to its function in mammals, though it is still unclear whether PRC1 has other activity[91].

Polycomb repressive complex 2

Core components of PRC2

PRC2 catalyzes methylation on H3K27[33,34,92,93]. The catalytic core of PRC2 is rather homogeneous compared to the variety of PRC1 complex. In Drosophila, the subunits that constitute the core PRC2 are methyltransferase E(z), extra sex comb and extra sex comb like (Esc/Escl), Su(z)12 and Nurf55; whereas the mammalian homologous proteins are EZH1/EZH2, EED, SUZ12 and RbAp46/RbAp48 (**Figure 5**)[94]. Quantitative proteomic analysis estimates roughly half of the H3 histone tails are di-methylated, whereas H3K27me1 and H3K27me3 each account for 10-20% of total H3 tails[95]. The high prevalence of H3K27me2 is regarded as a mechanism in prevention of stochastic enhancer activation by H3K27 acetylation, but its exact biological significance remains to be unveiled[96].

PRC2 and its catalytic product are required for normal development in Drosophila. Mutation in the subunit E(z) gives rises to de-repression of *Hox* genes and homeotic transformation[33,93]. In addition, a point mutation on H3K27 that no longer accommodates methylation phenocopies PRC2-mutant clones[97]. Unlike the indispensability of PRC1 in maintenance of pluripotency (to such that PRC1-null cells cannot be maintained), PRC2 is not required in early embryonic stage, in line with the fact that removal of canonical PRC1 that binds to H3K27me3 does not introduce gene activation[49]. Nevertheless, PRC2 is required in later stage to maintain transcriptional repression during ESCs differentiation[51].

EZHs contains a conserved catalytic SET [Su(var)3-9, Enhancer-of-zeste and Trithorax] domain, which is only active when associating with other core PRC2 subunits, therefore defining it as a holocomplex[92]. EZH2 and EZH1 function in a partial redundant manner: EZH2 is the primary homolog responsible for tri-methylation of H3K27. Deletion of EZH2 but not EZH1 leads to global reduction of

H3K27me2/3 in mESCs[50], in accordance with *in vitro* study showing EZH1 expresses weaker methyltransferase activity[40]. However, EZH1 complements EZH2 function as EZH1 depletion in EZH2-null context wipes out all forms of H3K27 methylation. More specifically, it exacerbates H3K27me3 diminution on Polycomb targets with ensuing gene de-repression[50]. Expression of EZH2 correlates with proliferative rate[40,98], EZH1's contribution would be more important in slow-proliferating or post-mitotic tissues[41,99,100]. Aside from enzymatic activity, EZH1 has been shown to mediate chromatin compaction[40,101].

EED contains an aromatic cage formed by the WD-40 repeats that recognizes H3K27me3, which allosterically activates enzymatic function of PRC2, serving as a feed-forward system in self-reinforcement[102], similar to vPRC1 subunit RYBP binds to its enzymatic product H2Aub. This mechanism is crucial in propagating their very own marks, as loss of EED dissociates and destabilizes the complex, leading to complete loss of H3K27 methylation. EED mutation is embryonic lethal, mutant conceptuses are reabsorbed at early gastrulation stage[103,104].

Likewise, SUZ12 ablation gives rise to full erasure of H3K27 methylation and embryonic lethality[105]. The C-terminal VEFS domain of SUZ12 interacts with EZH2 and EED, stabilizes the complex and is required for PRC2 catalytic function[106]. In a EED-KO and EZH1/2 dKO context, SUZ12 can be recruited to DNA, suggesting the recruitment of SUZ12 to CpG island is independent of other PRC2 components and H3K27 methylation. In addition, re-expression of SUZ12 in SUZ12-KO cells accurately restores the patterns of enrichment for PRC2 and H3K27me3 *in vitro,* and pluripotency *in vivo*. This *de novo* establishment of H3K27me3 pattern demonstrates that PRC2 recruitment is not exclusively dependent on autonomous epigenetic inheritance[107].

Lastly, RbAp46/48 are also necessary for full activation of PRC2 enzymatic activity[108], but they also partake in formation of other protein complexes such as LINC, NURF, NURD, and SIN3[79].

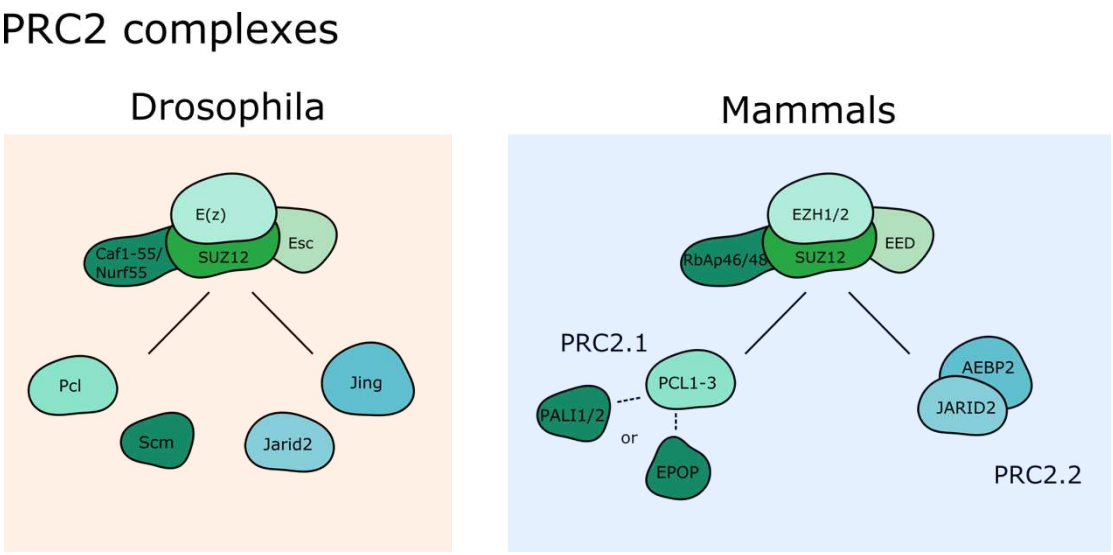

Figure 5. Polycomb repressive complex 2

Facultative subunits of PRC2

Much of the recent activities in the PcG field have focused on the characterization of new subunits, which contribute to the high modularity of PRC2 complex. Indeed, the PRC2 core complex interacts with several facultative factors creating two distinct complexes that are rather mutually exclusive termed PRC2.1 and PRC2.2. Association with these auxiliary subunits have been shown to modulate recruitment or enzymatic activity of PRC2[109]. PRC2.1 contains one of the three Polycomb-like paralogs (PCL1/2/3; also known as PHF1/MTF2/PHF19) and either PALI1/PALI2 or EPOP (C17ORF96). PRC2.2 consists of AEBP2 (Jing homolog) and JARID2 (**Figure 5**). While facultative PRC2.1 constituents interact with core PRC2 in a substoichiometric manner, the two co-factors of PRC2.2 are supposed to associate in a more stochiometric manner with the core complex[110].

<u>PRC2.1</u>

PCL proteins

The PCL paralogs contain one Tudor domain and two PHD domains in the N-terminal part. MTF2 and PHF19 are necessary for PRC2 recruitment to subsets of PcG target genes in mESCs. PCL2 was reported to bind to unmethylated DNA. Disruption of PCL2 leads to global reduction in PRC2 and H3K27me3

occupancy[111,112]. Short-hairpin RNA (shRNA)-mediated PCL2 depletion tend to stabilize the expression of pluripotent factors (Oct4, Nanog, and Sox2) in mESCs, and consequently disturbed cell fate commitment upon differentiation. However, the enhanced stem cell factor were likely secondary effects as chromatin immunoprecipitation (ChIP) demonstrated they were not direct target of PCL2[113]. Similarly, PHF19 is required for PRC2 binding to target genes, where its loss decreases occupancy of PRC2 and H3K27me3 on target sites[114,115]. PHF1 enhances catalytic activity of PRC2 as suggested by *in vitro* histone methyltransferase (HMTase) assay[116]. Knockdown of PHF1 results in de-repression a subset of HOX genes in culture cell system, indicating a contribution to gene silencing[116,117]. Recent data also suggested the winged-helix domain of PCLs extends the residence time of PRC2 on chromatin[118]. Interestingly, the Tudor domains of PHF1 and PHF19 express high affinity to H3K36me3, a mark associated with actively transcribed loci. However, the PCLs do not actually localize on H3K36me3-enriched regions in static states. It is likely that they establish transient contact with H3K36me3 that allows recruitment of PRC2 to initiate gene silencing[115,119].

EPOP

EPOP associates with PRC2 in a mutually exclusive manner with PALI1/PALI2[120]. EPOP interacts with PRC2 complex and the Elongin BC heterodimer. This module interacts with POL II and stimulates transcriptional elongation[121]. EPOP occupies a vast majority of PRC2 target sites and would bridge ElongBC to PRC2. Upon EPOP deletion, lowly-expressed PcG genes are downregulated in concert with decrease of H3K4me3 and POL II enrichment, consistent with a potential positive role for EPOP in transcriptional regulation. The physiological significance of maintaining low expression of PcG target genes is unknown. Of note, expression of key pluripotency factors and differentiation markers remains largely unchanged upon depletion of EPOP, indicating that EPOP does not disrupt cell identity in mESCs[122,123].

PALI1/PALI2

PRC2-associated LCOR isoform 1 (PALI1) and PALI2 are encoded by LCOR and LCORL gene loci, respectively. They share conserved PIP domains that interact with PRC2. PALI1 also interacts with G9a co-repressor complex and deubiquitinases USP11 and USP22. PALI1 promotes PRC2 catalytic activity *in vitro*

as shown by HMTase assay. Deletion of PALI1 reduces H3K27me3 on PRC2 targets in mESCs. Moreover, the homozygous mutant embryos show global reduction in H3K27me3. PALI1 deletion is embryonic lethal indicating that it is essential for mouse development[124].

PRC2.2

AEBP2

AEBP2 is a zinc-finger protein that exists in two isoforms corresponding to the embryonic and adult isoform, respectively[125]. Biochemical analysis shows that AEBP2 exclusively interacts with PRC2.2 components. AEBP2 promotes PRC2 catalysis, but counterintuitively, loss of AEBP2 increases H3K27me3 on target sites. In addition, AEBP2 defines the composition of PRC2.2, as its deletion creates a hybrid of PCL2-JARID2-PRC2. Mice harboring homozygous AEBP2 mutant display anterior transformation, a TrxG phenotype, with reported post-natal lethality[124,126].

JARID2

JARID2 contains a JmjC domain often associated to histone demethylase activity. However, the critical catalytic residues are not conserved, thereby the protein does not contain any enzymatic activity. Its ARID and zinc-finger domains are likemy involved in DNA binding, and the N-terminal region was reported to interact with nucleosomes. In vitro assays show that JARID2 stimulates the methyltransferase activity of PRC2, and that AEBP2 and JARID2 synergistically promote PRC2 activity[127,128]. Besides, JARID2 is methylated by PRC2. This mark is in turn recognized by EED to triggers an allosteric activation of PRC2 similarly to the mechanism similar described for H3K27me3[129]. JARID2 also contains a ubiquitin interaction motif that recognizes H2Aub, providing a crucial bridge that links to PRC1-mediated recruitment of PRC2. This interaction establishes the crosstalk between PRC1 and PRC2 and reinforce Polycomb domain formation[57,58]. Loss of JARID2 is embryonic lethal around E13.5[127]. It is proposed that JARID2 silencing inhibits the binding of PRC2 core components, thus reducing H3K27me3 and disrupting cell differentiation[130].

PRC2.1 and PRC2.2 synergistically coordinates PRC2 function

Genome-wide studies show global co-localization of PRC2.1 and PRC2.2 in mESCs, suggesting that they regulate common sets of genes. PRC2.1 recruitment would depend on the DNA binding activity of MTF2. In contrast, the occupancy of PRC2.2 is supposed to rely on PRC1 as loss of H2Aub profoundly reduces PRC2.2[131]. Deletion of PRC2.1 subunits PCLs evicts core PRC2 components as well as partial loss of PRC2.2 enrichment, leading to reduction of H3K27me3. On the contrary, removal of PRC2.2 has little effect on PRC2.1 localization and H3K27me3 enrichment. However, combinatorial silencing of both PRC2.1 and PRC2.2 leads to global dysregulation of PRC2 and H3K27me3, suggesting PRC2.1 and PRC2.2 interchangeably contribute to Polycomb silencing at same set of genes[132,133].

To understand the respective contribution of PRC2.1 and PRC2.2 during differentiation, recent study implements auxin-inducible degron system to rapidly deplete PRC2.1 subunit MTF2 or PRC2.2 subunit JARID2 throughout the course of cytokine-induced neural lineage commitment. Depletion of either component results in aberrant gene expression and thereby compromises differentiation. Although MTF2 and JARID2 extensively co-localize in neural progenitor cells (NPCs), distinct sets of genes are de-repressed in response to their respective degradation. Hence, while in pluripotent state the facultative PRC2 subunits act redundantly, both PRC2.1 and PRC2.2 are independently required for proper orchestration of gene expression during fate specification[134].

Other Polycomb complexes

Pho repressive complex (Pho-RC)

In Drosophila, current model suggests that PcG proteins are recruited to their target genes (e.g. Hox loci) by an array of specific cis-regulatory elements, which together form a Polycomb repressive elements (PRE)[135]. Pleiohomeotic (Pho) encodes a *bona fide* DNA-binding protein that binds to PREs[136]. Pho forms a complex with PcG protein Sfmbt termed Pho repressive complex (Pho-RC). Surprisingly, while homologs of the two subunits in this complex are found in

vertebrates, they were not reported to function together[137]. In flies, Sfmbt binds to Scm, and Scm interacts with Ph. This PhoRC-Scm-PRC1 interaction is proposed to mediate the recruitment of PRC1 to PREs[138]. Pho also interacts directly with PRC2 subunit E(z) and Esc[55]; and PRC1 subunits Ph and Pc[139]. Thus, it provides a molecular mechanism for the recruitment of PRC1 and PRC2 to target genes, and establishes a hierarchical recruitment model of PcG proteins.

PR-DUB

Genetic screen in flies identified a novel PcG gene, *Calypso*, whose mutation is associated to misexpression of *Ubx* genes, a canonical PcG phenotype[140]. Calypso encodes a deubiquitinase and forms a complex with Additional sex combs (Asx), hence it was termed Polycomb repressive deubiquitinase (PR-DUB)[141]. However, whether Calypso is a *bona fide* PcG protein remains debatable, as its mammalian ortholog suggests otherwise. A full chapter will be dedicated to molecular function of PR-DUB.

Recruitment and propagation of the repressive marks

Throughout the course of development, dynamic sophistication of transcriptional programs allows asymmetrical division, cell fate specification, commitment and end-point differentiation. This requires flexibility and reversibility of the activating and repressive system, and therefore includes *de novo* targeting of PcG system. Once established, the repressive marks need to be propagated and perpetuated to maintain cellular identity. This section will discuss the molecular mechanisms that recruit, reinforce and propagate the PcG system.

Recruitment by DNA elements

As abovementioned, PREs elements serve as docking sites for Pho-RC and consequently recruits PRC1 and PRC2 in Drosophila[27]. Within the broad H3K27me3 domains in *Hox* locus and other developmental genes, putative PREs can be recognized by strong and sharp binding of Polycomb components[142]. Based on the findings in flies, the search of PREs in mammalian system has embarked

but did not report convincing outcome despite perspired efforts. Besides, the transcription factor YY1, the mammalian homolog of Pho, was reported to have non-PcG related properties. YY1 binds to highly expressed genes and its inactivation leads predominately to gene downregulation[143].

Albeit PREs were not identified in mammals, sequence analysis on PRC1/PRC2 shows strong enrichment on CpG island that locates in the proximity of transcription starting sites[46,48,82-84]. PRC complexes display high affinity to hypomethylated CpG island, suggesting mutual exclusivity of DNA methylation and PcG system[144,145]. Indeed, ectopic insertion of short CG-rich motifs autonomously confers PRC2 binding and deposition of H3K27me3[146]. Regarding PRC1, it is shown to bind to DNA through its co-factors. Variant PRC1.1 was proposed to be recruited at unmethylated CpG island by its subunit KDM2B[83-85], though later work actually suggests that the role of KDM2B is to protect CpG island against DNA methylation[147]. The variant PRC1.6 can be recruited through TF E2F6[87,148] (**Figure 6a**)[95].

Recruitment by non-coding RNA

It is well established that the expression of the long non-coding (lnc) RNA Xist triggers the recruitment of PcG machinery on the chromosome where it coats. This process has constituted an important paradigm for lncRNA-mediated PcG regulation[149]. It was initially reported that A repeat of Xist recruits PRC2[150]. However, later studies show that A repeats are dispensable for PRC2 recruitment[151], and that the onset of Xist coating on inactivate X chromosome appear to precede PRC2 binding[152]. Instead, it was proposed that PCGF3/5-PRC1 complexes would be the early pioneers that are recruited by Xist, which trigger pervasive blanketing of H2Aub on inactive X, and subsequently, the recruitment of JARID2-containing PRC2 (**Figure 6b**)[86].

Recruitment and reinforcement by histone modifications

The high prevalence of co-localization of PRC1 and PRC2 on target sites suggests an active crosstalk between the two complexes. It was initially considered that PRC2 first deposits H3K27me3, and this mark is then recognized by CBX-PRC1 (cPRC1), and initiates PRC1 binding to chromatin[33,55]. Later findings suggested that PRC1 binding could take place in the absence of PRC2, showing the existence of alternative model than the hierarchical PRC2-driving-PRC1 recruitment [72,153]. Conversely, H2Aub deposited by vPRC1 provides a recognition site for JARID2-PRC2, which can in turn activate PRC2 independently of H3K27me3[57,58,64]. Together, in this framework, the enzymatic products of PRC1 and PRC2 can act as platforms for mutual recruitment. Of note, both PRC1 and PRC2 have self-reinforcing activities: variant PRC1 recognizes their very own H2Aub marks through the RYBP subunit[80]; PRC2 subunit EED triggers allosteric activation through H3K27me3 binding[102]. These positive "write-and-read" reinforcing activities are proposed to disseminate the marks alongside flanking chromatin (**Figure 6c**)[154].

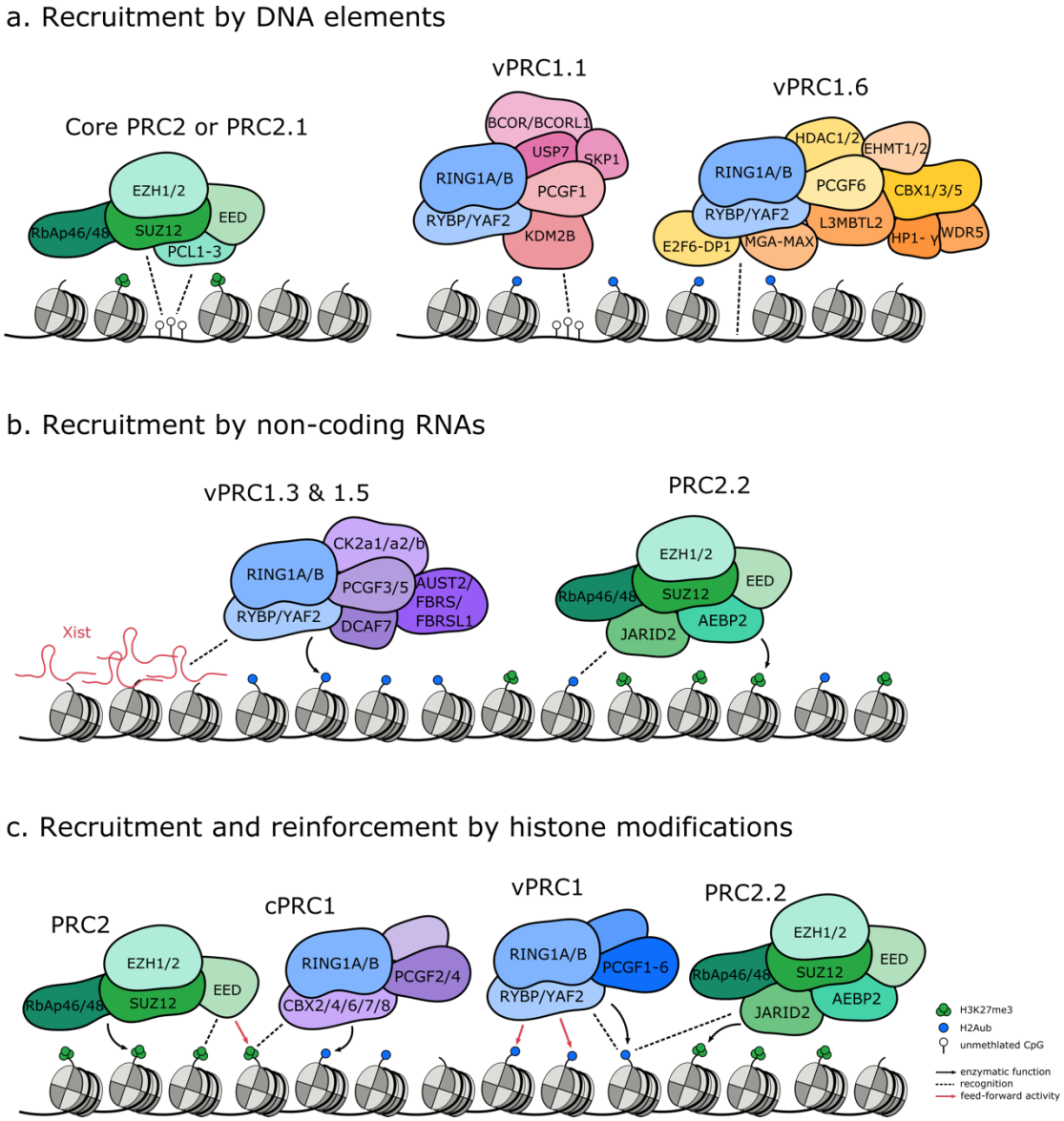

Figure 6. Recruitment and propagation of the repressive marks

Mechanisms refraining Polycomb machinery

While self-propagation proceeds, several mechanisms exist to constrain activities of PRC1 and PRC2 to prevent stochastic gene silencing. These include histone modifications, RNA interactions, DNA methylation and PRC2 interacting factors.

Histone marks

Genome-wide characterizations in flies and mammals show H3K27me3 is depleted at actively transcribed genes marked with H3K4me3 and H3K36me2/3. H3K4me3 is strictly localized at TSSs, while H3K36me2/3 spread throughout gene bodies[48,155]. In vitro studies show that the enzymatic activity of PRC2 is strongly inhibited by nucleosomal substrates H3K4me3 and H3K36me2/3. H3K27me3 and H3K4me3/H3K36me are mutually exclusive when presented on the same H3 tail[156,157]. Genetic modulation of H3K36 methylase shows that H3K36 methylation level negatively correlates with H3K27me3. For instance, loss-of-function (L-o-f) mutation of the Ash1 leads to global reduction of H3K36me2, resulting in increase of H3K27me3[158]. ShRNA-mediated depletion of H3K36 methylase NSD-1 leads to global expansion of H3K27me3 not only on PcG targets sites but also on flanking domains originally marked with H3K27me2 in mESCs[159]. Collectively, these observations suggest that active histone marks demarcate genome-wide domains of H3K27me3.

DNA methylation

PRC1 and PRC2 bind preferentially to unmethylated CpG motif, a frequent feature of promoters across vertebrate species[145]. Overall, CpG methylation and H3K27 methylation show mutual exclusive pattern, and that PRC2 has a reduced affinity to nucleosome that are methylated on CpG *in vitro*[144]. In EMSA assay, KDM2B-PRC1 also showed reduced binding with methylated CpG[85]. Artificial tethering of PRC2 promotes H3K27me3 deposition, but the activity is compromised with increased DNA methylation[146]. Consistently, H3K27me3 spreads *de novo* in DNA methyltransferase-deficient embryonic stem cells[160]. Conversely, recent finding

reports the requirement of PRC2 in maintenance of hypomethylated CpG island, revealing its active role in homeostasis of DNA methylation[161].

RNA

As described above, lncRNA can facilitate the recruitment of polycomb complexes. However, RNA can modulate polycomb function in other ways[162]. PRC2 interacts promiscuously with RNA, despite it shows higher affinity to G-quadruplex structures[163]. It has been proposed that nascent RNAs competes the binding with nucleosome, displacing PRC2 from its target genes, and thereby resulting in reduction of H3K27me3[163,164]. Biochemical analysis indicates that nascent transcripts from active genes binds to PRC2, of which RNA interacts with EZH2 and allosterically inhibit its activity[165,166]. Taken together, this view supports a model in which nascent transcripts locally repel PRC2 from active loci to guarantee proper transcriptional activity.

EZHIP (EZH1/2 inhibitory protein)

A recently identified co-factor of PRC2, CXORF67, inhibits PRC2 methyltransferase activity. It was therefore named EZH1/2 inhibitory protein (EZHIP). EZHIP is highly tissue-specific, being expressed predominantly in gonads, placenta and to a lesser extent, in brain. Deletion of EZHIP in mice leads to a global increase in H3K27me2/3 during gametogenesis, and compromises female fertility[167]. EZHIP is highly expressed in pediatric cancer posterior fossa type A (PFA) ependymomas and diffuse midline gliomas[168]. It remains nonetheless elusive how EZHIP inhibits PRC2 catalysis, potentially mimicking the H3K27M oncohistones that block EZH2/1 activity[169,170]. However, its precise mechanism of action and its role during gametogenesis require further investigation.

Interdependency and Independence of PRC1 and PRC2

Fundamental principles of PcG proteins recruitments have evolved and been revisited over the past years. It is now clear that, genomic occupancy patterns of PRC1 in flies and mammals are far less dependent on PRC2 than previously believed. In flies, PRC1 recruitment does not require PRC2 binding to PREs. On the contrary, PRC1 occupancy is a prerequisite for PRC2 localization[171]. Similar observation has been reported in mammalian cells: while PRC2 inactivation has little effects on PRC1 occupancy, vPRC1.1 deletion leads to displacement of PRC2, and consequently, reduction of H3K27me3 at a subset of targets[64]. More strikingly, recent *in vivo* studies report that preventing H2Aub deposition by PCGF1/PCGF6 depletion during oocyte maturation leads to loss of H3K27me3 in a series of Polycomb-imprinted genes, which are irreversibly inherited by the embryos and causes premature activation of developmental genes. Upon fertilization, dynamic of H2Aub domain formation precedes those of H3K27me3 during maternal-to-zygotic transition, suggesting that PRC2 follows the pattern established by PRC1. By contrast, conditional EED-KO in oocytes and maternal EED-KO embryos have little consequences on H2Aub domains even though both PRC1 and PRC2 depletion perturb proper zygotic activation[172,173]. Taken together, these observations indicate that, at least in early developmental stage, PRC2 manifests high dependency on PRC1, but not vice versa.

PRC1 is recruited to enhancers independently of PRC2

Since PRC1 can autonomously binds to its target sites, whether it entails specific functions arouses great interests. In order to tackle this question, Kloet and colleagues dissect the dynamic of PRC1 and PRC2 complexes during differentiation to NPCs. While binding of PRC2 subunits and H3K27me3 are greatly lost on target genes, PRC1 retains its binding on most of target sites. More importantly, an appreciable number of RING1B is gained in NPCs, with considerable new peaks occupying active promoters and enhancers devoid of H3K27me3, suggesting redistribution of RING1B occupancy during differentiation[174]. Similarly, RING1B is recruited to active enhancers in fly developing eye discs, whose enrichment potentially plays a role in the

establishment of long range enhancer-promoter contacts[175]. In various breast cancer cells, RING1B localizes largely on distinct active enhancers. This study actually shows that RING1B-bound enhancers associate with oncogenic activities[176]. All of the above studies suggest the existence of distinct PRC1 patterns in differentiating tissues or in cancer cells comparing to mESCs[176]. However, systematic analysis is required to investigate the physiological significance of enhancer-bound PRC1.

PRC1 represses genes independently of PRC2

Considering that PRC1 binds to distinct subsets of promoters and enhancers in somatic tissues, it is likely that PRC1 functions through PRC2-indepedent pathways. PRC1 is required for skin epithelium morphogenesis and stem cell specification[62], and conditional KO (cKO) of RING1A and RING1B in mouse epidermal tissue cause skin fragility in neonatal pups. However, the same phenotype is not observed in EED cKO pups, implying that PRC1 plays a more important role than PRC2 in maintaining epidermal integrity. PRC1 inactivation perturbs a larger number of genes compared to PRC2-KO, the upregulated genes are related to cell adhesion and cytoskeleton organization. All of the above observations convey the idea that PRC1 autonomously regulates transcription independently of PRC2, whether this process is a general phenomenon requires further study[177].

Can PRC2 repress transcription independently of PRC1?

While H3K27me3 is clearly necessary for the functionality of PRC2, the precise molecular mechanisms on how it confers the repressive signal remains only partially understood. A long-standing view is that H3K27me3 serves as docking site for cPRC1, which in turns compacts chromatin to restrict the access of transcriptional machinery. The problem of this hypothesis is that PRC2 occupies genomic regions lacking PRC1[46], and that inactivating cPRC1 does not cause target genes de-repression[49]. Recently, a couple of chromatin factors have been

proposed to possess modules that recognize H3K27me3 and could confer transcriptional inactivation.

CDYL/CDYL2 are H3K27me3 readers, which repress PRC2 target genes

Quantitative proteomic analysis using modified histone peptides or recombinant nucleosomes demonstrate that chromodomain proteins CDYL and CDYL2 recognize H3K9me3 and H3K27me2/3[144,178]. CDYL recognizes H3K9me3 and interacts with G9a, serving as a bridge between the H3K9 methylase and its enzymatic product. CDYL incorporates into REST and HDAC1 co-repressor complexes, suggesting a role in gene repression as further supported by luciferase reporter assay[179,180].

Potentially, CDYL can also act as a bridge between PRC2 and H3K27me3. Endogenous EZH2 and SUZ12, although with many other proteins, could be efficiently co-immunoprecipitated with CDYL in MCF-7, U2OS and HEK293T cells[181,182]. CDYL enhances PRC2 enzymatic activity toward oligonucleosomes *in vitro*. In MCF-7 cells, ChIP-qPCR shows that CDYL localizes at several PRC2 targets, and that short interfering RNA (siRNA)-mediated depletion of CDYL reduces PRC2 and H3K27me3 on target sites alongside with gene de-repression. These indicate that CDYL might participate in PRC2-mediated repression and it might form a positive feedback loop to facilitate propagation of H3K27me3 along the chromatin[181].

In addition, CDYL is crucial for reestablishment of the repressive marks during DNA replication. CDYL associates with chromatin assembly factor 1 (CAF1) and replicative helicase MCM complex, and is proposed to recruit EZH2 to replication forks to promote the deposition of H3K27me3 on newly assembled H3. U2OS cells bearing CDYL ablation are reported to have delayed H3K27me3 loading on new histones with impeded early S phase progression[182]. Nonetheless, it should be kept in mind that CDYL is rarely identified in immunoprecipitation from PRC2 components, and that U2OS cells is an arguable model as its H3K27me3 lvel by default is close to background level due to aberrant expression of EZHIP[167].

Interestingly, X chromosome inactivation is associated with coating of H3K27me3, and that CDYL is recruited to Xist-coated chromosome upon differentiation. CDYL

co-localizes with both H3K27me3 and H3K9me2, but its recruitment is lost in the absence of H3K27me3 (EED-KO cells). Mouse ESCs depleted for CDYL by shRNA display a perturbed cell differentiation, blockage of pluripotency with massive cell death, implying an important role in maintaining proper cellular function[183].

Taken together, these studies indicate that CDYL/CDYL2, readers of H3K27me2/3 and H3K9me3, participate in gene repression potentially by stimulating the methyltransferases themselves and/or by recruiting other co-repressor complexes.

BAHD1 and BAHD2 recognize H3K27me3 and repress PRC2 target genes

BAHD1 is first described as a heterochromatin factor that interacts with heterochromatin protein HP1, H3K9 methyltransferase SETDB1 and HDAC5. However, immunofluorescence (IF) in HEK293 cells shows that BAHD1 co-localizes with both inactive X chromosome and H3K27me3 puncta, but rarely with H3K9me3. BAHD1 contains a C-terminal BAH module whose removal compromises co-localization with H3K27me3[184]. Biochemical analysis reveals that the BAH domain binds to H3K27me3 with high affinity but not for H3K9me3. These suggest that BAH domain "reads" specifically H3K27me3[185]. Genome-wide studies report that Flag-tagged versions of BAHD1 and its paralog BAHD2 (also known as BAHCC1) bind to PRC2 targets. However, in those assays, strong enrichment of BAHD1/2 at house-keeping genes is also observed, prompting for follow-up validations to consolidate the detailed mechanistic pathways in BAHDs-mediated regulation. Nevertheless, depletion of either paralogs contributes to misexpression of PRC2 target genes, suggesting a potential participation in PRC2-mediated gene repression[184,186,187]. Tandem affinity chromatography purification following mass spectrometry revealed that BAHD1 interacts with histone deacetylase complexes HDAC1/2, MIER1/2/3 and KAP1; heterochromatin co-factors HP1γ/β, G9a and PPP2R1A; and chromatin remodeler RUVBL2. Intriguingly, BAHD1 also interacts with CDYL1/2[188]. However, it remains unclear whether BAHDs and CDYLs, these seemly irrelevant proteins, promote silencing in an independent or collaborative fashion.

More interestingly, forward genetic screen in filamentous fungus *Neurospora crassa* identifies a BAH-PHD-containing protein termed EPR-1 (effector of

Polycomb repression 1), which associates with H3K27-methylated chromatin. Loss of EPR-1 leads to the upregulation of H3K27-regulated genes without affecting H3K27 tri-methylation. Orthologs of EPR-1 are identified in an array of eukaryotes, suggesting evolutionarily conserved BAH-PHD domain protein in primitive Polycomb system. Remarkably, core PRC1 components are absent or so far not yet discovered in fungal lineages, providing a contextual evidence that ancient BAH-PHD-containing protein may mediate H3K27me3-dependent repression in a PRC1-independent manner[189].

PRC1 and PRC2 have autonomous yet overlapping functions in repression

It is long considered that PRC1 and PRC2 coordinate repression of target genes based on the evidence of (1) extensive overlap within target genes, (2) mutual recruitment of the two complexes and (3) similar large-scale de-repression in the absence of individual component. Leeb and colleagues observed that a subset of genes is derepressed only upon removal of both PRC1 (RING1B) and PRC2 (EED) in mESCs, and therefore proposed a model where PRC1 and PRC2 redundantly silence target genes[52]. Whether respective complexes independently regulate genes remains speculative in reason of the lack of appropriate model to address the question: deletion of PRC1 in cells impair cell proliferation and promotes differentiation.

A recent *in vivo* study focusing on mouse epidermal development reported observations similar to those of in mESCs. Conditional ablation of PRC1 or PRC2 in epidermal progenitors disturb the integrity of epidermis. However, PRC1/2 double cKO mice suffer severe skin defect which is not observed in neither of the respective KO mice. While PRC1-KO introduces higher level of de-repression that PRC2-KO (1353 genes vs. 551 genes), PRC1 and PRC2 ablation in combination elicits much higher extent of gene de-regulation (2492 genes), indicating the two complexes have overlapping functions in repressing target genes, consistent with a model of redundancy between PRC1 and PRC2 [190].

Interestingly, in both *in vitro* mESCs and *in vivo* mice studies, a set of genes is sensitive exclusively to the loss of one complex regardless of the other's status,

which strongly suggests both PRC1 and PRC2 can also autonomously represses their targets despite being coupled at the same genomic targets. This supports a model whereby PRC1 and PRC2 action in distinct mechanisms to achieve maintenance of gene repression. While it is not fully unexpected for PRC1[177], this is a relatively unexplored question for PRC2. A part of my thesis tackles the independency and interdependency of PRC1 and PRC2. By Integrating with genome-editing, transcriptomic analysis and epigenomic profiling, I have investigated the molecular mechanisms behind the interplay entwining the two complexes (see **Results**).

The BAP1 complex

The mammalian BAP1 complex consists of BAP1 and one of the ASXL paralogs (ASXLs). Their invertebrate orthologs are Calypso and Asx, respectively. On the functional level, BAP1 is a nuclear deubiquitinase that removes H2Aub. In Drosophila, the Calypso-Asx complex was proposed to be a component of the Polycomb machinery. However, as will be discussed below, recent observations have challenged this view. In the past years, abundant literatures have focused on deciphering its partners, functions, substrates, and more importantly, its contribution to cancer suppression. In this chapter, I will discuss BAP1 from the history of its discovery to current knowledge.

The discovery of BAP1 complex

BAP1 was initially described as a nuclear ubiquitin carboxyl-terminal hydrolase (UCH) that interacts with BRCA1 in a yeast two-hybrid system. Together with BRCA1, BAP1 exhibits anti-proliferative activity in breast cancer MCF-7 cells[191]. However, later studies did not confirm this piece of data as BAP1 does not regulate BRCA1 ubiquitination[192], and that the tumor suppressive activity of BAP1 is independent of BRCA1[193].

Approximately a decade after the identification of BAP1 in mammals, a genetic screen in *Drosophila* proposed the mutation of *calypso* as a Polycomb phenotype, ascribing from its widespread misexpression of *Hox* genes[140]. Calypso forms a deubiquitinase complex with Asx[141], a gene whose mutation shows posterior transformations but also enhances anterior transformation (TrxG phenotype), suggesting an involvement for both activation and repression of homeotic loci[194,195]. *Asx* has three mammalian homologs: Additional sex combs-like (ASXL1/2/3)[196-199]. ASXL1 and ASXL2 are ubiquitously expressed, whereas ASXL3 expression is restricted to brain tissues (Human Protein Atlas). Calypso-Asx deubiquitinates H2Aub *in vivo*, but not H2Bub, and is recruited altogether with Ph (PRC1) and Pho (Pho-RC) complexes to PREs of a large set of PcG target genes. Based on the phenotype of homeotic de-repression, the heterodimer Calypso-Asx

and its vertebrate ortholog BAP1-ASXLs were termed Polycomb repressive deubiquitinase (PR-DUB)[141].

The partners of BAP1 complex

BAP1 forms the mandatory complex with ASXLs: the presence of one protein is required for the stability of the other. For example, in U2OS osteosarcoma cells, siRNA-mediated BAP1 knockdown destabilizes ASXL2[200]. Reciprocally, ASXL1 or ASXL2 knockdown reduces BAP1 abundance, and concomitant depletion completely destabilizes BAP1, suggesting functional redundancy of the ASXL paralogs[200]. Indeed, we observed a dramatic reduction of BAP1 protein abundance in ASXL1/2 double-KO cells[201]. Of note, BAP1 modulates mono-ubiquitination of ASXLs which in turn promoted the deubiquitinase activity of the complex (see **Results**)[202].

Aside from its mandatory partner ASXLs, BAP1 has been found interacting with transcription factors FOXK1/2 and YY1; other chromatin factors KDM1B, MBD5/6, OGT; and cell cycle regulator HCF-1, suggesting a role of BAP1 in gene regulation, chromatin-associated processes and cell proliferation (**Figure 7**)[79,201,203-207].

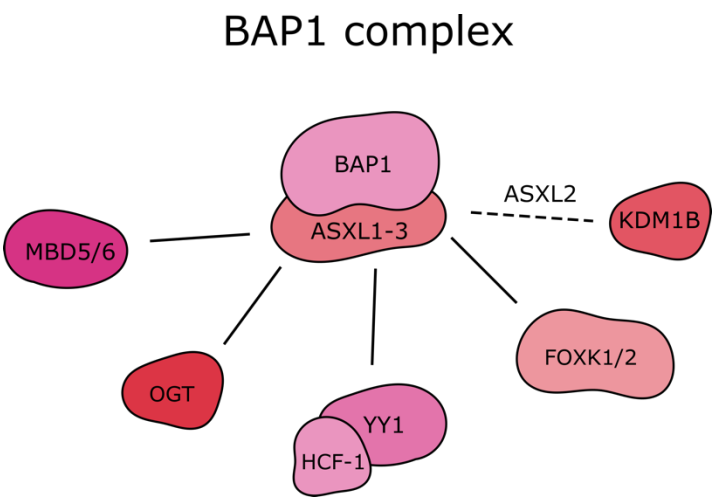

Figure 7. The BAP1 complex

The biological functions of BAP1

Either BAP1 or ASXL1 deletion is embryonic lethal. While BAP1 knockout embryos are reabsorbed around E9.5, ASXL1-KO embryos are no longer viable beyond E19.5 with reported posterior transformation and developmental disorders[207,208]. BAP1 is also frequently mutated in cancers. Viral introduction of BAP1 represses proliferation of breast cancers MCF-7 cells *in vitro*[191], and inhibits tumor growth of non–small cell lung cancer NCI-H226 cells both *in vitro* and *in vivo*[193]. The complex is essential for embryogenesis and it was also classified as a tumor suppressor. In this section, we will focus on the studies dissecting the function of BAP1, notably its H2A deubiquitinating activity.

BAP1 in transcriptional activation

As abovementioned, the nomenclature of BAP1 as a subunit of PR-DUB derived from the phenotype of *Calypso* mutant in Drosophila. Despite being described as part of the Polycomb machinery, it remains largely enigmatic how BAP1 and PRC1, its mechanistic antagonist, contribute concurrently to gene silencing. While this could be attributed to fine-tuned spatial and temporal regulation[141], recent genome-wide studies in vertebrates suggest an alternative point of view. Indeed, we have shown that in human HAP1 cells, BAP1 inactivation leads to visible accumulation of H2Aub in bulk, and that RNA-seq reveals downregulation of a plethora of genes that are related to multiple biological processes including development[201]. Downregulated genes show increase of H2Aub accompanied by concerted increase of H3K27me3. In contrast to previous study reporting that ASXL1 interacts with PCR2[209], we provide evidence that the increased enrichment of H3K27me3 was likely a secondary effect subsequent to transcriptional silencing, as removal of EZH2 did not rescue silencing of these genes. We showed that the enzymatic activity of BAP1 is crucial to activate transcription, as re-expression of the wild-type but not the catalytically dead C91S version of BAP1 rescued gene downregulation. Later studies report similar observations regarding the role of BAP1 in transcriptional regulation[210-212]. In Xenopus, morpholino-mediated BAP1 depletion arrests gastrulation with additional malformations, including axial foreshortening, microphthalmia or anophthalmia. BAP1 depletion

downregulates a cohort of genes related to lineage specification[213], providing evidence of functional conservation of BAP1 in vertebrates. Importantly, we showed that BAP1 is functionally inert in the absence of PRC1. In a PRC1-KO (therefore H2Aub-null) context, silencing BAP1 no longer triggers transcriptional alterations. This establishes the fact that PRC1 is epistatic to BAP1 and that BAP1 promotes transcription most likely through H2A deubiquitination.

BAP1's genomic localization

Considering the broad contribution of BAP1 to transcriptional regulation, several studies attempted to map genome-wide binding sites of BAP1. However, this task turned out to be challenging with little available resources being published. Whether this reflects a very dynamic behavior of BAP1 at chromatin, a distribution with little specificity along the genome or the lack of antibodies working for crosslinking-based chromatin profiling was unclear. Using Flag-tagged BAP1 knock-in system, Dey and colleagues reported BAP1 binds mostly at promoters in bone marrow-derived macrophages. Surprisingly, out of 5731 genes bound by BAP1, only 32 were downregulated in the absence of BAP1[207]. In mouse ESCs, Kolovos and colleagues used in-house generated polyclonal antibody and identify 1614 BAP1 peaks, which localized mostly on active or bivalent promoters. However, this is likely an underestimation, as Flag-tagged version reported 10-fold more peaks in the same study[210]. Wang and colleagues performed ChIP-seq analysis in triple-negative breast cancer CAL51 cells with homemade antibody, and showed that BAP1 is recruited both on promoters and enhancers. They further suggested that BAP1 is required for MLL3 recruitment and that BAP1 loss reduces the occupancy of MLL3 and therefore the deposition of H3K4me1[214]. Altogether those studies suggest an enrichment of BAP1 at a subset of promoters of transcriptionally active genes despite this conclusion was concluded on suboptimal experiments.

In this study, we implemented CUT&RUN-seq[215] and obtained a high resolution map of BAP1 binding sites. It indicates a predominant recruitment at enhancers. This result leads us to reconsider BAP1's mechanism of action (see **Results**).

Does BAP1 have other substrates that H2Aub?

Several proteins have been proposed to be BAP1's substrates, which in turn modulates their stability and function. For example, siRNA-mediated depletion of BAP1 was reported to result in increased ubiquitination of INO80 and consequently leading to its destabilization. Conversely, overexpression of WT but not catalytic mutant BAP1 increased INO80 at protein level. Altogether, the authors concluded that BAP1 participated in INO80-mediated DNA replication processes[216].

In another study, through a siRNA screen targeting human DUBs, BAP1-KD was shown to reduce the abundance of KLF5, a transcription factor highly expressed in ER-negative breast cancers. Re-expression of WT BAP1, but not its inactive mutant, restored KLF5 at protein level. The authors therefore claimed that by stabilizing KLF5, which is required for proper cell proliferation, BAP1 contributes to tumorigenicity and metastasis in breast cancer cells[217]. Similar approaches have led to the conclusion that gamma-tubulin and endoplasmic protein IP3R3 are also BAP1's substrates[218,219].

However, BAP1 inactivation results in substantial transcriptomic changes, and consequently differential protein abundance (including E3 ligases and DUBs)[220,221]. As a consequence, it is quite challenging to distinguish primary targets from secondary effects.

Pathologies associated to mutations of BAP1 or the ASXLs

Several types of cancer have been reported harboring BAP1 mutations. However, this does not overlap with cancers that ASXL1 mutation is reported. We will discuss diseases and malignancies linked to aberrant BAP1 and ASXLs functions.

BAP1

Somatic and germline BAP1 loss-of-function mutations are frequently reported in various cancers including uveal melanoma (UM), mesothelioma, renal cell carcinoma and cutaneous melanoma[222-224]. UM is a rare but aggressive intraocular tumor with high prevalence of metastasis[225,226]. Mutation of either GNAQ or GNA11 is the major driving event, and is followed by subsequent mutations of EIF1AX, SF3B1, and BAP1 in a mutually exclusive manner[227,228]. Local brachytherapy or enucleation provides good control, but half of the patients eventually develop metastasis[229]. Risks of metastasis is linked to monosomy of chromosome 3 and BAP1 inactivation[228]. Since no efficient treatment has been developed, the prognosis of metastasizing UM remains abysmal, with a median overall survival ranging from 4 to 15 months[226].

Mouse models have been established to study BAP1-related pathologies. BAP1 ablation is embryonic lethal around E9.5. Systemic conditional BAP1 deletion gave rise to splenomegaly, myeloproliferative disorder and myeloid transformation, resembling disease that commonly harbors ASXL1 mutation in human[207,230]. Pancreatic ductal adenocarcinoma is a lethal malignancy that have the highest incidence of Kras mutations[231], where heterozygous loss of BAP1 is reported in around a quarter of patients. Studies in mouse showed that deletion or heterozygous loss of BAP1 in pancreas caused tissue damage and pancreatitis with full penetrance[232]. Kras mutation drive pancreatic intraepithelial neoplasia that progress to pancreatic adenocarcinoma with concomitant BAP1 mutation or heterozygous loss, suggesting that BAP1 acts as a gate keeper to tumor development[232,233].

ASXLs

Heterozygous mutations of ASXL1 resulting in premature truncations are frequently reported in subtype of myeloid diseases with poor prognosis, including myelodysplastic syndromes, myeloproliferative neoplasms, chronic and acute myeloid leukemia[234]. The majority of the truncation occurs in exon 12, right

upstream of the PHD finger, a domain predicted to be a histone- or DNA-binding module[235]. It remains debatable how this truncation alters ASXL1 function: haploinsufficient loss-of-function or dominant negative gain-of-function (G-o-f).

On one hand, studies suggest homozygous ASXL1 truncations are L-o-f mutations in leukemia cell lines (such as KBM5 cells), where ASXL1 protein is no longer detectable. It was further reported that shRNA-mediated ASXL1 depletion leads to genome-wide reduction of H3K27me3. Hence, the authors concluded that ASXL1 is required to maintain PRC2 function, and that loss of H3K27me3 causes de-repression of HOXA genes, which subsequently contributes to myeloid transformation[209]. Of note, deletion of ASXL1 indeed results in myelodysplasia in mouse model[208].

On the other hand, recent studies suggested the mutations act as a G-o-f or dominant negative effect. They argue against the reported protein instability of the ASXL1 mutants by showing detectable protein level in mutant form of ASXL1[236]. In addition, ASXL1 mutants enhance the deubiquitinating activity of BAP1 *in vitro*. Expression of hyperactive BAP1-ASXL1^{mut} complexes results in dramatic reduction of H2Aub and subsequent gene de-repression in haematopoietic precursor cell line. H3K27me3 is also reduced but it is likely a secondary effect in response to transcriptional re-activation. In addition, multi-lineage differentiation is impaired, instead, it is skewed toward mast cells differentiation, promoting expansion of myeloid cells in a bone marrow transplant model[237,238]. Taken together, those latter studies suggest that ASXL1 mutations act as dominant G-o-f mutations that perturb cell identity.

ASXL2 mutation is rarely documented in carcinogenesis, with the exception of a specific type of acute myeloid leukemia that harbors unique RUNX1/RUNX1T1 chromosomal translocations and where high frequency of heterozygous ASXL2 frameshift mutation is reported[239].

ASXL1 and ASXL3 mutations have also been reported in developmental disorders. De novo truncation of ASXL1 has been reported to account for approximately half of the cases of “Bohring-Opitz syndrome” (OMIM 605039). This syndrome is characterized by craniofacial abnormalities, severe intellectual and growth

retardation, feeding difficulties and failure to thrive[240]. Bainbridge and colleagues showed that ASXL3 truncating mutation causes severe developmental disorders that share similar phenotypes with Bohring-Opitz syndrome. Clinical features of "Bainbridge-Ropers syndrome" (OMIM 615485) include severe feeding difficulties, failure to thrive, neurologic abnormalities and developmental delay[241,242]. Shared phenotypes between ASXL1 and ASXL3 L-o-f mutations suggest functional similarities of the paralogs in brain development.

Transcriptional regulation by Trithorax group of proteins

Identification of Trithorax group of proteins

The PcG and TrxG genes encode chromatin modifiers that show opposing effects on expression of *Hox* genes. In early 80s, a homeotic mutation was characterized under the name *trithorax* (*trx*) with reported severe segmental defects including halteres transforming to wings and general anterior transformation. The *trx* mutation phenocopies l-o-f mutation of *Hox* genes. Interestingly, other mutants also mimic this phenotype, including *fs(1)h*, *ash1* and *ash2*. These genes were therefore collectively termed as trithorax group of genes[243]. Many other proteins have now been classified as TrxG based on several other less stringent criteria such as sequence homology, biochemical activities, and effects on transcription. Given that transcription is a multi-step process that involves many factors, the TrxG proteins are expected to be more heterogeneous than PcG proteins[27]. Indeed, biochemical studies have revealed that TrxG activate transcription via a wide variety of mechanisms such as covalent modifications on histones, chromatin remodeling, formation of cohesion complex, assembly of mediator module and recruitment of POL II[7,27,244,245]. The histone modifying and chromatin remodeling activities that are better studied will be discussed below.

Histone modifying complexes

SET1/COMPASS-like family

Methylation of H3K4 is correlated with transcription. Tri-methylation of H3K4 is enriched at actively transcribed genes, and is also found to co-localize with H3K27me3 at poised promoters in pluripotent stage in mammals[53]. The mono-methylation of H3K4 (H3K4me1) has a different distribution that it demarcates enhancers[246]. The H3K4 methyltransferase complex was first isolated in yeast, but it is found also conserved across species. In yeast, the enzymatic subunit Set1 forms the COMPASS complex which is responsible for all forms of H3K4 methylation (H3K4me1/2/3). The yeast Set1 has 3 homologs (*dSet1*, *trx* and *trr*) in Drosophila and 6 homologs (SET1A, SET1B and MLL1-4) in mammals. SET1A/B

and dSet1-containing complexes are closely related to yeast COMPASS complex, while MLL1/2 (homolog of trx) and MLL3/4 (homolog of trr) assemble into COMPASS-like complexes but with distinct core subunits[247] (**Figure 8**).

SET1A/B-containing complexes are in charge of the bulk H3K4me2/3[248]. MLL1/2 complexes do not contribute to global H3K4me3, but are proposed to tri-methylate H3K4 at specific subset of genes such as bivalent genes and Hox loci in mESCs. Loss of MLL1/2-specific core complex MENIN leads to radical reduction of H3K4me3 across entire Hox loci[249,250]. MLL3/4 are the major methyltransferases mediating H3K4me1 at enhancers. MLL3/4 complex contains UTX, a H3K27-specific demethylase, suggesting a role in antagonizing PRC2 activity[251]. Similar functional partitioning of these distinct complexes has been found in Drosophila[245,252,253]. Loss of either homologs in drosophila and in mammals is embryonic lethal, indicating non-redundant roles of these methyltransferase[247].

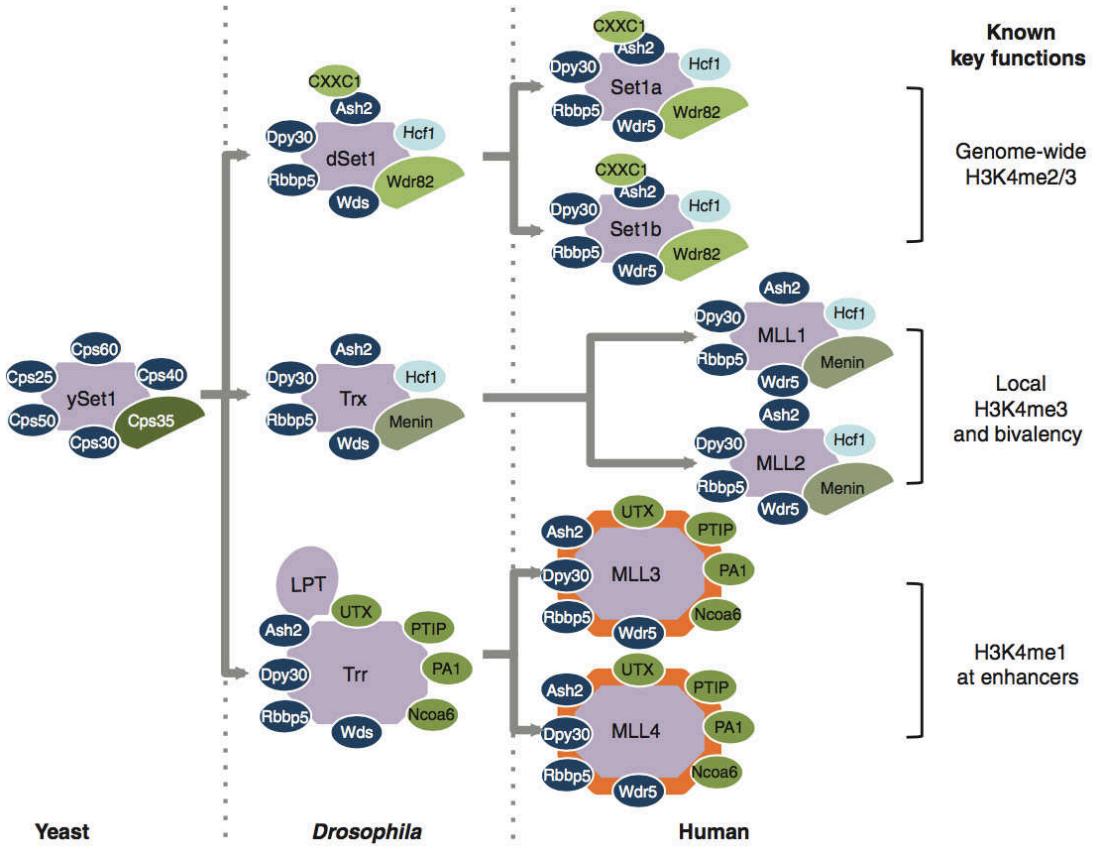

Figure 8. SET1/COMPASS-like family (Sze 2016)

Ash1

Ash1 is one of the first pioneers TrxG to be identified in genetic screens. Its mutant clones show similar phenotypes to those of Trx mutants. While ash1 is required for *Hox* gene expression, this seems to be PcG-dependent, as ash1 and trx are not required for *Hox* expression in PcG mutant. This suggest that some TrxG proteins such as Ash1 and Trx function is to block establishment of PcG-mediated repression[27]. Ash1 and its mammalian counterpart ASH1 specifically mono- and di- methylate H3K36[254]. These marks are enriched within the gene body of actively transcribed genes but also in a more pervasive manner across the genome. However, their function remains unclear to date. Biochemical studies suggest pre-existing H3K36 methylation inhibits PRC2 activity in vitro[156,157], and that genome-wide pattern of H3K27me3 anti-correlates those of H3K36 di- and tri-methylation[48,159]. This leads to the hypothesis that the H3K36-methyl marks restrict expansion of PRC2.

Nucleosome-remodeling complexes

Two TrxG proteins, *brahma* (*brm*) and *kismet* (*kis*), encode proteins that are associated with different ATP-dependent chromatin-remodeling complexes. Brm is homologous to yeast swi2 (snf2), which encodes a ATPase subunit in yeast SWI2/SNF2 complex, a complex is originally considered involved in yeast mating-type switching (*switch* [*swi*] genes) and sucrose-fermentation (*sucrose nonfermenting* [*snf*] genes), but which was later found participating in general transcriptional activity[255]. Kis belongs to the chromodomain-helicase-DNA binding CHD7 subfamily[244]. Both of the complexes modify nucleosomes but with different mechanisms: SWI/SNF family complexes alter the chromatin accessibility by repositioning nucleosomes, ejecting octamers or evicting histone dimers, allowing the binding of TFs and general transcriptional machinery to DNA. CHD family complexes partake in deposition of histones, maturation and spacing of nucleosomes [256].

BRG1 (SMARCA4) and BRM (SMARCA2) are the homologs of Brm[257,258]. They encode the catalytic subunits of mammalian SWI/SNF (mSWI/SNF) complexes (also known as BRG1/BRM-associated factor (BAF)) in a mutually exclusive manner[259]. Together with the two catalytic homologs, the products of 27 other genes encoding the subunits of mSWI/SNF complexes that can assemble into three distinct complexes termed canonical BAF, polybromo-associated BAF (PBAF) and noncanonical BAF (ncBAF). The three complexes comprise common as well as complex-specific subunits, allowing them to take parts to highly specialized functions[255,260]. Extensive exome and genome-wide sequencing have revealed that the subunits of SWI/SNF are mutated in approximately 20% of human cancers. Noteworthy, mutations of respective subunits are found in particular types of cancers, suggesting non-redundant roles in maintaining normal physiological functions (**Figure 9**)[261].

BRG1 deletion is embryonic lethal in mice, occurring around preimplantation stage[262]. In mESCs, depletion of BRG1 by shRNA reduces self-renew, changes colony morphology, alters expression level alkaline phosphatase, and leads to gradual loss of expression of pluripotency factors including OCT4, SOX2 and NANOG[263]. Another study in mESCs suggests that BRG1 maintains pluripotency by safeguarding STAT3 binding. STAT3 acts downstream of LIF cytokine and is important for pluripotency. Loss of BRG1 reduces chromatin accessibility at STAT3-binding sites and is associated to concomitant loss of STAT3 occupancy. Of note, H3K27me3 is increased at BRG1-binding sites in BRG1-cKO cells and the transcriptional silencing can be partially to fully restored upon knockdown of SUZ12. This suggests the potential antagonism between BRG1 and PRC2[264]. Chemically-induced recruitment of BAF complex on the OCT4 locus causes rapid eviction of both PRC1 and PRC2 and their histone marks within minutes in mouse embryonic fibroblasts (MEFs). This histone eviction requires the catalytic activity of BRG1[265]. Recent studies also suggest the importance of BAF complexes in persistently maintaining chromatin accessibility for TFs binding, as acute BRG1 dissociation from chromatin by either small molecule inhibitor or inducible protein depletion reduces accessibility on active promoters and enhancers. This results in decreased TFs occupancy and subsequent gene silencing. Together, these raise the idea that chromatin remodeling is continuously required for proper gene expression[266,267].

Kis encodes a large protein that is highly related to human CHD7 protein. CHD7 modifies chromatin accessibility and is the causative mutation of CHARGE syndrome: a syndrome with multiple developmental defects[268]. In Drosophila, Kis is required for transcriptional elongation. Loss-of-function of Kis mutation reduces POL II localization on salivary gland polytene chromosomes. In addition, Kis mutant shows reduced Ash1 occupancy, which is accompanied by loss of H3K36me2 and gain of H3K27me3 on polytene chromosomes. Therefore, this suggest that Kis counteracts PcG repression[158,269].

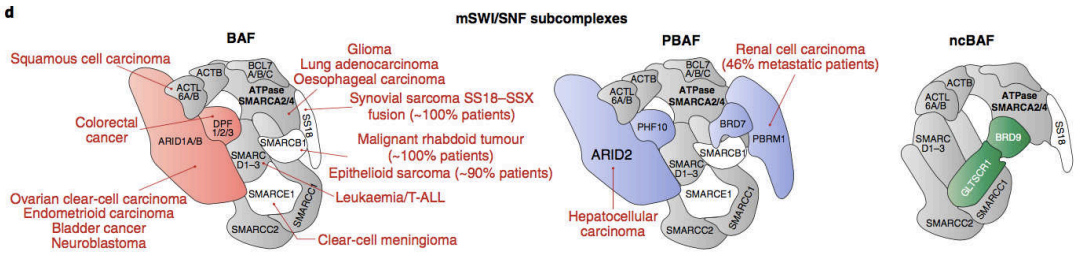

Figure 9. Mammaliam SWI/SNF complex (Valencia and Kadoch 2019)

Transcriptional regulation by enhancers

Characteristics of enhancers

Enhancers are segments of short DNA elements that activate gene expression at long-range distance. Typically, the core promoter *per se* is not efficient to sustain transcription and exhibits low basal activities[270]. Hence, arrays of specific enhancers contribute to gene control and fine-tune the spatiotemporal expression program. By engaging in physical contacts with their cognate promoters, enhancers facilitate transcription regardless of their distance, position and sequence orientation to promoters[271-273]. The first sequence identified as enhancer derived from a 72 bp-repeat found in SV40 genome, it stimulates the expression of human β-globin gene when inserted to arbitrary positions relative to the gene promoter[274]. Shortly after, similar enhancer sequences were found in metazoan genomes[275].

Typically, the size of enhancers ranges from 100 to 1000 bp. It is estimated that mammalian genomes contain hundreds of thousands of enhancers, controlling cell-specific gene expression program[273]. They are composed of various types of short motifs that are recognized by specific transcription factors. The distribution of diverse motifs across individual enhancer sequences articulates the combinatorial repertoires of miscellaneous TFs, including both general and lineage-specific TFs[10]. As chromatin is by default compacted[276], it acts as a hurdle for factors to access to DNA and to exert their activities. The "pioneer" TFs, notably master regulatory TFs, recruit chromatin remodelers (e.g. BRG1 and CHD7), which reinforce opening of local chromatin to allow accessibility for other chromatin factors[277,278]. Subsequently, open chromatin facilitates the recruitment of several other co-activators, in particular histone modifiers (e.g. MLL3/4 and acetyl transferase CBP/p300), bromodomain-containing protein BRD4, cyclin dependent kinase P-TEFB and Mediator complex[271,272]. MLL3/4 deposit H3K4me1 while the acetyltransferase CBP and its paralog p300 catalyze H3K27ac on flanking histones. Of note, H3K4me1 marks up general enhancers, whereas H3K27ac is a hallmark of active enhancers[246,279,280]. The highly cooperative "TFs and co-activators" network ultimately promotes a permissive chromatin environment, favoring the activity of the RNA polymerase II. Enhancers that are more active tend to be depleted in nucleosome and are associated more frequently

with co-activators[272]. Genome-wide mapping reveal substantial occupancy of POL II at both poised and active enhancers[281,282]. Active enhancers are able to drive local transcription, whose products, coined as enhancer RNAs (eRNAs), are typically short (~200nt), mostly non-polyadenylated and susceptible to exosomal degradation (**Figure 10**)[281,282]. These properties in turn have been employed to predict putative enhancers on genome-wide level in a variety of cellular contexts by high-throughput sequencing for detection of chromatin accessibility, enrichment of histone modifications, occupancy of TFs/co-activators binding and productivity of eRNAs (see below)[271,273].

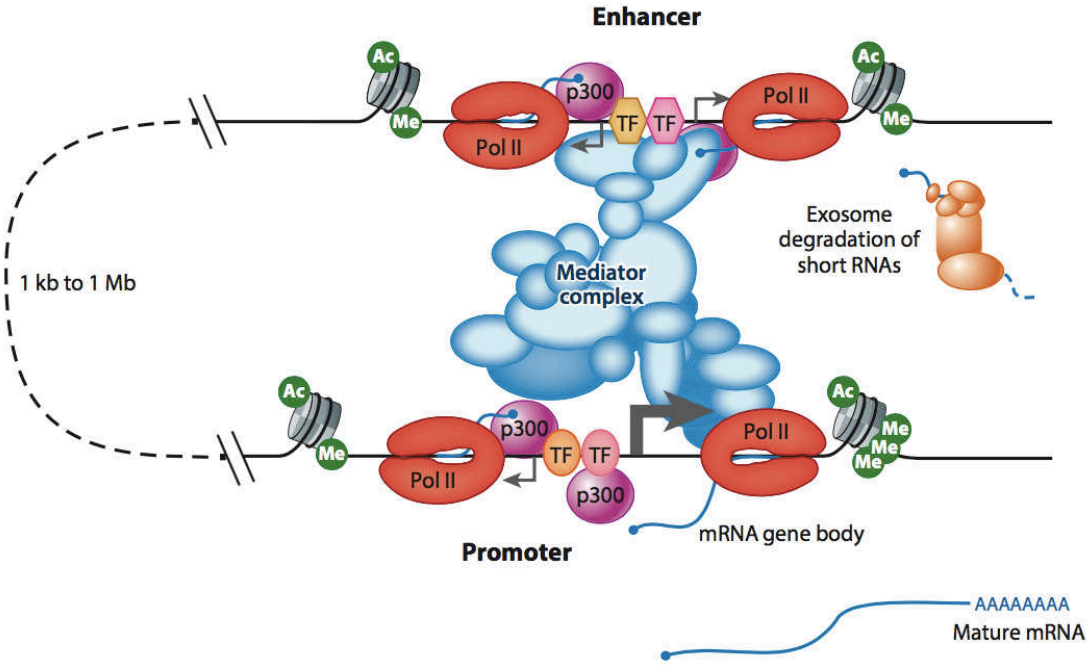

Figure 10. Characteristic of transcriptional active enhancers (Field 2020)

Identification and prediction of enhancers

Early works in identifying enhancers rely on reporter system, where a candidate enhancer sequence is positioned into the vicinity of a minimal promoter driving the expression of reporter genes such as β-galactosidase, luciferase or GFP. These approaches, usually not performed in the endogenous chromatin environment, is labor intensive and relatively low throughput[283].

A slew of new techniques based on next generation sequencing (NGS), which can predict functional regulatory DNA elements in eukaroytic genomes at a larger scale, have fueled the advancements in understanding the biological nature of enhancers[273,283,284]. Besides, the molecular features of enhancers have been extensively characterized by ChIP-seq datasets, notably histone modifications H3K4me1 and H3K27ac. These have been implemented to predict enhancer loci across species and cell types[246,279,280,285-287]. Systematic and integrative compilation of large compendia of enhancers is pioneered by Encyclopedia of DNA Elements (ENCODE) Consortium and Roadmap Epigenomics Consortium[288,289], which empower genomic annotation in myriad of genomic context. Algorithms (e.g. ChroHMM) can then be applied to classify characteristics of genomic segmentations[290].

In addition to occupancy of chromatin-associated factors and histone modifications, enhancers can also be predicted by chromatin accessibility, mapped by DNaseI hypersensitivity sites (DNaseI-seq), FAIRE-seq or transposase-accessible assay (ATAC-seq)[291-293]. Taking advantages of the self-transcribing activity of enhancers, various high-throughput methods identifying eRNAs can predict enhancer activities. These include GRO-seq, PRO-seq, and CAGE-analysis[284,294-296].The fact that enhancers are brought into close spatial proximity of their target promoters has also inspired two techniques to access enhancer-promoter pairing by coupling chromatin conformation capture-based techniques with ChIP to pull down enhancer-related marks, termed ChIA-PET and Hi-CHIP[297,298].

Despite the fact that a wide variety of criteria have been proposed to predict putative enhancers and estimate their activities, a growing body of literature suffer the “validation creep”–opinionated by Halfon– which signifies the tendency to move from considering a set of sequences as putative enhancers to accepting them as enhancers without functional validation. This renders a higher false-positive rate in the identification of enhancers[299]. Recently, reporter assay coupling with NGS technique has cranked up the scalability and inspires numerous massively parallel reporter assays (MPRAs)[271]. For instance, STARR-seq clones a library of genome-wide candidate enhancer sequences into fly cells. Considering that enhancers can be self-transcribed, the strength of individual

enhancers can be reflected by the propensity of respective eRNAs in the cellular RNA pool. This method therefore provides a genome-wide quantitative enhancer map[300]. Similarly, a collection of Drosophila TFs has been cloned to evaluate their regulatory function[301]. In addition, genome-wide enhancer modulation enabled by CRISPR-Cas9-based approach has been employed to decipher the function of putative enhancers. A library of guide RNAs targeting sequences of putative enhancers is designed and modulation is carried out by a variety of (d)Cas9 systems that either removes, activates or inactivates putative enhancers. Transcriptomic outcomes can then be interpreted as readout of enhancer activities. These Cas9 systems include (1) sequence deletion-based screen by Cas9[302]; (2) enhancer inhibition by recruiting repressor domain (dCas9-KRAB)[303,304] or H3K4 demethylase (dCas9-LSD1)[305]; and finally, (3) enhancer activation by recruiting domain of H3K27 acetyltransferase p300 (dCas9-p300)[306,307].

Enhancer regulation in 3D chromatin structure

Transcriptional enhancers are typically distal from the core promoters as such the linear genomic distance ranges from kilo-bases to 1MB[308]. Given the physical distances between the enhancers to their cognate promoters, current working models suggest chromatin folding brings the two parties into close proximity in three-dimensional space[309,310]. Nonetheless, large-scale assessment of enhancer activities in flies using reporter assay shows that the vast majority of enhancers activates targets of the closest genes[311]. Chromatin contact maps in human cells also depict favorable interactions within linearly short genomic ranges[312,313].

Chromatin is organized into TADs, these topologically distinct boundaries are proposed to segregate regulatory domains apart from each other. The “loop extrusion” model, driven by CTCF, cohesion complex and other co-factors, is considered responsible for shaping genome folding and establishing these TAD boundaries[4,5,314,315]. Chromatin looping interactions are favored at cis-regulatory elements within the same TADs, particularly active promoters, enhancers and CTCF binding sites but rarely at inactive genes, and that the looping interactions correlates with gene expression[312,316,317].

How the dynamic of 3D chromatin structure conducts proper enhancer-promoter contacts with respect to gene expression is highly debated[309,318,319]. One thing is that TAD structure is largely invariant across cell types, rendering the insufficiency to explain the highly modularity of enhancer activity across cell types[313,315]. In addition, promoter-enhancer interactions appear relative stable between tissue and across development, which is evidenced by the observation that looping interactions already take place at poised enhancers primed by H3K4me1, and that transient activation of enhancers do not alter chromatin structure. These implicate their wide pre-existence before activating the genes[312,320]. More strikingly, acute degradation of CTCF or cohesin, which leads loss of TADs, imposes only little impacts on transcription[321,322]. However, in a neural differentiation experiment in mESCs, smaller, less insulated interacting domains (called sub-TADs) undergo dynamic change throughout the course of neural differentiation [317]. Of note, recent studies on CTCF and cohesin suggest that TADs and chromatin loops are dynamic structures[323]. Hence, the contribution of chromatin topology to enhancer activity requires further studies.

Transcriptional co-activators mediate enhancer function

Co-activators are defined by their requirements in transcriptional activities, but as neither being a part of the basal transcriptional complex nor showing direct and selective binding of DNA[324]. Crosstalks between co-activators themselves and transcriptional machinery create a local environment that favors transcription. As a result, their localization on enhancers tend to positively correlate with gene expression. Perturbation of co-activators abrogates proper function of enhancers, suggesting them as the workhorses behind enhancer-mediated transcriptional activation[325]. Here, we concisely discuss several co-activators that have been demonstrated important in maintaining enhancer functionality, and how co-activators can be a popular druggable agents in treating various malignancies.

MLL3/4

MLL3/4 are members of TrxG proteins. They catalyze H3K4me1 on enhancers and participate in antagonism of PcG machinery[326]. Enhancers enriched solely by H3K4me1 while lacking H3K27ac have been associated with poised/primes states that awaits for activation[280]. Mapping of MLL3/4 binding site in human HCT116 cells and mESCs suggest their preferential localization at enhancers. Inactivation of MLL3/4 decreases H3K4me1 level both at bulk level and at enhancer loci, but rarely at promoters[326]. In mESCs, lack of MLL3/4 prevents proper differentiation as shown by experiments of teratoma formation. Mechanistically, MLL3/4 are required for recruitment of CBP/p300 on enhancers to properly deposit H3K27ac. Loss of MLL3/4 strongly reduces enrichment of H3K27ac in concert with reduced POL II occupancy and eRNAs production on active enhancers. This suggested that MLL3/4 are essential for priming and further activation of lineage-specific enhancers[327,328].

In Drosophila, genetic modulations of *trx* and *trr* activities (thus modulation of H3K4me1 enrichment) also affect H3K27ac level. For example, temperature-sensitive *trx* mutant reduces H3K4me1 in concert with reduction of H3K27ac and H3K18ac as demonstrated by IF on polytene chromosomes. On the contrary, overexpression of *trx* increases bulk H3K4me1, which drives higher accumulation of H3K27ac and concomitant loss of H3K27me3, suggesting that *trx* and *trr,* together with CBP/p300, counteract PcG system[329,330]. Later study demonstrates that the catalytic activity of *trx* is required for TrxG-mediated activation *in vivo*, and that H3K4me1 stimulates enzymatic activity of CBP in histone acetylation assay[330]. In contrast, a recent study in mESC suggests otherwise as catalytic-dead version of MLL3/4 are still able to stimulate gene activity, POL II loading and eRNAs production[328]. However, whether this specific mutation truly render a catalytic-dead form requires cautious scrutiny. Nevertheless, both Trx and Trr interact directly with CBP, implicating that the H3K4 methylases facilitate CBP recruitment to enhancers.

CBP

CBP/p300: multi-tasking in transcriptional activation

CBP and its highly similar paralog p300 encode histone acetyltransferases that catalyze acetylation of H3K27 as well as H3K18 and other sites[329,331]. The two proteins share several other conserved protein-binding domains including CH1, KIX, CH3, NCBD, bromodomain and PHD domain that interacts with a vast array of chromatin factors. CBP/p300 have at least 400 interacting protein partners, thereby acting as "hubs" in gene regulatory networks. Several mechanisms have been proposed to delineate CBP/p300-mediated transcriptional activation: (1) acting as scaffolds to bridge multiple chromatin factors and basal transcriptional machinery; (2) acetylating histones to promote open chromatin state; (3) acetylating TFs and other proteins to modulate their activities. The histone acetyltransferase activity, particularly the deposition of H3K27ac, have been widely investigated and will be discussed below[332,333].

H3K27ac is a hallmark of active enhancers

Work in Drosophila has shown that dCBP, the only homolog of mammal CBP/p300, catalyze the acetylation of H3K27 and H3K18. Acetylation and methylation at a given residue are chemically exclusive, as confirmed by genome-wide analysis in S2 cells. Drosophila CBP-null mutants are embryonic lethal. Moderate depletion of dCBP through siRNA remains viable but shows substantial loss of H3K27ac and elevation of H3K27me3. Importantly, these transgene clones show partial transformation with phenotypes similar to *trx* mutant. Conversely, overexpression of dCBP increases the level of H3K27ac, promoting gene expression, and reduces global H3K27me3 level. This suggests a direct opposition for dCBP/H3K27ac and the PcG machinery[329]. In mice, homozygous mutant of either CBP or p300 is embryonic lethal around E10.5, suggesting the H3K27ac dosage is crucial for normal development[334]. Conditional removal of CBP/p300 in MEFs wipes out H3K27ac and H3K18ac with concomitant increase of H3K27me3, and cell proliferation is severely compromised[335].

Acetylation of lysine residues is considered to effectively unfold chromatin as it neutralizes the basic charge of the lysines and thereby reduces the binding of

histones to negatively charged DNA[336]. Interestingly, H3K27 acetylation dissociates chromatin droplets *in vitro*, potentially creating an open chromatin state that is more permissive to transcription[276]. H3K27ac is enriched at transcriptionally active enhancers, whereas enhancers only enriched for H3K4me1 are considered as poised/primed. Of note, canonical active enhancers are co-occupied by H3K4me1 and H3K27ac but a unique subset of enhancers is only decorated by H3K27 acetylation[286]. Genes associated with active enhancers are actively transcribed comparing to their poised counterparts, and expression level highly correlates with level of H3K27ac enrichment[280].

CBP regulates enhancer integrity

Consistently with the roles of CBP/p300 in gene activation, shRNA-mediated KD of CBP or p300 results in global gene silencing, and notably genes associated with CBP-bound enhancers[337,338]. Targeted inhibition of CBP activity using chemical inhibitors C646 in mouse acute myeloid leukemia (AML) RN2 cells leads to profound reduction of H3K27ac level and eviction of BRD4 at promoters and enhancers. BRD4 is a co-activator that binds to acetyl lysine and is important for enhancer activities (see below). CBP inhibition consequently downregulates master genes related to cell identity in AML cells such as Myc, Cdk6 and Pecam1[337]. Another study proposes that CBP regulates also the occupancy of P-TEFb. As mentioned, P-TEFb complex enables transcription elongation through its enzymatic subunit CDK9 kinase. The kinase phosphorylates paused elongation complex to promote its activation. In addition, the kinase contributes to the phosphorylation of POL II CTD domain on serine 2, which stimulates elongation[339]. Treatment with another CBP inhibitors GNE-049 in AML MOLM-16 cells also results in reduction of H3K27ac in concert with reduced enrichment of CDK9 and POL II localizations at enhancers, and to a lesser extent, at promoters. This might explain the observed gene downregulation upon inhibitor treatment. Indeed, in a dynamic live cell tracking system, single-cell analysis illustrates that H3K27ac correlates with binding kinetics of TFs and release of paused POL II to promote transcription[340]. Taken together, the above studies suggest that CBP/p300 and H3K27ac are central to enhancer function [341].

Enhancer RNAs are proposed to stimulate CBP activity

Active enhancers are highly enriched for H3K27ac and produce bi-directional enhancer RNAs. Study shows that CBP binds directly to eRNAs both *in vitro* and *in vivo*. Binding of eRNAs promotes CBP catalysis, thus increasing H3K27ac and H4K5ac *in vitro*. Besides, acetylation of H3K27 and H4K5 are reported to decrease upon depletion of target eRNAs by anti-sense oligonucleotides. This study suggests a positive feedback loop between CBP and eRNAs. However, one cannot rule out that anti-sense oligos targeting eRNA could modulate enhancer through additional mechanisms[338].

BRD4

BRD4 is a reader of acetylated lysines

BRD4 is a member of bromodomain and extraterminal domain (BET) family. The BET family is characterized by tandem bromodomains. The family also includes also BRD2, BRD3 and a testis-specific member BRDT. The tandem bromodomains recognize acetylated H3 and H4. In vitro studies indicate that BRDs have high affinity toward hyperacetylated histones, but that they do not bind to mono- or unacetylated histones[342,343].

BRD4 associates with Mediator and elongation factors

The link between bromodomains and hyperacetylated histones explain the correlation of BRD4 with active transcription. Indeed, genome-wide studies illustrate BRD4's localization at active promoters and enhancers that are enriched in H3K27ac[337,344,345]. Depletion of BRD4 by shRNA strongly compromises pluripotency and self-renew in hESCs, and gives rise to massive transcriptional silencing[346]. In addition, mice embryos knockout for Brd4 die shortly after implantation, while the heterozygous counterparts suffer severe anatomical defects[347]. Mechanistically, BRD4 interacts with CyclinT1 and kinase CDK9, core subunits of P-TEFb[348,349]. Ectopic expression of BRD4 in HeLa cells increases RNA-Pol II-Ser2 CTD phosphorylation in bulk and stimulates transcription as shown by luciferase assay. In contrast, siRNA KD of BRD4 reduces occupancy of P-TEFb and leads to transcriptional silencing[348]. This suggests that BRD4 promotes transcription by regulating P-TEFb activity. Apart from P-TEFb, BRD4 also

interacts with Mediator complex[350], a mega-protein complex that is essential for active transcription (see section below). All of these evidences pinpoint the importance of BRD4 in transcriptional activation and in development.

BRD4 inhibition compromises Mediator and P-TEFb occupancy on chromatin

The localization at chromatin of Mediator (i.e. MED1) highly correlates with BRD4 signal[344,345]. Small molecule inhibitors of BET family proteins such as JQ-1 and iBET, acting as competitive agents of acetylated lysine, severely compromises BRD4 occupancy especially at enhancers. BRD4 binding to chromatin is lost upon JQ-1 treatment in Multiple myeloma (MM) MM1.S cells, and results in global gene silencing. Remarkably, a class of highly transcribed genes show more profound silencing, implying that highly active genes are more sensitive to BRD4 displacement. Mechanistically, inhibition of BRD4 triggers dislodgement of MED1 and CDK9 primarily from enhancers and consequently impairs POL II elongation through the gene bodies[344]. Of note, live cell super-resolution microscopy in mESCs enables to visualize co-localization of Mediator and POL II in stable clusters, but these clusters are sensitive to BRD4 inhibitors, further supporting the role of BRD4 in stabilizing Mediator localization[351].

Enhancer RNAs facilitates BRD4 binding

The role of eRNA was also investigated regarding BRD4 function. Rahnamoun and colleagues mapped RNA-BRD4 binding profile in human colon cancer SW480 cells, and report associations of eRNAs-BRD4 occur in BRD4-bound regions. This association requires the BRD4 tandem bromodomains, as truncation of these domains abolish the interaction *in vitro*. Furthermore, eRNAs facilitate BRD4 binding on acetylated histones peptides and nucleosomes in binding assays. KD of specific eRNA compromises BRD4 binding locally, reducing POL II binding, and ultimately limits the gene expression. These results again attribute an important regulatory role to eRNA, via a feed-forward mechanism promoting BRD4 occupancy [352]. Regardless, independent validations remain nonetheless necessary.

Mediator

Mediator is an evolutionarily conserved multi-subunit protein complex present in yeast, metazoans and plants. In mammals, the complex contains up to 30 subunits that assemble into the core Mediator and a CDK8 module (**Figure 11**). Mediator is ubiquitously required for transcription, as it bridges between TFs and pre-initiation complex[12,353]. Indeed, cryo-EM revealing the structure of yeast Mediator-PIC complex demonstrates multiple contacts of Mediator with POL II and GTFs[354]. As one of the limited number of co-activators that directly target POL II, it is proposed to transduce signals from TFs-bound enhancers to transcriptional machinery[12]. In addition, Mediator complex can recruit POL II, and stimulates CDK7 kinase within the PIC complex to phosphorylate Ser5 and Ser7 of the Pol II CTD domain, hence triggering Pol II release from promoters[355]. Knockout of each mammalian Mediator subunits is embryonic lethal, in accordance with its general requirement for transcription[356].

Modules of the Mediator complex

The core Mediator complex contains 26 subunits and is divided into 3 modules: head, middle and tail. Another four-subunit module, the CDK8 kinase module, contains the CDK8/CDK19, MED12/MED12L, MED13/MED13L and cyclin-C (CCNC). The CDK8 module can reversibly interact with the main Mediator complex and impacts the structure and function of the core Mediator[357]. The head and middle modules interact with PIC, while the tail module is associated with TFs[358]. Of which, MED14 functions as the central backbone of the complex that connects the three main modules[359]. It is shown that specific subunits of Mediator conduct their individual unique functions in response to developmental or environmental stimuli[12]. For example, as a target of various hormone receptors such as thyroid hormone, MED1 deletion fails to elicit expression of thyroid hormone-responsive genes in MEFs[360]. While Mediator positively regulates transcription, CDK8 module shows divergent impacts on transcription. Studies report that the module blocks the binding of Mediator-PIC, and therefore inhibits basal transcription. By contrast, the module has been implicated in positive regulation of elongation and may function as pause-release factor. It is likely that the CDK8 module has context specific function, a question that requires further investigation[353,361].

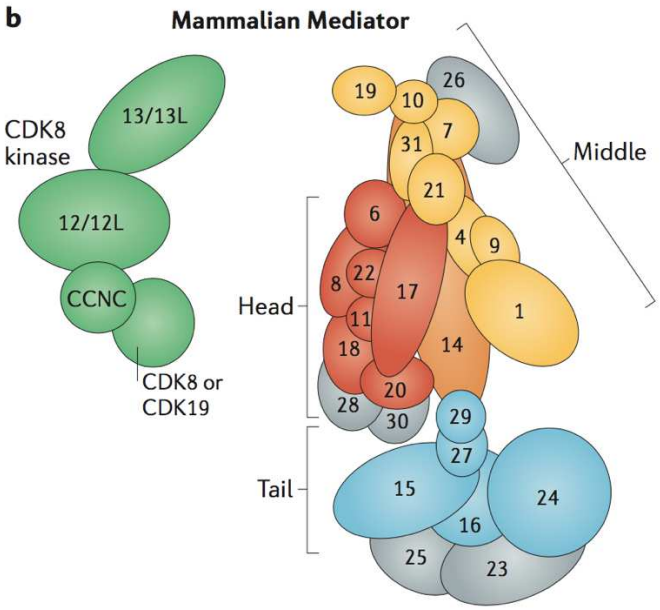

Figure 11. The Mediator complex (Soutoutrina 2018)

Collectively, recent studies dissecting the contributions of co-activators involved in enhancer function depicts a highly collaborative network in concert with auto-regulatory loop that reinforce their activity to strengthen the transcriptional processes (**Figure 12**).

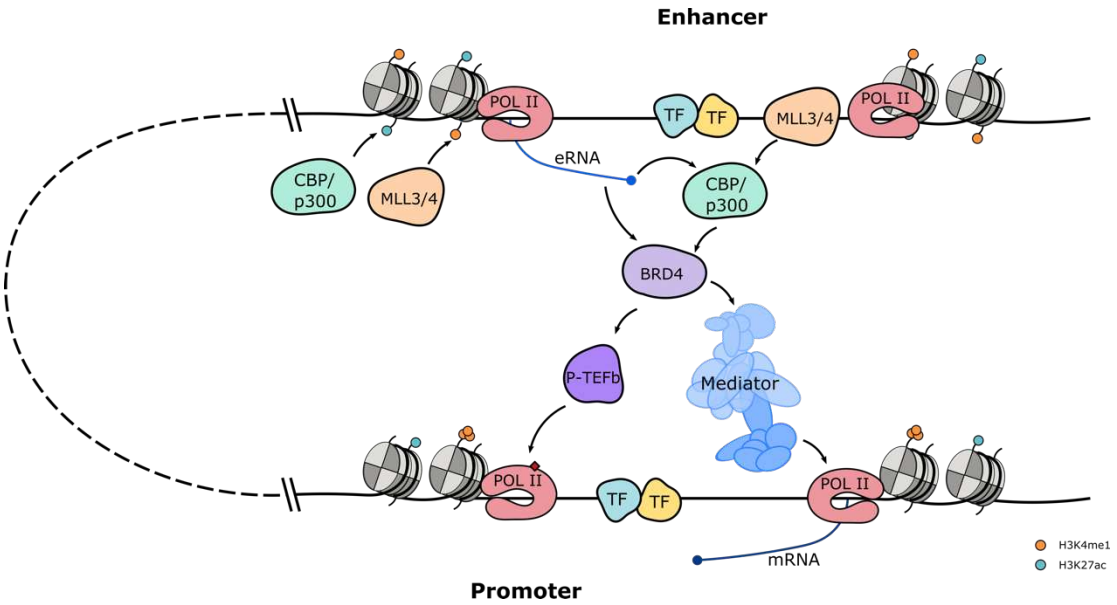

Figure 12. Transcriptional co-activators and their regulatory network

Cluster of enhancers (a.k.a. as Super-enhancers)

Key regulator genes for cell identity are controlled by cluster of enhancers.

Clusters of enhancers [or super-enhancers (SEs)] are regions —spanning tens of kbs in the genome—that are occupied by unusually high magnitude of master TFs and Mediator (i.e. MED1)[362]. These SEs drive high expression level of genes that often encode key regulators of cell identity. In mESCs, SEs locate, for instance, in proximity of pluripotent factors Oct4, Sox2, Nanog, Klf4 and Esrrb. These TFs in turn bind to their own SEs to drive high gene expression, forming interconnected feedback circuitry[362]. SEs are defined by computational methods through profiling the density of Mediator (MED1) or H3K27ac, and subsequent "stitching" of clusters of enhancers in vincinity[344,363]. Consequently, markers of active enhancer are highly enriched at SEs comparing to typical enhancers. These include POL II, MED1, CBP/p300, BRD4, CHD7, BRG1; and histone modifications H3K27ac and H3K4me1.

Analysis of 86 human cells and tissue samples show that a substantial fraction of SEs are cell-type specific[363]. In cancer cells, SE often control oncogenes and are therefore important in tumor pathogenesis. Taking proto-oncogene c-MYC as an example, it is frequently overexpressed and contribute to pathogenesis in various cancers[364]. SEs localized in gene dessert proximal to MYC are found in pancreatic cancers, T-cell leukemia, colorectal cancers cells but not in healthy tissues. Thus, SEs are likely to play key roles in cell identity in development and pathogenic progressions[363]. In conclusion, SE are a subclass of enhancers defined by their (1) genomic sizes, (2) high density of active factors (3) high transcriptional activities, and (4) determining roles in cell identity (**Figure 12**).

SEs vulnerability

Despite confering high transcriptional activities, SEs are extremely susceptible to enhancer perturbation. Reducing the level of MED1 by shRNA is reported to induce more profound downregulation on SEs-associated genes than on TEs-associated genes in mESCs[362]. In MM cells, BRD4 inhibition by chemical inhibitors

casts greater BRD4 dislodgement on SEs than typical enhancers, as evidenced by comparison of dose-dependent BRD4 eviction at IgH-MYC SEs and SMARCA4 typical enhancers. This disproportional eviction of BRD4 reflects on global transcription as showed by more pronounces reduction on RNA level in SEs-related genes than TEs counterparts. This disproportional eviction of BRD4 is translated into transcription, as global transcriptomic analysis confirmed greater reduction on RNA level in SEs-related genes than TEs counterparts[344]. Similar sensitivity can also be observed in cells treated with CDK7 inhibitor, where gene associated with SEs are more severely downregulated[365]. Given that SEs frequently drive expression of oncogenes, the sensitivity toward enhancer perturbation could be exploit for therapeutic targeting in various malignancies[344,345,366,367].

A phase separation model of transcriptional co-activators in gene control

In cellular organisms, the spatiotemporal control over biochemical reactions is modulated by the concentration of components which can increase reaction kinetics, whereas segregation can inhibit or refrain these reactions. Compartmentalization can be executed through lipid bilayer membrane where it creates impermeable physical barriers between the interior and exterior environments. However, many cellular compartments, such as Cajal bodies, are bound together by membrane-less structure, which exhibit liquid-like properties, highlighting the liquid-liquid phase separation is a common property in cells. These liquid-like condensates have been observed in various biological processes, and many names have been coined to elucidate their properties, including cellular bodies, nuclear bodies, membrane-less organelles, granules, speckles, aggregates, assemblages and membrane puncta[368]. Biomolecular condensates are often enriched with multivalent molecules, they harbor various elements that govern intra- or inter-molecular interactions. Indeed, reconstituted chromatin undergoes intrinsic condensation and forms phase separated droplets in physiologic salts[276]. Proteins containing large intrinsically disordered regions—a sequence with repeated amino acids and low complexity— are often prone to phase-separate under physiological conditions[369].

Many eukaryotic TFs and co-activators contain IDRs that are proposed to locally recruit POL II into dynamic condensates, which serve as hubs for transcription[370]. SEs accommodate high density of co-activators which could be associated with phase separation. Accordingly, BRD4 and MED1 form nuclear puncta that co-localized with SEs-regulated genes including Nanog, Klf4 and Trim29 in mESCs. In order to investigate if these condensates are important in gene control, 1,6-hexanediol, a nonspecific aliphatic alcohol known to disrupt liquid-like droplets, is incubated with cells. The treatment caused a reduction in BRD4 and MED1 condensates. ChIP-seq showed a general reduction of BRD4, MED1 and POL II occupancy on enhancers, but more pronounced at SEs, in coherence with SE vulnerability as described by chemical inhibition of BRD4[371]. It is nonetheless worth mentioning that a recent study suggests that 1,6-hexanediol has global effect on chromatin[372].

Summarizing both computational simulation and experimental procedures, a combination of DNA-TFs and IDRs-based (TF-coactivators; coactivators-coactivators) interactions on enhancers is proposed to be involved in the formation of TF-coactivator-POL II transcriptional condensates that concentrate the transcriptional machinery and reinforce its activity[373,374].

Opportunities for enhancer perturbation in cancers

Given the fact that enhancers are the key components that drive cell identity, dysregulations of enhancer functions have been reported in various developmental diseases and cancers. Here, we briefly discuss pathogenesis related to enhancers deregulation and potential therapeutic strategies.

Enhancer dysfunction in diseases

Several aspects can lead to enhancer perturbation, ranging from genetic alteration, TFs malfunction, mutations in chromatin modifiers, to 3D chromatin organization[375]. Genetic mutations or single nucleotide polymorphisms have frequently been found to occur at enhancers. They affect the binding affinity of TFs

and thereby interfere with transcription[363]. Various genetic alteration could drive gene overexpression. For instance, a translocation event in MM relocates the IgH enhancer in proximity to MYC locus, thus driving its overexpression[376]. A fusion BRD4-NUT oncogenic protein that leads to tether of NUT on DNA is the main driver of NUT-midline carcinoma, an aggressive subtype of squamous cell cancer[377]. Inactivating mutations in chromatin modifiers, such as MLL family[378], CBP[379], mSWI/SNF complex[255], have been frequently reported in cancers. Lastly, deletions that interrupts insulating neighborhoods introduces activation of oncogenes in T-cell acute lymphoblastic leukemia[380]. In addition, disruption of TAD boundary that establishes ectopic promoter-enhancer contacts can drive aberrant expression that cause limb developmental defects in human[381].

Drugging enhancer activities as therapeutic opportunities

The concept of transcription addition comes from the observation that cancerous cells progressively develop dependency on oncogene expression, whereby the withdrawal of these genes (e.g. MYC) prohibit proliferation of the considered tumor cells[376]. BET inhibitors have proven efficacy to treat NUT-midline carcinoma, where administration of the inhibitor gave rise to growth arrest in cell model[382]. Ever since, a growing number of BET inhibitors have been evaluated and some have entered early phase clinical studies. For example, OTX015 was in phase I study for hematologic malignancies, glioblastoma and solid tumors[366]. Aside from BET proteins inhibition, inhibitors targeting other co-activators were reported to have positive outcome in several tumor cells. For example, inhibitors targeting transcriptional kinases suppress SEs-genes and trigger cell death in heterogeneous cancer types[375]. CBP/p300 inhibitors disrupt genes associated with SEs and exerts anti-tumor activities in prostate cancer cells[383].

Results

Part I. Dissecting BAP'1 mechanism of action.

Given the tumor repressive role of BAP1 and its deubiquitinating activity which is tightly linked to Polycomb repressive systems, we are intrigued to unravel the very nature of the BAP1. The former PhD student in the lab has taken initiatives on the project and performed extensive characterization on BAP1. These include identification of BAP1's interacting partners and consequences in response to loss of BAP1. Using human HAP1 cell line as a model, he confirmed that genetic inactivation of BAP1 resulted in visible increase of bulk H2Aub level, in coherence with reported outcome of mutation in BAP1's ortholog *calypso* in flies[141]. Besides, he uncovered that loss of BAP1 gave rise to large-scale gene downregulation, which surprisingly, did not fit in to the *Hox* locus de-repression described in *calpyso* mutant in Drosophila[141]. In addition, a panel of downregulated genes could be rescued by overexpression of wild-type, but not catalytically dead version of BAP1, indicating the de-ubiquitiniating activity is crucial to BAP1-mediated transcriptional activation. Following, ChIP-seq analysis suggested BAP1 loss led to increased H2Aub and H3K27me3 enrichment on downregulated genes[210,211]. However, this increase of H2Aub increase was also found in upregulated genes upon BAP1 loss. The dynamic regulation of BAP1-mediated H2Aub regulation remains unclear and thus urging us to unravel BAP1's mechanism of action.

Continuing his work, I have taken on the task to dissect the link between BAP1 and transcriptional regulation during my PhD thesis. In order to further consolidate the idea that the role of BAP1 in to positively regulate transcription, I stimulated transcription by retinoic acid (RA)-based transactivation assay[384] in WT, BAP1-KO and two other co-activators (CBP and SMARCB1) KO lines in HAP1 cells. Gene expression levels evaluation by RT-qPCR indicated that deletion of BAP1 perturbed the expression of RA-responsive genes, similar to those reported in co-activators knockout conditions. However, it remains unclear, to what extent, BAP1 regulates transcription in a H2Aub-dependent manner. In order to address this specific question, I first generated a cell model that is devoid of H2A ubiquitination by deleting the redundant catalytic subunits RING1A and RING1B of PRC1 (PRC1-KO). PRC1 knockout led to de-repression a plethora of genes as expected. Interestingly, subsequent BAP1 inactivation in the context of PRC1-KO no longer rendered transcriptional alterations, suggesting that BAP1 is largely epistatic to

BAP1[211]. That is, BAP1 does not regulate gene expression in the absence of H2Aub. Our work provided evidences that BAP1 has a positive role in transcriptional regulation, which, importantly, is dependent on deubiquitination of histone H2A. This work has been published on Nature Communications (PMID: 30664650) in 2019, which I shared the co-first authorship with the former PhD student and a bioinformatician. (**Results Part 1.1**)

Following up this study, we sought to elaborate the detailed molecular mechanism on BAP1's mechanism of action. I first performed quantitative multi-omics analysis and demonstrated that BAP1 functions principally though H2A deubiquitination. Since BAP1 regulates H2Aub, I determined to unveil its genomic localization. Genome-wide mapping of BAP1 binding sites performed by CUT&RUN-seq[215] showed that BAP1 was primarily recruited to a subset of active enhancers. These BAP1-bound enhancers showed high enrichment of H2Aub, and that H2Aub greatly accumulated in the absence of BAP1. These implicates that PRC1 potentially has an inhibitory function at the enhancers, which is especially interesting as the activity of PRC1 at enhancers is less characterized[175,176]. Further, I showed that loss of BAP1 compromised proper recruitment of BRD4 at enhancers alongside with reduced BRD4 and MED1 condensates, suggesting that BAP1 maintains enhancer functions. However, whether H2Aub accumulation directly interferes with BRD4 binding requires further studies. Lastly, small-molecule inhibitors screen identified that BAP1-null cells were more sensitive to BET inhibitor than WT counterparts, which could potentially pave the way for therapeutic opportunities. This work has been prepared in manuscript for submission (**Results Part 1.2**).

Part II. PRC2 represses transcription independently of PRC1

PRC1 and PRC2 complexes are long considered cooperating to maintain gene repression[7]. While PRC1 is reported to repress a subset of genes independently of PRC2 in mouse epidermal tissues[177], it is less well-known if PRC2 confers repressive activities in a PRC1-independent fashion. The limitation to address this issues lies in lack of available cellular or animal models. As PRC1 and/or PRC2 complexes are often essential for maintenance in normal biological functions[7,25,30,71], this makes in challenging to dissect the respective function in a loss-of-function-based approach. Previously, I have generated cell line harboring PRC1 deletion[201]. I then further inactivated EED and rendered a cell model that is absent of both PRC1 and PRC2 complexes (PRC1/2 KO). This system granted me the opportunity to tackle the independency and interdependency between the two complexes.

Analysis on transcriptomic profiles of PRC1-null, PRC2-null and PRC1/2-null cells revealed that both PRC1 and PRC2 can autonomously repress transcription, which is in coherence with a recent in vivo study in mouse epidermal tissues (Cohen). Characterization of subsets of genes regulated by either PRC1 or PRC2 showed distinct patterns including H3K27me3 enrichment and basal transcriptional activities. In order to dissect the underlying molecular mechanism, we focused on identifying downstream effectors of PRC2-mediated silencing in the absence of PRC1 through both unbiased and candidate-based approaches. This includes investigating the roles of previously proposed H3K27me3 readers. This work this study is still ongoing, it is likely that it will reveal new actors for PRC2-mediated repression.

Part 1.1 BAP1 complex promotes transcription by H2A deubiquitination

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-018-08255-x OPEN

BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation

Antoine Campagne[1,2], Ming-Kang Lee[1,2], Dina Zielinski[1,2,3], Audrey Michaud[1,2], Stéphanie Le Corre[1,2], Florent Dingli[1], Hong Chen[1,2], Lara Z. Shahidian[4], Ivaylo Vassilev[1,2,3], Nicolas Servant[1,3], Damarys Loew[1], Eric Pasmant[5], Sophie Postel-Vinay[6], Michel Wassef[1,2] & Raphaël Margueron[1,2]

In *Drosophila*, a complex consisting of Calypso and ASX catalyzes H2A deubiquitination and has been reported to act as part of the Polycomb machinery in transcriptional silencing. The mammalian homologs of these proteins (BAP1 and ASXL1/2/3, respectively), are frequently mutated in various cancer types, yet their precise functions remain unclear. Using an integrative approach based on isogenic cell lines generated with CRISPR/Cas9, we uncover an unanticipated role for BAP1 in gene activation. This function requires the assembly of an enzymatically active BAP1-associated core complex (BAP1.com) containing one of the redundant ASXL proteins. We investigate the mechanism underlying BAP1.com-mediated transcriptional regulation and show that it does not participate in Polycomb-mediated silencing. Instead, our results establish that the function of BAP1.com is to safeguard transcriptionally active genes against silencing by the Polycomb Repressive Complex 1.

[1] Institut Curie, Paris Sciences et Lettres Research University, Sorbonne University, 75005 Paris, France. [2] INSERM U934/CNRS UMR3215, 75005 Paris, France. [3] INSERM U900, Mines ParisTech, 75005 Paris, France. [4] Institute of Functional Epigenetics, Helmholtz Zentrum München, Neuherberg 85764, Germany. [5] Department of Molecular Genetics Pathology, Cochin Hospital, HUPC AP-HP, EA7331, Faculty of Pharmacy, University of Paris Descartes, Paris 75014, France. [6] Département d'Innovation Thérapeutique et Essais Précoces, INSERM U981, Gustave Roussy, Université Paris-Saclay, Villejuif F-94805, France. These authors contributed equally: Antoine Campagne, Ming-Kang Lee, Dina Zielinski. Correspondence and requests for materials should be addressed to M.W. (email: michel.wassef@curie.fr) or to R.M. (email: raphael.margueron@curie.fr)

BRCA1-associated protein 1 (BAP1) was initially characterized as a nuclear deubiquitinase regulating the function of BRCA1[1]. Subsequent work suggested that BAP1 does in fact interact with BRCA1-associated RING domain 1 (BARD1) and regulates its ubiquitination[2]. A variety of proteins have since been reported to interact with BAP1, including transcription factors (YY1, FOXK1/2), chromatin binders and modifiers (ASXL1/2/3, KDM1B, OGT1), the cell cycle regulator HCFC1 and DNA repair proteins (MBD5/6)[3–9]. BAP1 enzymatic activity has been shown to regulate the ubiquitination of various proteins including gamma-tubulin[10], INO80[11], and BRCA1[1]. Accordingly, BAP1 participates in diverse cellular processes, such as transcriptional regulation and the DNA damage response. However, the precise function of BAP1 in transcriptional regulation remains elusive. Some studies have reported that BAP1 acts as a transcriptional activator while others have suggested that it is required for gene silencing[3,4,12].

In parallel to its characterization in mammals, studies in *Drosophila* identified the BAP1 ortholog Calypso as a novel Polycomb protein[13,14]. The Polycomb Group (PcG) of proteins is essential for the maintenance of gene repression, most prominently at developmentally regulated genes. Consequently, altering PcG function affects key cellular processes such as cell fate determination, cell proliferation, and genomic imprinting[15]. Two Polycomb complexes have been well characterized thus far: Polycomb Repressive Complex 1 (PRC1) and PRC2. PRC2 catalyzes di- and trimethylation of histone H3 on lysine 27 (H3K27me2/3), whereas PRC1 acts through chromatin compaction and monoubiquitination of histone H2A on lysine 119 (H2AK119ub1)[16,17]. Conserved from *Drosophila* to mammals, the activity of these complexes is necessary for maintaining transcriptional silencing of their target genes. The importance of H2AK119ub1 in Polycomb silencing has recently been called into question in *Drosophila*[18], as well as in mouse models[19]. Nonetheless, recent studies suggest that this mark participates in stabilizing PRC2 binding to chromatin[20,21].

Drosophila Calypso was found to partner with the Polycomb protein Additional Sex Combs (ASX) into a novel Polycomb complex termed Polycomb Repressive DeUBiquitinase (PR-DUB) complex. PR-DUB has been shown to catalyze deubiquitination of H2AK119ub1, opposite to the activity of PRC1[22]. The interaction between BAP1 and homologs of ASX (ASXL1/2/3 proteins) is conserved in mammals, as well as the H2AK119 DUB activity of BAP1[13,23]. How the antagonistic activities of Calypso/BAP1 and PRC1 converge to maintain transcriptional silencing remains enigmatic. Further, the link between PR-DUB and the Polycomb machinery is still controversial. Some studies have reported that ASXL1 interacts with PRC2 and is required for its recruitment[24–28] while others have suggested an antagonism between BAP1 and PRC2[29–31]. A clear picture of the function of BAP1 and ASXL proteins is still lacking. Understanding the function of PR-DUB is all the more important in view of the tumor-suppressive functions of BAP1 in several cancer types including uveal melanoma, mesothelioma, and clear-cell renal cell carcinoma and of ASXL proteins in hematologic malignancies[2,32].

In this study, we use biochemical, genome editing, and genome-wide methods to address the function of BAP1 in transcriptional regulation and its relationship to the Polycomb machinery. We show that the ASXLs are mandatory partners of BAP1 and are required for its stability and enzymatic activity. At the functional level, the complex formed with BAP1 (BAP1.com) is required for efficient transcription of many developmental genes. Accordingly, BAP1 appears to be largely dispensable for maintaining silencing of Polycomb target genes and in fact opposes PRC2-mediated silencing at a number of genes. The majority of BAP1-regulated genes, however, are not under regulation by PRC2, suggesting that the function of BAP1 in promoting transcription does not reflect an obligate antagonism with PRC2. We show that BAP1 is required upon transcriptional stimulation, as observed after retinoic acid (RA) treatment, a function that is shared with the CREBBP and SMARCB1 transcriptional co-activators. A general role in regulating gene expression is supported by the conspicuous colocalization between BAP1.com and RNA polymerase 2. Mechanistically, we show that BAP1's function depends on its deubiquitinase (DUB) activity and that BAP1 is functionally inert in the absence of H2AK119ub1. Our integrative analysis uncovers an essential function for BAP1.com as a transcriptional co-activator, acting by locally antagonizing PRC1 activity.

Results

Loss of BAP1 alters the expression of developmental genes. In order to comprehend the role of BAP1 in transcriptional regulation, we generated knockouts (KOs) for *BAP1*, *ASXL1*, *ASXL2*, and *EZH2*, the main catalytic subunit of PRC2, in human HAP1 cells, a model cell line that is nearly haploid and thus particularly amenable to genome editing[33]. KOs result from the insertion of a STOP cassette that interrupts transcription and translation of the target gene[34] and were validated by reverse transcription-quantitative polymerase chain reaction (RT-qPCR, Fig. 1a). In contrast to many cell lines where the knockdown of BAP1 severely compromises proliferation (Supplementary Fig. 1a), we found that BAP1 is dispensable for proliferation in HAP1 cells, thus providing a suitable system for studying its mechanism of action (Fig. 1b). Cellular fractionation confirmed the tight association of BAP1 with chromatin, as shown by its enrichment both in the soluble and insoluble fractions (Supplementary Fig. 1b), which prompted us to further investigate its chromatin-modifying activity. Western blot analysis of various histone marks showed that BAP1 loss is associated with an approximate twofold increase in total H2A ubiquitination levels, as well as a parallel increase in H2A.Z ubiquitination, with no effect on H2B ubiquitination (Fig. 1c, top panel). *ASXL1* and *ASXL2* KO led to a modest increase in H2A/H2A.Z ubiquitination. We did not observe any global effect upon KO of *BAP1*, *ASXL1*, or *ASXL2* on other histone marks such as H3K4me2 and H3K27me3 or on DNA modifications (Fig. 1c, bottom panel).

We then performed RNA-seq to analyze the transcriptome of wild-type and KO cell lines. The inactivation of BAP1 resulted in dramatic changes in gene expression ($n = 1893$ differentially expressed (DE) genes, false discovery rate (FDR) < 0.05 and absolute log2 fold-change > 1; Fig. 1d, left panel). Strikingly, the majority of affected genes were downregulated in *BAP1* KO cells ($n = 1370$), suggesting an involvement in gene activation rather than silencing. Gene ontology (GO) analysis revealed enrichment for a variety of biological processes ranging from broad terms such as cell communication, signaling, or regulation of proliferation to more specific terms such as blood vessel development (Fig. 1d, right panel). Several terms related to development were also enriched, consistent with the reported function of the BAP1 ortholog in *Drosophila*[13]. Inactivation of either ASXL protein also leads to preferential downregulation of genes (85 downregulated genes of 112 total DE genes in *ASXL1* KO cells and 262 downregulated genes of 406 DE genes in the *ASXL2* KO cells; Fig. 1e, also see Supplementary Fig. 1c for heatmaps of the top 100 DE genes in each KO condition) but has a much milder effect on gene expression than observed for the *BAP1* KO. Further, loss of ASXL1 or ASXL2 does not affect the chromatin localization of BAP1 (Supplementary Fig. 1b). This result suggests either that BAP1 can function independently of its interaction with ASXL proteins or that the ASXLs are largely redundant.

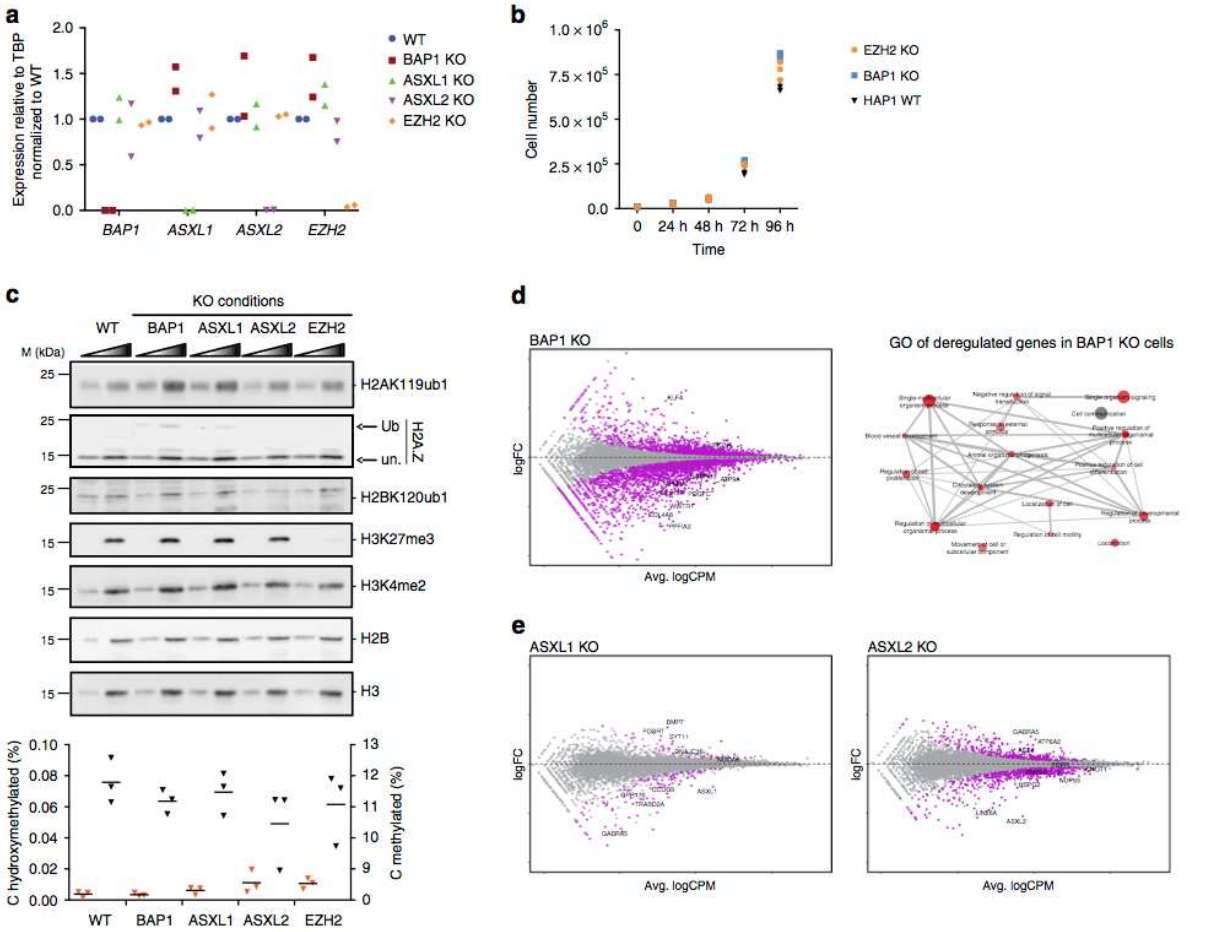

Fig. 1 Functional consequences of loss of BAP1, ASXL1, or ASXL2 on chromatin and gene expression. **a** RT-qPCR analysis of *BAP1*, *ASXL1*, *ASXL2*, and *EZH2* expression in the different KO conditions indicated on top. $n = 2$. **b** Proliferation curve of wild-type, *EZH2* KO and *BAP1* KO HAP1 cells. $n = 3$. **c** Top, western blot analysis of acid extracted histones with antibodies directed against various histone modifications (as indicated on the right) in the different cell lines indicated on top, M molecular weight. A two-point titration (1:2.5 ratio) is shown for each condition, ub ubiquitinated, un unmodified. Bottom, analysis of cytosine methylation (blue triangles) or hydroxymethylation (red triangles) in the KO conditions indicated at the bottom. Horizontal bars indicate the mean. $n = 3$. **d** Left panel: scatterplot showing log2 fold-change (logFC) expression between wild-type and *BAP1* KO cells versus average log2 counts per million (logCPM). Differentially expressed genes (DEGs) in *BAP1* KO cells are highlighted in purple. Right panel: representation of the non-redundant most enriched GO terms within the DEGs in *BAP1* KO cells. **e** Scatterplots as in **d**, showing gene expression changes in *ASXL1* and *ASXL2* KO cells

The ASXL/BAP1 core complex is conserved in mammals. To better understand the function of BAP1, we sought to determine whether it is part of a stable complex and, if so, with which partners. Purification of Calypso from *Drosophila* embryos revealed that it forms a heterodimeric complex with ASX[13]. Although the interaction between BAP1 and ASXL proteins is conserved[13,23,35], additional partners were reported in the mammalian complex[3,4,9]. We overexpressed FLAG-tagged versions of BAP1, ASXL1, or ASXL2 in HeLa cells (Supplementary Fig. 2a), followed by immunoprecipitation and mass spectrometry. The results confirmed previous reports, notably the identification of ASXL1/2, FOXK1/2, HCFC1, and KDM1B as partners of BAP1 (Fig. 2a). The ASXL1 and ASXL2 interactomes were similar with the exception of KDM1B, which is specifically pulled down by ASXL2. Of note, similar results were obtained for the BAP1 interactome in Uveal Melanoma MP41 cells (Supplementary Fig. 2b), suggesting that BAP1.com composition is not cell-type dependent.

We then sought to determine whether BAP1 partners are all present in a single complex or if BAP1 is engaged in distinct protein complexes. To this end, we analyzed the elution pattern of BAP1 partners by ion exchange chromatography followed by size-exclusion chromatography (SEC) (Fig. 2b). The first purification step (ion exchange) revealed that the HCFC1, FOXK1, and KDM1B generally elute independently of BAP1, which is almost exclusively found in the 500 mM salt fraction, suggesting that only a portion of each of these proteins is engaged in a complex with BAP1. In the second purification step (SEC), we observed that BAP1 elutes with a molecular weight of approximately 500 kDa, along with with HCFC1, FOXK1, and KDM1B. In contrast, YY1 elutes later and only partially overlaps with BAP1. To know whether the co-elution between BAP1 and its cofactors reflects the assembly of these proteins into a complex, we repeated the experiment with nuclear extract from BAP1 KO cells. The elution pattern from cation exchange did not reveal any major changes; we therefore continued with SEC. The elution patterns of HCFC1, KDM1B, and FOXK1 remain unchanged, indicating that the co-elutions observed with BAP1 in the wild-type extract do not reflect the formation of a stable complex (Fig. 2b). Considering the previously established tight

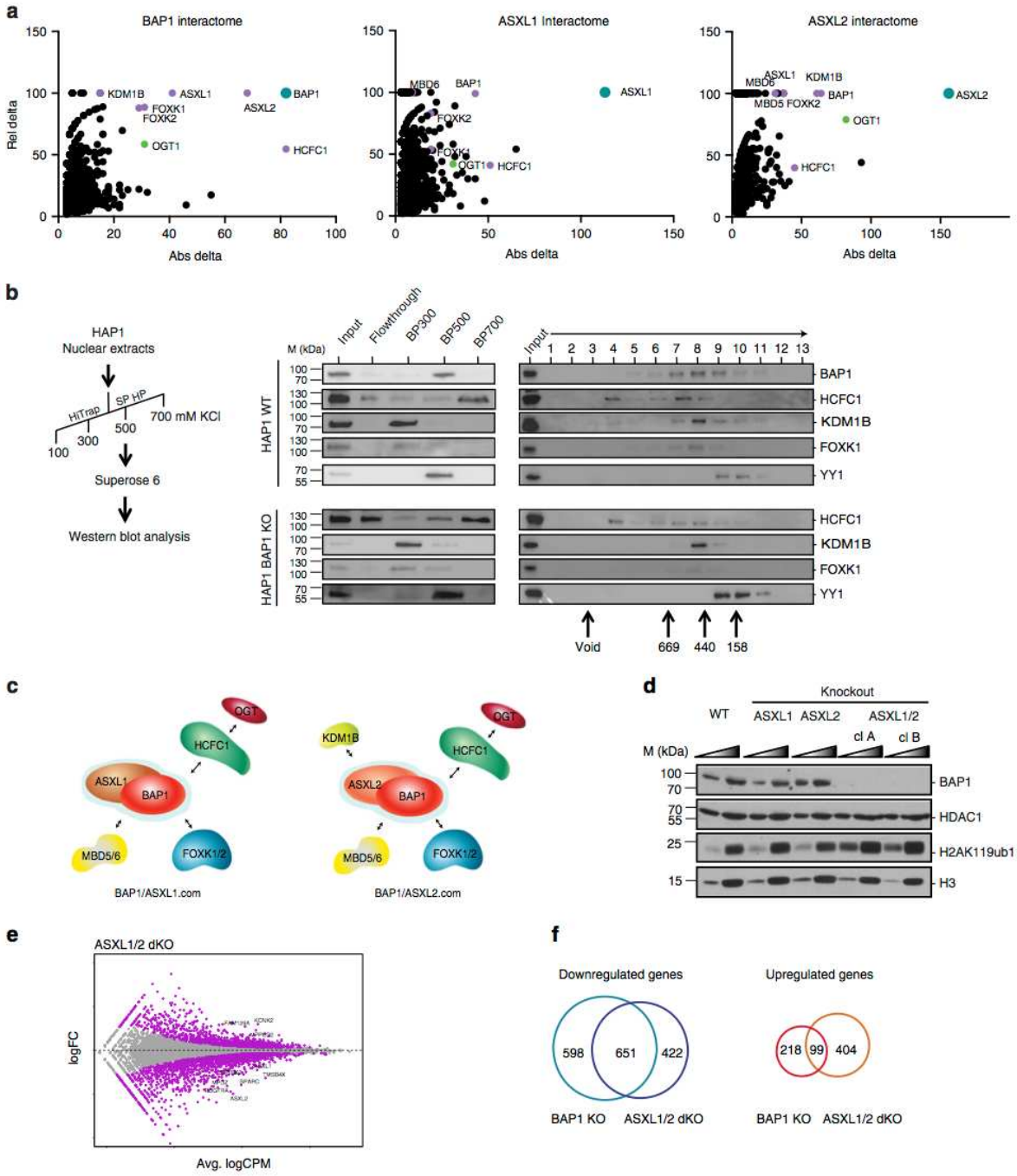

Fig. 2 BAP1.com core complex and associated factors. **a** Mass spectrometry analysis of HeLa cells overexpressing Flag-tagged versions of BAP1, ASXL1, and ASXL2. Graphs represent proteins relative to their absolute (Abs) and relative (Rel) delta compared with mass spectrometry analysis of empty vector expressing cells. Absolute delta is the absolute difference between distinct peptides identified in sample and control; relative delta is the ratio of absolute delta versus the sum of distinct peptides identified in sample and control. **b** Elution patterns of HAP1 WT and HAP1 *BAP1*-KO nuclear extracts following the purification scheme indicated in the left panel and monitored by western blot with the indicated antibodies. Middle pattern is a representative elution pattern (step elution with increased salt concentration) on a cation exchange column (SP-HP, GE). Right panel is a representative elution pattern on a size-exclusion column (Superose 6, PC3.2/30, GE). *Underneath is the correspondent*. **c** Schematic of BAP1.com core complex and associated factors depending on the ASXL paralog present. **d** Western blot analysis of BAP1 and H2AK119ub1 in the different KO conditions indicated above. Two independent clones of *ASXL1/2* dKO cells are shown. HDAC1 and H3 serve as nuclear and histone protein loading control respectively. A two-point titration (1:2.5 ratio) is shown for each condition. **e** Scatterplot showing log2 fold-change (logFC) expression between wild-type and *ASXL1/2* dKO cells as a function of average log2 counts per million (logCPM). Differentially expressed genes in *ASXL1/2* dKO cells are highlighted in purple. **f** Venn diagram showing the overlap between genes downregulated[60] or upregulated (right) in *BAP1* KO and *ASXL1/2* dKO cells

interaction between BAP1 and the ASXLs[13,23,35], our results suggest that BAP1 and ASXL proteins form a core complex (BAP1.com), which engages in transient interactions with additional partners such as FOXK1/2, HCFC1, or KDM1B (Fig. 2c). This model predicts that immunoprecipitation of any one of these transient partners would consistently retrieve the core complex but not necessarily other transient partners. Indeed, KDM1B immunoprecipitation from HeLa cells pulls down ASXL2 and BAP1 (as well as NSD3, which is part of a distinct complex with KDM1B) but none of the other transient partners (Supplementary Fig. 2c).

Given the role of ASXL1 and ASXL2 in driving BAP1-associated complex composition but the modest effect of their individual deletion on transcription, we sought to investigate their potential redundancy in BAP1-mediated H2A deubiquitination and transcriptional regulation. To address this question, we generated a double *ASXL1*/*ASXL2* KO. Of note, *ASXL3* is not expressed in wild-type HAP1 cells nor in *ASXL1/2* double KO cells (Supplementary Fig. 2d). We first evaluated the effect of this double KO on chromatin regulation and observed a robust increase in H2A ubiquitination levels similar to those observed in the *BAP1* KO (Fig. 2d). This result is consistent with the fact that interaction with the ASXLs is required for BAP1 enzymatic activity (Supplementary Fig. 2e). Notably, loss of ASXL1 and ASXL2 also led to a dramatic reduction in BAP1 protein levels (Fig. 2d), whereas *BAP1* transcript levels were unaffected (Supplementary Fig. 2d). Thus, ASXLs are not only necessary for the enzymatic activity of BAP1 but also for protein stability in vivo. Consistent with this effect on BAP1 protein accumulation, transcriptome analysis of *ASXL1/2* dKO cells revealed a major impact on gene expression (Fig. 2e). Most DE genes were downregulated (70%, 1073 downregulated genes out of 1576 total DE genes, also see Supplementary Fig. 2f for heatmaps of the top 100 DE genes) and there was a large overlap between genes downregulated in *BAP1* KO and *ASXL1/2* dKO cells (Fig. 2f, left). In comparison, the overlap between upregulated genes was much less pronounced (Fig. 2f, right), supporting the idea that the main role of BAP1.com is to promote transcription. Altogether, these results show that ASXL proteins are mandatory and redundant partners of BAP1.

BAP1 does not participate in Polycomb-mediated silencing. The data presented above suggest that the main role of BAP1 is to positively regulate transcription. While in agreement with several previously published studies, our findings contrast with a number of reports suggesting that BAP1 and ASXL proteins participate in Polycomb-mediated silencing[24–28]. To formally investigate the interplay between BAP1 and Polycomb proteins, we analyzed the consequences of BAP1 loss in conjunction with loss of RING1B and EZH2, key members of PRC1 and PRC2, respectively (Fig. 3a).

As expected, the main impact of inactivating either *RING1B* or *EZH2* is the transcriptional upregulation of a large set of genes (Supplementary Fig. 3a, b). The GO terms for the DE genes (either up or down regulated) in EZH2 KO cells partially overlap the categories observed upon BAP1 KO, including those related to signaling or development (Supplementary Fig. 3c and Fig. 1d, right panel). The terms associated with genes DE in the absence of RING1B are broader, which might reflect a less developmental specific function of PRC1 (Supplementary Fig. 3d). To further compare transcriptional changes between all KO conditions, we performed principal component analysis (PCA, Fig. 3b). Along PC1, most of the variance is driven by differences in BAP1 and polycomb machinery, again suggestive of distinct gene regulatory functions. As the Polycomb machinery is involved in gene silencing, we investigated the overlap between genes upregulated upon KO of *EZH2* or *RING1B* and KO of *BAP1*. As shown in Fig. 3c, only a minority of the genes regulated by PRC1 and/or PRC2 becomes upregulated upon loss of BAP1. To determine whether this limited overlap reflects a synergistic action of BAP1 with the Polycomb machinery, we investigated chromatin changes occurring at genes upregulated in *BAP1* KO. If BAP1 functions together with the Polycomb machinery to maintain transcriptional silencing at a subset of Polycomb target genes, loss of BAP1 is expected to result in a decrease in the Polycomb-mediated chromatin signature. However, chromatin immunoprecipitation (ChIP) followed by deep sequencing (ChIP-seq) did not reveal a decrease in Polycomb histone marks H2AK119ub1 or H3K27me3 at these genes in *BAP1* KO cells (Fig. 3d). In fact, H2AK119ub1 increased, possibly reflecting the more global increase of the mark caused by loss of BAP1 (see further below). As expected, transcriptional upregulation corresponded with marked increase in H3K4me3, a histone mark deposited preferentially near the 5′ ends of transcriptionally active genes. These results suggest that transcriptional upregulation occurring upon BAP1 loss is not caused by impaired Polycomb-mediated silencing. Instead, these gene expression changes may be secondary effects of widespread transcriptional downregulation.

We then analyzed the genome-wide distribution of the Polycomb-specific histone marks H3K27me3 and H2AK119ub1. Previous studies have reported a crucial role for the ASXLs in H3K27me3 deposition[24–28], but analysis of H3K27me3 revealed a high correlation in the genome-wide localization of the mark between wild-type, *BAP1*, *ASXL1*, and *ASXL2* KO cells, suggesting that loss of the ASXL proteins does not globally affect H3K27me3 distribution (Fig. 3e see also Supplementary Fig. 3e). This analysis, together with the lack of global change in H3K27me3 abundance in *BAP1*, *ASXL1*, or *ASXL2* KO cells as gauged by western blot (Fig. 1c), rules out an essential role for BAP1 and ASXL proteins in PRC2 function. In contrast, the distribution of H2AK119ub1 was significantly altered in BAP1 KO cells (Fig. 3f). Differential analysis of H2AK119ub1 signal between BAP1 KO and wild-type cells revealed widespread gains upon loss of BAP1 (12,388 regions) and much less depleted regions (3456), in keeping with the global increase of the mark seen by western blot analysis (Fig. 1c). H2AK119ub1 gains localized throughout the genome with an approximate twofold enrichment at putative promoter and enhancer regions over random peaks (Supplementary Fig. 3f). Interestingly, consistent with the known interplay between PRC1 and PRC2, gains of H2AK119ub1 were accompanied by an increase of H3K27me3 (Fig. 3g). Altogether, these data establish that BAP1.com does not act in synergy with the Polycomb machinery to maintain gene silencing but, instead, that its activity might restrain Polycomb enzymatic activity.

PRC2-antagonistic and -independent role of BAP1. As shown above, the major impact of loss of BAP1 is the downregulation of gene expression accompanied by widespread gains of H2AK119ub1 and H3K27me3. To investigate whether gains of Polycomb marks and transcriptional changes following BAP1 loss are linked, we first analyzed changes in chromatin composition at genes that are downregulated in the absence of BAP1 (Fig. 4a). As expected, transcription-associated H3K4me3 decreased concomitantly with transcriptional downregulation. In principle, increased levels of the H2AK119ub1 and H3K27me3 repressive marks could be a direct consequence of loss of BAP1 deubiquitinase activity or could be a secondary event caused by transcriptional downregulation[36]. To discern between these two possibilities, we genetically inactivated *EZH2* in *BAP1* KO cells

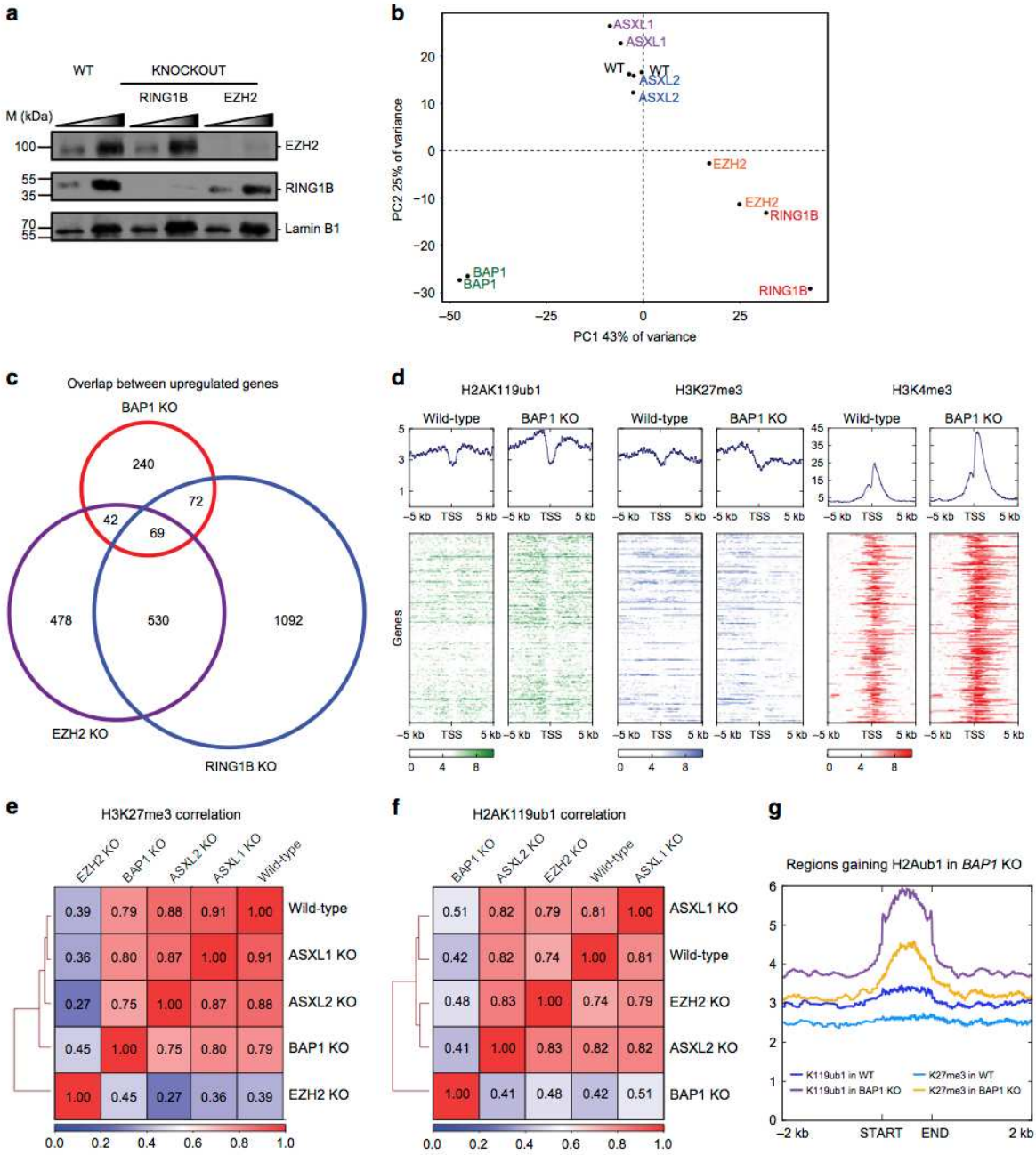

Fig. 3 BAP1.com is dispensable for Polycomb-mediated silencing. **a** Western blot analysis of EZH2 and RING1B in wild-type (WT), *RING1B* KO, and *EZH2* KO cells. Lamin B1 is used as a loading control. A two-point titration (1:2 ratio) is shown for each condition. **b** PCA analysis of WT, *RING1B*, *EZH2*, *ASXL1*, *ASXL2*, and *BAP1* KO transcriptome. **c** Venn diagram showing the overlap between genes upregulated in *BAP1*, *EZH2*, or *RING1B* KO cells. **d** Heatmaps showing H2AK119ub1, H3K27me3, and H3K4me3 distribution in a −5/ + 5 kb window around the transcription start site (TSS) of genes upregulated in *BAP1* KO cells in wild-type and *BAP1* KO cells. Corresponding average profiles are plotted on top of each heatmap. **e**, **f** Correlation heatmap of H3K27me3 (**e**) and H2AK119ub1 (**f**) distribution between wild-type cells and the different KO conditions indicated. **g** Plots showing average enrichment of H2AK119ub1 and H3K27me3 in wild-type and *BAP1* KO cells at regions that gain H2AK119ub1 upon BAP1 loss

(Fig. 4b) and assessed whether *EZH2* deletion can restore expression of BAP1-regulated genes. Of note, *EZH2* deletion in cells already KO for *BAP1* did not impair proliferation (Fig. 4c), consistent with recent evidence challenging the reported synthetic lethal relationship between EZH2 inhibition and BAP1 inactivation[29,37]. Of the 913 genes downregulated upon *BAP1* KO, a large majority (741 genes) remain silent in the *BAP1/EZH2* double KO (Fig. 4d, top heatmap), whereas the expression of a minority (172 genes) is increased in the double KO compared with BAP1 single KO cells (Fig. 4d, bottom heatmap). This set of genes is also upregulated upon deletion of *EZH2* in a *BAP1* wild-type context, indicative of a balanced antagonistic regulation

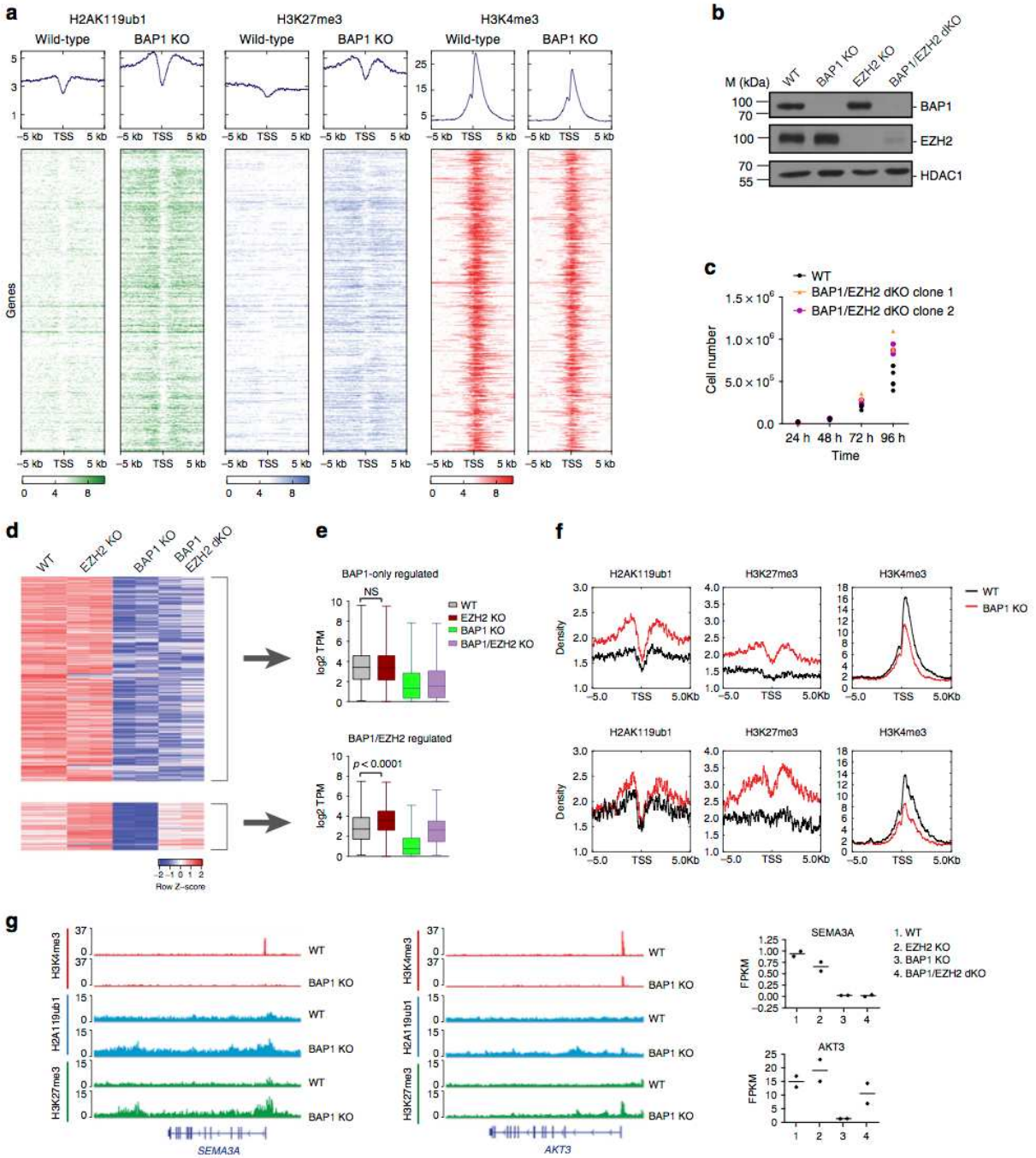

Fig. 4 BAP1.com promotes transcription independently of an antagonism with the PRC2 complex. **a** Heatmaps showing H2AK119ub1, H3K27me3, and H3K4me3 distribution in a −5/ + 5 kb window around the transcription start site (TSS) of genes downregulated in *BAP1* KO cells in wild-type and BAP1 KO cells. Corresponding average profiles are plotted on top of each heatmap. **b** Western blot analysis of BAP1 and EZH2 in single and double *BAP1/EZH2* KO cells. HDAC1 is used as a loading control. **c** Dot plot showing proliferation of wild-type cells and two independent clones of *BAP1/EZH2* dKO cells. $n = 3$. **d** Heatmap of gene expression (*Z*-scores) in the different genotypes. Top: BAP1-only-regulated genes, and bottom: BAP1- and EZH2-regulated genes, TSS = transcription start site. **e** Box-plots (median, lower, and upper quartiles, lowest and highest values) of log2 transcript per million (TPM) expression values of BAP1-only-regulated genes (top) and BAP1/EZH2-regulated genes[61] in the different conditions as indicated. Result of the Mann–Whitney test on the *EZH2* KO versus wild-type comparison is indicated. **f** Plot showing average enrichment of H2AK119ub1, H3K27me3, and H3K4me3 in a −5/ + 5 kb around the TSS for BAP1-only-regulated genes (top) and BAP1/EZH2-regulated genes[61]. **g** Example snapshots of H2AK119ub1, H3K27me3, and H3K4me3 enrichment in WT and *BAP1* KO cells at a BAP1-only-regulated gene and a BAP1/EZH2-regulated gene (middle). Expression values of the corresponding genes across WT, *EZH2*, *BAP1*, and *BAP1/EZH2* KO conditions as detected in corresponding RNA-seq data are shown on the right. Horizontal bars indicate the mean expression

between BAP1 and EZH2 (Fig. 4d, e). Moreover, all genes downregulated in the absence of BAP1 gain H3K27me3 upon *BAP1* KO (Fig. 4f, see also Fig. 4g and Supplementary Fig. 4a, b for specific examples), suggesting that such a gain is a consequence rather than a cause of transcriptional downregulation. Thus, although BAP1 and PRC2 act in an opposite fashion at a number of genes, BAP1 promotes gene expression in a manner that is largely independent of an antagonism with the PRC2 complex.

Similarities between BAP1.com and general co-activators. Having analyzed the impact of loss of BAP1.com on steady-state gene transcription, we next sought to examine its role in transcriptional activation in response to a transcriptional stimulus. Considering previous reports suggesting that BAP1 may modulate nuclear receptor-mediated gene regulation[38,39], we investigated whether deletion of *BAP1* would affect response to RA treatment. We initially performed a time-course analysis of two of the best-characterized direct RA transcriptional targets: *RARβ* and *CYP26A1* (Fig. 5a). In contrast to the above-mentioned reports that suggest a repressive role for BAP1.com at RA target genes, we observed that activation of both *RARβ* and *CYP26A1* was severely compromised in *BAP1* KO cells. Of note, loss of BAP1 did not affect the protein levels of RARα, a major RA-binding nuclear receptor (Supplementary Fig. 5a). To determine whether these results reflect a general role for BAP1 in the transcriptional response to RA, we analyzed the transcriptome of WT and *BAP1* KO cells in response to RA treatment (after 24 h). Nearly twice as many genes were significantly activated upon RA treatment in WT than in *BAP1* KO cells ($n = 88$ versus $n = 47$, FDR < 0.05 and absolute log2 fold-change > 1, Supplementary Fig. 5b). Furthermore, analyzing the entire set of genes activated in either condition, we found that the response to RA is significantly attenuated in *BAP1* KO cells (Fig. 5b), demonstrating that BAP1 is required for optimal RA-mediated transcriptional activation.

To further investigate the hypothesis that BAP1 functions as a transcriptional co-activator, we compared transcriptional defects occurring in *BAP1* KO cells and KOs of *SMARCB1* (Supplementary Fig. 5c), which encodes an essential member of the BAF chromatin-remodeling complex, and *CREBBP*, which encodes a histone acetyltransferase. Both KO cell lines were viable, although they appeared less healthy than the wild-type counterpart. As with loss of BAP1, loss of SMARCB1 and CREBBP severely impaired RA-mediated transcriptional activation of *RARβ* and *CYP26A1* (Fig. 5a). To determine to what extent BAP1, CREBBP, and SMARCB1 affect transcription, we compared the transcriptome of HAP1 cells mutated for each of these genes (this study and[40]). As shown in Fig. 5c, each KO leads to the downregulation of a similar number of genes (BAP1: 1498 genes, CREBBP: 1106 genes and SMARCB1: 862 genes). We next compared the overlap between genes regulated by the three proteins. Although RNA-seq for each KO condition was performed in different laboratories, there was a significant overlap between genes downregulated upon loss of BAP1, CREBBP, or SMARCB1 (Fig. 5d, $p < 0.0001$ for all three comparisons). Nonetheless, the majority of downregulated genes remain specific to each respective KO, indicating that BAP1, CREBBP, and SMARCB1 each regulate a distinct set of genes.

To determine the level of specificity of gene recruitment, we compared BAP1.com localization at chromatin to the localization of transcription machinery. We obtained previously published ChIP-seq data for ASXL1[9] and RNA-PolII (ENCODE) in human HEK-293 cells, and BAP1[4] and RNA-PolII (ENCODE) in mouse bone marrow-derived macrophages. Analysis of ASXL1/PolII and BAP1/PolII enrichment revealed a marked correlation between RNA-PolII and BAP1.com profiles, both in terms of enrichment intensity and profile along the gene body (Fig. 5e, f and Supplementary Fig. 5d, e and ChIP-seq screenshots in Fig. 5g and Supplementary Fig. 5f). This result strongly supports a role for BAP1.com as a general co-activator and suggests that functional differences between BAP1 and other co-activators are due to gene-specific requirements for their respective enzymatic activities during transcriptional activation, rather than gene-specific targeting. Together, these data provide compelling evidence that BAP1.com functions as a general transcriptional co-activator.

PRC1 is epistatic to BAP1. Finally, we sought to understand how BAP1 exerts its function on transcription. In principle, BAP1 could act through its enzymatic activity to modify proteins that modulate transcription, and/or by recruiting factors that participate in promoting transcription. Several reported BAP1 substrates such as H2AK119ub1, BARD1, HCFC1, and OGT could potentially mediate its function in gene activation[4–6,41]. HCFC1 is a transcriptional regulator and, in addition to modulating its ubiquitination, BAP1 has been suggested to contribute to its recruitment and/or regulation of its stability. More recently, BAP1 was suggested to participate in the recruitment of the MLL3 histone methyltransferase at enhancers by direct interaction with MLL3 PHD repeats[31].

To determine which of these functions is critical for BAP1 activity, we first assessed if BAP1 function requires its DUB activity. For this purpose, we performed rescue experiments in *BAP1* KO cells, reintroducing either wild-type BAP1 or a catalytically dead version (BAP1 C91S). Both versions of BAP1 were re-expressed at a similar level, slightly higher than the original endogenous level (Fig. 6a, upper panel). Focusing on a selection of genes whose expression is dramatically reduced in the absence of BAP1, re-expression of wild-type BAP1 protein restores up to 75% of the wild-type levels of the transcripts while the C91S mutant is unable to rescue transcription of the tested genes (Fig. 6a, lower panel). Although based on a subset of genes, this analysis suggests that BAP1 catalytic activity is required for its function, consistent with previous studies[3,13].

We next sought to determine the relative contribution of H2AK119ub1 compared with other substrates to BAP1's function. H2AK119ub1 is solely deposited by the PRC1 complex through its two paralogous enzymes RING1A and RING1B. We thus genetically inactivated both enzymes in HAP1 cells and then mutated BAP1 in the context of *RING1A/B* dKO cells. As expected, H2AK119ub1 was completely absent from both *RING1A/B* dKO and *RING1A/B;BAP1* tKO cells (Fig. 6b). Of note, proliferation of the two mutant conditions was unimpaired (Supplementary Fig. 6). We first assessed the overlap between genes positively regulated by BAP1 and negatively regulated by PRC1. As shown in Fig. 6c, a minority of genes, 17%, were under such opposite regulation. Analysis of BAP1-only-regulated genes as a group confirmed the absence of regulation by PRC1 (Fig. 6d) and also showed that this set of genes has an overall higher level of expression compared with BAP1/PRC1-regulated genes (Fig. 6d). To determine whether this BAP1-mediated transcriptional regulation nevertheless depends on PRC1-mediated H2AK119ub1, we assessed transcriptional changes in *RING1A/B;BAP1* tKO versus *RING1A/B* dKO. *BAP1* was the only DE transcript between the two mutant conditions (Fig. 6e), indicating that BAP1 is no longer functional in the absence of PRC1 activity. Consistently, transcriptional response to RA treatment was equivalent in *RING1A/B* dKO and *RING1A/B*; *BAP1* tKO cells (Fig. 6f, transcriptional response to RA differs in *RING1A/B* dKO

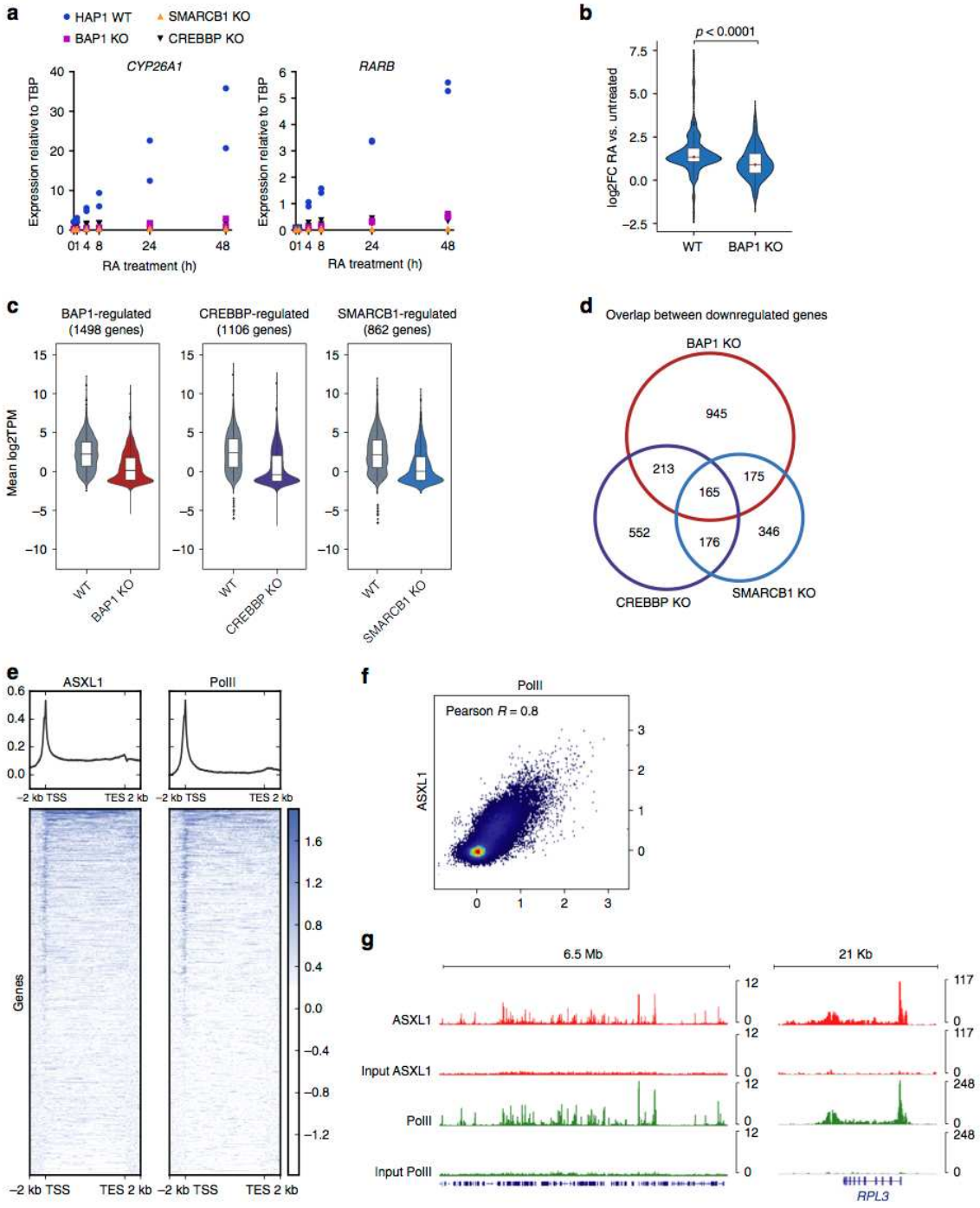

Fig. 5 Comparison of BAP1 and the CREBBP and SMARCB1 co-activator proteins. **a** RT-qPCR analysis of *CYP26A1* and *RARB* expression following RA treatment at different time-points in wild-type or *BAP1*, *CREBBP*, and *SMARCB1* KO cells. $n = 2$. **b** Violin plots showing log2 fold-change expression of RA-responsive genes ($n = 114$ genes, see text for details) in wild-type and BAP1 KO cells. *P*-value from the Mann–Whitney test is shown. **c** Violin plots showing log2TPM expression of BAP1- CREBBP-, and SMARCB1-regulated genes in wild-type (WT) or in the respective KO conditions. **d** Venn diagram showing overlap between genes that are downregulated in *BAP1*, *CREBBP*, and *SMARCB1* KO cells. **e** Heatmaps showing ASXL1 and RNA PolII density around the TSS and termination end site (TES) (including 2-kb upstream and downstream) scaled to an equivalent 10 kb in human HEK-293 cells. Corresponding average profiles are plotted above each heatmap. **f** Scatterplot showing PolII versus ASXL1 enrichment around the TSS (including 2 kb upstream and downstream) at all annotated genes. Pearson correlation coefficient is displayed. **g** Snapshots of ASXL1 and RNA PolII enrichment at representative regions. The input is displayed below each corresponding ChIP-seq experiment

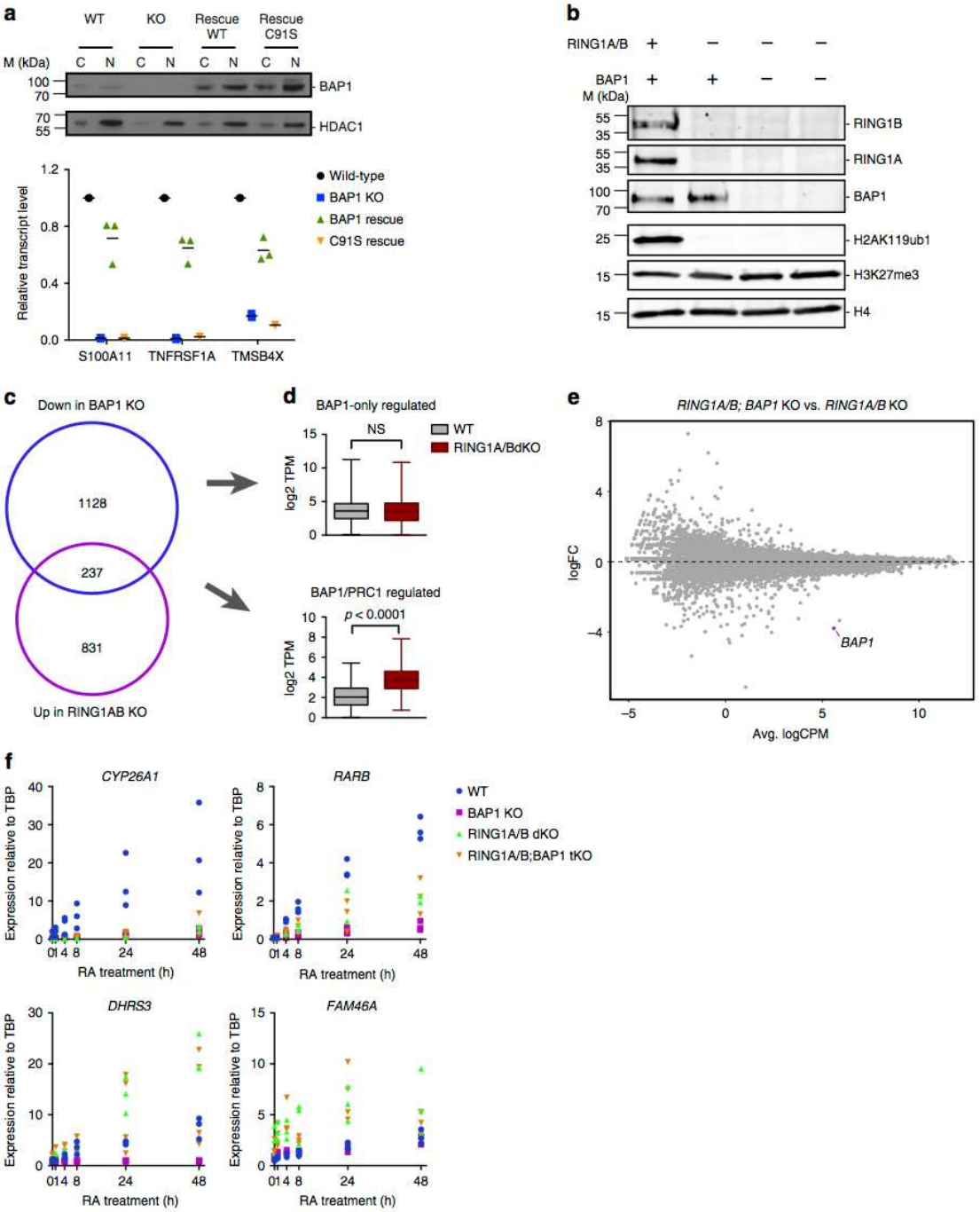

Fig. 6 BAP1 function relies on its deubiquitinase activity against H2AK119ub1. **a** Reintroduction of wild-type or catalytically dead (C91S) BAP1 in *BAP1* KO cells. Top panel: western blot analysis of BAP1 expression in cytoplasmic (C) or nuclear (N) fractions. HDAC1 is used as a loading control. Bottom panel: RT-qPCR analysis of three BAP1-regulated genes in wild-type, *BAP1* KO and the two rescue conditions. Horizontal bars indicate the mean expression. $n = 3$. **b** Western blot analysis of RING1A, RING1B, BAP1, H2AK119ub1, and H3K27me3 in wild-type, *RING1A/B* double KO, or *RING1A/B; BAP1* triple KO HAP1 cells. H4 is used as a loading control. **c** Venn diagram showing the overlap between genes downregulated in *BAP1* KO or upregulated in *RING1A/B* dKO cells. **d** Box-plots (median, lower, and upper quartiles, lowest and highest values) of log2 transcript per million (TPM) expression values of BAP1-only-regulated genes (top) and BAP1/PRC1-regulated genes[61] in WT and *RING1A/B* dKO cells. *P*-value of the Mann–Whitney test on WT versus *RING1A/B* dKO comparison is indicated. **e** Scatterplot showing log2 fold-change (logFC) expression between *RING1A/B* dKO and *RING1A/B; BAP1* tKO cells as a function of average log2 counts per million (logCPM). *BAP1*, the only differentially expressed genes is highlighted in purple. **f** RT-qPCR analysis of *CYP26A1*, *RARB*, *DHRS3*, and *FAM46A* expression following RA treatment at different time-points in wild-type or *BAP1*, *RING1A/B*, *RING1A/B;BAP1* KO cells

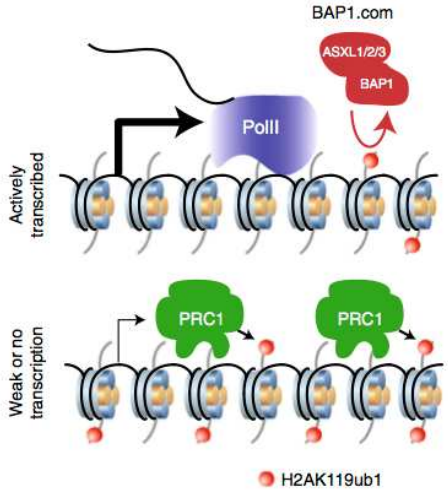

Fig. 7 Model of BAP1.com function

cells compared with WT cells). This result contrasts with the limited dependency of loss of BAP1 on the PRC2 complex (Fig. 4), and suggests that gain of H2AK119ub1 might inhibit transcription in a way that is partly independent of PRC2. Taken together, these results establish that PRC1 is epistatic to BAP1.com and that BAP1 promotes transcription by counteracting PRC1-mediated H2A ubiquitination (Fig. 7).

Discussion

Despite their prominent roles as tumor suppressors, the biological functions of BAP1 and ASXL proteins remain poorly characterized. In this study, we performed extensive biochemical and genetic analyses in isogenic mutant cell lines to address the role of BAP1 in transcriptional regulation, the respective contributions of BAP1 and the ASXL proteins to the regulation of gene expression and the interplay between BAP1.com and the Polycomb machinery.

Both biochemical and genetic evidence indicate that BAP1 and ASXL proteins function together to regulate H2AK119ub1 and gene expression. Our biochemical analyses largely confirm the previously reported interactome of BAP1[9] and further enable distinguishing transient interactors (FOXK1/2, MBD5/6, HCFC1, etc.) from the core complex composed of BAP1 and one ASXL paralog. With the exception of KDM1B, which appears to interact specifically with BAP1-ASXL2, the BAP1 interactome is essentially identical whether the complex forms around ASXL1 or ASXL2. This might reflect the fact that most of the interactions are mediated directly by BAP1, as suggested for HCFC1, YY1, OGT, and FOXK1/2[3,35]. Nonetheless, interaction with the ASXL paralogs is required for BAP1 enzymatic activity and protein stability. The overall interchangeability between the ASXLs explains the observation that knocking out a single paralog only mildly affects gene expression. An immediate implication of this finding concerns the tumor-type-specific spectrum of BAP1 and ASXL mutations. While previous studies have argued that this non-overlapping mutation spectrum might be the result of independent and sometimes opposite functions of BAP1 and ASXL proteins[24,29], our findings instead suggest that, due to the redundancy among ASXL proteins, the loss of only one of them results in a much less severe disruption of BAP1.com function than loss of BAP1. Hence, we propose that the predominance of ASXL mutations in myeloid malignancies may be the result of a selective pressure aimed at only partially ablating BAP1 function, whereas loss of BAP1 in other malignancies would reflect a need for complete inactivation of the complex.

In contrast to the reported role of Calypso in *Drosophila*, our results argue against an implication of BAP1.com in the Polycomb machinery. Instead, the picture that emerges from this work is that the function of BAP1.com is to promote transcription by limiting PRC1 repressive activity. This property may not be surprising given that the activity of the complex is opposite to that of PRC1 in the regulation of H2A ubiquitination and that PRC1 and ASXL1 show rather distinct localization patterns[9]. The more limited dependency on PRC2 activity that we observed is likely the result of partly divergent functions between the two Polycomb complexes, and in particular the ability of PRC1 to be recruited and silence genes independently of PRC2[42,43].

Our findings are also in line with previous studies using artificial recruitment of BAP1 to a transgene[3], overexpressing hyperactive forms of ASXL1[44] or focusing on individual loci[30], which all support a role for BAP1 in gene activation. An obvious question arising from these results is whether BAP1 and Calypso have divergent functions or whether we can reconcile their proposed contributions to gene regulation in both species. First, it is noteworthy that there is no clear consensus as to whether the Calypso–Asx complex is generally involved in Polycomb-mediated silencing. Indeed, mutation of Asx in *Drosophila* leads to a complex phenotype that exhibits features of both Polycomb and Trithorax mutants[45], a situation that is also found in *Asxl1* mutant mouse embryos[46]. Second, only a subset of Polycomb target genes was found to be aberrantly activated upon loss of Asx or Calypso[47]. Third, part of the difference might result from the extent to which BAP1 and Calypso regulate H2A ubiquitination. Although loss of Asx leads to an approximate 10-fold increase in the total levels of H2AK118ub1[22], we only observe an approximate twofold increase in the level of the mark upon inactivation of either of BAP1 or ASXL1/2. This suggests that in *Drosophila*, the Calypso–Asx complex might have a more critical function in restricting PRC1 activity to its normal site of action. Aberrant deposition of H2AK119ub1 and consequently of H3K27me3 could indirectly impair Polycomb-mediated transcriptional silencing. Genome-wide functional studies analyzing the global consequences of loss of Asx or Calypso in flies should help clarify these questions and address whether PR-DUB and BAP1.com have similar functions.

Interestingly, the function of BAP1 in stimulating transcription is comparable to that of well-known general co-activators such as SMARCB1 and CREBBP, not only to ensure steady-state gene transcription of hundreds of genes but also in response to stimulus as exemplified by RA treatment. Hence, efficient transcriptional stimulation entails the action of enzymatic activities that converge at creating a permissive chromatin environment, either through histone eviction and/or repositioning (e.g., SMARCB1), reduction of histone charge through acetylation (e.g., CREBBP) or removal of repressive chromatin marks (e.g., BAP1). Interestingly, these different activities can be needed together or separately, as suggested by our observation that SMARCB1, CREBBP, and BAP1 regulate common, as well as specific sets of genes. As BAP1.com generally colocalizes with the transcriptional machinery, we speculate that while present at most transcribed regions, BAP1.com impacts gene expression selectively depending on the chromatin environment. Further investigation will be necessary to decipher what determines the transcriptional response to BAP1 deletion. It will also be interesting to investigate what controls BAP1.com targeting to transcribed regions. We envision that the ASXL proteins, through their PHD finger, a domain that can potentially bind methylated lysine or arginine residues[48], could participate in reading post-translational marks associated with active transcription. Our work provides insight into the function of BAP1.com and paves the way

for novel strategies to target tumors harboring alterations in this chromatin-modifying complex.

Methods

Cell lines. HAP1 cells were kindly provided by T. Brummelkamp and cultured in Iscove's Modified Dulbecco's Medium media supplemented with 10% fetal bovine serum (FBS) and 1% L-glutamine. HeLa-S3 cells were kindly provided by S. Ait-Si-Ali. They were cultured in adherence in Dulbecco's modified Eagle's media (DMEM) supplemented with 10% FBS and 1% L-glutamine. Non-adherent culture of HeLa cells was performed in DMEM supplemented with 5% FBS and 1% L-glutamine following guidelines from Nakatani and Ogryzko[49]. SF9 cells were cultured in SF-900 II serum-free medium (Invitrogen) supplemented with 5% FBS, 1% penicillin/streptomycin (Invitrogen) and Amphotericin B at 28 °C. All cell lines were tested for mycoplasma contamination on a regular basis.

Constitutive KOs in HAP1 cell line. Mutations of *BAP1*, *ASXL1*, *ASXL2, EZH2, RING1A*, and *SMARCB1* in HAP1 cells were performed using CRISPR/CAS9 technology using the strategy described in[34]. Briefly, a STOP cassette containing a antibiotic resistance gene followed by a polyadenylation sequence from SV40 was inserted by homologous recombination in intronic or exonic sequences of the target genes. Intronic targeting vectors include the EN2 splice acceptor sequence for proper splicing of the antibiotic resistance gene. Antibiotic resistance clones were then picked in 96-well plates and genotyped. KOs were validated by RT-qPCR and western blot when the corresponding antibody was available. The selected clones were thus used as constitutive KOs, using the parental cell line as a control in all experiments. *RING1B* and *CREBBP* KO HAP1 cells were purchased from Horizon Discovery.

Stable expression in HeLa cells. For mass spectrometry analysis of Flag-tagged BAP1, ASXL1, ASXL2, and KDM1B in HeLa cells, complementary DNAs (cDNAs) encoding the different proteins were first subcloned in pRev retroviral plasmid (gift from S. Ait-Si-Ali), downstream a 2xFlag-2xHA sequence and upstream an internal ribosome entry site sequence followed by CD25 cDNA. Retroviruses were produced by transfection of a 293 Phoenix cell line (gift from S. Ait-Si-Ali) and HeLa-S3 cells were infected by incubation with viral supernatants for 3 h at 37 °C. Infected cells were then selected by fluorescence activated cell sorting against CD25 expression using CD25-FITC antibody and following manufacturer's instructions (BD Biosciences 553866). Expression was assessed by western blot analysis of nuclear extracts.

Rescue experiment (BAP1 WT and C91S). Reintroduction of wild-type or enzymatically dead (C91S) BAP1 was performed by infection of BAP1 KO cells with a pBABE retrovirus[50]. Production of retroviral particles was performed in 293T cells. Transduction was performed by incubating the cells with viral particles mixed with Polybrene (final concentration, 8 µg/ml) for 3 h at 37 °C and subsequently selected with puromycin.

Proliferation assays. In all, 10,000 cells were plated in six-well plates in triplicates and counted every 24 h over 4 days using a Vi-cells counter (Beckman-Coulter).

RT-qPCR. Total RNA was isolated using Trizol-Chloroform extraction and isopropanol precipitation. cDNA was synthesized using High Capacity cDNA RT kit (4368814-Applied Biosystems) and quantitative PCR was performed with technical triplicate using SYBR green reagent (Roche) on a ViiA7 equipment (Applied Biosystems). At least two independent experiments (biological replicates) were performed for each assay and RT negative controls were always included. Primer sequences for qPCR analysis are provided in Supplementary Table 1.

RNA sequencing. RNA sequencing were performed for two independent biological replicates for each condition. In total, 100-bp single-end reads were generated for the RA analysis and 100-bp paired-end reads for all other samples using the HiSeq 2500 platform. Raw reads were trimmed with cutadapt (1.12;[51]) using the Trim Galore! (0.4.4; bioinformatics.babraham.ac.uk) wrapper (default settings) and subsequently aligned to the complete human ribosomal RNA sequence with bowtie (1.2;[52]). Reads that did not align to rRNA were then mapped to the human reference genome (GRCh37/hg19) and gene counts generated with STAR (2.5.2b;[53]) using the following parameters: --quantMode GeneCounts --outSAMtype BAM SortedByCoordinate --runMode alignReads --outFilterMismatchNmax 6 --outFilterMultimapNmax 20 --outSAMmultNmax 20 --outSAMprimaryFlag OneBestScore. Counts were generated using properly paired (for paired-end data) and uniquely mapped reads that overlap the exon boundaries of each gene. More than 94% of reads mapped uniquely for all paired-end sequencing samples and >92% for single end. Per sample read counts are provided in Supplementary Table 2.

BAM files for the CREBBP-KO and HAP1-WT samples were obtained from the Institut de Cancérologie Gustave Roussy and gene counts for these samples were generated using featureCounts (1.5.1;[54]) with the following parameters: -C -p -s 2 -T 8 -F GTF -t exon. Raw reads were trimmed as part of bcl2fastq for Illumina adapters and aligned with RSEM (1.2.25;[55]) and Bowtie2 (2.2.6;[52]) to GRCh37/hg19 using default parameters. Total uniquely mapping reads were calculated using RSeQC (2.6.4;[56]).

The reference FASTA was downloaded from UCSC (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/) and the annotation (GTF) file from gencodegenes.org (comprehensive gene annotation Release 19/GRCh37.p13).

Differential expression analysis. Genes were filtered to include those with counts per million (CPM) > 0.5 in at least two samples. Raw count data were transformed to log2-CPM and normalized with the TMM method using edgeR (3.18.1; Robinson 2010 and McCarthy 2012). A linear model was fit to the normalized expression values for each gene and empirical Bayes statistics were computed for each KO versus wild-type or rescue with limma (3.32.7; Ritchie, 2015). DE genes were identified from the linear fit after adjusting for multiple comparisons and filtered to include those with FDR < 0.05 and absolute logFC > 1.

GO enrichment analysis was performed with goseq (1.28.0) using the Wallenius method to calculate a probability weighting function for top differential genes as a function of a gene's median transcript length. GO terms with FDR < 0.01 were collapsed using REVIGO[57].

CREBBP, SMARCB1, BAP1 analysis. Raw data for SMARCB1 and WT were downloaded from GEO (GSE75515). RNA-seq analysis was performed as described previously. Only genes common to all datasets were used in the analysis. Significance of overlap for significantly upregulated and downregulated genes between all three KOs was determined applying a one-sided Fisher's exact test (alternative hypothesis = greater).

Chromatin immunoprecipitation. For ChIPs experiment, cell confluence and amount of starting material were kept constant by plating defined number of cells 2 days before cross-linking. Briefly, cells were fixed in 1% formaldehyde for 10 min at room temperature, quenched by adding glycine to a final concentration of 0.125 M, rinsed with phosphate-buffered saline (PBS) and resuspended in buffer LB1 (Hepes-KOH pH 7.5 50 mM, NaCl 140 mM, EDTA 1 mM, glycerol 10%, NP-40 0.5%, Triton X-100 0.25% + Protease inhibitors). Cells were rocked at 4 °C for 10 min, pelleted and resuspended in buffer LB2 (NaCl 200 mM, EDTA 1 mM, EGTA 0.5 mM, Tris pH 8 10 mM + Protease inhibitors), pelleted again and resuspended in buffer LB3 (EDTA 1 mM, EGTA 0.5 mM, Tris pH 8 10 mM + Protease inhibitors). Sonication was performed on a Bioruptor (Diagenode), 0.5% *N*-lauroyl-sarcosine added and after rocking at room temperature for 10 min supernatant was kept. For the immunoprecipitation, chromatin (10 µg) was incubated antibodies (around 2 µg) overnight in presence of 1% Triton and 0.1% sodium deoxycholate. Beads blocked with bovine serum albumin were added the day after and incubated at 4 °C for 3 h before processing to the washes in RIPA buffer six times (50 mM Hepes pH 7.6, 10 mM EDTA, 0.7% DOC, 1% NP-40, 0.5 M LiCl + Protease inhibitors) and once in buffer TEN (10 mM Tris pH 8.0, 1 mM EDTA, 50 mM NaCl). Elution was done in buffer TES (50 mM Tris pH 8.0, 10 mM EDTA, 1% sodium dodecyl sulfate), before reversing the crosslink overnight and incubating the samples successively with RNAse A and proteinase K prior to phenol/chloroform/isoamyl-alcohol DNA extraction. ChIPs were analyzed by qPCR using the primers described in Supplementary Table 1.

ChIP sequencing. In total, 100-bp single-end reads were generated using the HiSeq2500 platform. Reads were mapped to the human reference genome (GRCh37/hg19) with Bowtie2 (2.2.9) using default parameters. PCR duplicates were removed with Picard Tools MarkDuplicates (1.97; http://broadinstitute.github.io/picard). Total uniquely mapping reads were calculated using RSeQC bam_stat.py (2.6.4). BAM files were filtered to exclude common artifact regions (merged consensus artifact regions: http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg19-human/). Reads were counted in bins of length 25, RPKM normalized, and converted to bigWig format using DeepTools bamCoverage (2.4.1) for all heatmaps.

Scores upstream and downstream of transcription start sites (TSSs) were computed from normalized bigWig files with deepTools computeMatrix (2.4.1) using reference-point mode. TSS plots were generated with deepTools plotHeatmap (2.4.1).

To identify regions that gain or lose H2AK119ub1, differential analysis was performed using SICER (1.1) between the BAP1 KO and WT H2AK119ub1 using random background to determine statistically enriched regions. To assess the window and gap sizes used in the analysis, we plotted the aggregate score versus the gap size for window sizes of 200, 400, 600, and 800 and gap sizes 1w to 5w. Increasing the window size to 800 showed saturation near a gap of 4w. A gap size of 2400 (3w) corresponded with a score that was sufficiently close to saturation and was chosen for the final analysis. For gap sizes 1w up to 5w, BAP1 KO showed an average of 3.6× more H2AK119ub1 enriched regions than the WT. With w = 800 and g = 2400, 12,388 regions (average length = 24,714.2) were significantly increased (E = 1000, L = 0.74) in the H2AK119ub1 BAP1 KO v. WT and 3456 regions (average length = 30,712.9) were significantly decreased.

PolII, BAP1 and ASXL1 ChIP-seq analysis. Raw data were downloaded from GEO (GSE40723, GSE36027, GSE51673, and GSE31477). FASTQ files were merged for samples with multiple runs and mapping was performed as previously described to hg19 or mm10. bigWig files were generated with deepTools bamCoverage (2.4.1).

Artifact regions were excluded and read counts were normalized to log2(sample/input).

Average values 2 kb around the TSS were computed using deepTools multiBigwigSummary and correlation plots for these regions were generated with deepTools plotCorrelation (--corMethod pearson –removeOutliers –skipZeros).

Merged consensus blacklists for hg19 and mm10 were obtained from the Kundaje lab at Stanford University (Stanford, CA, USA).

Antibodies. BAP1 (1/500 dilution; C-4; sc-28383), FOXK1 (1/500 dilution; G-4; sc-373810), YY1 (1/500 dilution; sc-7341) and RAR-alpha (1/500 dilution; sc-551 ×) antibodies were purchased from Santa-Cruz; FLAG (1/1000 dilution; M2; F1804) was purchased from Sigma; HCFC1 (1/1000 dilution; A301-400A) and RNF2 (1/1000 dilution; A302-869A) antibodies were purchased from Bethyl Laboratories; Lamin B1 (1/3000 dilution; ab16048); RING1A (1/1000 dilution; 2820S), RING1B (1/1000 dilution; D22F2; 5694S) H2AK119ub1 (1/3000 dilution; D27C4; 8240S), H3K27me3 (1/3000 dilution; C36B11, 9733S), H3K4me3 (1/3000 dilution; C42D8, 9751) and H4 (1/3000 dilution; 2935S) antibodies, were purchased from Cell Signaling Technology; H2A.Z (1/1000 dilution; 39113), H2B (1/1000 dilution; 5HH2-2A8; 61037); H2BK120ub (1/1000 dilution; C56; 39623), H3 (1/3000 dilution; C-terminal; 39163), and KDM1B (1/1000 dilution; 61457) antibodies were purchased from Active Motif; H3K4me2 (1/3000 dilution; MCA-MAB10003-100-Ex) antibody was purchased from Cosmo Bio; alpha-Tubulin (1/3000 dilution; 1F4E3; A01410) was purchased from Genscript.

Histone extraction. Cells were lysed in a hypotonic lysis buffer (10 mM Tris pH 6.8, 50 mM Na_2SO_4, 1%Triton X-100, 10 mM $MgCl_2$, 8.6% sucrose, and protease inhibitors) and carefully homogenized using a Dounce A homogenizer. After a centrifugation step at 6000 *g* for 10 min at 4 °C, the pellet was washed with 10 mM Tris pH 7.5, 13 mM EDTA and resuspended in ice-cold water. Protein precipitation was performed by addition of sulfuric acid 0.4 N final concentration and 1-h incubation on ice. The samples were centrifugated at 20,000 *g* for 10 min at 4 °C and the histone-containing supernatant collected and neutralized by addition of 0.5 volume of 1.5 M Tris pH 8.8. Quantification was performed by Bradford assay and confirmed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) stained with Coomassie.

Nuclear extracts (high salt). For nuclear extract preparation, cells were incubated with buffer A (10 mM Hepes pH 7.9, 2.5 mM $MgCl_2$, 0.25 M sucrose, 0.1% NP-40 and protease inhibitors) for 10 min on ice, centrifuged at 8000 rpm for 10 min, resuspended in buffer B (25 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 700 mM NaCl, 0.1 mM EDTA, 20% glycerol and protease inhibitors), sonicated and centrifuged at 14,000 rpm for 15 min. Uncropped western blot data are provided in Supplementary Figs. 7–14.

Chromatography analysis. For analysis of endogenous protein profiles in HAP1 cells, nuclear extracts (high salt) were first dialyzed against BP100 (50 mM potassium phosphate pH 6.8, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, protease inhibitors) and clarified by high-speed centrifugation. Samples were then purified by ion exchange chromatography using a HiTrap SP HP 5 ml column (GE Healthcare). Elution was performed by step elution with increasing NaCl concentration. In all, 500 mM elution was concentrated 5× time on centricon (Millipore, cut-off 10 kDa) and analyzed on Superose 6 PC3.2 increase column (GE Healthcare). The native molecular size markers used for column calibration were thyroglobulin (669 kDa), ferritin (440 kDa), and aldolase (158 kDa).

Mass spectrometry analysis. For mass spectrometry analysis of Flag-tagged constructs overexpressed in HeLa cells, 100 mg of nuclear extracts were used. Nuclear extracts were first dialyzed in BC250 (50 mM Tris pH 8.0, 250 mM KCl, 1 mM EDTA, 10% Glycerol and protease inhibitors). Precipitates were removed by centrifugation and the supernatant was incubated with anti-FLAG M2 affinity gel overnight. The beads were then washed three times with BC250 + 0.05% NP-40 and eluted with 0.2 mg/ml Flag peptide, precipitated with ice-cold acetone and resuspended in 1× Laemmli Sample Buffer. Of note, for mass spectrometry analysis of Flag-tagged BAP1 overexpressed in MP41 cells, 50 mg of nuclear extracts was used and first purified on an ion exchange chromatography HiTrap Q 1 ml column (GE Healthcare).

After IP and elution of enriched proteins, SDS-PAGE (Invitrogen) was used without separation as a cleanup step to remove lipids, metabolites, salts, and denaturing agents from the samples. After colloidal blue staining (LabSafe GEL BlueTM GBiosciences), four gel slices were excised and proteins were reduced with 10 mM DTT prior to alkylation with 55 mM iodoacetamide. After washing and shrinking the gel pieces with 100% MeCN, in-gel digestion was performed using trypsin/Lys-C (Promega) overnight in 25 mM NH_4HCO_3 at 30 °C.

Peptides were extracted and analyzed by nano liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) using an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled to an Orbitrap Fusion mass spectrometer (Q-OT-qIT, Thermo Fisher Scientific). Samples were loaded on a C18 precolumn (300 µm inner diameter × 5 mm; Dionex) at 20 µl/min in 2% MeCN, 0.05% TFA. After a desalting for 3 min, the precolumn was switched on the C18 column (75 µm i.d. × 50 cm, packed with C18 PepMap™, 3 µm, 100 Å; LC Packings) equilibrated in solvent A (2% MeCN, 0.1% HCOOH). Bound peptides were eluted using a two-step linear gradient of 147 min (from 1 to 20% (v/v)) of solvent B (100% MeCN, 0.085% HCOOH) and 65 min (from 20 to 40% (v/v)) of solvent B, at a 400 nl/min flow rate and an oven temperature of 40 °C. We acquired Survey MS scans in the Orbitrap on the 400–1200 m/z range with the resolution set to a value of 120,000 and a 4×10^5 ion count target. Each scan was recalibrated in real time by co-injecting an internal standard from ambient air into the C-trap. Tandem MS was performed by isolation at 1.6 Th with the quadrupole, higher collisional dissociation fragmentation with normalized collision energy of 35, and rapid scan MS analysis in the ion trap. The MS2 ion count target was set to 10^4 and the max injection time was 100 ms. Only those precursors with charge state 2–7 were sampled for MS2. The dynamic exclusion duration was set to 60 s with a 10 ppm tolerance around the selected precursor and its isotopes. The instrument was run in top speed mode with 3-s cycles.

Data were searched against the uniprot-Human database, using Sequest HT from Proteome Discoverer 1.4 (thermo Scientific). Enzyme specificity was set to trypsin and a maximum of two miss cleavages was allowed. Oxidized methionine, N-terminal acetylation, and carbamidomethyl cysteine were set as variable modifications. The mass tolerances in MS and MS/MS were set to 10 ppm and 0.6 Da, respectively. The resulting files were further processed using myProMS[58]. The Sequest HT target and decoy search results were validated at 1% FDR with Percolator. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD011808[59].

Methylation analysis. Genomic DNA was extracted following manufacturer protocol (DNA easy, Qiagen), treated with RNAse A and RNAse T1, purified again and quantified using nanodrop. One microgram of DNA was then digested by the degradase plus (Zymoresearch), ethanol precipitated and the supernatant was then evaporated on speedvac. Samples were then reconstituted in 10 µl of solution A' (2% methanol, 0.1% HCOOH), vortex-mixed, centrifuged and transferred to a high-performance liquid chromatography vial for micoLC-MS/MS analysis. Fractions were used directly in solution A' and analyzed (5 µl) using the RSLCnano system connected to the Orbitrap Fusion mass spectrometer. Sample separation was achieved on a C18 column (4.6 × 100 mm, packed with ZORBAX Eclipse XDB C18, 1.8 µm particles, Agilent Technologies) after 5 min loading in solvent A', with a linear gradient of 10 min (from 0 to 30% (v/v) of solvent B' (80% MeCN, 0.1% HCOOH)) at 500 µl/min. Data acquisition was performed in the Orbitrap on the 200–300 m/z range with the resolution set to a value of 240,000 at m/z 200. To determine the intensity of each nucleosides, we extracted from the MS survey of microLC-MS/MS raw files the extracted ion chromatogram (XIC) signal by using the retention time and m/z values of the well-characterized synthetic nucleoside ions using the Xcalibur softwares (manually). XIC areas were integrated in Xcalibur under the QualBrowser interface using the ICIS algorithm.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

ChIP-seq and RNA-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus under the accession codes GSE110133 (ChIP-seq) and GSE110142 (RNA-seq) [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE110143]. The raw protomics data are available via ProteomeXchange with identifier PXD011808. Raw data are provided for all western blots in Supplementary Figs. 7–14. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding authors upon reasonable request. A Reporting Summary for this Article is available as a Supplementary Information file.

Received: 2 January 2018 Accepted: 21 December 2018
Published online: 21 January 2019

References

1. Jensen, D. E. et al. BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* **16**, 1097–1112 (1998).
2. Carbone, M. et al. BAP1 and cancer. *Nat. Rev. Cancer* **13**, 153–159 (2013).
3. Yu, H. et al. The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol. Cell. Biol.* **30**, 5071–5085 (2010).
4. Dey, A. et al. Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science* **337**, 1541–1546 (2012).
5. Machida, Y. J., Machida, Y., Vashisht, A. A., Wohlschlegel, J. A. & Dutta, A. The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *J. Biol. Chem.* **284**, 34179–34188 (2009).

6. Nishikawa, H. et al. BRCA1-associated protein 1 interferes with BRCA1/BARD1 RING heterodimer activity. *Cancer Res.* **69**, 111–119 (2009).
7. Ji, Z. et al. The forkhead transcription factor FOXK2 acts as a chromatin targeting factor for the BAP1-containing histone deubiquitinase complex. *Nucleic Acids Res.* **42**, 6232–6242 (2014).
8. Baymaz, H. I. et al. MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics* **14**, 2179–2189 (2014).
9. Hauri, S. et al. A high-density map for navigating the human polycomb complexome. *Cell Rep.* **17**, 583–595 (2016).
10. Zarrizi, R., Menard, J. A., Belting, M. & Massoumi, R. Deubiquitination of gamma-tubulin by BAP1 prevents chromosome instability in breast cancer cells. *Cancer Res.* **74**, 6499–6508 (2014).
11. Lee, H. S., Lee, S. A., Hur, S. K., Seo, J. W. & Kwon, J. Stabilization and targeting of INO80 to replication forks by BAP1 during normal DNA synthesis. *Nat. Commun.* **5**, 5128 (2014).
12. Matatall, K. A. et al. BAP1 deficiency causes loss of melanocytic cell identity in uveal melanoma. *BMC Cancer* **13**, 371 (2013).
13. Scheuermann, J. C. et al. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**, 243–247 (2010).
14. Gaytan de Ayala Alonso, A. et al. A genetic screen identifies novel polycomb group genes in *Drosophila*. *Genetics* **176**, 2099–2108 (2007).
15. Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **10**, 697–708 (2009).
16. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
17. Wang, H. et al. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
18. Pengelly, A. R., Kalb, R., Finkl, K. & Muller, J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev.* **29**, 1487–1492 (2015).
19. Illingworth, R. S. et al. The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. *Genes Dev.* **29**, 1897–1902 (2015).
20. Blackledge, N. P. et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and Polycomb domain formation. *Cell* **157**, 1445–1459 (2014).
21. Cooper, S. et al. Targeting Polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep.* **7**, 1456–1470 (2014).
22. Scheuermann, J. C., Gutierrez, L. & Muller, J. Histone H2A monoubiquitination and Polycomb repression: the missing pieces of the puzzle. *Fly. (Austin)* **6**, 162–168 (2012).
23. Sahtoe, D. D., van Dijk, W. J., Ekkebus, R., Ovaa, H. & Sixma, T. K. BAP1/ASXL1 recruitment and activation for H2A deubiquitination. *Nat. Commun.* **7**, 10292 (2016).
24. Abdel-Wahab, O. et al. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell* **22**, 180–193 (2012).
25. Abdel-Wahab, O. et al. Deletion of Asxl1 results in myelodysplasia and severe developmental defects in vivo. *J. Exp. Med.* **210**, 2641–2659 (2013).
26. Inoue, D. et al. Myelodysplastic syndromes are induced by histone methylation-altering ASXL1 mutations. *J. Clin. Invest.* **123**, 4627–4640 (2013).
27. Li, T., Hodgson, J. W., Petruk, S., Mazo, A. & Brock, H. W. Additional sex combs interacts with enhancer of zeste and trithorax and modulates levels of trimethylation on histone H3K4 and H3K27 during transcription of hsp70. *Epigenetics Chromatin* **10**, 43 (2017).
28. Valletta, S. et al. ASXL1 mutation correction by CRISPR/Cas9 restores gene function in leukemia cells and increases survival in mouse xenografts. *Oncotarget* **6**, 44061–44071 (2015).
29. LaFave, L. M. et al. Loss of BAP1 function leads to EZH2-dependent transformation. *Nat. Med.* **21**, 1344–1349 (2015).
30. Wu, X. et al. Tumor suppressor ASXL1 is essential for the activation of INK4B expression in response to oncogene activity and anti-proliferative signals. *Cell Res.* **25**, 1205–1218 (2015).
31. Wang, L. et al. Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. *Nat. Med.* **24**, 758–769 (2018).
32. Oak, J. S. & Ohgami, R. S. Focusing on frequent ASXL1 mutations in myeloid neoplasms, and considering rarer ASXL2 and ASXL3 mutations. *Curr. Med. Res. Opin.* **33**, 781–782 (2017).
33. Carette, J. E. et al. Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat. Biotechnol.* **29**, 542–546 (2011).
34. Wassef, M. et al. Versatile and precise gene-targeting strategies for functional studies in mammalian cell lines. *Methods* **121-122**, 45–54 (2017).
35. Daou, S. et al. The BAP1/ASXL2 histone H2A deubiquitinase complex regulates cell proliferation and is disrupted in cancer. *J. Biol. Chem.* **290**, 28643–28663 (2015).
36. Riising, E. M. et al. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol. Cell* **55**, 347–360 (2014).
37. Schoumacher, M. et al. Uveal melanoma cells are resistant to EZH2 inhibition regardless of BAP1 status. *Nat. Med.* **22**, 577–578 (2016).
38. Lee, S. W. et al. ASXL1 represses retinoic acid receptor-mediated transcription through associating with HP1 and LSD1. *J. Biol. Chem.* **285**, 18–29 (2010).
39. Park, U. H., Yoon, S. K., Park, T., Kim, E. J. & Um, S. J. Additional sex comb-like (ASXL) proteins 1 and 2 play opposite roles in adipogenesis via reciprocal regulation of peroxisome proliferator-activated receptor {gamma}. *J. Biol. Chem.* **286**, 1354–1363 (2011).
40. Dubey, R. et al. Chromatin-remodeling complex SWI/SNF controls multidrug resistance by transcriptionally regulating the drug efflux pump ABCB1. *Cancer Res.* **76**, 5810–5821 (2016).
41. Misaghi, S. et al. Association of C-terminal ubiquitin hydrolase BRCA1-associated protein 1 with cell cycle regulator host cell factor 1. *Mol. Cell. Biol.* **29**, 2181–2192 (2009).
42. Gao, Z. et al. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* **45**, 344–356 (2012).
43. Tavares, L. et al. RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664–678 (2012).
44. Balasubramani, A. et al. Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1-BAP1 complex. *Nat. Commun.* **6**, 7307 (2015).
45. Sinclair, D. A., Campbell, R. B., Nicholls, F., Slade, E. & Brock, H. W. Genetic analysis of the additional sex combs locus of *Drosophila melanogaster*. *Genetics* **130**, 817–825 (1992).
46. Fisher, C. L. et al. Additional sex combs-like 1 belongs to the enhancer of trithorax and polycomb group and genetically interacts with Cbx2 in mice. *Dev. Biol.* **337**, 9–15 (2010).
47. Gutierrez, L. et al. The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development* **139**, 117–127 (2012).
48. Sanchez, R. & Zhou, M. M. The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.* **36**, 364–372 (2011).
49. Nakatani, Y. & Ogryzko, V. Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol.* **370**, 430–444 (2003).
50. Hebert, L. et al. Modulating BAP1 expression affects ROS homeostasis, cell motility and mitochondrial function. *Oncotarget* **8**, 72513–72527 (2017).
51. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
52. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
53. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
54. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
55. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
56. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
57. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).
58. Poullet, P., Carpentier, S. & Barillot, E. myProMS, a web server for management and validation of mass spectrometry-based proteomic data. *Proteomics* **7**, 2553–2556 (2007).
59. Vizcaino, J. A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, 11033 (2016).
60. Steger, D. J. et al. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol. Cell. Biol.* **28**, 2825–2839 (2008).
61. Schwartzman, J. et al. A DNA methylation microarray-based study identifies ERG as a gene commonly methylated in prostate cancer. *Epigenetics* **6**, 1248–1256 (2011).

Acknowledgements

The Labex DEEP, AAP EpiG and Cancer, SIRIC, "Programme Incitatif et Coopératif" (PIC) Uveal Melanoma and the Institut Curie supported work in R.M.'s laboratory. A.C. and M.W. were recipients of fellowships from the "Association pour la Recherche contre le Cancer (ARC)". This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 666003. Mass spectrometry analyses were performed by the Institut Curie Protein Mass Spectrometry Laboratory, supported by grants from "Region Ile-de-France" and the FRM. High-throughput sequencing was performed by the NGS platform of the Institut Curie, supported by grants ANR-10-EQPX-03 and ANR-10-INBS-09-08 from the Agence Nationale de le Recherche (investissements d'avenir) and by the Canceropôle

Ile-de-France. We thank Tatiana Popova for help with the RNA-seq and, members of the Margueron Lab, M-H Stern and P. Gilardi for comments on the manuscript.

Author contributions

A.C., M.-K.L., A.M., S.L.C., H.C., and L.Z.S. performed most of the experiments. D.Z., I. V., and N.S. performed bioinformatic analysis. F.D. and D.L. performed mass spectrometry analyses. E.P. and S.P.-V. provided unpublished reagents and data. A.C., M.W., and R.M. prepared the manuscript. M.W. and R.M. conceived and supervised the study. All authors contributed to experimental design and edited the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-08255-x.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Part 1.2 Genomic localization of BAP1 reveals an unexpected role for H2AK119ub in the regulation of enhancer function.

(Manuscript in preparation)

Title: Genomic localization of BAP1 reveals an unexpected role for H2AK119ub in the regulation of enhancer function.

Ming-Kang Lee[1], Laia Richart[1], Sétareh Aflaki[1], Aurélien Dauphin[2], Elaine Del Nery[3], Florent Dingli[4], Damarys Loew[4], Sophie Postel-Vinay[5], Raphaël Margueron[6]

[1] Institut Curie, CNRS UMR3215, INSERM U934, Paris Sciences et Lettres Research University, Sorbonne University, 75005 Paris, France
[2] Institut Curie, Plateforme Imagerie PICT-IBiSA, Paris Sciences et Lettres Research University, 75005 Paris, France
[3] BioPhenics platform, Institut Curie, Paris Sciences et Lettres Research University, 75005 Paris, France
[4] Institut Curie, Paris Sciences et Lettres Research University, Centre de Recherche, Laboratoire de Spectrométrie de Masse Protéomique, 75005 Paris, France
[5] Institut Gustave Roussy, CNRS UMR 981, Université Paris-Saclay, 94805 Villejuif, France
[6] Institut Curie, CNRS UMR3215, INSERM U934, Paris Sciences et Lettres Research University, Sorbonne University, 75005 Paris, France. raphael.margueron@curie.fr

Abstract

In mammals, BRCA1-associated protein 1 (BAP1) is reported to positively regulate transcription through the deubiquitination of histone H2A on lysine 119 (H2Aub), a repressive histone mark deposited by the Polycomb Repressive Complex 1 (PRC1). BAP1 functions therefore downstream of PRC1, however, its precise mechanism of action remains largely unclear. Here, we identified a specific recruitment of BAP1 to active enhancers, where H2Aub selectively and broadly accumulated upon BAP1 silencing. We showed that the recruitment of BAP1 is dependent on the acetyltransferase CBP, and that BAP1 silencing mitigates BRD4 occupancy at enhancers. Consistently, super-resolution microscopy revealed that loss of BAP1 led to reduced phase-separated condensates of BRD4 and MED1 both in size and number. These results indicate that aberrant accumulation of H2Aub at enhancer impairs their function. Finally, we show that BAP1-null cells are more sensitive to small-molecule inhibitor targeting BRD4, which could potentially pave the way for the development of new therapeutic strategies targeting BAP1-null malignancies.

Introduction

Orchestrated gene expression program determines cell identity and function. It requires the coordination between DNA regulatory elements and chromatin-associated factors[1]. Initiation of messenger RNA transcription requires the recruitment of RNA polymerase II (POL II) at promoters, where POL II assembles into pre-initiation complex (PIC) with general transcription factors. Once RNA synthesis reaches a critical point, an elongation complex assembles but is paused downstream of transcription starting site (TSS) and waits for further signals prior to entering productive transcription[2]. Pause-release of POL II is a critical step in regulation of transcription frequency[3], which involves several co-activators including positive transcription elongation factor b (P-TEFb), whose subunit phosphorylates serine 2 on CTD domain of POL II to activate the elongation complex[4]. Once released, the extension of RNA chain proceeds until reaching the termination signal[1].

A class of distal regulatory elements, called enhancers, can establish physical contacts with gene promoters to facilitate transcription[5]. Enhancers contain sequence-specific motifs that can be recognized by transcription factors (TFs)[6]. Binding of TFs on enhancers is accompanied by recruitment of transcriptional co-activators including chromatin remodelers, histone modifiers, bromodomain-containing protein BRD4, P-TEFb, and Mediator complex[6,7]. Together, they foster a local environment that favors recruitment of POL II and promote transcriptional processes[5,7,8].

Histone modifications on the flanking nucleosomes of enhancers characterize their activities[9]. Mono-methylation on histone H3 lysine 4 (H3K4me1) is deposited by MLL3/4[10], whereas acetylation of H3 on lysine 27 (H3K27ac) is catalyzed by acetyltransferase CBP and its paralog p300[11]. Deposition of H3K4me1 and H3K27ac on enhancers are essential for transcriptional activation. For example, H3K4me1 controls CBP binding to enhancers[12], and loss of either H3K4me1 or H3K27ac results in global gene silencing[12-14]. Genome-wide studies indicate that the enrichment of these two marks can be employed to accurately predict enhancers in a given context[15,16]. H3K4me1 features both poised and active enhancers while H3K27ac marks active enhancers specifically. Of note, the level of H3K27ac directly correlates with the transcriptional activity of associated genes[17,18].

High level of H3K27ac on enhancers facilitates binding of co-activator BRD4,

a member of the BET family of proteins that recognizes acetyl lysine thanks to its tandem bromodomain module[19]. BRD4 recruits Mediator complex[20], a multimeric complex that directly interacts with TFs and POL II and is supposed to transduce activation signals from TFs-bound enhancers to the transcriptional machinery[21-23]. Mediator complex recruits POL II and triggers Pol II release from promoters[24]. Besides, BRD4 also recruits P-TEFb to promote transcription elongation[25,26]. Recruitment of these co-activators are indispensible for proper transcriptional activites, as inhibition of BRD4 binding by selective BET inhibitors evicts BRD4, Mediator (e.g. MED1) and POL II from enhancers, and results in loss of expression of neighboring genes[23,27].

The Polycomb machinery plays an essential in the maintenance of transcriptional silencing. In mammals, it is mainly composed of Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2), both being endowed with histone modifying activity essential for their function[28]. PRC1 catalyzes mono-ubiquitination on histone H2A lysine 119[29], and PRC2 deposits mono-, di- and tri-methylation on histone H3 at lysine 27 (H3K27me1/2/3)[30]. Both histone marks anti-correlates with transcription[31,32]. PRC1 and PRC2 can be recruited to CpG island at gene promoters[31,33-35], and target specifically genes involving in development, pluripotency and lineage specification[32,36,37].

The histone deubiquitinase BAP1 was originally described as a Polycomb protein in Drosophila, where mutation of its ortholog, *calypso*, gives rise to de-repression of *Hox* genes[38]. Calypso forms a complex with Additional sex combs (Asx) and catalyzes H2A deubiquitination. This activity is conserved within the mammalian homolog complex composed of BAP1 and ASXL1/2/3[39]. However, several studies in vertebrates have now established that BAP1 is involved in transcriptional activation and that it acts through its deubiquitinase activity downstream of PRC1[40-42]. Indeed, BAP1 no longer regulates transcription when PRC1 is absent[40,43]. Nevertheless, what are the target of BAP1 and how the balance between BAP1 and PRC1 is orchestrated remain poorly understood. These questions are of particular importance as they could shed light on the contribution of BAP1 to tumorigenesis[44].

By integrating multi-omics analyses, we demonstrated that H2Aub is the primary substrate of BAP1. We mapped BAP1 localization at chromatin and uncover a preferential binding at active enhancers which was not reported previously probably for technical reasons. BAP1's recruitment requires the

activity of the acetyltransferase CBP. Inactivation of BAP1 resulted in extensive accumulation of H2Aub on active enhancers accompanied by impairment of co-activators BRD4 and Mediator functions, as reported by decreased BRD4 binding and reduction in BRD4 and MED1 condensates. Here, we provided evidence that H2Aub plays an important role in the regulation of enhancers activity and that its deposition should be kept under tight control. Finally, small-molecule inhibitor screen uncovered increased sensitivity to BET inhibitor upon BAP1 deletion. We propose that loss of BAP1 sensitizes cells to further perturbation of enhancer function and that this weakness can be exploited for the development of therapeutic strategies.

Results

Lack of evidence for PRC1-independent substrate of BAP1.

It was recently shown that BAP1 is unable to regulate transcriptional regulation in the absence of PRC1[40,43]. Yet, BAP1 was also reported to regulate the ubiquitination of other substrates not known to be modified by PRC1[45], and involve in pathways not necessary related to chromatin. In order to evaluate their importance, we performed quantitative mass spectrometry to analyze both the ubiquitinome and the proteome of cells either wild-type (WT) or knockout for BAP1 (BAP1-KO) using the human HAP1 cell line as model (Fig. S1a). We enriched our samples for ubiquitinated peptides through the capture of di-glycine remnant (K-ε-GG), peptides that are formed after trypsinization of protein containing ubiquitinated lysine (Fig. 1a)[46]. We could detect about 9000 unique ubiquitination sites, both in WT and BAP1-KO condition, a number in the range of previous report[46]. We then compared the ubiquitinome in presence or absence of BAP1 and identified a bit over 500 differentially enriched sites (p-adj. < 0.05, log2 FC ≥ 1 or ≤ -1; Fig. 1b), including as expected H2A. Counterintuitively, a majority of peptides showed reduced enrichment in the BAP1-KO context (Fig. 1b). This result could either reflect reduced ubiquitination or reduced protein abundance. To address this question, we determine whether there was a correlation between proteome and ubiquitinome for this set of proteins and observed a mild correlation (r=0.418; Fig. S1b), suggesting that at least part of the change of ubiquitination are the consequence of differential protein abundance.

Interestingly, the proteome suggests a biased toward reduced protein abundance in BAP1-KO (Fig. 1c), which reminds the trend observed at the transcriptomic level[40]. Indeed, there is a significant correlation between transcriptome and proteome (Fig. S1c), and about two third of the proteins that are less abundant in the absence of BAP1 are also less abundant at the mRNA levels (Fig. 1d).

Since it was difficult to disentangle the direct from the indirect effect of BAP1 inactivation on the ubiquitome, we repeated the experiment in a context where PRC1 is absent (RING1A and RING1B double KO; Fig. S1a) reasoning that it will limit BAP1-dependent transcriptional regulation. Strikingly, only two proteins show now a significant increase of ubiquitination upon BAP1 deletion (Fig. 1e). Besides, two proteins, ASXL1 and ASXL2, show reduced ubiquitination regardless of the genetic background (Fig. 1 b,e). ASXL1 is also downregulated at the protein

level but not mRNA level (Fig. 1f and Fig. S1d), consistent with the idea that BAP1 and the ASXLs form a stable complex necessary for their stability[47]. On a whole, the proteome, similarly to the ubiquitome, is mostly unaffected by the deletion of BAP1 when PRC1 is deleted (Fig. 1f).

Hence, the ubiquitinome, proteome and transcriptome all support the idea that the cellular functions of BAP1 are almost exclusively epistatic to PRC1, and that the regulation of the transcription through the removal of H2Aub is BAP1 principal, if not exclusive, mechanism of action.

BAP1 is recruited to a subset of active enhancers

Previous studies have suggested a strong correlation between recruitment of BAP1 or ASXL1 and transcription as evaluated by RNA-Pol II ChIP-seq[40,48,49], but at the same time a rather pervasive increased of H2Aub in the absence of BAP1 was reported[40,43]. This prompted us to further investigate BAP1 genomic localization. We used CUT&RUN sequencing[50] since this method does not require cross-linking potentially limiting the detection of indirect peaks. We obtained rather sharp peaks for BAP1 that were lost in the BAP1 knockout context, and confirmed the presence of RNA-Pol II at those peaks (Fig. 2a). However, the reverse was not true (*i.e*. some RNA-Pol II peaks are not necessary enriched for BAP1). This is consistent with the genomic ontology analysis of BAP1 peaks revealing a specific localization at intergenic regions and within introns but a rather infrequent localization at promoters (Fig. 2b). Besides, we noticed that BAP1 peaks coincide with enrichment for H3K27ac and H3K4me1 altogether leading to the hypothesis that BAP1 localizes at enhancers (Fig. 2 a,c).

In order to systematically evaluate the chromatin signatures of BAP1-occupied regions, we implemented ChromHMM modeling and define 10 chromatin states[51]. It annotated two Polycomb repressed states, bivalent and active TSS, quiescence, two classes of weak enhancers (EnhWk1/2) and three classes of active enhancers (EnhA1/2/3) (Fig. 2d). Weak enhancers were marked by H3K4me1 (EnhWk1), with a particular cluster discerned by H2Aub enrichment (EnhWk2). H3K27ac alone (EnhA1) or in concert with H3K4me1 (EnhA2) marked active enhancers. The third class (EnhA3) is probably even more active, being enriched for H3K4me1, H3K27ac, RNA-POL II and BRD4 (Fig. 2d). We then questioned at which chromatin state BAP1 peaks are associated and found a strong enrichment for enhancers, in particular the third and most active flavor (Fig.

2e). Actually, about 40% of the genomic regions classified as EnhA3 are enriched for BAP1 (Fig. 2f). Intriguingly, this chromatin state is also enriched for H2Aub suggesting that the control of H2Aub enrichment could be important for this class of active enhancers (Fig. 2a,d). A role for PRC1 at enhancer was actually already proposed in the literature[52,53], and indeed we observed that a substantial fraction of RING1B peaks are found at intergenic regions and introns that overlap with RNA-Pol II (Fig. 2a,g,h).

The specific localization of BAP1 at active enhancers contrast with a previous report in triple-negative breast cancer CAL51 cells showing a large proportion of BAP1 recruited to promoters[49]. To determine whether the differences reflect the cell type, we knocked out BAP1 in CAL51 cells and repeated CUT&RUN-seq (Fig. S2a,b). The genomic ontology of the peaks was similar to the one obtained in HAP1 cells with more than three quarter of BAP1-occupied sites were either introns or intergenic (Fig. 2b). Interestingly, ChIP-seq data (GSE97326) and CUT&RUN-seq data display a strong overlap when performed in the same cell line (Fig. S2b,c). However, CUT&RUN-seq retrieved far fewer peaks but with a better signal-to-noise ratio. Of note, the comparison of BAP1 peaks in HAP1 *versus* CAL51 cells reveals a relatively modest overlap (Fig S2d), which is consistent with the cell-specific usage of enhancers. Altogether, we conclude that BAP1 is recruited primarily at a subset of active enhancers.

H2Aub selective accumulation on enhancers upon BAP1 loss

It is now well-documented that the inactivation of BAP1 leads to a robust upregulation of H2Aub (Fig. S3a) which, up to now, was reported to be uniform throughout the genome[40,43]. However, the selective recruitment of BAP1 at enhancers led us to re-investigate this question. We first plotted H2Aub enrichment across the chromatin states defined previously (Fig. 3a), and then the fold change (FC) comparing wild-type to BAP1-KO cells (Fig. 3b). We confirmed the general elevation of H2Aub enrichment but noticed also that the three chromatin states whose FC is above 2 (log2 FC>1) are subclasses of enhancers (EnhWk2, EnhA2 & EnhA3). Of note, the gains of H2Aub upon BAP1 inactivation are more pronounced at enhancers that are bound by BAP1 than the ones of the same subclass that are not as shown for the EnhA3 subclass (Fig. 3c).

We then investigate whether BAP1 inactivation has any impact on chromatin states by applying ChromDiff analysis[54], a group-wise comparison method that

revealed relevant changes in epigenomic characteristics. BAP1-bound enhancers remained largely constant in chromatin state even when it is deleted: only four percent of BAP1 peaks (n=335 out of 7830) became more active, and thirteen (n=994 out of 7830) became less active (Fig. 3d). Activating and inactivating transitions are accompanied by the expected gain and loss of H3K27ac and H3K4me1 (Fig 3e,f). However, in both cases, H2Aub is gained and transcription of the associated genes tend to be reduced although to different extend (Fig. S3b).

In the case of BAP1-unbound enhancers, a similar number of enhancers undergoes activating and inactivatong transitions (9 % each; Fig S3c). Most of the transitions occur between the quiescent and EnhWk1 (both ways) and are therefore associated to changes of H3K4me1, H3K27ac and POL II although with modest enrichment (Fig. S3d,e).

Altogether, our results suggest that the deletion of BAP1 results in subtle changes of chromatin state at enhancers. Focusing on the enhancers initially bound by BAP1, we observed a biased toward transitions reflecting enhancer inactivation, but this remains a minor subset. We conclude that BAP1-mediated H2Aub regulation is not critical for the deposition of histone modifications that mark enhancer.

The acetyltransferase CBP is crucial for BAP1 recruitment

Since BAP1 appears dispensable for H3K27ac and H3K4me1 deposition at enhancers, we wondered whether conversely these marks could contribute to BAP1 recruitment. H3K27ac is seen as a general marker for active enhancers, that is deposited by the acetyltransferase paralogs p300 and CBP. While deletion of both enzymes often impairs proliferation, deletion of one of them remains viable while resulting in a global reduction of the acetylation (Fig. 4a and S4a). We therefore compared BAP1 recruitment by CUT&RUN-seq in presence or absence of CBP. Visual examination of tracks revealed marked changes in BAP1 localization (Fig. 4b). Indeed, peaks calling indicated that most of BAP1 peaks are lost upon inactivation of CBP (6684 out of 7844; Fig. 4c). The remaining BAP1-bound regions showed diminished occupancy (Fig. 4d). Importantly, reduced binding does not reflect reduced protein abundance as western blot showed comparable BAP1 protein accumulation in WT and CBP-KO cells (Fig. S4b). A number of regions appears to have gained BAP1 binding in the absence of CBP (1835; Fig. 4c), these regions correspond to weak enhancers. Accordingly, they are weakly

enriched for H3K4me1 and H3K27ac, but their respective enrichment slightly increase upon CBP inactivation (Fig. 4e).

These results confirm the essential activity of histone acetyltransferases, in particular CBP, for enhancer function and reveal that BAP1 recruitment is dramatically impaired in the absence of CBP. This suggests that BAP1 acts downstream of the acetylation of H3K27ac at enhancers.

BAP1 inactivation compromises BRD4 and MED1 function

Having shown that BAP1 deletion, and consequently the increased enrichment for H2Aub, does not have a foremost impact on enhancer chromatin state, but considering that it impacts transcription, we hypothesized that it could affect downstream effectors of enhancer function. This led us to analyze whether the recruitment of transcriptional co-activators such as BRD4 or the Mediator complex, is altered in the absence of BAP1. As expected, CUT&RUN-seq indicates a specific enrichment of BRD4 at active enhancers (Fig. 2d and S5a), in particular at the chromatin state EnhA3. We actually observed a broad co-localization between BRD4 and BAP1 (Fig. 5a,b). Protein stability of BRD4 and of mediator subunits do not appear to be affected by the deletion of BAP1 as evaluated both by western blot (Fig. 5c) and through the proteome (Fig. S5b). However, BRD4 CUT & RUN-seq analysis revealed a dramatic loss of BRD4 recruitment at its target peaks upon deletion of BAP1 (Fig 5a,d and S5c). Interestingly, the reduction in BRD4 recruitment was a bit less pronounced than upon deletion of CBP but, *in fine*, it led to the loss of a similar number of peaks (Fig. 5d, S5c).

Microscopic analysis of BRD4 and Mediator complex within the cells shows an organization in discrete loci, referred to as condensates, which are proposed to be important for gene control[55]. The impaired recruitment of BRD4 upon deletion of BAP1 led us to investigate whether these condensates are also globally affected. Toward this end, we performed super-resolution microscopy and accordingly observed a dotted pattern for BRD4 and MED1 (Fig. 5e). Quantification of these dots reveals a reduction in size both for BRD4 and for MED1 (Fig. 5e,f), and also a lower number of MED1 condensates (Fig. 5g). To further support the destabilization of enhancer integrity upon deletion of BAP1 and increase level of H2Aub, we evaluate how tightly BRD4 and mediator are associated to chromatin using gradient salt fractionation[56]. This experiment indicates a tendency of both BRDs and mediator subunits to elute at lower salt concentration in the BAP1-KO

cells than in the wild type condition. For instance, BRD4 which elutes mainly in the 1M NaCl fraction in the wild-type cells is found more abundant in the 750mM Nacl faction in the absence of BAP1 (Fig. 5h).

In summary, through three different methods, we showed that BAP1 inactivation destabilize the downstream effectors of enhancer activity that are BRD4 and the mediator complexes. It is therefore tempting to postulate that the inactivation of BAP1 might interfere with transcription through the impairment of enhancer/promoter communication.

Increased sensitivity to BET inhibitors of BAP1-mutant cells

Inhibitors targeting BET proteins disrupts binding of BRD4, which has been shown to affect in particular the transcription of MYC oncogenes. These inhibitors are seen as therapeutic targets for hematological cancers[57-59]. We reasoned that BAP1-KO cells could respond better to BET inhibitors considering that enhancer functions are already partially compromised. Indeed, BAP1-KO cells treated with OTX015, a BET inhibitor, showed more pronounced erasure of BRD4 and MED1 condensates (Fig. 6a-d)[27,57]. In addition, cell viability assay reported 24-fold (ED50 78 nM *versus* 1928 nM) increase of sensitivity to OTX015 of BAP1-KO cells as compare to WT cells (Fig. 6e).

We then determined whether this observation is specific to the studied model cell line or whether it represents a more general trend. The underlying question is whether BAP1 status could predict the response to BETi in tumors harboring frequent BAP1 loss-of-function (LOF) mutation, such as uveal melanoma (UM)[60]. To address this question, we studied the UM-derived isogenic UPMD2 WT/ BAP1-KO cells and PDX-derived cells with BAP1 wild-type (MP41, MM66) and mutant status (MP38, MP46)[61,62]. Similar to HAP1 cells, BAP1-mutant cells displayed increased sensitivity to OTX015 in both isogenic and PDX-derived models (Fig. 6f-g). To further support our result, we also studied the response to another chemically unrelated BET inhibitor, namely BI894999. This inhibitor was also more effective in UPMD2 BAP1-KO cells (Fig. 6f), thus confirming our hypothesis that BAP1KO cells were more vulnerable to enhancer disruption.

Discussion

Taken together, our study establishes that BAP1 localizes on the most active enhancers, where it prevents excessive H2Aub deposition by Polycomb complex PRC1, a mark that probably destabilize co-activator function. Moreover, BAP1-inactivating cells are more susceptible to co-activators disruption by BET inhibitors, which could constitute a therapeutic strategy for BAP1-null malignancies such as UM.

BAP1 has retained lots of attention recently due to its putative tumor suppressive role in various cancers[44]. However, despite some progresses, several aspects of its mechanism of action remains enigmatic. What are the substrates of BAP1 deubiquitinase activity has been a long standing question. Indeed, the ubiquitome is substantially modified in response to the deletion of BAP1, however, it was difficult to determine whether this was due to direct regulation by BAP1 or whether this was an indirect consequence of BAP1-mediated transcriptional regulation. By cross-referring transcriptomic, proteomic and ubiquitomic analyses, we establish that most of the changes in the ubiquitome reflect transcriptomic alteration since they are no longer observed in the absence of PRC1. It is therefore tempting to conclude that H2Aub is the main substrate of BAP1 even though we cannot rule out that other proteins, whose ubiquitination is catalyzed by PRC1, could also be BAP1's substrate.

Another important matter is BAP1 localization at chromatin. So far, only few studies published ChIP-seq which suggest enrichment of BAP1 at transcriptionally active region[41,63]. However, our attempts to repeat those approaches were unsuccessful, this prompts us to implement another method (CUT&RUN). We obtained a high resolution mapping of BAP1 localization whose specificity was controlled by performing the same experiment in a BAP1-KO cell line. This experiment revealed a specific recruitment of BAP1 at a subset of active enhancers and was confirmed in two distinct cell types. It is difficult to explain the discrepancy between our result and previous studies with certainty, but we hypothesized that it could be due to the fact that ChIP-seq required cross-linking whereas CUT & RUN does not. In agreement with this hypothesis, there is a significant overlap between our CUT&RUN peaks for BAP1 in CAL51 and the peaks identified previously by ChIP-seq[49]. The main differences are that there are much less peaks identified by CUT & RUN and that most of them are found at enhancers.

An important implication of this localization of BAP1 at enhancers and of its

role through the regulation of H2Aub is that PRC1 should somehow be able to also be recruited there. Yet, studies on mammalian Polycomb machinery using mouse embryonic stem cells (mESCs) have established the recruitment of PRC1 to promoters[31,36]. But in the HAP1 cells, a considerable subset of RING1B peaks are enriched at enhancers (Fig. 2 a,g), an observation that was also reported in fly developing imaginal disc and in human breast cancers cells[52,64]. This suggest that PRC1 localization at enhancers might be a rather frequent phenomenon that was overlooked until now. The subsequent question is what the role of H2Aub at enhancers. Our study suggests that excessive deposition of H2Aub is associated to BRD4 eviction and reduced co-activators condensates. This would therefore constitute an alternative mechanism for PRC1-mediated transcriptional repression to the proposed sequestration of the transcriptional machinery and restrain on transcriptional elongation previously described[65,66]. It remains to be determined what is the molecular mechanism. It will be particularly interesting, in this regard, to study whether H2A ubiquitination could directly prevent BRDs binding to chromatin. Alternatively, we speculate that H2Aub could modulate the tendency of active enhancers, enriched for BRD4 and MED1, to be part of phase-separated transcriptional condensates[55,67].

Finally, a question that remain opened is why global interference with enhancer function would contribute to tumor progression. The hyper active enhancers targeted by BAP1 are linked to high expression level of lineage-specific genes that control cell identity, but also drive expression of oncogenes (e.g. MYC) in neoplastic tissues[58,68]. We speculate that BAP1 deletion might impair cell identity and favor potentially less differentiated stage that could have contributed to the neoplastic transformation observed in a mouse model where systemic conditional deletion of BAP1 caused myeloid transformation[63,69]. This altered enhancer landscape nonetheless seems to create a weakness to the affected cells by rendering them more vulnerable to enhancer perturbations like the ones achieved through the use of BET inhibitors[70]. Further investigations will be required to evaluate the potential of this therapeutic strategy in clinic.

Material and Methods

Cell culture

HAP1 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10 % fetal bovine serum (FBS, Gibco), 2 mM L-glutamine (Gibco) and 1 × non-essential amino acid (Gibco). CAL51 cells were cultured in RPMI 1640 (Gibco) supplemented with 10 % FBS. Uveal melanoma cell lines MP41, MM66, MP38 and MP46 were established from primary patient tumors or patient-derived xenografts as documented[40]. UPMD2 cells were kindly provide by Dr. Richard Marais. Cells were cultured in Ham'12 Nutrient Mix (Gibco) supplemented with 10 % FBS.

Generation of KO cell lines

Generation of KO cells lines was performed using CRISPR-Cas9 technology as previously described[49]. In brief, a STOP cassette containing an antibiotic resistance gene followed by a polyadenylation sequence was inserted into early exons of target genes by homologous recombination. After antibiotic selection, clones were genotyped and complete KO was validated by western blot.

Preparation of nuclear extract and immunoblotting

Cells were washed once with PBS and then resuspended with 5 volumes of Buffer A (10 mM HEPES pH 7.9, 5 mM $MgCl_2$, 0.25 M sucrose and 0.1 % NP-40, 1 mM DTT, 200 µM PMSF, and protease inhibitors). After 10 min incubation on ice, cells were pelleted by centrifugation at 8000 g for 10 min. Supernatant was removed and pellets were resuspended with 5 volumes of Buffer B (25 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 0.1 mM EDTA pH 8.0, 20 % glycerol, 700 mM NaCl, 1 mM DTT, 200 µM PMSF and protease inhibitors. After 10 min incubation on ice, nuclei were sonicated for 45 s with 10 % amplitude, then centrifuged at 14000 g for 15 min at 4 °C. The supernatant was transferred to a new tube and taken as nuclear extract. Protein concentration was measured by Bradford assay (Biorad). Western Blot analysis of protein extracts was performed by StarBright Blue 700 fluorescent secondary antibodies (Biorad) and DyLight 800 secondary antibody (Biorad). Imaging was carried out by ChemiDoc System (Biorad).

Gradient salt fractionation for nuclear extract

70 million cells were harvest and resuspended in 500 µl of Buffer A (10 mM HEPES

pH 7.9, 5 mM $MgCl_2$, 0.25 M sucrose and 0.1 % NP-40, 1 mM DTT, 200 μM PMSF, and protease inhibitors). After 10 min incubation on ice, cells were pelleted by centrifugation at 8000 g for 10 min. Supernatant was collected as cytosolic fraction. Nuclei pellet was resuspended with 500 μl of Buffer C (250 mM NaCl with 20 mM HEPES-KOH pH7.9, 25 % Glycerol, 0.1 % NP-40, 0.5 mM $MgCl_2$, 0.2 mM EDTA pH 8.0, 1 mM DTT, 0.5 mM PMSF and protease inhibitors). After 2 hr incubation at 4 °C with rotation, nuclei were pelleted and supernatant was collected as 250 mM NaCl fraction. Pellets were then incubated with 500 μl of Buffer C (500 mM NaCl) for 2 hr at 4 °C with rotation. After centrifugation, supernatant was collected as the 500 mM NaCl franction. Lastly, chromatin pellet was incubated with 500 μl of Buffer C (1M NaCl) overnight at 4 °C with rotation and samples were proceeded for sonication. Western blot analysis was performed using 1 % (cytosolic and 250 mM franction) or 3 % (500 mM and 1M fraction) of fraction proportional to total volume.

Mass Spectrometry

Around 1-2 × 10^8 cells were harvested for acquisition of 20 mg of proteins, which corresponds roughly to 10× 150 mm culture dishes. After removing culture media, plates were washed once with cold PBS. PBS was then removed by careful decanting. 10 mL of Urea Lysis Buffer (9 M Urea, 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA pH 8.0, 50 μM PR-619, 1 mM chloroacetamide) was added to the first plate. After scraping cells into the lysis buffer, the buffer was transferred into the second plate. Sequentially scraping all the plates and cells were collected in 50 mL conical tube. Lysate was sonicated for 3 times for 15 s at 20 % amplitude. 1/278 volume of 1.25 M DTT was added. After 45 min incubation at RT, 1 mL of 100 mM iodoacetamide was added and proceed with 30 min incubation at RT in dark. 3 volumes of 50 mM Tris-HCl was introduced to acquire final concentration of 2 M Urea. Lysate was digested by trypsin overnight at RT. Digested lysate was purified by Sep-PaK column and lyophilized. Ubiquitinated peptides were enriched by immunoaffinity purification (IAP) using PTMScan Ubiquitin Remnant K-ε-GG Motif Kit (Cell Signalling Technology) following manufacturer's instruction. Samples were analysed with liquid chromatography (LC) tandem mass spectrometry (MS/MS) for quantitative proteome and ubiquitome profiling. For protein and ubiquitination sites quantification, an extracted ion chromatogram (XICs) from proteotypic peptides shared between compared conditions (TOPn

matching) and missed cleavages were allowed. Median and scale normalization were applied on the total signal to correct the XICs for each biological replicate. To estimate the significance of the change in ubiquitination sites and protein abundance, a linear model adjusted on peptides ions and biological replicates was performed. *P*-value was adjusted with Benjamini–Hochberg procedure with FDR threshold set to be 0.05.

CUT&RUN-seq

CUT&RUN was performed as previously described with minor modifications[61,62]. In brief, 1 million cells were pelleted at 600 g for 3 min at RT. After washing twice with 1 mL of wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine (Sigma) and protease inhibitors), cells were resuspended in wash buffer and ready for binding with beads. 10 µl of Convanavalin A beads (Bang Laboratories) was washed once with 1 mL binding buffer (20 mM HEPES pH 7.9, 10 mM KCl, 1 mM $CaCl_2$ and 1 mM $MnCl_2$) and placed on magnet stand to remove the liquid. 10 µl of binding buffer was used to resuspend the beads then the slurry was transferred to cells and incubated for 10 min at RT with rotation. After brief spin-down, tubes were placed on magnet to quickly withdraw the liquid. 50 µl of antibody buffer (wash buffer supplemented with 0.1 % digitonin (Millipore), 2 mM EDTA and 1:100 dilution of antibody of interest) was pipetted and cells were incubated for 10 min at RT with mild agitation. Permeabilized cells were decanted carefully and washed once with 1 mL dig-wash buffer (0.1 % digitonin in wash buffer). A secondary rabbit anti-mouse antibody (ab6709, abcam) binding step was carried out if the host species of primary antibodies are mouse. 50 µl of pA-MNase in dig-wash buffer (final concentration of 700 ng/ mL) was incubated with cells for 10 min at RT with agitation. After 2 washes with 1 mL dig-wash buffer, beads were resuspended with 100 µl dig-wash buffer and placed on heat block immersed in wet ice to chill down to 0 °C. 2 µl of 100 mM $CaCl_2$ was added to activate pA-MNase and incubated on heat block for 30 min. 100 µl of 2 × stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02 % digitonin, 1:200 RNase A, glycogen (50 mg/ mL) and heterologous spike-in DNA (2 pg/ml) was added to quench pA-MNAse, and fragments were released by 10 min incubation at 37 °C with rotation. After centrifugation at 14000 g for 5 min at 4 °C, DNA fragments were recovered by NucleoSpin (Macherey Nagel) or phenol-chloroform purification. Library was prepared by Accel-NGS 2S plus DNA library Kits (Swift Biosciences) for Illumina

barcoded system. PCR were set to 16 cycles. After post-library size selection, library size distribution and concentration were validated by Tapestation 4200 (Agilent). Libraries were sequenced as paired-ended 100bp reads on Illumina Novaseq platform.

Immunofluorescence (IF)

All procedures were performed at RT unless otherwise specified. Cells were seeded on glass coverslip and were fixed in 4 % paraformaldehyde in PBS for 10 minutes. After 3 washes with PBS for 5 min, cells were permeabilized with 0.5 % triton in PBS for 5 min. After 3 washes in PBS, cells were blocked with 4 % BSA in PBS for 1 hour, then proceeded with primary antibody incubation(1:500 dilution in blocking buffer) overnight at 4 °C. Following 3 washes with PBS for 5 min, cells were incubated with secondary antibody Alexa 488 in 1:500 dilution for 1 hour. Cells were washed 3 times in PBS for 5 min, and nuclei were counterstain with DAPI in 2 µg/ mL (Sigma) for 2 min. Following 3 washes, glass slides were mounted with Vectashield (Vector laboratories), sealed with nail polish and stored at 4 °C in dark. Images were acquired by OMX super-resolution microscopy with z-stacking. Acquisition were performed from at least 2 independent experiments.

Epigenetic drug screening

A library of 56 epigenetic drugs was ordered from Selleckchem Chemical (Extended Data Table 2). 100 cells (WT or BAP1 KO) were seeded into 384-well plates. 24 hours after seeding, cells were treated with compounds or vehicle (DMSO) with titration in an 8- point, 3-fold dilutions starting at a concentration of 10 µM to test the dose-response. After 12 days of incubation with culture media renewal, cell viability assay was carried out by luminescence detection using CellTiter-Glo 2.0 Assay kit (Promega) following manufacturer's instruction. Cell viability values were normalized by DMSO control on a per-plate basis. Compound activity was determined by computation median effective dose (ED50). A differential score was computed as the ratio and the difference by comparison to WT cells.

Cell viability assay in UM cells

Dose response was performed at an 8-point titration, with 3-fold serial dilution starting at the concentration of 2 µM. Cells were plated on 96-well plates. For each

cell line, seeding densities were as following: 4000 cells per well for UPMD WT; 2000 cells for UPMD BAP1 KO, MP41, MM66; 6000 cells per well for MP38 and MP46. 24 h after seeding, cells were treated with either the compound or vehicle (DMSO). After 12 days of incubation with media renewal every 3 days, cell viability was assessed by MTT assay. Breifly, cells were incubated with 100 µl of MTT medium (0.5 mg/ mL, Sigma) for 4 hr. The insoluble formazan was dissolved by addition of 100 µl of lysis buffer (10 % SDS, 0.01 N HCI) with overnight incubation at RT in dark. The absorbance was determined by SPARK microplate reader (TECAN). The wavelength to measure the formazan product was set at 570 nm, and reference was set at 650 nm. ED50 was analyzed by GraphPad Prism. The experiments were repeat at least three times independently.

CUT&RUN-seq data analysis

Reads were mapped to the human reference genome (GRCh37/hg19) with Bowtie2 using default parameters. Aligned reads were sorted by SAM tools. PCR duplicates were removed with Picard Tools MarkDuplicates (https://github.com/bioinfo-pf-curie/ChIP-seq). Generated BAM files were filtered to exclude common artifact regions. (artefact regions: https://github.com/Boyle-Lab/Blacklist/tree/master/lists). Exploratory data analyses were performed using Galaxy Europe interface (https://usegalaxy.eu/). Biological replicates were merged with MergeSamFiles for downstream analysis. Reads were counted in bins of length 50, RPKM normalized, and converted to bigWig format using bamCoverage (v3.3.2.0.0). Peaks were called with MACS2 (v 2.1.1.20160309.6) with default parameters. For control files, Igg.bam files are used, expects for BAP1 and RING1B peaks calling, where the control files were IP performed in respective KO-cells. Minimum FDR (q-value) cutoff for peak detection is modified in each epitope probed. The FDR cutoff was set as 0.01 for BAP1 and RING1B; 0.0001 for POL ii, and BRD4; 0.05 for BAP1 in CAL51 cells. Broad peak detection FDR cutoff was set as 0.01 for H2Aub and H3K27ac; 0.0001 for H3K4me1. Genomic annotation was performed using the Bioconductor package ChIPseeker (v1.18.0). Windows of 2 kb centered on transcription start sites (TSSs) were defined as promoter regions. Metaplot and heatmap analyses were performed using deepTools (v3.3.2.0.0): RPKM normalized log2 ratio between treated files and control files (Igg or BAP1/RING1B IP in KO-cells) were calculated by bamCompare. Matrix was prepared by computeMatrix (v3.3.2.0.0)

for metaplot and heatmap visualization. ChromHMM (v1.22) and ChromDiff analyses were perform using concatenation of WT and BAP1-KO sequencing files. Biological replicates of CUT&RUN-seq for H3K4me1, H3K27ac, H2Aub, POL ii and BRD4; and ChIP-seq for H3K27me3, H3K4me3 were fed to the algorithm for state emission following ChromDiff differential analysis. Only consistent transitions in both replicates were considered. Enrichment of H2Aub across chromatin states was analyzed by featureCounts (v2.0.1). Pair-ended fragments with both reads aligned were considered. Chimeric fragments were discarded. Counts were then normalized by genomic sizes.

Data availability

CUT&RUN-seq data for this study will be deposited in the Gene Expression Omnibus. The proteome and ubiquitome data will be uploaded.

Figures and figure legends

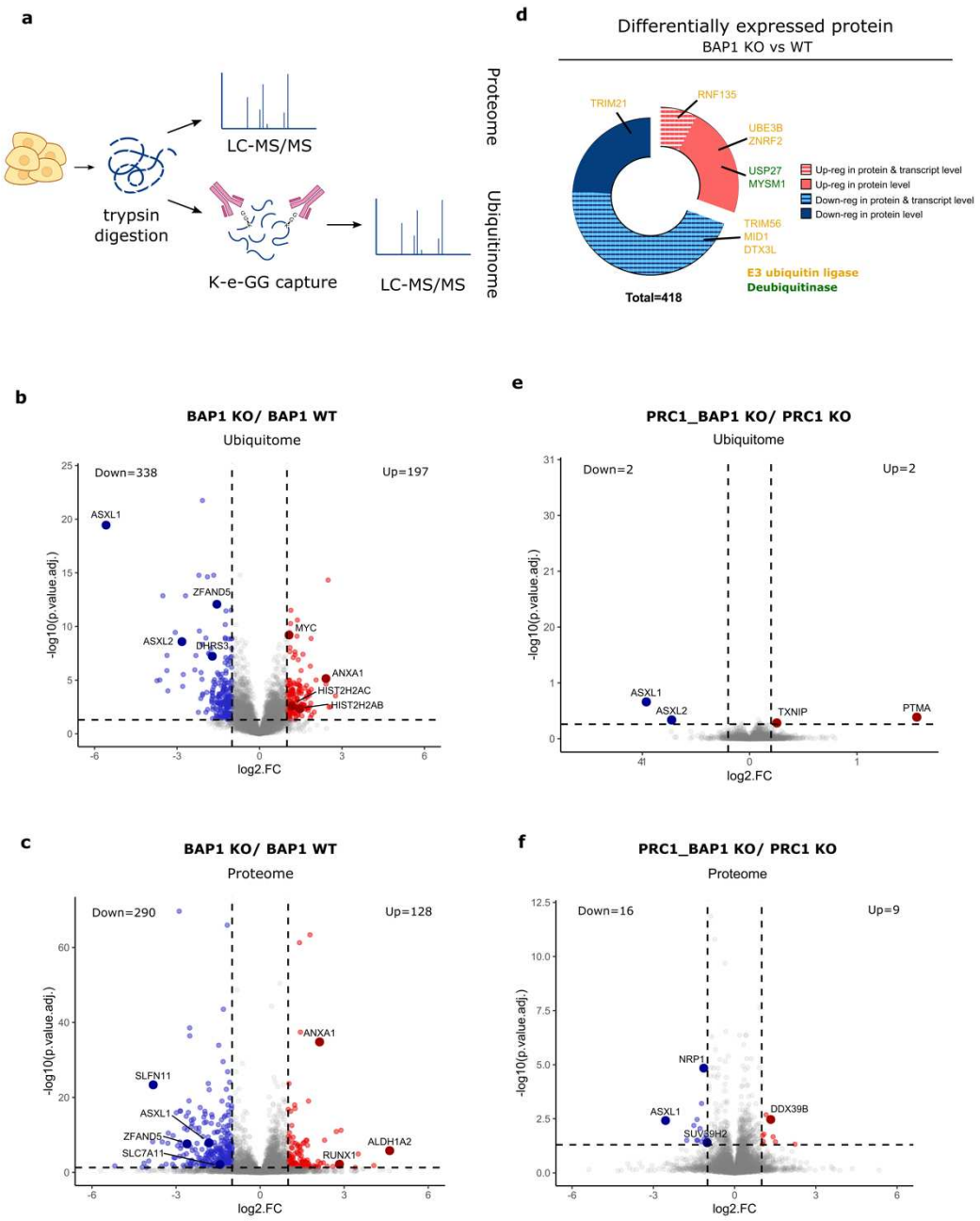

Figure 1

Figure 1: Lack of evidence for PRC1-independent substrate of BAP1.

(a) Schematic representation of Mass Spectrometry analysis of ubiquitome and proteome. IAP: immunoaffinity purification, LC-MS/MS= liquid chromatography tandem mass spectrometry.

(b) Volcano plot showing log2-fold changes in ubiquitinated peptides (K-ε-GG) of BAP1 KO *versus* WT cells. Significant enrichment changes were color coded (p-adj. < 0.05 with absolute value of log2 FC ≥ 1). Red represented increase in enrichment and blue represented decrease in enrichment.

(c) Volcano plot showing log2 fold changes in protein abundance of BAP1-KO *versus* WT cells. Significant enrichment changes were color coded (p-adj. < 0.05 with absolute value of log2 fold changes ≥ 1). Red represented increase in enrichment and blue represented decrease in enrichment.

(d) Pie chart showing differentially expressed proteins in BAP1-KO *versus* WT cells. Upregulated proteins were shown in red and downregulataed proteins were shown in blue. Dashed-line pattern represented proteins that were also differentially expressed on transcript level. E3 ligases were labeled in yellow and deubiquitinases were labeled in green.

(e) As (b) for PRC1_BAP1-KO *versus* PRC1-KO cells.

(f) As (c) for PRC1_BAP1-KO *versus* PRC1-KO cells.

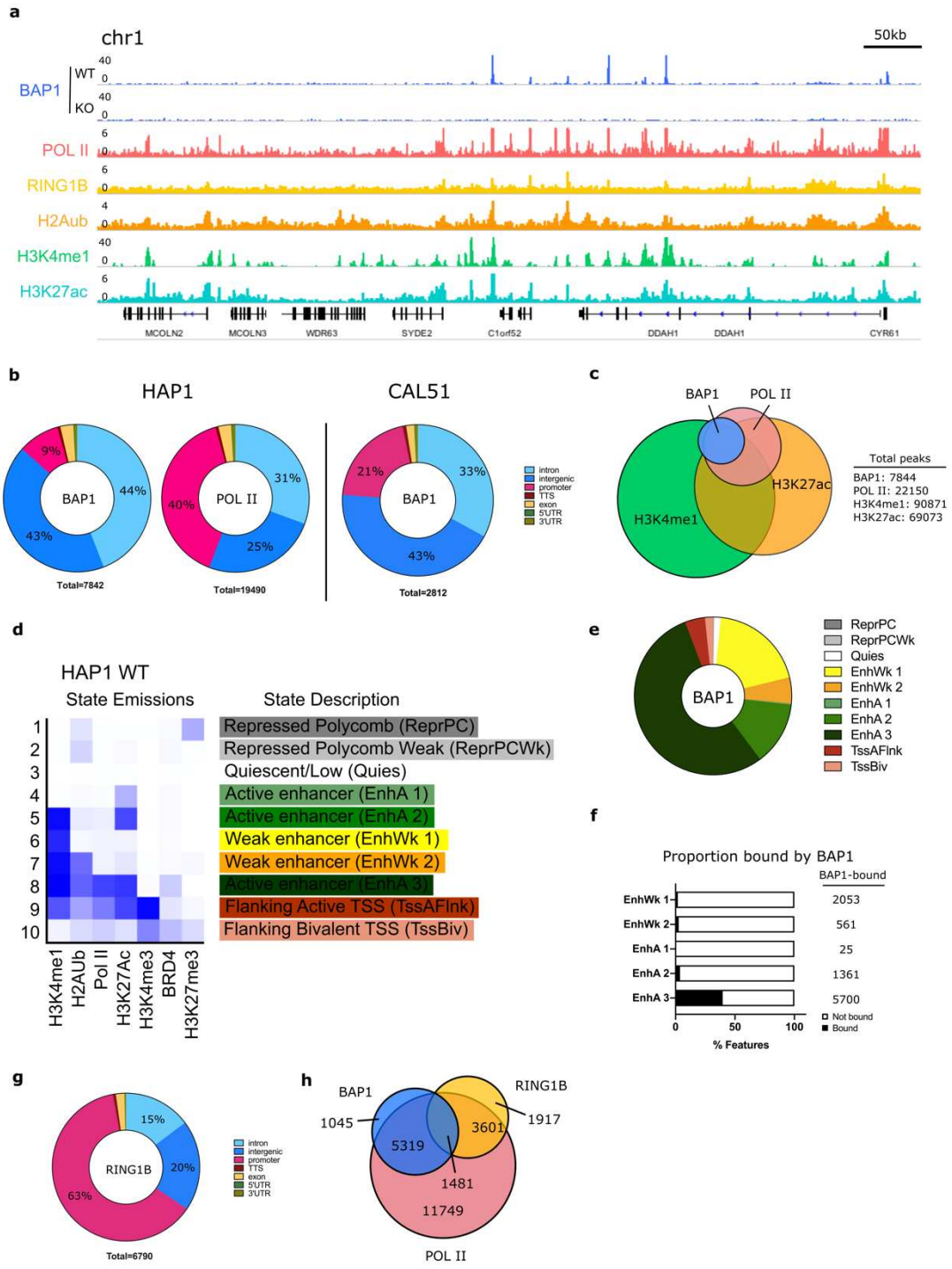
a
chr1
50kb
BAP1
WT
KO
POL II
RING1B
H2Aub
H3K4me1
H3K27ac
MCOLN2
MCOLN3
WDR63
SYDE2
C1orf52
DDAH1
DDAH1
CYR61
b
HAP1
CAL51
BAP1
POL II
BAP1
Total=7842
Total=19490
Total=2812
intron
intergenic
promoter
TTS
exon
5'UTR
3'UTR
c
BAP1
POL II
H3K4me1
H3K27ac
Total peaks
BAP1: 7844
POL II: 22150
H3K4me1: 90871
H3K27ac: 69073
d
HAP1 WT
State Emissions
State Description
Repressed Polycomb (ReprPC)
Repressed Polycomb Weak (ReprPCWk)
Quiescent/Low (Quies)
Active enhancer (EnhA 1)
Active enhancer (EnhA 2)
Weak enhancer (EnhWk 1)
Weak enhancer (EnhWk 2)
Active enhancer (EnhA 3)
Flanking Active TSS (TssAFlnk)
Flanking Bivalent TSS (TssBiv)
H3K4me1
H2AUb
Pol II
H3K27Ac
H3K4me3
BRD4
H3K27me3
e
BAP1
ReprPC
ReprPCWk
Quies
EnhWk 1
EnhWk 2
EnhA 1
EnhA 2
EnhA 3
TssAFlnk
TssBiv
f
Proportion bound by BAP1
BAP1-bound
EnhWk 1
EnhWk 2
EnhA 1
EnhA 2
EnhA 3
2053
561
25
1361
5700
Not bound
Bound
% Features
g
RING1B
Total=6790
h
BAP1
RING1B
POL II
1045
3601
1917
5319
1481
11749

Figure 2

Figure 2: BAP1 is recruited to a subset of active enhancers

(a) Genomic traces displaying CUT&RUN-seq tracks for BAP1, POL II, RING1B, H2Aub, H3K4me1 and H3K27ac.

(b) Pie charts showing genomic annotation for BAP1 and POL II peaks in HAP1 cells and BAP1 peaks in CAL51 cells.

(c) Venn diagram illustrating overlap between BAP1, POL II, H3K4me1 and H3K27ac peaks.

(d) Heatmaps summarizing ChromHMM model of 10-state emission parameters of concatenated set of HAP1 WT and BAP1-KO cells. H3K27me3, H2Aub, H3K4me1, H3K4me3, H3K27ac, POL II and BRD4 were used as input files. Each row corresponded to an emission state and each column referred as observed data of histone marks or proteins. Emission states were color-coded. Polycomb repressive states were labeled in greys. Active/ bivalent states were in red and pink, respectively. Weak enhancers were in yellow and orange. Active enhancers were in greens. Quiescence state was labeled in white. Input data were CUT&RUN-seq performed in this study, except H3K27me3 and H3K4me3 which were ChIP-seq obtaibed from GSE110133[71].

(e) Pie chart displaying annotation of BAP1 peaks across chromatin states.

(f) Bar chart showing proportion of classes of enhancers bound by BAP1.

(g) Pie charts showing genomic annotation for RING1B peaks in HAP1 cells.

(h) Venn diagram illustrating overlap between BAP1, RING1B and POL II peaks in HAP1 WT cells.

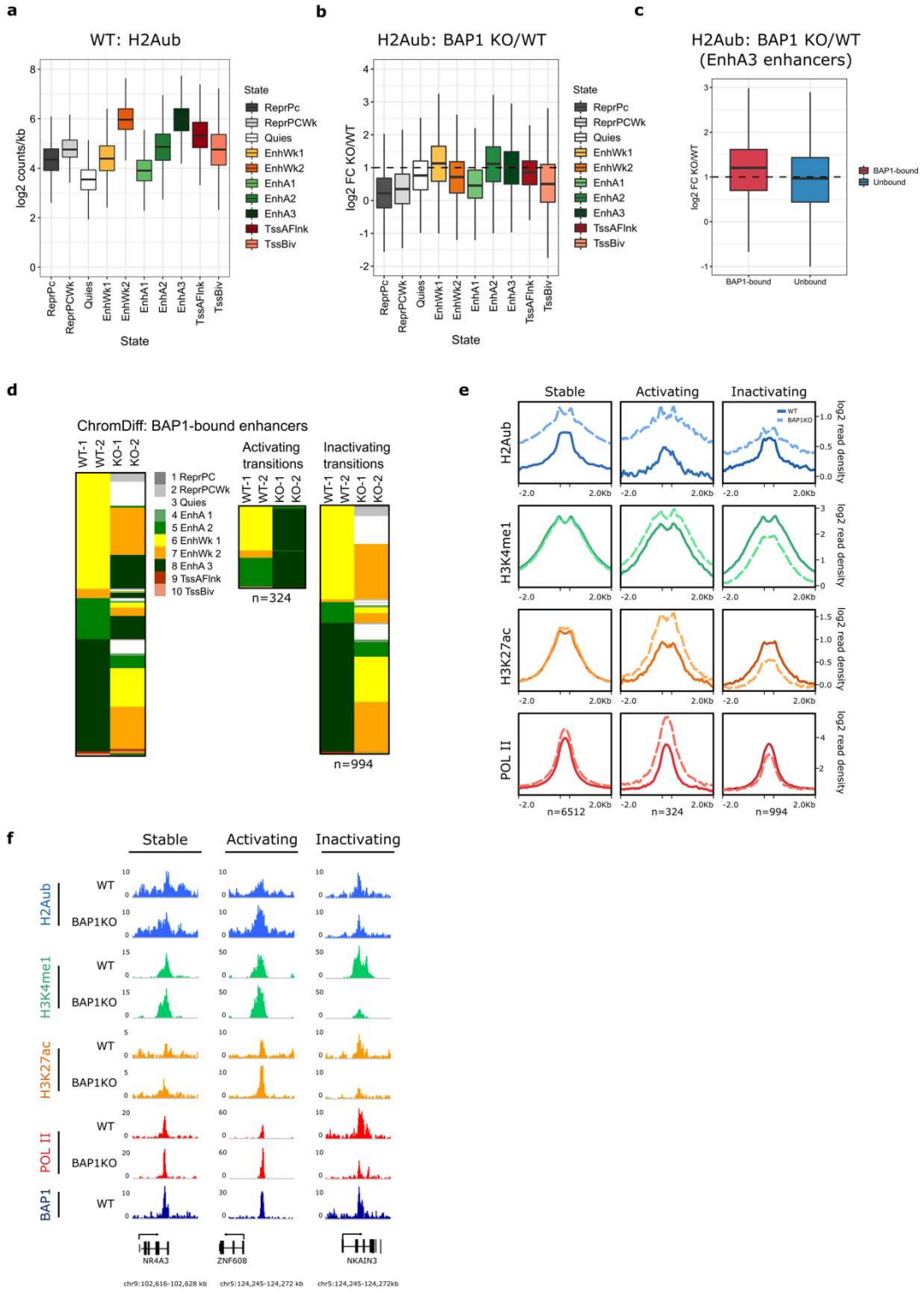
a
WT: H2Aub
log2 counts/kb
State
ReprPc
ReprPCWk
Quies
EnhWk1
EnhWk2
EnhA1
EnhA2
EnhA3
TssAFlnk
TssBiv
b
H2Aub: BAP1 KO/WT
log2 FC KO/WT
c
H2Aub: BAP1 KO/WT
(EnhA3 enhancers)
BAP1-bound
Unbound
d
ChromDiff: BAP1-bound enhancers
WT-1
WT-2
KO-1
KO-2
1 ReprPC
2 ReprPCWk
3 Quies
4 EnhA 1
5 EnhA 2
6 EnhWk 1
7 EnhWk 2
8 EnhA 3
9 TssAFlnk
10 TssBiv
Activating transitions
Inactivating transitions
n=324
n=994
e
Stable
Activating
Inactivating
H2Aub
H3K4me1
H3K27ac
POL II
log2 read density
n=6512
n=324
n=994
f
Stable
Activating
Inactivating
H2Aub
H3K4me1
H3K27ac
POL II
BAP1
WT
BAP1KO
NR4A3
ZNF608
NKAIN3

Figure 3

Figure 3: Selective accumulation of H2Aub on enhancers upon BAP1 loss

(a) Box plot showing H2Aub signal in WT cells across chromatin states.

(b) Box plot showing log2 fold changes of H2Aub signal between BAP1-KO and WT status across the chromatin states.

(c) Box plot comparing log2-fold changes of H2Aub enrichment on BAP1-bound EnhA3 or unbound EnhA3 in BAP1-KO *versus* WT cells.

(d) ChrommDiff analysis showing transitions of chromatin states on BAP1-bound enhancers in the absence of BAP1. Only consistent transitions in both replicates were considered.

(e) Metaplot analysis displaying enrichment of H2Aub, H3K4me1, H3K27ac and POL II in WT and BAP1-KO cells across stable, activating and inactivating states at BAP1-bound enhancers.

(f) Genomic snapshots showing examples of H2Aub, H3K4me1, H3K27ac, POL II and BAP1 signals on BAP1-bound enhancers with stable, activating and inactivating transitions.

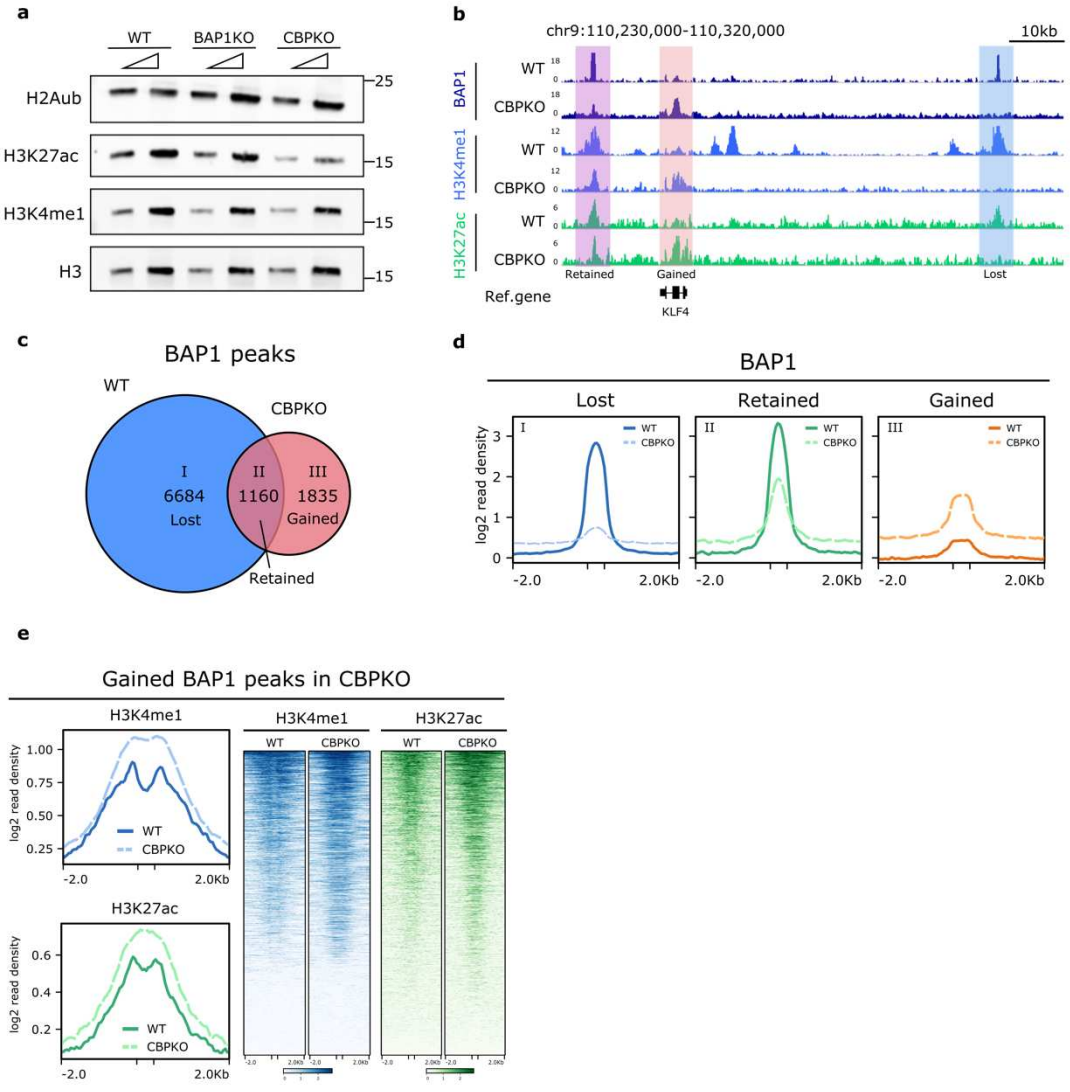
a
WT
BAP1KO
CBPKO
H2Aub
H3K27ac
H3K4me1
H3
25
15
15
15
b
chr9:110,230,000-110,320,000
10kb
BAP1
H3K4me1
H3K27ac
WT
CBPKO
Retained
Gained
Lost
Ref.gene
KLF4
c
BAP1 peaks
WT
CBPKO
I
6684
Lost
II
1160
III
1835
Gained
Retained
d
BAP1
Lost
Retained
Gained
log2 read density
-2.0
2.0Kb
e
Gained BAP1 peaks in CBPKO
H3K4me1
H3K27ac
log2 read density
WT
CBPKO
-2.0
2.0Kb

Figure 4

Figure 4: The acetyltransferase CBP is crucial for BAP1 recruitment

(a) Western blot analysis of H2Aub, H3K27ac and H3K4me1 in WT, BAP1-KO and CBP-KO cells. Total H3 was used as loading control.

(b) Genomic tracks showing CUT&RUN-seq for BAP1, H3K4me1 and H3K27ac in WT and CBP-KO cells. Blue square highlighted the BAP1 peak that was lost in CBP-KO cells. Purple square highlighted the retained peaks and pink square highlighted the gained peak.

(c) Venn diagram illustrating overlapping BAP1 peaks in WT and CBP-KO cells.

(d) Metaplot analysis comparing BAP1 signal in WT and CBP-KO cells on BAP1 peaks that were either lost, retained or gained in the absence of CBP.

(e) Metaplot and heatmap analysis displaying H3K4me1 and H3K27ac signal in WT and CBP-KO cells on gained BAP1 peaks in the absence of CBP.

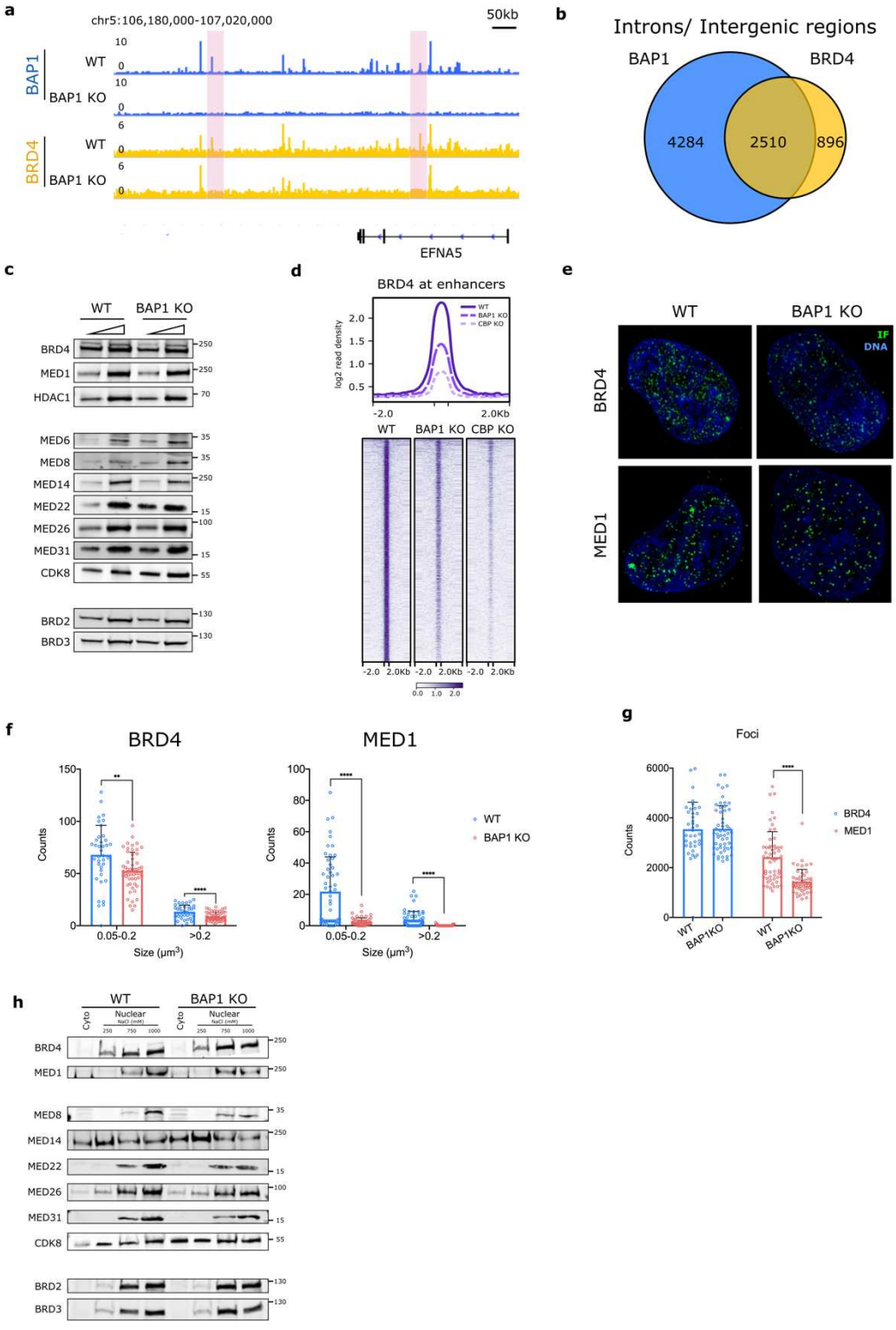
a
chr5:106,180,000-107,020,000
50kb
BAP1
WT
BAP1 KO
BRD4
WT
BAP1 KO
EFNA5
b
Introns/ Intergenic regions
BAP1
BRD4
4284
2510
896
c
WT
BAP1 KO
BRD4
MED1
HDAC1
MED6
MED8
MED14
MED22
MED26
MED31
CDK8
BRD2
BRD3
d
BRD4 at enhancers
log2 read density
WT
BAP1 KO
CBP KO
-2.0
2.0Kb
e
WT
BAP1 KO
BRD4
MED1
IF
DNA
f
BRD4
MED1
Counts
Size (μm3)
0.05-0.2
>0.2
WT
BAP1 KO
g
Foci
Counts
BRD4
MED1
WT
BAP1KO
h
WT
BAP1 KO
Cyto
Nuclear
NaCl (mM)
BRD4
MED1
MED8
MED14
MED22
MED26
MED31
CDK8
BRD2
BRD3

Figure 5

Figure 5: BAP1 inactivation compromises BRD4 and MED1 function

(a) Snapshot of genomic tracks for BAP1 and BRD4 in WT and BAP1-KO cells. Regions where BRD4 binding were in BAP1-KO cells were highlighted in pink.

(b) Venn diagram showing overlap of BAP1 and BRD4 peaks on intergenic regions and introns.

(c) Western blot analysis of total nuclear extract for BRDs and Mediator subunits in WT and BAP1-KO cells. HDAC1 was used as loading control.

(d) Metaplot and heatmap analyses displaying BRD4 enrichment in WT, BAP1-KO and CBP-KO cells. Enhancers enriched in BRD4 peaks defined in WT cells were plotted.

(e) Super-resolution microscopy showing immunofluorescence against BRD4 and MED1 with DAPI counterstain in WT and BAP1-KO cells.

(f) Bar chart displaying quantification of large BRD4 and MED1 foci in WT and BAP1-KO cells. For each condition, more than 50 cells were analyzed for foci quantification except BRD4 IF in WT cells (n=38). Threshold for large BRD4 foci was set as in the range of 0.05-0.2 (μm^3) and above 0.2 (μm^3). Thresholds for large MED1 foci was set in the range of 0.05-0.2 (μm^3) and above 0.2 (μm^3). Statistical analysis were carried out by normality tests following Student's t-test or Mann-Whitney test (** represents $p< 0.01$, and **** represents $p<0.0001$.)

(g) Bar chart displaying quantification of total BRD4 and MED1 foci in WT and BAP1-KO cells. Statistical analyses were carried out by normality tests following Mann-Whitney test (**** represents $p<0.0001$.)

(h) Western blot analysis of nuclear fractionations for BRDs and Mediator subunits in WT and BAP1-KO cells. Cyto= cytosolic fraction.

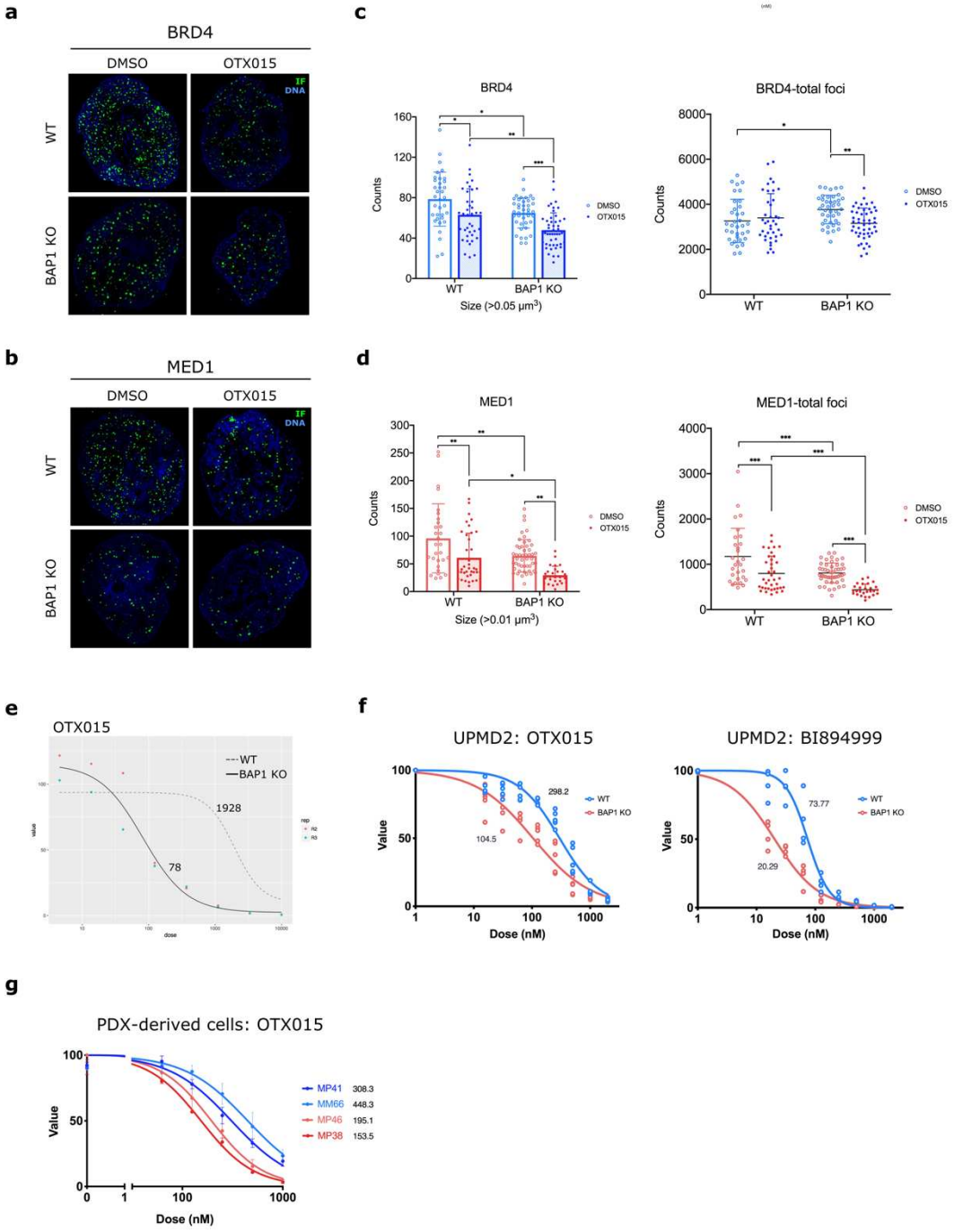
a
BRD4
DMSO
OTX015
IF
DNA
WT
BAP1 KO
b
MED1
DMSO
OTX015
IF
DNA
WT
BAP1 KO
c
BRD4
Counts
DMSO
OTX015
WT
BAP1 KO
Size (>0.05 µm³)
BRD4-total foci
Counts
DMSO
OTX015
WT
BAP1 KO
d
MED1
Counts
DMSO
OTX015
WT
BAP1 KO
Size (>0.01 µm³)
MED1-total foci
Counts
DMSO
OTX015
WT
BAP1 KO
e
OTX015
WT
BAP1 KO
1928
78
dose
f
UPMD2: OTX015
Value
WT
BAP1 KO
298.2
104.5
Dose (nM)
UPMD2: BI894999
Value
WT
BAP1 KO
73.77
20.29
Dose (nM)
g
PDX-derived cells: OTX015
Value
MP41 308.3
MM66 448.3
MP46 195.1
MP38 153.5
Dose (nM)

Figure 6

Figure 6: Loss of BAP1 increases sensitivity to BRD4 inhibition

(a) Super-resolution microscopy showing immunofluorescence against BRD4 with DAPI counterstain in WT and BAP1-KO cells treated with OTX015 or DMSO control.

(b) As in (c) for MED1 staining.

(c) Bar chart displaying quantification of large and total BRD4 foci in WT and BAP1-KO cells treated with OTX015 or DMSO control. Threshold for large foci was defined by a volume greater than 0.05 μm3.

(d) As in (c) for MED1 foci, whereas the threshold for large foci was defined by a volume above 0.01 μm3.

(e) Dose response curve of OTX015 in WT and BAP1-KO HAP1 cells. Gray dashed-line represented WT condition and black line represented BAP1-KO condition. The experiments were performed 3 times independently. Second and third replicates were represented in red and blue dots, respectively.

(f) Dose response curves of OTX015 (left panel) and BI894999 (right panel) in WT and BAP1-KO UPMD2 cells. Blue line represented WT condition and red line represented BAP1-KO condition. The experiments were performed at least 3 times independently. Each circle represented one observation on dose-response.

(g) Dose response curves of OTX015 in PDX-derived cells with BAP1 wild-type or mutant status. Blue line represented cell lines with BAP1-WT status. Red line represented cell lines with BAP1-mutant status. The experiments were performed 3 times independently. Each dot represented the average of observed dose-responses from the three replicates. Error bars represented standard error of the mean (SEM).

Supplementary figures and figure legends

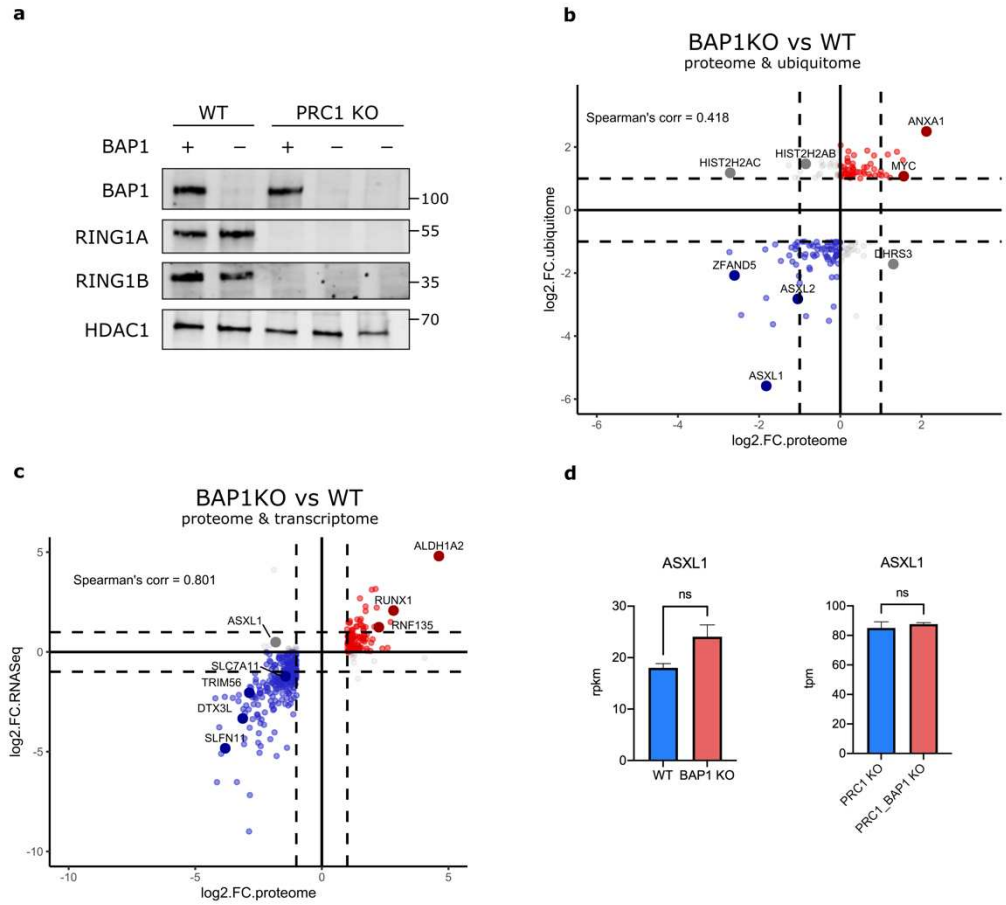

Figure S1

Figure S1

(a) Western blot analysis of BAP1, RING1A and RING1B. HDAC1 was used as loading control.

(b) Correlation plot showing proteomic and ubiquitomic analysis in BAP1-KO *versus* WT cells. Differentially enriched peptides on ubiquitination level were labeled in red (more enriched) or blue (less enriched). Correlation between ubiquitome and proteome was determined by Spearman's ranking.

(c) Correlation plot showing proteomic and ubiquitomic analysis in BAP1-KO *versus* WT cells. Differentially enriched peptides on ubiquitination level were labeled in red (more enriched) or blue (less enriched). Correlation between ubiquitome and proteome was determined by Spearman's ranking.

(d) Bar charts showing transcript level of ASXL1 representing in rpkm (reads per kilo base per million) or tpm (transcripts per million).

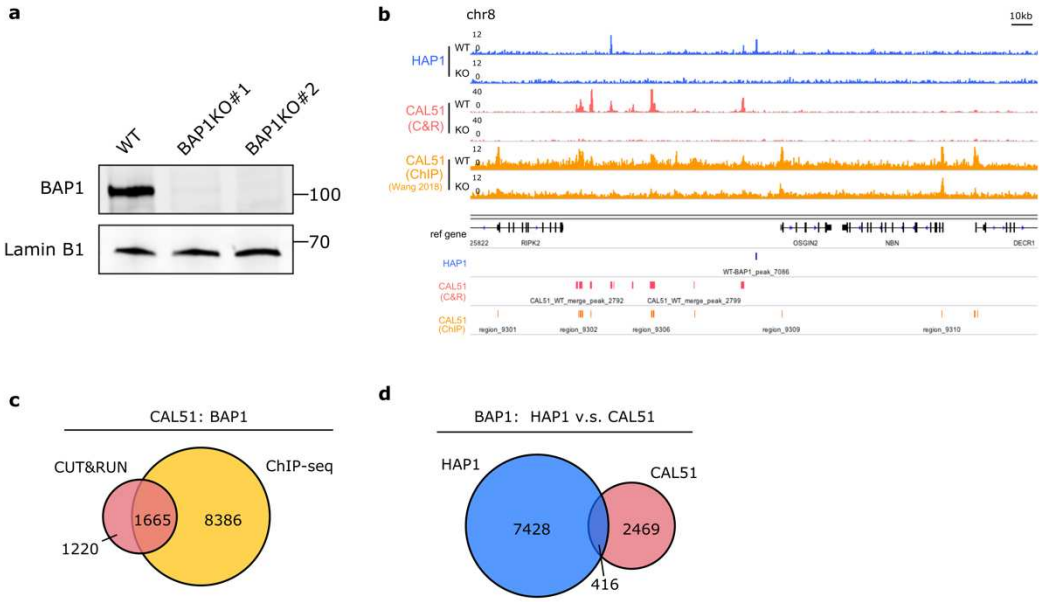

Figure S2

Figure S2

(a) Western blot analysis of BAP1 in CAL51 cells. LaminB1 was used as loading control.

(b) Genomic tracks showing CUT&RUN and ChIP-seq for BAP1 in HAP1 and CAL51 cells. BAP1 ChIP-seq wass acquired from published data (GEO: GSE97326[55,58,67,72])

(c) Venn diagram illustrating overlapping BAP1 peaks performed by CUT&RUN-seq and ChIP-seq in CAL51 cells.

(d) Venn diagram displaying overlapping BAP1 peaks between HAP1 and CAL51 cells.

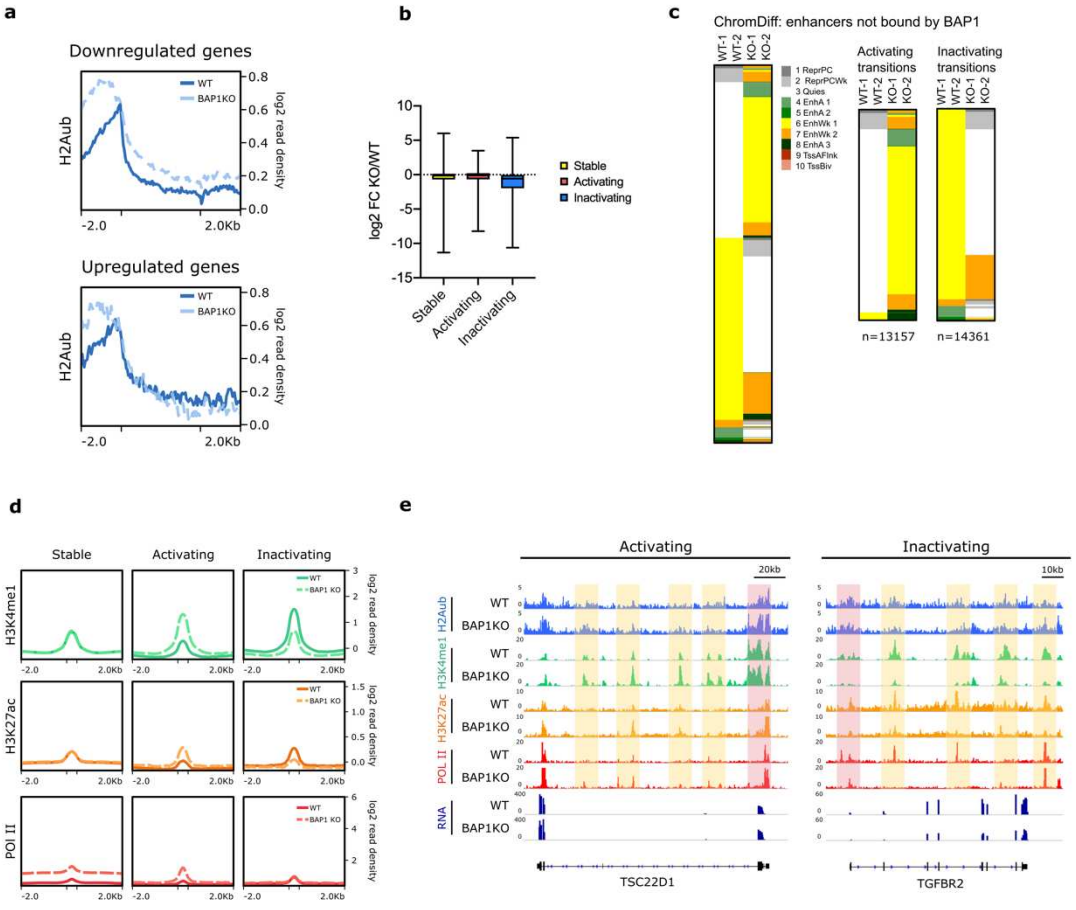
a
Downregulated genes
H2Aub
WT
BAP1KO
log2 read density
-2.0
2.0Kb
Upregulated genes
b
log2 FC KO/WT
Stable
Activating
Inactivating
c
ChromDiff: enhancers not bound by BAP1
WT-1
WT-2
KO-1
KO-2
1 ReprPC
2 ReprPCWk
3 Quies
4 EnhA 1
5 EnhA 2
6 EnhWk 1
7 EnhWk 2
8 EnhA 3
9 TssAFlnk
10 TssBiv
Activating transitions
Inactivating transitions
n=13157
n=14361
d
Stable
Activating
Inactivating
H3K4me1
H3K27ac
POl II
e
Activating
Inactivating
20kb
10kb
H2Aub
H3K4me1
H3K27ac
POL II
RNA
WT
BAP1KO
TSC22D1
TGFBR2

Figure S3

Figure S3

(a) Metaplot analysis comparing H2Aub signal between WT and BAP1-KO cells at TSSs of upregulated and downregulated genes.

(b) Box plot depicting log2 fold changes of expression level on genes associated with stable, activating and inactivating transition of BAP1-bound enhancers comparing BAP1-KO and WT cells. RNA-seq data of this study was acquired from online database GSE110142.

(c) ChrommDiff analysis showing differential chromatin states between BAP1-KO and WT cells on enhancers that were not bound by BAP1. Only consistent transitions in both replicates were considered.

(d) Metaplot analysis showing H3K4me1, H3K27ac and POL II signals in WT and BAP1-KO cells across stable, activating and inactivating states of BAP1-unbound enhancers.

(e) Genomic snapshots showing examples of H2Aub, H3K4me1, H3K27ac, POL II and RNA transcript level on enhancers undergo activating or inactivating transitions.

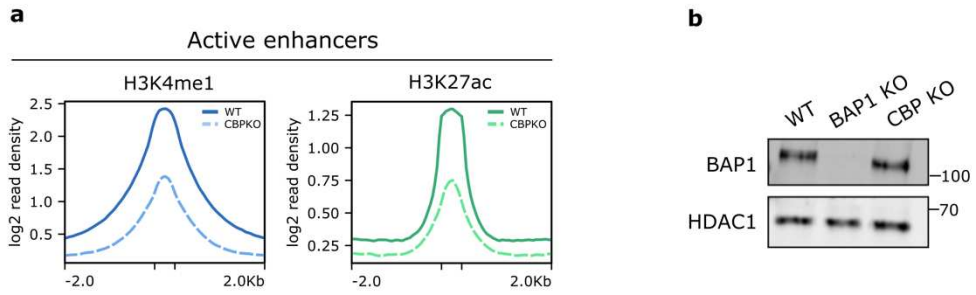

Figure S4

Figure S4

(a) Metaplot analysis displaying CUT&RUN-seq for H3K4me1 and H3K27ac in WT and CBP-KO cells. Active enhancers were defined by enrichment of H3K4me1 and H3K27ac in WT cells.

(b) Western blot analysis of BAP1 in WT, BAP1-KO and CBP-KO cells. HDAC1 was used as loading control.

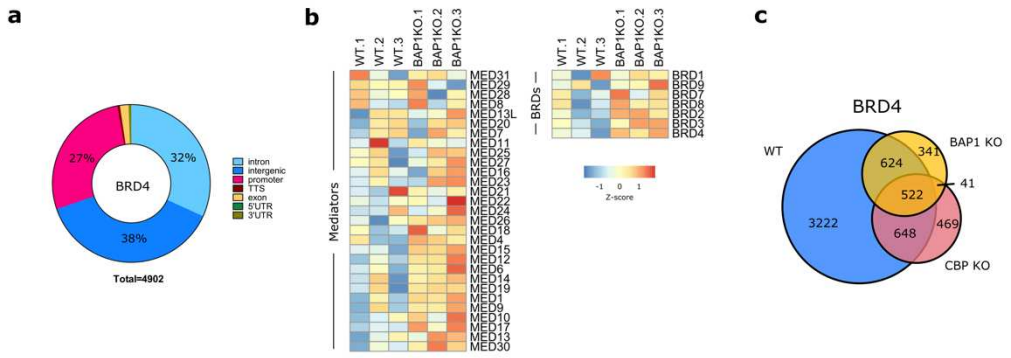

Figure S5

Figure S5

(a) Pie chart showing genomic annotation of BRD4 peaks in WT cells.

(b) Heatmap showing proteomic analysis of protein abundance for Mediator subunits and BRDs in WT and BAP1-KO cells.

(c) Venn diagram of overlapping BRD4 peaks in WT, BAP1-KO and CBP-KO cells.

Extended Data Table 1

List of antibodies used in this study

Antibodies	Host	Application	Source	Identifier
BAP1	mouse monoclonal	WB, C&R	Santa Cruz	sc-28383
BRD2	mouse monoclonal	WB	Santa Cruz	sc-393720
BRD3	mouse monoclonal	WB	Santa Cruz	sc-81202
BRD4	rabbit monoclonal	WB, C&R, IF	CST	13440S
CDK8	rabbit monoclonal	WB	CST	17395S
H2Aub	rabbit monoclonal	WB, C&R	CST	8240S
H3	rabbit polyclonal	WB	abcam	ab1791
H3K27ac	rabbit polyclonal	WB, C&R	abcam	ab4729
H3K4me1	mouse monoclonal	WB, C&R	CST	5326S
H4	mouse monoclonal	WB	CST	2935S
HDAC1	mouse monoclonal	WB	CST	5356S
LaminB1	rabbit polyclonal	WB	abcam	ab61048
MED1	rabbit polyclonal	IF	abcam	ab64965
MED1	rabbit polyclonal	WB	Bethyl	A300-793A
MED6	mouse monoclonal	WB	Santa Cruz	sc-390474
MED8	mouse monoclonal	WB	Santa Cruz	sc-365960
MED14	rabbit polyclonal	WB	Invitrogen	PA5-44864
MED22	mouse monoclonal	WB	Santa Cruz	sc-393738
MED26	rabbit monoclonal	WB	CST	14950S
MED31	rabbit polyclonal	WB	abcam	ab98142
RING1A	rabbit polyclonal	WB	CST	2820S
RING1B	rabbit monoclonal	WB, C&R	CST	5694S
RNA POL II	mouse monoclonal	C&R	active motif	61081

References

1. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45-54 (2019).
2. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* **33**, 960-982 (2019).
3. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet* **49**, 1045-1051 (2017).
4. Peterlin, B.M. & Price, D.H. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**, 297-305 (2006).
5. Field, A. & Adelman, K. Evaluating Enhancer Function and Transcription. *Annu Rev Biochem* **89**, 213-234 (2020).
6. Spitz, F. & Furlong, E.E. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613-26 (2012).
7. Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet* **21**, 71-87 (2020).
8. Long, H.K., Prescott, S.L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170-1187 (2016).
9. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**, 272-86 (2014).
10. Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem* **81**, 65-95 (2012).
11. Tie, F. *et al.* CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* **136**, 3131-41 (2009).
12. Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc Natl Acad Sci U S A* **113**, 11871-11876 (2016).
13. Tie, F. *et al.* Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development* **141**, 1129-39 (2014).
14. Roe, J.S., Mercan, F., Rivera, K., Pappin, D.J. & Vakoc, C.R. BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia. *Mol Cell* **58**, 1028-39 (2015).
15. Heintzman, N.D. *et al.* Distinct and predictive chromatin signatures of

transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-8 (2007).

16. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
17. Heintzman, N.D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-12 (2009).
18. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-6 (2010).
19. Dey, A., Chitsaz, F., Abbasi, A., Misteli, T. & Ozato, K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proc Natl Acad Sci U S A* **100**, 8758-63 (2003).
20. Jiang, Y.W. *et al.* Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc Natl Acad Sci U S A* **95**, 8538-43 (1998).
21. Soutourina, J. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol* **19**, 262-274 (2018).
22. Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C. & Werner, M. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* **331**, 1451-4 (2011).
23. Cho, W.K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412-415 (2018).
24. Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. & Kornberg, R.D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608 (1994).
25. Jang, M.K. *et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol Cell* **19**, 523-34 (2005).
26. Yang, Z. *et al.* Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* **19**, 535-45 (2005).
27. Bhagwat, A.S. *et al.* BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements. *Cell Rep* **15**, 519-530 (2016).
28. Blackledge, N.P., Rose, N.R. & Klose, R.J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol* **16**, 643-649 (2015).

29. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873-8 (2004).
30. Holoch, D. & Margueron, R. Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends Biochem Sci* **42**, 531-542 (2017).
31. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242 (2008).
32. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* **24**, 265-76 (2010).
33. Farcas, A.M. *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* **1**, e00205 (2012).
34. Wu, X., Johansen, J.V. & Helin, K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol Cell* **49**, 1134-46 (2013).
35. Mendenhall, E.M. *et al.* GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* **6**, e1001244 (2010).
36. Boyer, L.A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-53 (2006).
37. Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**, 1123-36 (2006).
38. Gaytan de Ayala Alonso, A. *et al.* A genetic screen identifies novel polycomb group genes in Drosophila. *Genetics* **176**, 2099-108 (2007).
39. Scheuermann, J.C. *et al.* Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**, 243-7 (2010).
40. Campagne, A. *et al.* BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation. *Nat Commun* **10**, 348 (2019).
41. Kolovos, P. *et al.* PR-DUB maintains the expression of critical genes through FOXK1/2- and ASXL1/2/3-dependent recruitment to chromatin and H2AK119ub1 deubiquitination. *Genome Res* **30**, 1119-1130 (2020).
42. Kuznetsov, J.N. *et al.* BAP1 regulates epigenetic switch from pluripotency to differentiation in developmental lineages giving rise to BAP1-mutant cancers. *Sci Adv* **5**, eaax1738 (2019).
43. Fursova, N.A. *et al.* BAP1 constrains pervasive H2AK119ub1 to control the transcriptional potential of the genome. *Genes Dev* (2021).
44. Carbone, M. *et al.* BAP1 and cancer. *Nat Rev Cancer* **13**, 153-9 (2013).

45. Carbone, M. *et al.* Biological Mechanisms and Clinical Significance of BAP1 Mutations in Human Cancer. *Cancer Discov* **10**, 1103-1120 (2020).
46. Vere, G., Kealy, R., Kessler, B.M. & Pinto-Fernandez, A. Ubiquitomics: An Overview and Future. *Biomolecules* **10**(2020).
47. Daou, S. *et al.* The BAP1/ASXL2 Histone H2A Deubiquitinase Complex Regulates Cell Proliferation and Is Disrupted in Cancer. *J Biol Chem* **290**, 28643-63 (2015).
48. Zhang, P. *et al.* Loss of ASXL1 in the bone marrow niche dysregulates hematopoietic stem and progenitor cell fates. *Cell Discov* **4**, 4 (2018).
49. Wang, L. *et al.* Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. *Nat Med* **24**, 758-769 (2018).
50. Skene, P.J., Henikoff, J.G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019 (2018).
51. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).
52. Loubiere, V., Papadopoulos, G.L., Szabo, Q., Martinez, A.M. & Cavalli, G. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci Adv* **6**, eaax4001 (2020).
53. Clapier, C.R., Iwasa, J., Cairns, B.R. & Peterson, C.L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407-422 (2017).
54. Yen, A. & Kellis, M. Systematic chromatin state comparison of epigenomes associated with diverse properties including sex and tissue type. *Nat Commun* **6**, 7973 (2015).
55. Sabari, B.R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**(2018).
56. Teves, S.S. & Henikoff, S. Salt fractionation of nucleosomes for genome-wide profiling. *Methods Mol Biol* **833**, 421-32 (2012).
57. Chapuy, B. *et al.* Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777-90 (2013).
58. Loven, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-34 (2013).
59. Dawson, M.A. *et al.* Inhibition of BET recruitment to chromatin as an

effective treatment for MLL-fusion leukaemia. *Nature* **478**, 529-33 (2011).
60. Harbour, J.W. *et al.* Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330**, 1410-3 (2010).
61. Amirouchene-Angelozzi, N. *et al.* Establishment of novel cell lines recapitulating the genetic landscape of uveal melanoma and preclinical validation of mTOR as a therapeutic target. *Mol Oncol* **8**, 1508-20 (2014).
62. Nemati, F. *et al.* Establishment and characterization of a panel of human uveal melanoma xenografts derived from primary and/or metastatic tumors. *Clin Cancer Res* **16**, 2352-62 (2010).
63. Dey, A. *et al.* Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science* **337**, 1541-6 (2012).
64. Chan, H.L. *et al.* Polycomb complexes associate with enhancers and promote oncogenic transcriptional programs in cancer through multiple mechanisms. *Nat Commun* **9**, 3377 (2018).
65. Stock, J.K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**, 1428-35 (2007).
66. Nakagawa, T. *et al.* Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes Dev* **22**, 37-49 (2008).
67. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. & Sharp, P.A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23 (2017).
68. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454-1458 (2016).
69. Yuan, S., Norgard, R.J. & Stanger, B.Z. Cellular Plasticity in Cancer. *Cancer Discov* **9**, 837-851 (2019).
70. Richart, L., Bidard, F.C. & Margueron, R. Enhancer rewiring in tumors: an opportunity for therapeutic intervention. *Oncogene* (2021).
71. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
72. Pelish, H.E. *et al.* Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* **526**, 273-276 (2015).

Part 2: PRC2 complex can repress transcription independently of PRC1 complex

(Preliminary results)

Ming-Kang Lee, Samuel Le Cam, Dina Zielinski, Michel Wassef, Andrey Tvardovskiy, Till Bartke and Raphaël Margueron

Introduction

Polycomb repressive complex 1 and 2 (PRC1 and PRC2) regulate important biological processes in most eukaryotes[1]. In mice, loss of either PRC1 and PRC2 component lead to fatal defects at implantation or early gastrulation stages[2,3]. Genome-wide mapping suggests that binding sites for PRC1 and PRC2 are strongly enriched for CpG island at promoters, and that both complexes and their marks largely co-occupy the same set of target genes[4-7]. PRC1 and PRC2 are long considered cooperating to maintain gene repression[8]. In mouse embryonic stem cells (mESCs), simultaneous deletion of RING1B and EED introduced de-repression of a large pool of genes that was not expressed in either RING1B or EED deletion[9]. Similar observation is also reported recently during epidermis in mice[10]. Hence, it fits with the long-standing view that PRC1 and PRC2 coordinate transcriptional silencing.

Emerging studies suggest that PRC1 can act autonomously to regulate gene expression. For example, PRC1 but not PRC2 represses a population of genes that are crucial in epidermal development in mice[11]. PRC1 can also be recruited to enhancers in developing eye disc in flies and in human breast cancer cells[12,13]. However, whether PRC2 can function independently of PRC1 remains enigmatic. Recently, a couple of chromatin factors such as CDYL, BAHD1 and their paralogs have been proposed to possess modules that recognize H3K27me3 and contribute to transcriptional repression[14-19]. This prompts us to investigate the independency and interdependency of PRC1 and PRC2. By integrating genome-editing, transcriptomic analysis and epigenomic profiling, we uncovered a subset of genes that were repressed exclusively by PRC2. The ongoing study aims to unravel the molecular mechanisms underlying this PRC2-mediated-PRC1-independent transcriptional silencing.

Results

PRC1 and PRC2 repress distinct subsets of genes

PRC1 and PRC2 complexes commonly share target genes in mESC, suggesting they broadly co-regulate transcription. Given that removal of PRC1 impairs stemness in mESC, it is difficult to study the direct consequences of its deletion[20-22]. We have previously established stable RING1A/B double KO and EED KO cells in human HAP1 cells (herein referred to as PRC1-KO and-PRC2 KO, respectively)[23]. Silencing either PRC1 or PRC2 leads to gene derepression but does not impair cell proliferation, nor does it result in any obvious cellular phenotype. Taking advantage of this model, we wondered if we could define the respective contributions of PRC1 and PRC2 to gene repression. We first inactivated EED in PRC1-KO cells (Fig. 1a), thus generating stable clones fully devoid of Polycomb machinery (herein referred to as PRC1/2-KO). Consequently, H3K27me3 and H2Aub were no longer detectable shown by western blot (Fig. 1b). Of note, deletion of EED in a PRC1-null context did not show severe proliferative defects (Fig. 1c), and no morphological changes were observed (data not shown).

Next, we performed RNA-seq on PRC1-KO, PRC2-KO and PRC1/2-KO cells to understand the transcriptomic outcomes in the loss of PRCs comparing to WT counterparts. Differential analysis indicated that loss of PRC1 or PRC2 resulted in derepression of large subsets of genes as anticipated, but the transcriptomic consequences were more pronounced in PRC1/2-KO cells (Fig. 1d). PRC1-KO and PRC2-KO led to upregulation of around 1000 and 1300 genes, respectively; whereas approximately 1900 genes were upregulated in PRC1/2-KO cells, in agreement with previous reports suggesting PRC1 and PRC2 redundantly repress gene expression[9,10].

Out of these 1901 genes, one set of 633 genes was de-repressed by loss of either PRC1 or PRC2, while another class consisting of 479 genes was de-repressed only by concomitant loss of PRC1 and PRC2 (Fig. 1e). We defined the former set of genes that was synergistically repressed by PRC1 and PRC2 (PRC1/2-synergistic), as loss of either components gave rise to its de-repression. The latter class of genes was only derepressed in the case where both PRC1/2 were lost, that is, was redundantly silenced by PRC1 and PRC2 (termed as PRC1/2-redundant.) In addition, subsets of genes that were sensitive only to deletion of either PRC1 or

PRC2, meaning that deletion of the other complex did not affect their expression levels. These sets of genes that were exclusively regulated by PRC1 or PRC2 were defined as “PRC1-sensitive” and “PRC2-sensitive”.

Interestingly, while PRC1 has been reported to repress genes independently of PRC2[11], however, to our knowledge, whether the same is true for PRC2 is unknown. In order to address this question, we first evaluated how robust were these four categories by evaluating transcription in the different genotypes (Fig. 1f). In PRC1-KO cells, PRC1-sensitive and PRC1/2-synergistic genes were upregulated on average more than 2-fold, but not those of PRC2-sensitive and PRC1/2-redundant genes (Fig. 1f, left panel). Coherently, the expression levels of PRC2-sensitive and PRC1/2-synergistic genes had more than 2-fold increase in PRC2-KO cells (Fig. 1f, middle panel). Lastly, all four categories of genes increased at least 2-fold in PRC1/2-KO cells (Fig. 1f, right panel). In addition, expression levels of a couple of genes in each category were further validated by RT-qPCR to confirm the observed characteristics (Fig. 1g). Importantly, this result also showed that the set of genes regulated by either PRC1 or PRC2 are truly distinct and the difference are not artificial division reflecting threshold effect.

By characterizing the transcriptomic outputs, we identified different set of genes that were regulated by PRC1/2 in distinct fashions. Remarkably, we observed PRC2 could silence some of its targets in a PRC1-independent manner. This is especially intriguing as PRC2 have been traditionally considered to coordinate repression with PRC1[5]. Our various PRCs-KO models enable to tackle directly this question.

H3K27me3 mediates PRC2 repressive activities

PRC2 represses transcription largely through its enzymatic activity[24], even though other mechanisms were proposed such as PRC2-EZH1-mediated chromatin compaction[25]. In order to investigate whether PRC2-mediated-PRC1-independent gene repression requires H3K27me3 deposition, we inhibited the enzymatic activities of EZH1/2, the catalytic subunits of PRC2, by treatment with EZHs-specific inhibitor UNC1999 for 10 days. Bulk H3K27me2/3 were readily detected in the control mock treatment (inactive analog UNC2400). However, these marks were no longer detectable upon treatment with UNC1999 as it would be seen upon

genetic deletion of PRC2 (Fig. 2a). Of note, the bulk level of H3K27me1 was reduced but not fully abolished under UNC1999 inhibition, suggesting of residual EZHs activities under chemical inhibition (Fig. 2a). Next, we performed RNA-seq analysis in inhibitor-treated and mock control cells, and identified 696 differentially expressed genes (DEGs), the majority of which were de-repressed (Fig. 2b). We noticed that the population of derepressed genes under PRC2 inhibition were smaller than the set of genes identified upon PRC2 deletion (618 genes v.s. 1274 genes). The relatively smaller pool of de-repressed genes in PRC2 inhibition could either originate from technical explanation (*i.e.* variability between replicates) or reflect the fact that the chemical inhibition was not as efficient as the genetic deletion. Nonetheless, the two sets of genes show substantial overlap (494 out of 618 genes; Fig. 2c), and the expression levels were highly correlated within PRC2-KO de-repressed genes (r=0.86; Fig. 2d). This indicates that PRC2 enzymatic activity, and therefore most likely H3K27me3, are required for this repressive activity.

PRC2-sensitive genes are enriched for H3K27me3 and are weakly expressed

We performed CUT&RUN-seq for H3K27me3 and H2Aub in WT, PRC1-KO and PRC2-KO cells to better define the epigenetic landscape of PRC2 and PRC1 target genes. We retrieved 38316 H3K27me3 peaks in WT cells. Notably, enrichment of H3K27me3 on these peaks markedly diminished in the absence of PRC1, suggesting a role of PRC1 in proper PRC2 recruitment and H3K27me3 deposition (Fig. 2e). Loss of large H3K27me3 domain could be observed in multiple genomic sites including HOXC loci in the absence of PRC1. In contrast, H2Aub accumulation was barely compromised by PRC2 deletion (Fig. 2f). Together, these suggest that PRC2 occupancy partially relied on PRC1; but that PRC1 recruitment was mostly independent of PRC2. These results are consistent with the observations reported in mESC model[21,26,27].

Focusing on PRC2-sensitive genes, we quantified enrichment for H3K27me3 and H2Aub at the transcription starting sites (TSSs) and flanking regions of the four categories of PRC1/2 target genes. Interestingly, PRC2-sensitive genes had the highest enrichment of H3K27me3, while its enrichment in PRC1-sensitive genes were close to the background level. H2Aub levels remained comparable across each category (Fig. 2g). This result is consistent with the tight correlation between

PRC2 function and H3K27me3 deposition but also the more debated link between PRC1 function and H2Aub deposition. Considering our ongoing study on BAP1, H2Aub and enhancer, it will be interesting to check whether a better correlation between H2Aub and transcriptomic response to PRC1 deletion could be explained by analyzing enrichment of this mark at enhancers.

Previous study suggested that the density of H3K27me3 negatively correlates with gene expression[6,7]. Consistently, we observed that PRC2-sensitive genes had the lowest basal transcriptional activities among the four categories. By contrast, PRC1-sensitive genes, which display the lowest enrichment for H3K27me3, have the highest transcript level. Genes co-regulated by PRC1 and PRC2 show intermediate levels of transcription (Fig. 2h). As expected, RNA polymerase ii (POL ii) enrichment (evaluated by CUT&RUN) parallels transcription measured through mature RNA accumulation (Fig. 2i), and anti-correlates with H3K27me3.

PRC1-independent readers of H3K27me3 could contribute to PRC2-mediated repression

We have identified a subset of genes that are silenced exclusively by PRC2 and that are characterized by a high enrichment for H3K27me3. So far, this mark has been to function through the recruitment of additional factor instead of through the direct modulation of chromatin structure. In addition to PRC1, several proteins, including BAHD1/2 and CDYL, have been reported to contain domains recognizing methylated H3K27me3[14,17]. Moreover, depletion of those proteins might promote derepression of PRC2 target genes[18,19,28]. Altogether, these results lead us to investigate whether these chromatin factors could be involved in the transcriptional silencing maintained by PRC2 independently of PRC1.

In a PRC1-KO context, we first inactivated CDYL and BAHD1 by CRISPR-Cas9 technology. Inactivation was verified by genomic DNA genotyping and by RT-qPCR (data not shown). We then further deleted BAHD1 deletion in the PRC1_CDYL-KO cells to generate PRC1_CDYL_BAHD1-KO clones (herein referred to as PRC1_qKO) (Fig. 3a). We performed RNA-seq to examine whether loss of BADH1 and CDYL contribute to PRC2 silencing. Silencing of either gene independently has mild consequences on gene expression (approximately 100 DEG). In contrast the combined deletion of CDYL and BAHD1 has a more

pronounced effect on the gene expression profile with about 651 DEG (Fig. 3b). This additive effect is consistent with that CDYL and BAHD1 both contribute to gene repression. Importantly, these genes showed substantial but partial overlap with the set of genes derepressed upon PRC2-KO in a PRC1-KO context (Fig. 3c, d). In addition to being a partial overlap, we also note that the magnitude of derepression in the PRC1_qKO vs PRC1_PRC2_KO are quite different. Taken together, these results suggest that while BAHD1 and CDYL could contribute to PRC2 silencing, other readers are also involved. This could be other unknown factors or homologs of CDYL for instance.

Search for unknown readers of H3K27me3

Unbiased approach aiming at identifying H3K27me3 readers were already implemented and resulted mostly in the identification of PRC1 subunits [15]. We therefore reasoned that repeating such approach in the absence of PRC1 complex could help identifying less abundant interactors. We prepared nuclear extract of WT and PRC1-KO cells and performed pull-down assay on unmodified and H3K27me3-modified recombinant nucleosomes. The enriched peptides were then proceeded to mass spectrometry (MS) analysis. As expected, we were able to recover PRC2 subunits including EZH2, EED, MTF2 and PHF1, and canonical PRC1 subunits such as RING1B, CBX2/8 and PHC2/3 in H3K27me3-enriched nucleosomes (Fig. 4a). Upon inactivation of RING1A and RING1B, several PRC1 subunits (e.g. YAF2 and PHC2/3) were destabilized and binding of PHCs were lost (Fig, 4b,c). However, CBX2 and CBX8 remained bound to H3K27me3-nucleosomes, suggesting their binding can occur outside of PRC1 complex (Fig. 4b).

Beyond the “usual suspects”, we discovered high enrichment for SIN3 complex and MXI1/MNT MYC-antagonists in both WT and PRC1-KO cells (Fig.4 a, b). SIN3/HDAC complex interacts with an array of (co-)repressors and functions to remove acetylation mark on histone, thereby promoting gene silencing[29]. MXI1 and MNT compete with MYC to form heterodimers with transcription factor MAX. Therefore, they repress transcription by acting as MYC antagonists to prevent MYC-mediated transcriptional activation. Interestingly, MNT and MXI1 have been shown to interact with the SIN3 complex[30]. Together, we identified SIN3 complex and MNT/MXI1 as potential H3K27me3 interacting proteins by pull-down assay. This result still needs confirmation though independent methods and additional

investigation to understand the possible underlying mechanism.

Perspectives and Discussion

In this study, we have established model cell lines allowing to completely manipulate the Polycomb machinery, and therefore disentangle the respective contribution of PRC1 and PRC2. We provided clear evidence that PRC2 could repress transcription of a subset of genes independently of PRC1 (Fig. 1e). The underlying mechanism might turn out to be relevant beyond mammals since PRC1 complex is reported to be absent from some species such as fungi [8].

Through a candidate-based approach, we investigated the contribution of CDYL and BAHD1 to PRC2-mediated gene silencing. Our results show that the combined deletion of CDYL and BAHD1 only partially explained PRC2-mediated transcriptional repression. One of the possible explanation is the existence of the paralogs: CDYL2 and BAHD2 (also known as BAHCC1). Studies showed that removal of BAHD2 resulted in de-repression of PRC2 target genes[19], while CDYL2 is less well-characterized. In HAP1 cells, CDYL2 is not expressed, and no compensatory effect was observed upon CDYL deletion. Yet, transcripts of CDYL2 become detectable in the quadruple knockouts (data not shown). Hence, we cannot rule out that CDYL2 contribute to some level of a PRC2-mediated silencing.

Apart from that, we observed a group of genes that was exclusively de-repressed in PRC1_qKO but no PRC1/2-KO cells (Fig. 3c). This indicates that BAHDs/CDYLs are potentially involved in H3K27me3-independent pathways. Indeed, BAHD1 and CDYL were also shown to bind H3K9 methylation *in vitro*[16,31].

Our pull-down assay following MS analysis revealed that the SIN3/HDAC co-repressor complex and MXI1/MNT are specifically bound to H3K27 tri-methylated nucleosomes (Fig. 4). Interestingly, CDYL and BAHD1 interact also with HDAC1/2 in vitro[32,33]. It would be interesting to understand if and how SIN3, MXI1/MNT, CDYL, and BAHD1 coordinate H3K27me3-mediated repression. Different biochemical, genetic, genomic approaches would be required to unravel the detailed mechanism.

Material and Methods

Cell culture

HAP1 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10 % fetal bovine serum (FBS, Gibco), 2 mM L-glutamine (Gibco) and 1 × non-essential amino acid (Gibco).

Generation of KO cell lines

Generation of KO cells lines was performed using CRISPR-Cas9 technology as previously described[34]. In brief, a STOP cassette containing an antibiotic resistance gene followed by a polyadenylation sequence was inserted into early exons of target genes by homologous recombination. After antibiotic selection, clones were genotyped and complete KO was validated by western blot.

Cell proliferation assay

10^5 Cells were seeded on each well of a 6-well plate in technical triplicates and were counted with Vi-CELL XR system (Beckman COULTER) at 24 hr interval for a period of 4 days. The experiments were repeated at least 3 times.

Chemical inhibition of PRC2 activity

Cells were treated with 1 μm of UNC1999 or mock control (UNC2400) for 10 days. Culture medium was renewed every 3 days. Cells were passaged constantly when reached 80 % of confluence. Cells were harvest for RNA and protein extraction after 10 days of treatment.

Preparation of nuclear extract and immunoblotting

Cells were washed once with PBS and then resuspended with 5 volumes of Buffer A (10 mM HEPES pH 7.9, 5 mM $MgCl_2$, 0.25 M sucrose and 0.1 % NP-40, 1 mM DTT, 200 μM PMSF, and protease inhibitors). After 10 min incubation on ice, cells were pelleted by centrifugation at 8000 g for 10 min. Supernatant was removed and pellets were resuspended with 5 volumes of Buffer B (25 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 0.1 mM EDTA pH 8.0, 20 % glycerol, 700 mM NaCl, 1 mM DTT, 200 μM PMSF and protease inhibitors. After 10 min incubation on ice, nuclei were sonicated for 45 s with 10 % amplitude, then centrifuged at 14000 g for 15 min at 4 °C. The supernatant was transferred to a new tube and taken as nuclear extract. Protein concentration was measured by Bradford assay (Biorad). Western Blot

analysis of protein extracts was performed by StarBright Blue 700 fluorescent secondary antibodies (Biorad) and DyLight 800 secondary antibody (Biorad). Imaging was carried out by ChemiDoc System (Biorad).

RNA extraction and RT-qPCR

Total RNA was isolated using Trizol-Chloroform extraction and iso-propanol precipitation. cDNA was synthesized using High Capacity cDNA RT kit -Applied Biosystems) and quantitative PCR was performed with technical triplicate using SYBR green reagent (Roche) on a ViiA7 equipment (Applied Biosystems). At least two independent experiments (biological replicates) were performed for each assay and RT negative controls were always included. Primer sequences for qPCR analysis are provided in Supplementary Table 1.

RNA Sequencing

50 bp single end reads were generated using the HiSeq2500 in Rapid Run mode. Raw reads were trimmed for adapters with cutadapt (1.12) using the Trim Galore! (0.4.4) wrapper (default settings) and subse- quently mapped to the complete human rRNA sequence with Bowtie2 (2.2.9). Reads that did not map to rRNA were then mapped with STAR (2.5.2b) to the full reference genome (hg19) using the following parameters: –outSAMtype BAM SortedByCoordinate –runMode alignReads –outFilterMismatchNmax 6 – outFilterMultimapNmax 20 – outSAMmultNmax 20 –outSAMprimaryFlag OneBestScore. Gene counts were generated using STAR –quant_mode (uniquely mapped, properly paired reads that overlap the exon boundaries of each gene).

Differential Expression Analysis

Genes with CPM > 0.5 in at least 2 samples were kept for differential analysis. This threshold was chosen based on the average log2CPM per gene to separate expressed genes from unexpressed genes. Raw count data was normalized with the TMM method and transformed to log2-CPM. A linear model was fit to the normalized data and empirical Bayes statistics were computed. Differentially expressed genes for each comparison were identified from the linear fit after adjusting for multiple testing and filtered to include those with FDR < 0.05 and absolute logFC > 1.

CUT&RUN-seq

CUT&RUN was performed as previously described with minor modifications[35]. In brief, 1 million cells were pelleted at 600 g for 3 min at RT. After washing twice with 1 mL of wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine (Sigma) and protease inhibitors), cells were resuspended in wash buffer and ready for binding with beads. 10 µl of Convanavalin A beads (Bang Laboratories) was washed once with 1 mL binding buffer (20 mM HEPES pH 7.9, 10 mM KCl, 1 mM CaCl2 and 1 mM MnCl2) and placed on magnet stand to remove the liquid. 10 µl of binding buffer was used to resuspend the beads then the slurry was transferred to cells and incubated for 10 min at RT with rotation. After brief spin-down, tubes were placed on magnet to quickly withdraw the liquid. 50 µl of antibody buffer (wash buffer supplemented with 0.1 % digitonin (Millipore), 2 mM EDTA and 1:100 dilution of antibody of interest) was pipetted and cells were incubated for 10 min at RT with mild agitation. Permeabilized cells were decanted carefully and washed once with 1 mL dig-wash buffer (0.1 % digitonin in wash buffer). After 2 washes with 1 mL dig-wash buffer, beads were resuspended with 100 µl dig-wash buffer and placed on heat block immersed in wet ice to chill down to 0 °C. 2 µl of 100 mM CaCl2 was added to activate pA-MNase and incubated on heat block for 30 min. 100 µl of 2 × stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02 % digitonin, 1:200 RNase A, glycogen (50 mg/ mL) and heterologous spike-in DNA (2 pg/ml) was added to quench pA-MNAse, and fragments were released by 10 min incubation at 37 °C with rotation. After centrifugation at 14000 g for 5 min at 4 °C, DNA fragments were recovered by NucleoSpin (Macherey Nagel) or phenol-chloroform purification. Library was prepared by Accel-NGS 2S plus DNA library Kits (Swift Biosciences) for Illumina barcoded system. PCR were set to 16 cycles. After post-library size selection, library size distribution and concentration were validated by Tapestation 4200 (Agilent). Libraries were sequenced as paired-ended 100bp reads on Illumina Novaseq platform.

Nucleosome pull-down assay

Approximately 150 million cells were harvest and resuspended in 500 µl of Buffer A (10 mM HEPES pH 7.9, 5 mM $MgCl_2$, 0.25 M sucrose and 0.1 % NP-40, 1 mM DTT, 200 µM PMSF, and protease inhibitors). After 10 min incubation on ice, cells were pelleted by centrifugation at 8000 g for 10 min. Supernatant was collected as cytosolic fraction. Nuclei pellet was resuspended with 500 µl of Buffer C (350 mM

NaCl with 20 mM HEPES-KOH pH7.9, 25 % Glycerol, 0.1 % NP-40, 0.5 mM $MgCl_2$, 0.2 mM EDTA pH 8.0, 1 mM DTT, 0.5 mM PMSF and protease inhibitors). After 2 hr incubation at 4 °C with rotation, nuclei were pelleted and supernatant was collected as 350 mM NaCl fraction. Pellets were then incubated with 500 µl of Buffer C (500 mM NaCl) for 2 hr at 4 °C with rotation. After centrifugation, supernatant was collected as the 500 mM NaCl fraction. Pellets were then incubated with 500 µl of Buffer C (750 mM NaCl) for 2 hr at 4 °C with rotation. After centrifugation, supernatant was collected as the 750 mM NaCl fraction. Lastly, chromatin pellet was incubated with 500 µl of Buffer C (1M NaCl) overnight at 4 °C with rotation and samples were proceeded for sonication. All fractions were then combined. Protein concentration was measure by Bradford assay. 500 µg of total protein was prepared for pull-down assay and proceded with Mass Spectrometry analysis. The experiments were performed with 3 biological replicates and each biological repeat was performed under 3 technical replicates.

CUT&RUN-seq data analysis

Reads were mapped to the human reference genome (GRCh37/hg19) with Bowtie2 using default parameters. Aligned reads were sorted by SAM tools. PCR duplicates were removed with Picard Tools MarkDuplicates (https://github.com/bioinfo-pf-curie/ChIP-seq). Generated BAM files were filtered to exclude common artifact regions. (artefact regions: https://github.com/Boyle-Lab/Blacklist/tree/master/lists). Exploratory data analyses were performed using Galaxy Europe interface (https://usegalaxy.eu/). Biological replicates were merged with MergeSamFiles for downstream analysis. Reads were counted in bins of length 50, RPKM normalized, and converted to bigWig format using bamCoverage (v3.3.2.0.0). H3K27mes peaks were called with MACS2 (v 2.1.1.20160309.6) with default parameters. Igg.bam files are used as control files. Minimum FDR (q-value) cutoff for broad peak detection was set as 0.0001. Peaks with the proximity within a 2kb window were merged. Metaplot and heatmap analyses were performed using deepTools (v3.3.2.0.0): RPKM normalized log2 ratio between H3K27me3 files and Igg control files were calculated by bamCompare. Matrix was prepared by computeMatrix (v3.3.2.0.0) for metaplot and heatmap visualization.

Supplementary Table 1: primers used for RT-qPCR in this study

Target	Forward sequence	Reverse sequence	Amplicon size (bp)
BMP7	TCGGCACCCATGTTCATGC	GAGGAAATGGCTATCTTGCAGG	150
CHD3	TGGAGATCCTTGATGCCAATGA	GCGTCCAGATCAGTGACCG	121
GATA4	GTGTCCCAGACGTTCTCAGTC	GGGAGACGCATAGCCTTGT	102
PAX7	ACCCCTGCCTAACCACATC	ACCCCTGCCTAACCACATC	121
ASXL3	ATTAGCCTGTCTGAATGCAATGC	GACTAAATCCAACGTGCCATCT	142
RUNX2	TGGTTACTGTCATGGCGGGTA	TCTCAGATCGTTGAACCTTGCTA	101
HES5	GGAGCGTCAGGAACTGCAC	AAGAGCCTGCACCAGGACTA	73
TNR	AAGAGCCTGCACCAGGACTA	AAGAGCCTGCACCAGGACTA	73

References

1. Kuroda, M.I., Kang, H., De, S. & Kassis, J.A. Dynamic Competition of Polycomb and Trithorax in Transcriptional Programming. *Annu Rev Biochem* **89**, 235-253 (2020).
2. Voncken, J.W. *et al.* Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A* **100**, 2468-73 (2003).
3. O'Carroll, D. *et al.* The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol* **21**, 4330-6 (2001).
4. Boyer, L.A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-53 (2006).
5. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242 (2008).
6. Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**, 1123-36 (2006).
7. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60 (2007).
8. Schuettengruber, B., Bourbon, H.M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34-57 (2017).
9. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* **24**, 265-76 (2010).

10. Cohen, I. *et al.* Polycomb complexes redundantly maintain epidermal stem cell identity during development. *Genes Dev* **35**, 354-366 (2021).
11. Cohen, I. *et al.* PRC1 preserves epidermal tissue integrity independently of PRC2. *Genes Dev* **33**, 55-60 (2019).
12. Loubiere, V., Papadopoulos, G.L., Szabo, Q., Martinez, A.M. & Cavalli, G. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci Adv* **6**, eaax4001 (2020).
13. Chan, H.L. *et al.* Polycomb complexes associate with enhancers and promote oncogenic transcriptional programs in cancer through multiple mechanisms. *Nat Commun* **9**, 3377 (2018).
14. Vermeulen, M. *et al.* Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967-80 (2010).
15. Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470-84 (2010).
16. Escamilla-Del-Arenal, M. *et al.* Cdyl, a new partner of the inactive X chromosome and potential reader of H3K27me3 and H3K9me2. *Mol Cell Biol* **33**, 5005-20 (2013).
17. Zhao, D. *et al.* The BAH domain of BAHD1 is a histone H3K27me3 reader. *Protein Cell* **7**, 222-6 (2016).
18. Fan, H. *et al.* A conserved BAH module within mammalian BAHD1 connects H3K27me3 to Polycomb gene silencing. *Nucleic Acids Res* **49**, 4441-4455 (2021).
19. Fan, H. *et al.* BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. *Nat Genet* **52**, 1384-1396 (2020).
20. Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**, 1513-24 (2008).
21. Blackledge, N.P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445-1459 (2014).
22. Chamberlain, S.J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **26**, 1496-505 (2008).
23. Campagne, A. *et al.* BAP1 complex promotes transcription by opposing

PRC1-mediated H2A ubiquitylation. *Nat Commun* **10**, 348 (2019).

24. Wassef, M. *et al.* EZH1/2 function mostly within canonical PRC2 and exhibit proliferation-dependent redundancy that shapes mutational signatures in cancer. *Proc Natl Acad Sci U S A* **116**, 6075-6080 (2019).
25. Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell* **32**, 503-18 (2008).
26. Fursova, N.A. *et al.* Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol Cell* **74**, 1020-1036 e8 (2019).
27. Blackledge, N.P. *et al.* PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol Cell* **77**, 857-874 e9 (2020).
28. Zhang, Y. *et al.* Corepressor protein CDYL functions as a molecular bridge between polycomb repressor complex 2 and repressive chromatin mark trimethylated histone lysine 27. *J Biol Chem* **286**, 42414-42425 (2011).
29. Grzenda, A., Lomberk, G., Zhang, J.S. & Urrutia, R. Sin3: master scaffold and transcriptional corepressor. *Biochim Biophys Acta* **1789**, 443-50 (2009).
30. Hurlin, P.J. & Huang, J. The MAX-interacting transcription factor network. *Semin Cancer Biol* **16**, 265-74 (2006).
31. Bierne, H. *et al.* Human BAHD1 promotes heterochromatic gene silencing. *Proc Natl Acad Sci U S A* **106**, 13826-31 (2009).
32. Mulligan, P. *et al.* CDYL bridges REST and histone methyltransferases for gene repression and suppression of cellular transformation. *Mol Cell* **32**, 718-26 (2008).
33. Lakisic, G. *et al.* Role of the BAHD1 Chromatin-Repressive Complex in Placental Development and Regulation of Steroid Metabolism. *PLoS Genet* **12**, e1005898 (2016).
34. Wassef, M. *et al.* Versatile and precise gene-targeting strategies for functional studies in mammalian cell lines. *Methods* **121-122**, 45-54 (2017).
35. Skene, P.J., Henikoff, J.G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019 (2018).

Figures and figure legends

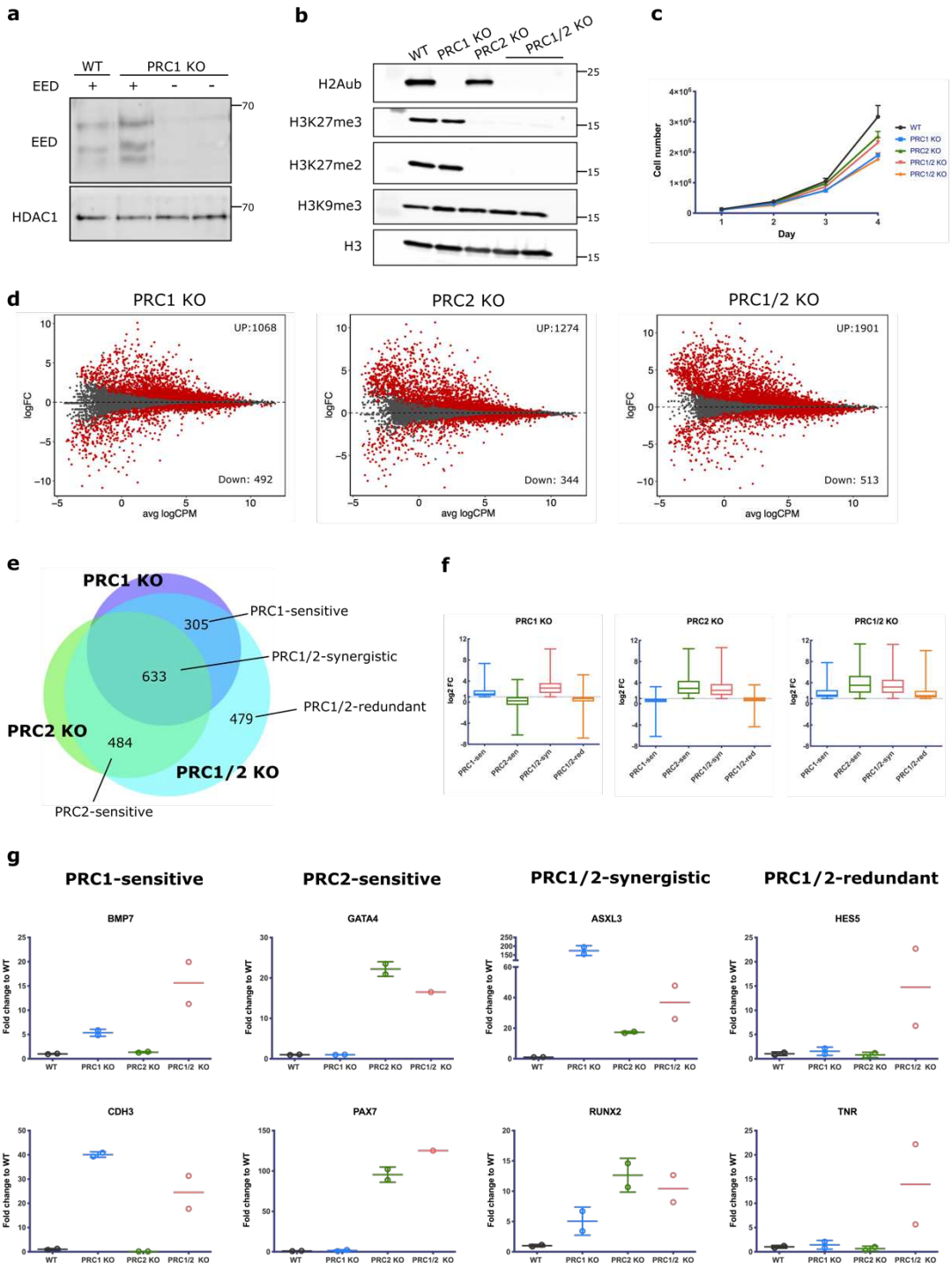

Figure 1

Figure 1: PRC1 and PRC2 repress distinct subsets of genes

(a) Western blot analysis of EED in WT and PRC1-KO cells. HDAC1 was used as loading control.

(b) Western blot analysis of H2Aub, H3K27me3, H3K27me3 and H3K9me3 in WT, PRC1-KO, PRC2-KO and PRC1/2-KO cells. Total H3 was used a loading control.

(c) Growth curve of WT, PRC1-KO, PRC2-KO and PRC1/2-KO cells.

(d) MA-plot showing DEGs in PRC1-KO, PRC2-KO and PRC1/2-KO cells comparing to WT cells.

(e) Venn diagram displaying overlap of DEGs in the PRC1-KO, PRC2-KO and PRC1/2-KO context.

(f) Box plot showing average log2 fold changes of the four classes of genes in PRC1-KO, PRC2-KO and PRC1/2-KO cells.

(g) RT-qPCR showing mRNA expression levels of PRC1-sensitive, PRC2-sensitive, PRC1/2-redundant and PRC1/2-synergistic genes. Gene expression was first normalized to TBP, FC was then normalized comparing WT condition.

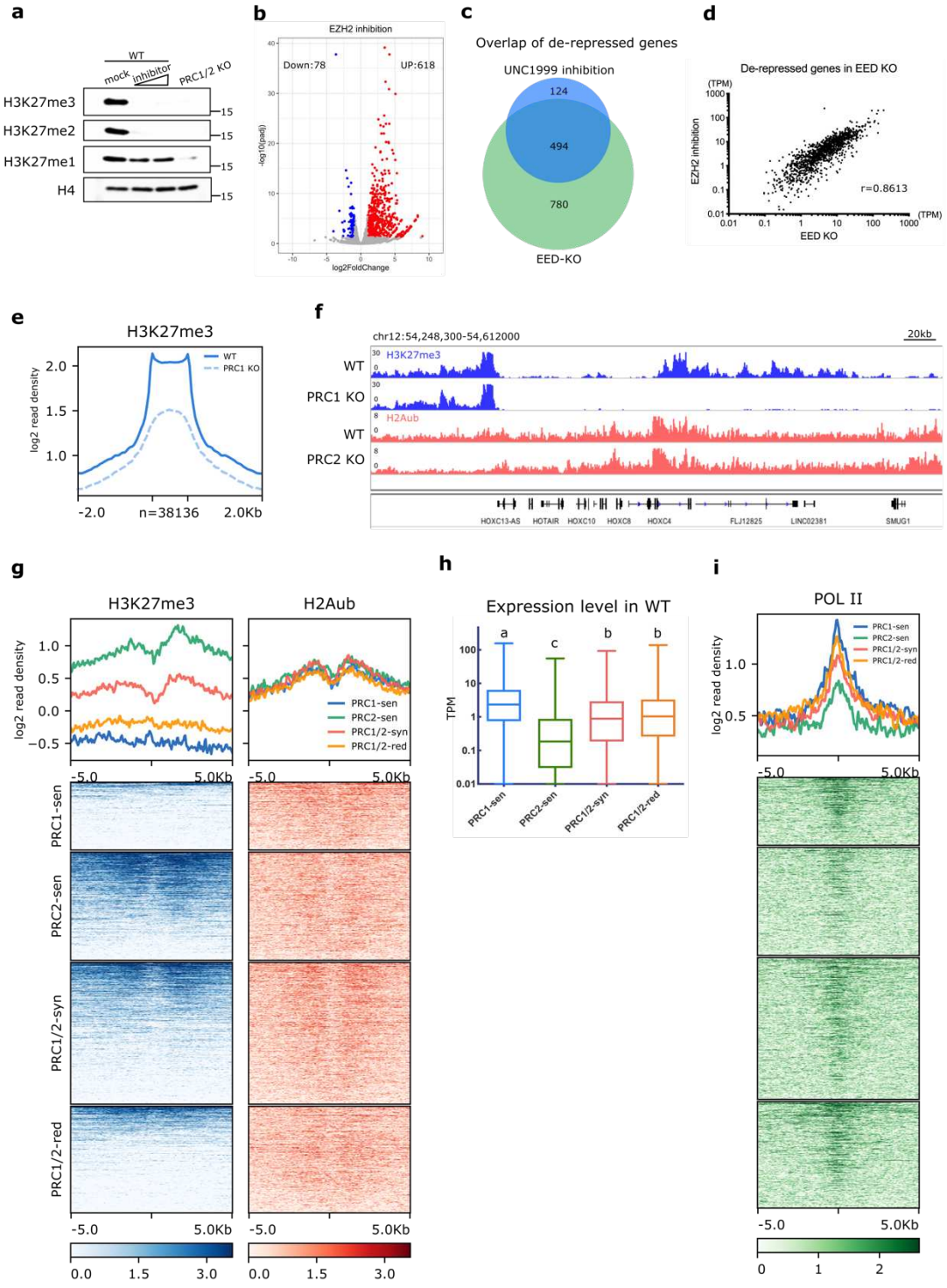
a
WT
mock
inhibitor
PRC1/2 KO
H3K27me3
H3K27me2
H3K27me1
H4
15
b
EZH2 inhibition
Down:78
UP:618
log2FoldChange
c
Overlap of de-repressed genes
UNC1999 inhibition
124
494
780
EED-KO
d
De-repressed genes in EED KO
EZH2 inhibition
EED KO
r=0.8613
(TPM)
e
H3K27me3
WT
PRC1 KO
log2 read density
-2.0
n=38136
2.0Kb
f
chr12:54,248,300-54,612000
20kb
WT
H3K27me3
PRC1 KO
H2Aub
WT
PRC2 KO
HOXC13-AS
HOTAIR
HOXC10
HOXC8
HOXC4
FLJ12825
LINC02381
SMUG1
g
H3K27me3
H2Aub
log2 read density
PRC1-sen
PRC2-sen
PRC1/2-syn
PRC1/2-red
-5.0
5.0Kb
h
Expression level in WT
TPM
a
c
b
b
PRC1-sen
PRC2-sen
PRC1/2-syn
PRC1/2-red
i
POL II
log2 read density
PRC1-sen
PRC2-sen
PRC1/2-syn
PRC1/2-red

Figure 2

Figure 2: PRC2-sensitive genes are enriched for H3K27me3 and are weakly expressed

(a) Western blot analysis of H3K27me3, H3K27me3 and H3K27me1 in WT cells after UNC1999 or UNC2400 treatment. Nuclear extract for PRC1/2-KO cells was used as negative control for background signal. H4 is used as loading control.

(b) Volcano plot showing the DEGs in cells treated with UNC1999 comparing to mock control. Red dots represented upregulated genes and blue dots represented downregulated genes.

(c) Venn diagram displaying overlap of DEGs in EZHs-inhibitor treated WT cells and PRC2-KO context.

(d) Plot showing correlation of expression level of the DEGs genes in PRC2-KO cells comparing inhibitor-treated cells *versus* PRC2-KO cells.

(e) Meta-plot displaying CUT&RUN-seq for H3K27me3 enrichment in WT and PRC1-KO cells.

(f) Genomic traces showing CUT&RUN-seq for H3K27me3 enrichment in WT and PRC1-KO cells, and H2Aub level in WT and PRC2-KO cells.

(g) Meta-plots and heatmaps illustrating H3K27me3 and H2Aub density in PRC1-sensitive, PRC2-sensitive, PRC1/2-redundant and PRC1/2-synergistic genes.

(h) Box plot illustrating the average expression levels of PRC1-sensitive, PRC2-sensitive, PRC1/2-redundant and PRC1/2-synergistic genes in WT cells. Statistical analysis was carried out by normality tests following Krustal-Wallis test.

(i) Meta-plots and heatmaps illustrating POL ii density in PRC1-sensitive, PRC2-sensitive, PRC1/2-redundant and PRC1/2-synergistic genes.

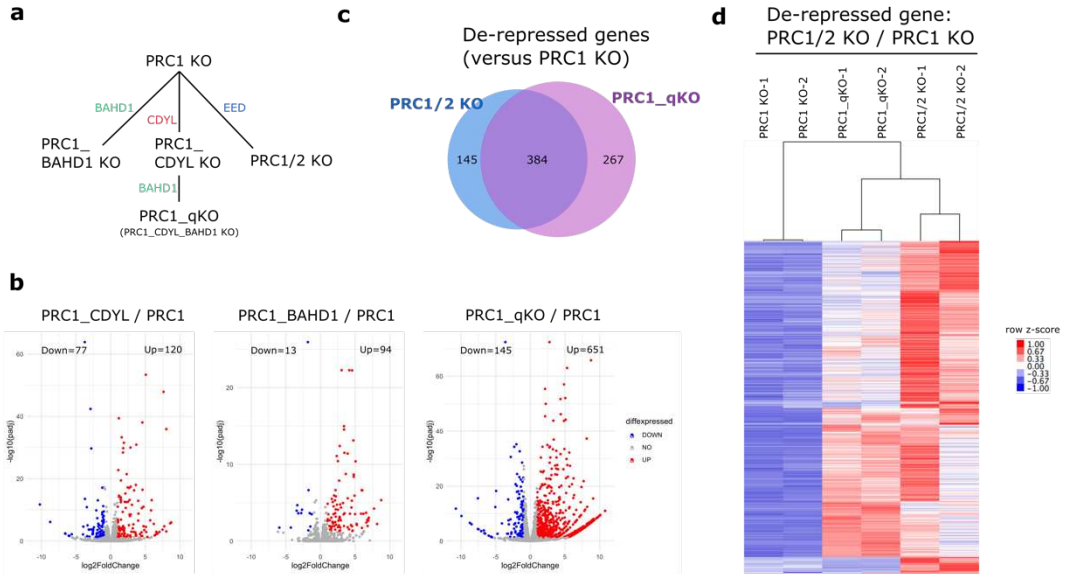

Figure 3

Figure 3: PRC1-independent readers of H3K27me3 could contribute to PRC2-mediated repression

(a) Schematic representation on strategy generating various KO cells.

(b) Volcano plots displaying DEGs in PRC1_CDYL-KO, PRC1-BAHD1-KO and PRC1_qKO comparing to PRC1-KO cells.

(c) Venn diagram displaying overlap of DEGs in the PRC1_qKO and PRC1/2-KO cells.

(d) Heatmap showing clustering of upregulated genes in PRC1/2-KO condition comparing to PRC1-KO condition. 2 biological replicates of each genotype (PRC1-KO, PRC1_qKO and PRC1/2-KO) were included.

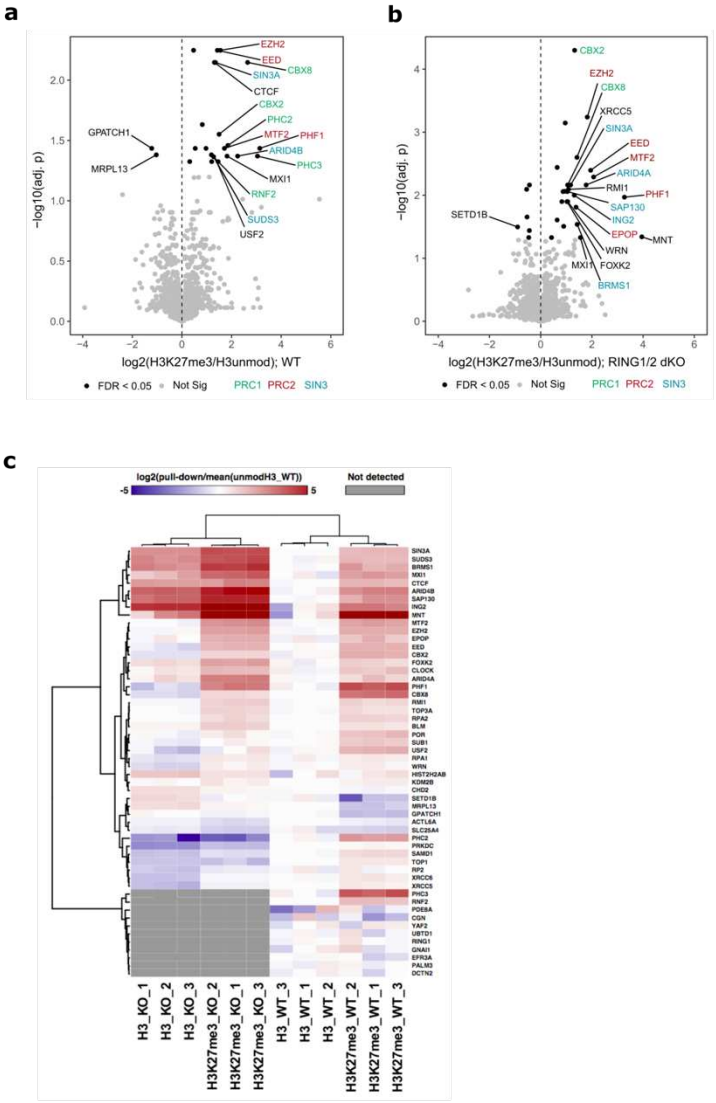

Figure 4

Figure 4: Search for unknown readers of H3K27me3

(a) MA-plots showing differential protein abundance enriched in H3K27me3-modified nucleosome pulldown comparing to unmodified counterparts. The experiments were performed in the WT and PRC1-KO contexts.

(b) Heatmap clustering displaying enrichment of proteins that were enriched in H3K27me3-modified nucleosome pulldown comparing to unmodified counterparts. The experiments were performed in the WT and PRC1-KO contexts.

Discussion

By integration of transcriptomic, proteomic and ubiquitomic analyses performed in various genetic knockout cell models, we consolidated that BAP1 exerts its function mainly through regulating H2A ubiquitination. Based on our CUT&RUN-seq, we identified that BAP1 is recruited predominantly to active enhancers highly enriched in POL ii, BRD4 and enhancer-related histone marks. We observed that loss of BAP1 introduced disproportional accumulation of H2Aub specifically enriched on BAP1-bound active enhancers. While BAP1 silencing did not affect the level H3K4me1, H3K27ac and POL II, it compromised the occupancy of BRD4 at its target loci. We hypothesized that BAP1 is likely to contribute to the maintenance of enhancer functionality (**Figure 13**).

Several aspects concerning BAP1's action at enhancers are yet remains largely unknown, further investigations would be of potential interest in elaborating the nature of BAP1.

(1) Driving forces of BAP1 recruitment to enhancers.

On one hand, we uncovered an unexpected hierarchical recruitment where CBP is generally required for BAP1 localization. However, the link between CBP and BAP1 is far from well-characterized. CBP and its paralog p300 catalyze acetylation on histone marks as well as TFs[329,385]. Besides, these paralogs harbor various domains that can interact with miscellaneous chromatin-associated factors, and thereby acting as a transcriptional hub to foster pervasive transcription[332]. In line with this idea, loss of CBP could triggers a cascade of effects that potentially impairs BAP1 recruitment. We showed that BAP1 is recruited to H3K27ac-enriched enhancers. However, whether H3K27ac *per se*, the consequence of H3K27ac deposition (e.g. increased chromatin accessibility, recruitment of other co-activators…etc.) or CBP itself is required for BAP1 recruitment requires further investigation. Artificial tether of CBP (WT or catalytically inactive form) to a given genomic loci using dCas9 system[306,307] could be of interest to dissect if CBP alone is sufficient to initiate *de novo* recruitment of BAP1. If, so, whether it requires the enzymatic activities.

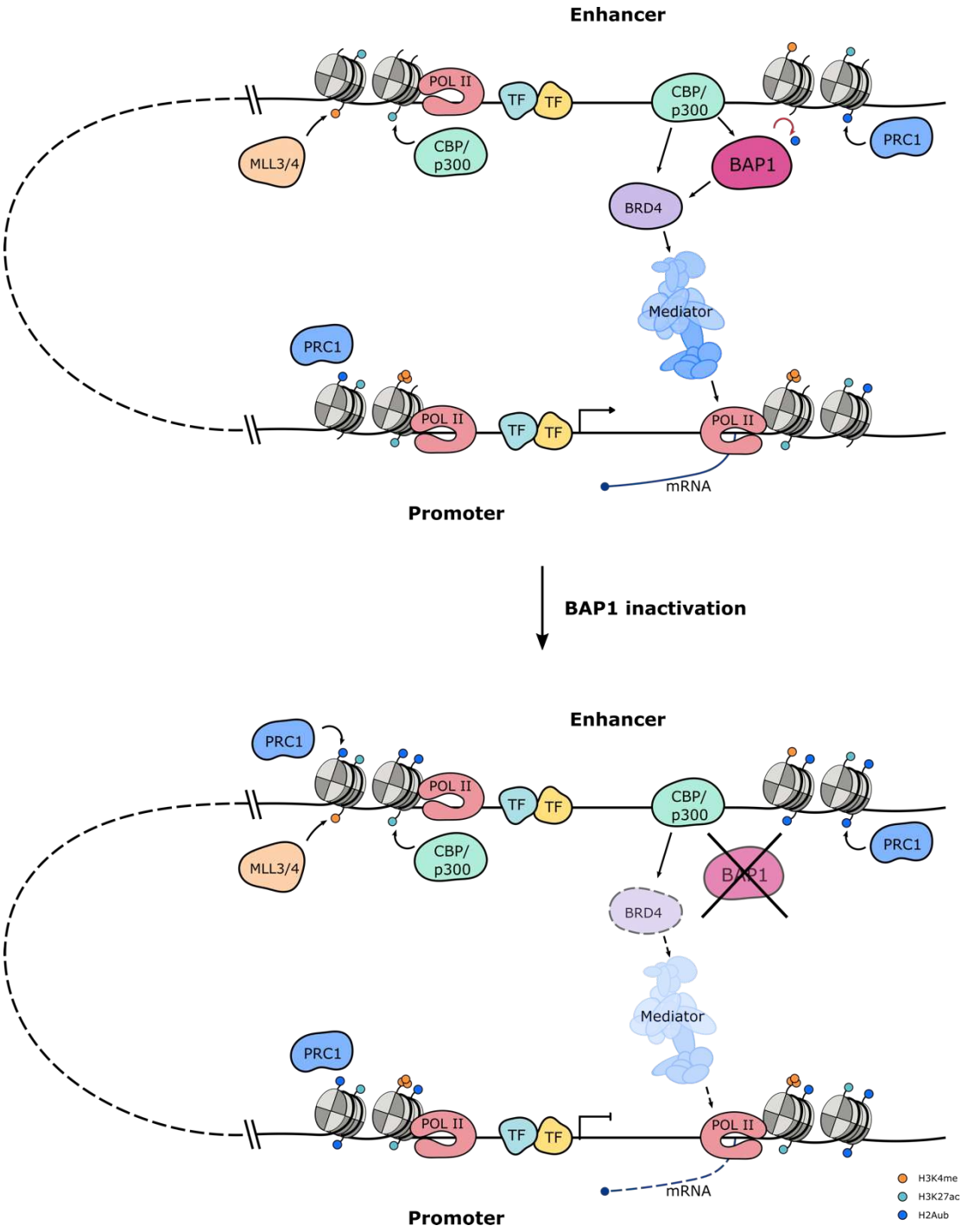

Figure 13. Summary of BAP1's action

On the other, BAP1's interacting partners might contain domains that confers it's recruitment to chromatin. ASXL1-3 contain PHD domains, a histone- or DNA-binding module[386], which could potentially drive BAP1 binding to DNA. BAP1 also interacts with TFs YY1 and FOXK1/2 that recognize specific DNA motifs[204,210]. But

so far there is lacking of supporting evidence and it is highly likely that they also have respective BAP1-independent functions[143,387].

In the interest to understand if BAP1 selectively binds to subsets of enhancers that harbor particular motifs. We performed motif discovery using HOMER[388] and compared BAP1-bound enhancers with unbound counterparts. We found that BAP1-bound enhancers were more enriched in NANOG, SOX3/6/10 binding motifs (data not shown). Interestingly, we did not find enrichment of FOXK1/2 motifs. However, whether this enrichment entails any biological significance requires more studies.

(2) BAP1-mediated enhancer regulations.

To this end, we observed that loss of BAP1 impaired integrity of co-activators demonstrated by reduced genomic occupancy and condensates formation. This perturbed enhancer function reflected on global gene downregulation. Apart from that, the consequences of BAP1 inactivation remain to be determined. Considering apparent reduction of MED1 condensates was observed and that BRD4 is crucial to recruit Mediator complex[118,345], it is rather foreseeable that BAP1 inactivation could give rise to reduced MED1 occupancy at enhancers. Following this rationale, we have performed CUT&RUN-seq for MED1 but failed to obtain robust result as we acquired presumably underestimated number of peaks (~200 sites, data not shown). This could potentially be ascribed to the nature of Mediator complex where direct contacts with DNA or histones are not common[358]. Alternative approach could be including the cross-linking step in the procedure to capture transient interactions such as ChIP-seq[362,363].

Aside from Mediator complex, BRD4 also recruits P-TEFb[349], a complex important for transcriptional elongation[19]. While POL II enrichment on BAP1-bound enhancers remained comparable in the absence of BAP1, we observed a decreased density at the TSSs and across the gene bodies of downregulated genes (data not shown). It is reported that POL II pause-release is the critical step controlling the frequency of transcription[8]. Taken together, I hypothesized that interrupted transcriptional elongation could be the main cause leading to large-scale gene silencing. Several non-exhaustive methods could be implemented to support this idea. For example, genome-wide profiling for CDK-9 kinase (subunit of p-TEFb)

and different forms of phosphorylated POL II[211,346]. In addition, live-cell imaging tracking of transcriptional events and POL II clusters could also be feasible to provide supportive evidence[389-391].

A couple of studies show that BAP1 inactivation does not decrease but actually increase genome-wide chromatin accessibility[211,220]. This is oddly counterintuitive as enhancer inactivation accompanied by eviction of co-activators often reduces chromatin accessibility[283]. If any to support this piece of observation, our proteomic assay suggested upregulation of chromatin remodeler BRG1 (SMARCA4) on protein level in the absence of BAP1 (data not shown). However, whether BRG1 also showed increased chromatin occupancy requires cautious validation. It could also be that this increased accessibility was derived from other chromatin factors. Nevertheless, it would be interesting to illustrate the link between loss of BAP1 (or H2Aub elevation) to increased accessibility.

(3) BAP1-bound enhancers and their cognate gene promoters

One important key to identify BAP1-regulated biological processes is to decipher which genes are BAP1-bound enhancers associated to. While it holds true that most enhancers controls expression of their proximal promoters[311], experimental approaches (i.e. Hi-CHIP[298]) enable systematic visualization of enhancer-promoter contacts and can accurately allocate "pairing" of enhancers to their cognate target promoters. Not only this technique can determine genes regulated BAP1-bound enhancers, it also provides the opportunity to scrutinize the dynamic of enhancer-promoter contacts upon loss of BAP1. For example, based on observed reduction in BRD4 and MED1 condensates, a hypothesis could be that BAP1 inactivation results in decreased promoter-enhancer contact frequency[371,392]. One interesting observation in our study is that while the majority of BAP1 peaks localize at the enhancers, RING1B is largely recruited to promoters but with substantial fraction to enhancers. By implementing Hi-ChIP technique to pair BAP1-bound enhancers with their target promoters, it would be intriguing to investigate if BAP1(enhancer)-RING1B(promoter) coordinate transcription of a particular gene, but in a distant fashion. Understanding these can deepen our knowledge on BAP1-regulated processes.

(4) Cell-type specific recruitment of BAP1?

In pluripotent mESCs, around 90 % of RING1B peaks is located at CpG island proximal to promoters[45,46]. Similarly, recent studies used HA- or Flag-tagged version of BAP1 to map its binding sites in mESCs, and reveal a strong enrichment at promoters while showing mild overlap with RING1B peaks[210,212]. While these results are not coherent in our and other's observations in HAP1 and CAL51 cells[214], it could be possible that BAP1 binding is re-distributed from promoter to enhancers in more differentiated cells or cancer cells, which is reminiscent of RING1B occupancy at enhancers observed in flies developing tissues and human breast cancer cells[175,176]. In the attempt to address this question, it could be worthwhile to re-characterize genome-wide BAP1 binding profile by CUT&RUN-seq in both mESCs and induced differentiating cells (such as NPCs).

(5) BAP1 and phase-separated condensates, super-enhancers and cell identity control

A recent study shows that reconstituted chromatin intrinsically undergoes phase separation in physiologic salt in cell nuclei, producing dense and dynamic droplets. These droplets then disassociate with the presence of H3K27ac, in coherence with the that acetylation renders a more open chromatin state[276]. Interestingly, BRD4 can form a new phase-separated state with droplets, which can be immiscible with unmodified chromatin droplets, mimicking nuclear chromatin subdomains[276]. Similar observation is reported in study on Mediator (MED1) condensates[371]. In our study, we discovered that BAP1 loss resulted in reduced co-activators condensates in size and in numbers, which potentially ascribes to H2Aub accumulation at the enhancers. Considering that the canonical PRC1 complex also has self-aggregation properties that forms distinct phase-separated Polycomb domains[393], it is speculative that increased PRC1 activities on enhancers would compete to assemble its own Polycomb condensates with co-activators which restrains the formation of co-activator domains, and consequently leading to a reduction of droplets. This hypothesis can be evaluated by observations on dynamics of BRD4 droplets formation using reconstituted chromatin with or without the presence of PRC1 complex. In parallel, as eviction of BRD4 is concomitantly observed in the absence of BAP1, it could also be interesting to understand if H2Aub interfere with the binding of BRD4 to acetylated nucleosomes using *in vitro* biochemical studies.

Reference

1. McGhee, J.D. & Felsenfeld, G. Nucleosome structure. *Annu Rev Biochem* **49**, 1115-56 (1980).
2. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol* **20**, 535-550 (2019).
3. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**, 292-301 (2001).
4. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
5. Nora, E.P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-5 (2012).
6. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* **111**, 996-1001 (2014).
7. Schuettengruber, B., Bourbon, H.M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34-57 (2017).
8. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45-54 (2019).
9. Lorch, Y. & Kornberg, R.D. Chromatin-remodeling for transcription. *Q Rev Biophys* **50**, e5 (2017).
10. Spitz, F. & Furlong, E.E. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613-26 (2012).
11. Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C. & Werner, M. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* **331**, 1451-4 (2011).
12. Malik, S. & Roeder, R.G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet* **11**, 761-72 (2010).
13. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**, 541-6 (2009).
14. Kostrewa, D. *et al.* RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323-30 (2009).

15. Kornberg, R.D. Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* **30**, 235-9 (2005).
16. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* **33**, 960-982 (2019).
17. Henriques, T. *et al.* Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol Cell* **52**, 517-28 (2013).
18. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet* **49**, 1045-1051 (2017).
19. Peterlin, B.M. & Price, D.H. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**, 297-305 (2006).
20. Vos, S.M. *et al.* Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607-612 (2018).
21. Jonkers, I., Kwak, H. & Lis, J.T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**, e02407 (2014).
22. Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* **16**, 190-202 (2015).
23. Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science* **330**, 612-6 (2010).
24. Attwood, J.T., Yung, R.L. & Richardson, B.C. DNA methylation and the regulation of gene transcription. *Cell Mol Life Sci* **59**, 241-57 (2002).
25. Kuroda, M.I., Kang, H., De, S. & Kassis, J.A. Dynamic Competition of Polycomb and Trithorax in Transcriptional Programming. *Annu Rev Biochem* **89**, 235-253 (2020).
26. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* **16**, 71-84 (2015).
27. Kassis, J.A., Kennison, J.A. & Tamkun, J.W. Polycomb and Trithorax Group Genes in Drosophila. *Genetics* **206**, 1699-1725 (2017).
28. Lewis, E.B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-70 (1978).
29. Whitcomb, S.J., Basu, A., Allis, C.D. & Bernstein, E. Polycomb Group proteins: an evolutionary perspective. *Trends Genet* **23**, 494-502 (2007).
30. Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by

Polycomb proteins. *Sci Adv* **1**, e1500737 (2015).

31. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873-8 (2004).
32. Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**, 845-54 (2005).
33. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-43 (2002).
34. Czermin, B. *et al.* Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185-96 (2002).
35. Sung, S. & Amasino, R.M. Vernalization and epigenetics: how plants remember winter. *Curr Opin Plant Biol* **7**, 4-10 (2004).
36. Frapporti, A. *et al.* The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in Paramecium. *Nat Commun* **10**, 2710 (2019).
37. Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**, 1513-24 (2008).
38. de Napoles, M. *et al.* Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* **7**, 663-76 (2004).
39. Morey, L. *et al.* Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47-62 (2012).
40. Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell* **32**, 503-18 (2008).
41. Stojic, L. *et al.* Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells. *Epigenetics Chromatin* **4**, 16 (2011).
42. Voncken, J.W. *et al.* Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A* **100**, 2468-73 (2003).
43. O'Carroll, D. *et al.* The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol* **21**, 4330-6 (2001).
44. Lee, T.I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-13 (2006).

45. Boyer, L.A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-53 (2006).
46. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242 (2008).
47. Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**, 1123-36 (2006).
48. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60 (2007).
49. Fursova, N.A. *et al.* Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol Cell* **74**, 1020-1036 e8 (2019).
50. Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* **32**, 491-502 (2008).
51. Riising, E.M. *et al.* Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* **55**, 347-60 (2014).
52. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* **24**, 265-76 (2010).
53. Bernstein, B.E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-26 (2006).
54. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134-48 (2013).
55. Wang, L. *et al.* Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* **14**, 637-46 (2004).
56. Min, J., Zhang, Y. & Xu, R.M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* **17**, 1823-8 (2003).
57. Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* **21**, 569-71 (2014).
58. Cooper, S. *et al.* Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat Commun* **7**, 13661 (2016).
59. Cui, K. *et al.* Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* **4**, 80-93 (2009).

60. Ye, L. *et al.* Histone demethylases KDM4B and KDM6B promotes osteogenic differentiation of human MSCs. *Cell Stem Cell* **11**, 50-61 (2012).
61. Ezhkova, E. *et al.* EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair. *Genes Dev* **25**, 485-98 (2011).
62. Cohen, I. *et al.* PRC1 Fine-tunes Gene Repression and Activation to Safeguard Skin Development and Stem Cell Specification. *Cell Stem Cell* **22**, 726-739 e7 (2018).
63. Endoh, M. *et al.* Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. *PLoS Genet* **8**, e1002774 (2012).
64. Blackledge, N.P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445-1459 (2014).
65. Zhang, Z. *et al.* Role of remodeling and spacing factor 1 in histone H2A ubiquitination-mediated gene silencing. *Proc Natl Acad Sci U S A* **114**, E7949-E7958 (2017).
66. Nakagawa, T. *et al.* Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes Dev* **22**, 37-49 (2008).
67. Stock, J.K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**, 1428-35 (2007).
68. Francis, N.J., Kingston, R.E. & Woodcock, C.L. Chromatin compaction by a polycomb group protein complex. *Science* **306**, 1574-7 (2004).
69. Illingworth, R.S. *et al.* The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. *Genes Dev* **29**, 1897-902 (2015).
70. Gao, Z. *et al.* PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* **45**, 344-56 (2012).
71. Blackledge, N.P., Rose, N.R. & Klose, R.J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol* **16**, 643-649 (2015).
72. Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-78 (2012).

73. Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* **38**, 452-64 (2010).
74. Isono, K. *et al.* SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell* **26**, 565-77 (2013).
75. Grau, D.J. *et al.* Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev* **25**, 2210-21 (2011).
76. Plys, A.J. *et al.* Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev* **33**, 799-813 (2019).
77. Tatavosian, R. *et al.* Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *J Biol Chem* **294**, 1451-1463 (2019).
78. Huseyin, M.K. & Klose, R.J. Live-cell single particle tracking of PRC1 reveals a highly dynamic system with low target site occupancy. *Nat Commun* **12**, 887 (2021).
79. Hauri, S. *et al.* A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep* **17**, 583-595 (2016).
80. Arrigoni, R. *et al.* The Polycomb-associated protein Rybp is a ubiquitin binding protein. *FEBS Lett* **580**, 6233-41 (2006).
81. Rose, N.R. *et al.* RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. *Elife* **5**(2016).
82. Mendenhall, E.M. *et al.* GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* **6**, e1001244 (2010).
83. Farcas, A.M. *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* **1**, e00205 (2012).
84. Wu, X., Johansen, J.V. & Helin, K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol Cell* **49**, 1134-46 (2013).
85. He, J. *et al.* Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat Cell Biol* **15**, 373-84 (2013).
86. Almeida, M. *et al.* PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081-1084 (2017).

87. Endoh, M. *et al.* PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *Elife* **6**(2017).
88. Scelfo, A. *et al.* Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities. *Mol Cell* **74**, 1037-1052 e7 (2019).
89. Pengelly, A.R., Kalb, R., Finkl, K. & Muller, J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev* **29**, 1487-92 (2015).
90. Tsuboi, M. *et al.* Ubiquitination-Independent Repression of PRC1 Targets during Neuronal Fate Restriction in the Developing Mouse Neocortex. *Dev Cell* **47**, 758-772 e5 (2018).
91. Blackledge, N.P. *et al.* PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol Cell* **77**, 857-874 e9 (2020).
92. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**, 2893-905 (2002).
93. Muller, J. *et al.* Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**, 197-208 (2002).
94. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-9 (2011).
95. Oksuz, O. *et al.* Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol Cell* **70**, 1149-1162 e5 (2018).
96. Ferrari, K.J. *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol Cell* **53**, 49-62 (2014).
97. Pengelly, A.R., Copur, O., Jackle, H., Herzig, A. & Muller, J. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science* **339**, 698-9 (2013).
98. Bracken, A.P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* **22**, 5323-35 (2003).
99. Bae, W.K. *et al.* The methyltransferases enhancer of zeste homolog (EZH) 1 and EZH2 control hepatocyte homeostasis and regeneration. *FASEB J* **29**, 1653-62 (2015).

100. Mu, W., Starmer, J., Shibata, Y., Yee, D. & Magnuson, T. EZH1 in germ cells safeguards the function of PRC2 during spermatogenesis. *Dev Biol* **424**, 198-207 (2017).
101. Grau, D. *et al.* Structures of monomeric and dimeric PRC2:EZH1 reveal flexible modules involved in chromatin compaction. *Nat Commun* **12**, 714 (2021).
102. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762-7 (2009).
103. Schumacher, A., Faust, C. & Magnuson, T. Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature* **384**, 648 (1996).
104. Montgomery, N.D. *et al.* The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr Biol* **15**, 942-7 (2005).
105. Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J* **23**, 4061-71 (2004).
106. Rai, A.N. *et al.* Elements of the polycomb repressor SU(Z)12 needed for histone H3-K27 methylation, the interface with E(Z), and in vivo function. *Mol Cell Biol* **33**, 4844-56 (2013).
107. Hojfeldt, J.W. *et al.* Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nat Struct Mol Biol* **25**, 225-232 (2018).
108. Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* **15**, 57-67 (2004).
109. Holoch, D. & Margueron, R. Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends Biochem Sci* **42**, 531-542 (2017).
110. van Mierlo, G., Veenstra, G.J.C., Vermeulen, M. & Marks, H. The Complexity of PRC2 Subcomplexes. *Trends Cell Biol* **29**, 660-671 (2019).
111. Perino, M. *et al.* MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat Genet* **50**, 1002-1010 (2018).
112. Casanova, M. *et al.* Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. *Development* **138**, 1471-82 (2011).
113. Walker, E. *et al.* Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal

and differentiation. *Cell Stem Cell* **6**, 153-66 (2010).

114. Ballare, C. *et al.* Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat Struct Mol Biol* **19**, 1257-65 (2012).
115. Brien, G.L. *et al.* Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat Struct Mol Biol* **19**, 1273-81 (2012).
116. Cao, R. *et al.* Role of hPHF1 in H3K27 methylation and Hox gene silencing. *Mol Cell Biol* **28**, 1862-72 (2008).
117. Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V. & Reinberg, D. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol* **28**, 2718-31 (2008).
118. Choi, J. *et al.* DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. *Nat Struct Mol Biol* **24**, 1039-1047 (2017).
119. Cai, L. *et al.* An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol Cell* **49**, 571-82 (2013).
120. Alekseyenko, A.A., Gorchakov, A.A., Kharchenko, P.V. & Kuroda, M.I. Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc Natl Acad Sci U S A* **111**, 2488-93 (2014).
121. Okumura, F., Matsuzaki, M., Nakatsukasa, K. & Kamura, T. The Role of Elongin BC-Containing Ubiquitin Ligases. *Front Oncol* **2**, 10 (2012).
122. Liefke, R., Karwacki-Neisius, V. & Shi, Y. EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape. *Mol Cell* **64**, 659-672 (2016).
123. Beringer, M. *et al.* EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. *Mol Cell* **64**, 645-658 (2016).
124. Conway, E. *et al.* A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. *Mol Cell* **70**, 408-421 e8 (2018).
125. Kim, H., Ekram, M.B., Bakshi, A. & Kim, J. AEBP2 as a transcriptional activator and its role in cell migration. *Genomics* **105**, 108-15 (2015).
126. Grijzenhout, A. *et al.* Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. *Development* **143**, 2716-23 (2016).
127. Takeuchi, T. *et al.* Gene trap capture of a novel mouse gene, jumonji,

required for neural tube formation. *Genes Dev* **9**, 1211-22 (1995).

128. Son, J., Shen, S.S., Margueron, R. & Reinberg, D. Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev* **27**, 2663-77 (2013).
129. Sanulli, S. *et al.* Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol Cell* **57**, 769-783 (2015).
130. Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306-10 (2010).
131. Perino, M. *et al.* Two Functional Axes of Feedback-Enforced PRC2 Recruitment in Mouse Embryonic Stem Cells. *Stem Cell Reports* **15**, 1287-1300 (2020).
132. Healy, E. *et al.* PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol Cell* **76**, 437-452 e6 (2019).
133. Hojfeldt, J.W. *et al.* Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. *Mol Cell* **76**, 423-436 e3 (2019).
134. Petracovici, A. & Bonasio, R. Distinct PRC2 subunits regulate maintenance and establishment of Polycomb repression during differentiation. *Mol Cell* (2021).
135. Sengupta, A.K., Kuhrs, A. & Muller, J. General transcriptional silencing by a Polycomb response element in Drosophila. *Development* **131**, 1959-65 (2004).
136. Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.L. & Kassis, J.A. The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell* **1**, 1057-64 (1998).
137. Klymenko, T. *et al.* A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev* **20**, 1110-22 (2006).
138. Peterson, A.J. *et al.* Requirement for sex comb on midleg protein interactions in Drosophila polycomb group repression. *Genetics* **167**, 1225-39 (2004).
139. Mohd-Sarip, A., Venturini, F., Chalkley, G.E. & Verrijzer, C.P. Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. *Mol Cell Biol* **22**, 7473-83 (2002).

140. Gaytan de Ayala Alonso, A. *et al.* A genetic screen identifies novel polycomb group genes in Drosophila. *Genetics* **176**, 2099-108 (2007).
141. Scheuermann, J.C. *et al.* Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**, 243-7 (2010).
142. Kassis, J.A. & Brown, J.L. Polycomb group response elements in Drosophila and vertebrates. *Adv Genet* **81**, 83-118 (2013).
143. Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T. & Pasini, D. Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res* **40**, 3403-18 (2012).
144. Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470-84 (2010).
145. Long, H.K. *et al.* Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife* **2**, e00348 (2013).
146. Jermann, P., Hoerner, L., Burger, L. & Schubeler, D. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proc Natl Acad Sci U S A* **111**, E3415-21 (2014).
147. Boulard, M., Edwards, J.R. & Bestor, T.H. FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat Genet* **47**, 479-85 (2015).
148. Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D.M. & Nakatani, Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296**, 1132-6 (2002).
149. Brockdorff, N. Polycomb complexes in X chromosome inactivation. *Philos Trans R Soc Lond B Biol Sci* **372**(2017).
150. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-6 (2008).
151. da Rocha, S.T. *et al.* Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell* **53**, 301-16 (2014).
152. Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D. & Heard, E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644-9 (2004).
153. Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J* **25**, 3110-22

(2006).

154. Yu, J.R., Lee, C.H., Oksuz, O., Stafford, J.M. & Reinberg, D. PRC2 is high maintenance. *Genes Dev* **33**, 903-935 (2019).
155. Kharchenko, P.V. *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480-5 (2011).
156. Schmitges, F.W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-41 (2011).
157. Yuan, W. *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem* **286**, 7983-7989 (2011).
158. Dorighi, K.M. & Tamkun, J.W. The trithorax group proteins Kismet and ASH1 promote H3K36 dimethylation to counteract Polycomb group repression in Drosophila. *Development* **140**, 4182-92 (2013).
159. Streubel, G. *et al.* The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol Cell* **70**, 371-379 e5 (2018).
160. Lynch, M.D. *et al.* An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J* **31**, 317-29 (2012).
161. Li, Y. *et al.* Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biol* **19**, 18 (2018).
162. Almeida, M., Bowness, J.S. & Brockdorff, N. The many faces of Polycomb regulation by RNA. *Curr Opin Genet Dev* **61**, 53-61 (2020).
163. Beltran, M. *et al.* G-tract RNA removes Polycomb repressive complex 2 from genes. *Nat Struct Mol Biol* **26**, 899-909 (2019).
164. Beltran, M. *et al.* The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res* **26**, 896-907 (2016).
165. Kaneko, S., Son, J., Bonasio, R., Shen, S.S. & Reinberg, D. Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes Dev* **28**, 1983-8 (2014).
166. Zhang, Q. *et al.* RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nat Struct Mol Biol* **26**, 237-247 (2019).
167. Ragazzini, R. *et al.* EZHIP constrains Polycomb Repressive Complex 2 activity in germ cells. *Nat Commun* **10**, 3858 (2019).
168. Antin, C. *et al.* EZHIP is a specific diagnostic biomarker for posterior fossa ependymomas, group PFA and diffuse midline gliomas H3-WT with EZHIP

overexpression. *Acta Neuropathol Commun* **8**, 183 (2020).

169. Hubner, J.M. *et al.* EZHIP/CXorf67 mimics K27M mutated oncohistones and functions as an intrinsic inhibitor of PRC2 function in aggressive posterior fossa ependymoma. *Neuro Oncol* **21**, 878-889 (2019).
170. Jain, S.U. *et al.* PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism. *Nat Commun* **10**, 2146 (2019).
171. Kahn, T.G. *et al.* Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. *Nucleic Acids Res* **44**, 10132-10149 (2016).
172. Mei, H. *et al.* H2AK119ub1 guides maternal inheritance and zygotic deposition of H3K27me3 in mouse embryos. *Nat Genet* **53**, 539-550 (2021).
173. Chen, Z., Djekidel, M.N. & Zhang, Y. Distinct dynamics and functions of H2AK119ub1 and H3K27me3 in mouse preimplantation embryos. *Nat Genet* **53**, 551-563 (2021).
174. Kloet, S.L. *et al.* The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol* **23**, 682-690 (2016).
175. Loubiere, V., Papadopoulos, G.L., Szabo, Q., Martinez, A.M. & Cavalli, G. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci Adv* **6**, eaax4001 (2020).
176. Chan, H.L. *et al.* Polycomb complexes associate with enhancers and promote oncogenic transcriptional programs in cancer through multiple mechanisms. *Nat Commun* **9**, 3377 (2018).
177. Cohen, I. *et al.* PRC1 preserves epidermal tissue integrity independently of PRC2. *Genes Dev* **33**, 55-60 (2019).
178. Vermeulen, M. *et al.* Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967-80 (2010).
179. Caron, C. *et al.* Cdyl: a new transcriptional co-repressor. *EMBO Rep* **4**, 877-82 (2003).
180. Mulligan, P. *et al.* CDYL bridges REST and histone methyltransferases for gene repression and suppression of cellular transformation. *Mol Cell* **32**, 718-26 (2008).
181. Zhang, Y. *et al.* Corepressor protein CDYL functions as a molecular bridge between polycomb repressor complex 2 and repressive chromatin mark trimethylated histone lysine 27. *J Biol Chem* **286**, 42414-42425 (2011).

182. Liu, Y. *et al.* Chromodomain protein CDYL is required for transmission/restoration of repressive histone marks. *J Mol Cell Biol* **9**, 178-194 (2017).
183. Escamilla-Del-Arenal, M. *et al.* Cdyl, a new partner of the inactive X chromosome and potential reader of H3K27me3 and H3K9me2. *Mol Cell Biol* **33**, 5005-20 (2013).
184. Bierne, H. *et al.* Human BAHD1 promotes heterochromatic gene silencing. *Proc Natl Acad Sci U S A* **106**, 13826-31 (2009).
185. Zhao, D. *et al.* The BAH domain of BAHD1 is a histone H3K27me3 reader. *Protein Cell* **7**, 222-6 (2016).
186. Fan, H. *et al.* BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. *Nat Genet* **52**, 1384-1396 (2020).
187. Fan, H. *et al.* A conserved BAH module within mammalian BAHD1 connects H3K27me3 to Polycomb gene silencing. *Nucleic Acids Res* **49**, 4441-4455 (2021).
188. Lakisic, G. *et al.* Role of the BAHD1 Chromatin-Repressive Complex in Placental Development and Regulation of Steroid Metabolism. *PLoS Genet* **12**, e1005898 (2016).
189. Wiles, E.T. *et al.* Evolutionarily ancient BAH-PHD protein mediates Polycomb silencing. *Proc Natl Acad Sci U S A* **117**, 11614-11623 (2020).
190. Cohen, I. *et al.* Polycomb complexes redundantly maintain epidermal stem cell identity during development. *Genes Dev* **35**, 354-366 (2021).
191. Jensen, D.E. *et al.* BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* **16**, 1097-112 (1998).
192. Mallery, D.L., Vandenberg, C.J. & Hiom, K. Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *EMBO J* **21**, 6755-62 (2002).
193. Ventii, K.H. *et al.* BRCA1-associated protein-1 is a tumor suppressor that requires deubiquitinating activity and nuclear localization. *Cancer Res* **68**, 6953-62 (2008).
194. Sinclair, D.A. *et al.* The Additional sex combs gene of Drosophila encodes a chromatin protein that binds to shared and unique Polycomb group sites on polytene chromosomes. *Development* **125**, 1207-16 (1998).
195. Milne, T.A., Sinclair, D.A. & Brock, H.W. The Additional sex combs gene of

Drosophila is required for activation and repression of homeotic loci, and interacts specifically with Polycomb and super sex combs. *Mol Gen Genet* **261**, 753-61 (1999).

196. Fisher, C.L., Berger, J., Randazzo, F. & Brock, H.W. A human homolog of Additional sex combs, ADDITIONAL SEX COMBS-LIKE 1, maps to chromosome 20q11. *Gene* **306**, 115-26 (2003).
197. Katoh, M. & Katoh, M. Identification and characterization of ASXL2 gene in silico. *Int J Oncol* **23**, 845-50 (2003).
198. Katoh, M. & Katoh, M. Identification and characterization of ASXL3 gene in silico. *Int J Oncol* **24**, 1617-22 (2004).
199. Fisher, C.L., Randazzo, F., Humphries, R.K. & Brock, H.W. Characterization of Asxl1, a murine homolog of Additional sex combs, and analysis of the Asx-like gene family. *Gene* **369**, 109-18 (2006).
200. Daou, S. *et al.* The BAP1/ASXL2 Histone H2A Deubiquitinase Complex Regulates Cell Proliferation and Is Disrupted in Cancer. *J Biol Chem* **290**, 28643-63 (2015).
201. Campagne, A. *et al.* BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation. *Nat Commun* **10**, 348 (2019).
202. Daou, S. *et al.* Monoubiquitination of ASXLs controls the deubiquitinase activity of the tumor suppressor BAP1. *Nat Commun* **9**, 4385 (2018).
203. Ji, Z. *et al.* The forkhead transcription factor FOXK2 acts as a chromatin targeting factor for the BAP1-containing histone deubiquitinase complex. *Nucleic Acids Res* **42**, 6232-42 (2014).
204. Yu, H. *et al.* The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol Cell Biol* **30**, 5071-85 (2010).
205. Baymaz, H.I. *et al.* MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics* **14**, 2179-89 (2014).
206. Machida, Y.J., Machida, Y., Vashisht, A.A., Wohlschlegel, J.A. & Dutta, A. The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *J Biol Chem* **284**, 34179-88 (2009).
207. Dey, A. *et al.* Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science* **337**, 1541-6 (2012).
208. Abdel-Wahab, O. *et al.* Deletion of Asxl1 results in myelodysplasia and

severe developmental defects in vivo. *J Exp Med* **210**, 2641-59 (2013).

209. Abdel-Wahab, O. *et al.* ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell* **22**, 180-93 (2012).
210. Kolovos, P. *et al.* PR-DUB maintains the expression of critical genes through FOXK1/2- and ASXL1/2/3-dependent recruitment to chromatin and H2AK119ub1 deubiquitination. *Genome Res* **30**, 1119-1130 (2020).
211. Fursova, N.A. *et al.* BAP1 constrains pervasive H2AK119ub1 to control the transcriptional potential of the genome. *Genes Dev* (2021).
212. Conway, E. *et al.* BAP1 enhances Polycomb repression by counteracting widespread H2AK119ub1 deposition and chromatin condensation. *Mol Cell* (2021).
213. Kuznetsov, J.N. *et al.* BAP1 regulates epigenetic switch from pluripotency to differentiation in developmental lineages giving rise to BAP1-mutant cancers. *Sci Adv* **5**, eaax1738 (2019).
214. Wang, L. *et al.* Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. *Nat Med* **24**, 758-769 (2018).
215. Skene, P.J., Henikoff, J.G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019 (2018).
216. Lee, H.S., Lee, S.A., Hur, S.K., Seo, J.W. & Kwon, J. Stabilization and targeting of INO80 to replication forks by BAP1 during normal DNA synthesis. *Nat Commun* **5**, 5128 (2014).
217. Qin, J. *et al.* BAP1 promotes breast cancer cell proliferation and metastasis by deubiquitinating KLF5. *Nat Commun* **6**, 8471 (2015).
218. Zarrizi, R., Menard, J.A., Belting, M. & Massoumi, R. Deubiquitination of gamma-tubulin by BAP1 prevents chromosome instability in breast cancer cells. *Cancer Res* **74**, 6499-508 (2014).
219. Bononi, A. *et al.* BAP1 regulates IP3R3-mediated Ca(2+) flux to mitochondria suppressing cell transformation. *Nature* **546**, 549-553 (2017).
220. Artegiani, B. *et al.* Probing the Tumor Suppressor Function of BAP1 in CRISPR-Engineered Human Liver Organoids. *Cell Stem Cell* **24**, 927-943 e6 (2019).
221. Baughman, J.M. *et al.* NeuCode Proteomics Reveals Bap1 Regulation of

Metabolism. *Cell Rep* **16**, 583-595 (2016).

222. Abdel-Rahman, M.H. *et al.* Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma, and other cancers. *J Med Genet* **48**, 856-9 (2011).
223. Rai, K., Pilarski, R., Cebulla, C.M. & Abdel-Rahman, M.H. Comprehensive review of BAP1 tumor predisposition syndrome with report of two new cases. *Clin Genet* **89**, 285-94 (2016).
224. Carbone, M. *et al.* BAP1 and cancer. *Nat Rev Cancer* **13**, 153-9 (2013).
225. Carvajal, R.D. *et al.* Metastatic disease from uveal melanoma: treatment options and future prospects. *Br J Ophthalmol* **101**, 38-44 (2017).
226. Khoja, L. *et al.* Meta-analysis in metastatic uveal melanoma to determine progression free and overall survival benchmarks: an international rare cancers initiative (IRCI) ocular melanoma study. *Ann Oncol* **30**, 1370-1380 (2019).
227. Martin, M. *et al.* Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nat Genet* **45**, 933-6 (2013).
228. Harbour, J.W. *et al.* Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330**, 1410-3 (2010).
229. Yang, J., Manson, D.K., Marr, B.P. & Carvajal, R.D. Treatment of uveal melanoma: where are we now? *Ther Adv Med Oncol* **10**, 1758834018757175 (2018).
230. Bejar, R. *et al.* Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med* **364**, 2496-506 (2011).
231. Waters, A.M. & Der, C.J. KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer. *Cold Spring Harb Perspect Med* **8**(2018).
232. Perkail, S., Andricovich, J., Kai, Y. & Tzatsos, A. BAP1 is a haploinsufficient tumor suppressor linking chronic pancreatitis to pancreatic cancer in mice. *Nat Commun* **11**, 3018 (2020).
233. Lee, H.J. *et al.* The Tumor Suppressor BAP1 Regulates the Hippo Pathway in Pancreatic Ductal Adenocarcinoma. *Cancer Res* **80**, 1656-1668 (2020).
234. Gelsi-Boyer, V. *et al.* Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. *J Hematol Oncol* **5**, 12 (2012).
235. Abdel-Wahab, O. *et al.* Concomitant analysis of EZH2 and ASXL1 mutations

in myelofibrosis, chronic myelomonocytic leukemia and blast-phase myeloproliferative neoplasms. *Leukemia* **25**, 1200-2 (2011).

236. Inoue, D. *et al.* Truncation mutants of ASXL1 observed in myeloid malignancies are expressed at detectable protein levels. *Exp Hematol* **44**, 172-6 e1 (2016).
237. Balasubramani, A. *et al.* Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1-BAP1 complex. *Nat Commun* **6**, 7307 (2015).
238. Asada, S. *et al.* Mutant ASXL1 cooperates with BAP1 to promote myeloid leukaemogenesis. *Nat Commun* **9**, 2733 (2018).
239. Micol, J.B. *et al.* Frequent ASXL2 mutations in acute myeloid leukemia patients with t(8;21)/RUNX1-RUNX1T1 chromosomal translocations. *Blood* **124**, 1445-9 (2014).
240. Hoischen, A. *et al.* De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat Genet* **43**, 729-31 (2011).
241. Bainbridge, M.N. *et al.* De novo truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. *Genome Med* **5**, 11 (2013).
242. Kuechler, A. *et al.* Bainbridge-Ropers syndrome caused by loss-of-function variants in ASXL3: a recognizable condition. *Eur J Hum Genet* **25**, 183-191 (2017).
243. Ingham, P.W. trithorax and the regulation of homeotic gene expression in Drosophila: a historical perspective. *Int J Dev Biol* **42**, 423-9 (1998).
244. Kingston, R.E. & Tamkun, J.W. Transcriptional regulation by trithorax-group proteins. *Cold Spring Harb Perspect Biol* **6**, a019349 (2014).
245. Schuettengruber, B., Martinez, A.M., Iovino, N. & Cavalli, G. Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* **12**, 799-814 (2011).
246. Heintzman, N.D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-12 (2009).
247. Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem* **81**, 65-95 (2012).
248. Wu, M. *et al.* Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol* **28**, 7337-44 (2008).

249. Denissov, S. *et al.* Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141**, 526-37 (2014).
250. Wang, P. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol* **29**, 6074-85 (2009).
251. Sze, C.C. & Shilatifard, A. MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer. *Cold Spring Harb Perspect Med* **6**(2016).
252. Ardehali, M.B. *et al.* Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *EMBO J* **30**, 2817-28 (2011).
253. Mohan, M. *et al.* The COMPASS family of H3K4 methylases in Drosophila. *Mol Cell Biol* **31**, 4310-8 (2011).
254. Tanaka, Y., Katagiri, Z., Kawahashi, K., Kioussis, D. & Kitajima, S. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene* **397**, 161-8 (2007).
255. Centore, R.C., Sandoval, G.J., Soares, L.M.M., Kadoch, C. & Chan, H.M. Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. *Trends Genet* **36**, 936-950 (2020).
256. Clapier, C.R., Iwasa, J., Cairns, B.R. & Peterson, C.L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407-422 (2017).
257. Kwon, H., Imbalzano, A.N., Khavari, P.A., Kingston, R.E. & Green, M.R. Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* **370**, 477-81 (1994).
258. Wang, W. *et al.* Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *EMBO J* **15**, 5370-82 (1996).
259. Kadoch, C. & Crabtree, G.R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv* **1**, e1500447 (2015).
260. Valencia, A.M. & Kadoch, C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. *Nat Cell Biol* **21**, 152-161 (2019).
261. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat*

Genet **45**, 592-601 (2013).

262. Bultman, S. *et al.* A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol Cell* **6**, 1287-95 (2000).
263. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* **106**, 5181-6 (2009).
264. Ho, L. *et al.* esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat Cell Biol* **13**, 903-13 (2011).
265. Kadoch, C. *et al.* Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat Genet* **49**, 213-222 (2017).
266. Iurlaro, M. *et al.* Mammalian SWI/SNF continuously restores local accessibility to chromatin. *Nat Genet* **53**, 279-287 (2021).
267. Schick, S. *et al.* Acute BAF perturbation causes immediate changes in chromatin accessibility. *Nat Genet* **53**, 269-278 (2021).
268. Bouazoune, K. & Kingston, R.E. Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders. *Proc Natl Acad Sci U S A* **109**, 19238-43 (2012).
269. Srinivasan, S., Dorighi, K.M. & Tamkun, J.W. Drosophila Kismet regulates histone H3 lysine 27 methylation and early elongation by RNA polymerase II. *PLoS Genet* **4**, e1000217 (2008).
270. Zabidi, M.A. & Stark, A. Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet* **32**, 801-814 (2016).
271. Field, A. & Adelman, K. Evaluating Enhancer Function and Transcription. *Annu Rev Biochem* **89**, 213-234 (2020).
272. Long, H.K., Prescott, S.L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170-1187 (2016).
273. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**, 272-86 (2014).
274. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299-308 (1981).

275. Schaffner, W. Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem* **396**, 311-27 (2015).
276. Gibson, B.A. *et al.* Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484 e21 (2019).
277. Donaghey, J. *et al.* Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat Genet* **50**, 250-258 (2018).
278. Fernandez Garcia, M. *et al.* Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol Cell* **75**, 921-932 e6 (2019).
279. Heintzman, N.D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-8 (2007).
280. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-6 (2010).
281. Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-7 (2010).
282. De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* **8**, e1000384 (2010).
283. Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet* **21**, 71-87 (2020).
284. Catarino, R.R. & Stark, A. Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev* **32**, 202-223 (2018).
285. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-8 (2009).
286. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-83 (2011).
287. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-20 (2012).
288. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
289. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
290. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).

291. Thurman, R.E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82 (2012).
292. Cusanovich, D.A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324 e18 (2018).
293. Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R. & Lieb, J.D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**, 877-85 (2007).
294. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-8 (2008).
295. Kwak, H., Fuda, N.J., Core, L.J. & Lis, J.T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950-3 (2013).
296. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461 (2014).
297. Li, G. *et al.* Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. *BMC Genomics* **15 Suppl 12**, S11 (2014).
298. Mumbach, M.R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**, 919-922 (2016).
299. Halfon, M.S. Studying Transcriptional Enhancers: The Founder Fallacy, Validation Creep, and Other Biases. *Trends Genet* **35**, 93-103 (2019).
300. Arnold, C.D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-7 (2013).
301. Stampfel, G. *et al.* Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**, 147-51 (2015).
302. Diao, Y. *et al.* A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* **14**, 629-635 (2017).
303. Thakore, P.I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143-9 (2015).
304. Fulco, C.P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769-773 (2016).

305. Kearns, N.A. *et al.* Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* **12**, 401-403 (2015).
306. Hilton, I.B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510-7 (2015).
307. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G.C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol Cell* **66**, 285-299 e5 (2017).
308. Lettice, L.A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-35 (2003).
309. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* **20**, 437-455 (2019).
310. Robson, M.I., Ringel, A.R. & Mundlos, S. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol Cell* **74**, 1110-1122 (2019).
311. Kvon, E.Z. *et al.* Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91-5 (2014).
312. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-4 (2013).
313. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).
314. Merkenschlager, M. & Nora, E.P. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu Rev Genomics Hum Genet* **17**, 17-43 (2016).
315. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).
316. Sanyal, A., Lajoie, B.R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109-13 (2012).
317. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572 e24 (2017).
318. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**, 2038-49 (2016).
319. Furlong, E.E.M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341-1345 (2018).

320. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96-100 (2014).
321. Nora, E.P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e22 (2017).
322. Rao, S.S.P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e24 (2017).
323. Hansen, A.S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20-32 (2018).
324. Spiegelman, B.M. & Heinrich, R. Biological control through regulated transcriptional coactivators. *Cell* **119**, 157-67 (2004).
325. Krasnov, A.N., Mazina, M.Y., Nikolenko, J.V. & Vorobyeva, N.E. On the way of revealing coactivator complexes cross-talk during transcriptional activation. *Cell Biosci* **6**, 15 (2016).
326. Hu, D. *et al.* The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol Cell Biol* **33**, 4745-54 (2013).
327. Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc Natl Acad Sci U S A* **113**, 11871-11876 (2016).
328. Dorighi, K.M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell* **66**, 568-576 e4 (2017).
329. Tie, F. *et al.* CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* **136**, 3131-41 (2009).
330. Tie, F. *et al.* Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development* **141**, 1129-39 (2014).
331. Bannister, A.J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384**, 641-3 (1996).
332. Bedford, D.C., Kasper, L.H., Fukuyama, T. & Brindle, P.K. Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. *Epigenetics* **5**, 9-15 (2010).
333. Wang, F., Marshall, C.B. & Ikura, M. Transcriptional/epigenetic regulator CBP/p300 in tumorigenesis: structural and functional versatility in target

recognition. *Cell Mol Life Sci* **70**, 3989-4008 (2013).

334. Yao, T.P. *et al.* Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* **93**, 361-72 (1998).
335. Jin, Q. *et al.* Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J* **30**, 249-62 (2011).
336. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
337. Roe, J.S., Mercan, F., Rivera, K., Pappin, D.J. & Vakoc, C.R. BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia. *Mol Cell* **58**, 1028-39 (2015).
338. Bose, D.A. *et al.* RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell* **168**, 135-149 e22 (2017).
339. Marshall, N.F., Peng, J., Xie, Z. & Price, D.H. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* **271**, 27176-83 (1996).
340. Stasevich, T.J. *et al.* Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **516**, 272-5 (2014).
341. Raisner, R. *et al.* Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. *Cell Rep* **24**, 1722-1729 (2018).
342. Dey, A., Chitsaz, F., Abbasi, A., Misteli, T. & Ozato, K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proc Natl Acad Sci U S A* **100**, 8758-63 (2003).
343. LeRoy, G., Rickards, B. & Flint, S.J. The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. *Mol Cell* **30**, 51-60 (2008).
344. Loven, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-34 (2013).
345. Bhagwat, A.S. *et al.* BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements. *Cell Rep* **15**, 519-530 (2016).
346. Di Micco, R. *et al.* Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep* **9**, 234-247 (2014).
347. Houzelstein, D. *et al.* Growth and early postimplantation defects in mice

deficient for the bromodomain-containing protein Brd4. *Mol Cell Biol* **22**, 3794-802 (2002).

348. Jang, M.K. *et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol Cell* **19**, 523-34 (2005).
349. Yang, Z. *et al.* Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* **19**, 535-45 (2005).
350. Jiang, Y.W. *et al.* Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc Natl Acad Sci U S A* **95**, 8538-43 (1998).
351. Cho, W.K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412-415 (2018).
352. Rahnamoun, H. *et al.* RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. *Nat Struct Mol Biol* **25**, 687-697 (2018).
353. Allen, B.L. & Taatjes, D.J. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16**, 155-66 (2015).
354. Robinson, P.J. *et al.* Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166**, 1411-1422 e16 (2016).
355. Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. & Kornberg, R.D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608 (1994).
356. Poss, Z.C., Ebmeier, C.C. & Taatjes, D.J. The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* **48**, 575-608 (2013).
357. Knuesel, M.T., Meyer, K.D., Bernecky, C. & Taatjes, D.J. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev* **23**, 439-51 (2009).
358. Soutourina, J. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol* **19**, 262-274 (2018).
359. Robinson, P.J. *et al.* Molecular architecture of the yeast Mediator complex. *Elife* **4**(2015).
360. Ito, M., Yuan, C.X., Okano, H.J., Darnell, R.B. & Roeder, R.G. Involvement of the TRAP220 component of the TRAP/SMCC coactivator complex in embryonic development and thyroid hormone action. *Mol Cell* **5**, 683-93 (2000).

361. Fant, C.B. & Taatjes, D.J. Regulatory functions of the Mediator kinases CDK8 and CDK19. *Transcription* **10**, 76-90 (2019).
362. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-19 (2013).
363. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
364. Dang, C.V. MYC on the path to cancer. *Cell* **149**, 22-35 (2012).
365. Chipumuro, E. *et al.* CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* **159**, 1126-1139 (2014).
366. Stathis, A. & Bertoni, F. BET Proteins as Targets for Anticancer Treatment. *Cancer Discov* **8**, 24-36 (2018).
367. Chapuy, B. *et al.* Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777-90 (2013).
368. Banani, S.F., Lee, H.O., Hyman, A.A. & Rosen, M.K. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
369. Pak, C.W. *et al.* Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol Cell* **63**, 72-85 (2016).
370. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28-45 (2020).
371. Sabari, B.R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**(2018).
372. Itoh, Y. *et al.* 1,6-hexanediol rapidly immobilizes and condenses chromatin in living human cells. *Life Sci Alliance* **4**(2021).
373. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. & Sharp, P.A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23 (2017).
374. Shrinivas, K. *et al.* Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol Cell* **75**, 549-561 e7 (2019).
375. Richart, L., Bidard, F.C. & Margueron, R. Enhancer rewiring in tumors: an opportunity for therapeutic intervention. *Oncogene* (2021).
376. Bradner, J.E., Hnisz, D. & Young, R.A. Transcriptional Addiction in Cancer. *Cell* **168**, 629-643 (2017).
377. Alekseyenko, A.A. *et al.* The oncogenic BRD4-NUT chromatin regulator

drives aberrant transcription within large topological domains. *Genes Dev* **29**, 1507-23 (2015).

378. Herz, H.M., Hu, D. & Shilatifard, A. Enhancer malfunction in cancer. *Mol Cell* **53**, 859-66 (2014).
379. Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189-95 (2011).
380. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454-1458 (2016).
381. Lupianez, D.G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).
382. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067-73 (2010).
383. Lasko, L.M. *et al.* Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. *Nature* **550**, 128-132 (2017).
384. Allegretto, E.A. *et al.* Transactivation properties of retinoic acid and retinoid X receptors in mammalian cells and yeast. Correlation with hormone binding and effects of metabolism. *J Biol Chem* **268**, 26625-33 (1993).
385. Chan, H.M. & La Thangue, N.B. p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J Cell Sci* **114**, 2363-73 (2001).
386. Katoh, M. Functional and cancer genomics of ASXL family members. *Br J Cancer* **109**, 299-306 (2013).
387. Bowman, C.J., Ayer, D.E. & Dynlacht, B.D. Foxk proteins repress the initiation of starvation-induced atrophy and autophagy programs. *Nat Cell Biol* **16**, 1202-14 (2014).
388. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
389. Cho, W.K. *et al.* RNA Polymerase II cluster dynamics predict mRNA output in living cells. *Elife* **5**(2016).
390. Ali, M.Z., Choubey, S., Das, D. & Brewster, R.C. Probing Mechanisms of Transcription Elongation Through Cell-to-Cell Variability of RNA Polymerase. *Biophys J* **118**, 1769-1781 (2020).
391. Guo, Y.E. *et al.* Pol II phosphorylation regulates a switch between

transcriptional and splicing condensates. *Nature* **572**, 543-548 (2019).

392. Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ay, F. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat Commun* **10**, 4221 (2019).

393. Seif, E. *et al.* Phase separation by the polyhomeotic sterile alpha motif compartmentalizes Polycomb Group proteins and enhances their activity. *Nat Commun* **11**, 5609 (2020).

RÉSUMÉ

Chez les eucaryotes, la maintenance de l'identité cellulaire implique le contrôle précis de l'expression des gènes. Ceci résulte de l'action concertée des facteurs de transcription et des facteurs contrôlant la structure de la chromatine. Les complexes répresseurs Polycomb (PRC1 et PRC2) sont des modificateurs de la chromatine qui orchestrent la répression transcriptionnelle en catalysant respectivement l'ubiquitinylation de H2A (H2Aub) et la méthylation de H3K27. A l'inverse, BAP1 (BRCA1-Associated Protein 1) favorise la transcription en retirant H2Aub, agissant donc comme un antagoniste de PRC1. Toutefois, les détails du mécanisme par lequel BAP1 régule la transcription restent mal compris. L'interaction entre PRC1 et PRC2 est également un sujet encore débattu. Mon projet de thèse visait à étudier ces deux importantes questions.

(1) La protéine BAP1 est localisée à une fraction des enhancers où elle stabilise le recrutement de BRD4**.**

Dans ces études, nous avons montré que BAP1 favorise la transcription en s'opposant au complexe PRC1 et que BAP1 est inerte en son absence. Des analyses à l'échelle du génome entier ont révélé que la protéine BAP1 est recrutée à une fraction des enhancers. Par ailleurs, l'inactivation de BAP1 amène à l'accumulation de H2Aub et à l'altération du recrutement de BRD4. En accord avec ces résultats, des expériences de microscopie à super résolution indiquent une réduction des condensées de BRD4 et de MED1 dans les cellules knockout pour BAP1. Cela suggère que BAP1 a un rôle crucial pour l'intégrité de certains enhancers. De façon importante, en traitant des cellules isogèniques avec des inhibiteurs de BET, nous avons montré que les cellules mutantes pour BAP1 montrent une sensibilité particulière à l'inhibition de la prolifération. Ces résultats suggèrent que promouvoir les perturbations des enhancers pourrait constituer une stratégie thérapeutique dans les pathologies où le gène BAP1 est muté.

(2) PRC2 réprime la transcription indépendamment de PRC1**.**

PRC1 et PRC2 ont été considérés depuis longtemps comme agissant de concert pour maintenir la répression. Toutefois, en analysant les profiles transcriptomiques de cellules où soit PRC1, soit PRC2, soit les deux sont inactivés, nous avons démontré que PRC1 et PRC2 peuvent agir de façon autonome pour réprimer la transcription. Au travers d'approches non-biaisées et d'approches basées sur une sélection de gènes candidats, nous essayons d'identifier les effecteurs de cette répression dépendant exclusivement de PRC2. Cela implique l'étude de protéines préalablement proposées comme interagissant avec H3K27me3. Cette étude est en cours mais il est probable qu'elle va révéler de nouveaux acteurs de la répression dépendant de PRC2.

MOTS CLÉS

régulation transcriptionnelle, machinerie Polycomb, BAP1, ubiquitinylation, enhancers

ABSTRACT

In eukaryotes, the maintenance of cell identity entails the precise control of gene expression, which results from the concerted actions of transcription factors and factors controlling chromatin structure. Polycomb repressive complex 1 and 2 (PRC1 and PRC2) are chromatin modifiers that orchestrate transcriptional repression by catalyzing H2Aub and H3K27me3, respectively. By contrast, BRCA1-associated protein 1 (BAP1) promotes transcription by removing H2Aub, acting as an antagonist of PRC1. However, the detailed mechanism of how BAP1 regulates transcription remains largely elusive. The interplay between PRC1 and PRC2 is also far from being fully understood. My PhD study aimed at investigating the underlying mechanisms for these two important questions.

(1) BAP1 is recruited to a subset of active enhancers where it stabilizes BRD4 occupancy.

In these studies, we showed that BAP1 promotes transcription by opposing PRC1 activity, and that BAP1 is mostly inert in its absence. Genome-wide analysis revealed that BAP1 is recruited to a subset of active enhancers. Besides, inactivation of BAP1 led to accumulation of H2Aub and impaired BRD4 recruitment. Consistently, super-resolution microscopy demonstrated reduced condensates of BRD4 and MED1 in BAP1-KO cells. This suggests that BAP1 has a crucial function for the integrity of a subset of enhancers. Importantly, by treating isogenic cells with BET inhibitors, we showed that cells mutant for BAP1 display a more pronounced proliferative response. This result suggests that further perturbation of enhancers function could be a therapeutic strategy for BAP1-null malignancies.

(2) PRC2 represses transcription independently of PRC1.

PRC1 and PRC2 are long considered cooperating to maintain gene repression. However, analyzing transcriptomic profiles of PRC1-null, PRC2-null and PRC1/2-null cells, we demonstrated that both PRC1 and PRC2 can autonomously repress transcription. Through both unbiased and candidate-based approaches, we focus on identifying downstream effectors of PRC2-mediated silencing in the absence of PRC1. This includes investigating the roles of previously proposed H3K27me3 readers. While this study is still ongoing, it is likely that it will reveal new actor for PRC2-mediated repression.

KEYWORDS

transcriptional regulation, Polycomb machinery, BAP1, ubiquitination, enhancers